UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
INSTITUTE OF SOUND AND VIBRATION RESEARCH

ACOUSTIC MODELS OF CONSONANT RECOGNITION IN COCHLEAR IMPLANT USERS

BY

CARL VERSCHUUR

THESIS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

MARCH 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
INSTITUTE OF SOUND AND VIBRATION RESEARCH

Doctor of Philosophy

ACOUSTIC MODELS OF CONSONANT RECOGNITION BY COCHLEAR IMPLANT USERS

by Carl Verschuur

Normal-hearing adults have no difficulty in recognising consonants accurately, even in moderately adverse listening conditions. By contrast, users of multichannel cochlear implants have difficulty with the accurate perception of consonants, even in good listening conditions. Cochlear implant users are known to show systematic deficits in recognition of consonant features, with perception of the place feature, which relies on spectral information, being worst. These deficits may be attributed both to signal distortions introduced by the processing of the implants and to other factors, in particular the spectrotemporal distortions which occur at the interface between electrode array and auditory nervous system, including cross-channel interaction. The objective of the work reported here was to attempt to partial out the relative contribution of these different factors to consonant recognition. This was achieved by comparing cochlear implant users' perceptual errors, analysed in terms of information transmission, with errors made by normal-hearing subjects listening to acoustic models of implant processing, in various conditions.

Two initial experiments were undertaken to develop and refine an acoustic model of the Nucleus 24 cochlear implant. Findings from these two experiments informed the design of the main acoustic model experiment, which was undertaken in parallel with a further experiment involving users of the Nucleus 24 device. In both experiments, subjects listened to nonsense syllables with and without the addition of stationary background noise, in three different configurations of implant processing parameters. Additionally, in the acoustic model experiment, a simulation of cross-channel spread of excitation, or "channel interaction", was varied. Results showed that acoustic model experiments were predictive of the pattern of consonant feature transmission in cochlear implant users with better baseline consonant recognition scores. Deficits in consonant recognition in this subgroup could be explained by the loss of phonemically relevant acoustic information in speech due to the nature of cochlear implant processing, while channel interaction appeared to play a smaller role in accounting for problems in consonant recognition. The work also evaluated the effect of changes in channel number and stimulation rate and failed to find any changes in consonant recognition as these parameters were varied. The lack of a stimulation rate effect was consistent with acoustic measurements of the temporal modulation transfer function of the processor, which showed almost no change across stimulation rates.

Contents

List of figures

List of tables

List of equations

Declaration of authorship

I, Carl Verschuur, declare that the thesis entitled
"Acoustic models of consonant recognition in cochlear implant users"
and the work presented in it are my own. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted the published work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed:……………………………………………………………………

Date:………………………………………………………………………

Acknowledgements

I would like to thank my PhD supervisor, Mark Lutman, whose insights moved my thinking on so much during the time I was working on this thesis. I would also like to thank him more generally for his support in developing my academic skills and my academic career. I would also like to thank my thesis reviewers, David Simpson, Bob Damper and Christine Schadle, for their time and hard work. I also wish to thank other colleagues at the ISVR for advice and support, including Daniel Rowan and Julie Brinton in particular. The thesis could not have been completed successfully without the help and support of the patients and staff of the South of England Cochlear Implant Centre. I would also like to acknowledge Michael Bűchler and Philip Laneau who were invaluable in developing the "AMO" MATLAB code and also Brett Swanson who developed the NIC-STREAM software. Finally, I would like to dedicate this thesis to my wife, Bronagh Walsh, without whose support and advice, both practical and intellectual, I could not have completed this thesis.

List of abbreviations

ACE	Advanced Combination Encoder
AFC	Alternative forced choice
AM	Acoustic model
ANOVA	Analysis of variance
CI	Cochlear implant
CIS	Continuous interleaved sampling
FFT	Fast Fourier transform
IIR	Infinite impulse response
MANOVA	Multivariate analysis of variance
NH	Normal hearing
PPS/CH	Pulses per second per channel
RMS	Root mean square
SAM	Sinusoidal amplitude modulation
SD	Standard deviation
SNR	Signal to noise ratio
SPL	Sound pressure level
TMTF	Temporal modulation transfer function

Chapter 1. Introduction

A cochlear implant (CI) is a surgically implanted auditory prosthesis that bypasses an impaired peripheral auditory system by means of direct stimulation of the residual neural elements in the auditory system. Cochlear implantation has become widely accepted as a cost-effective and beneficial treatment for profound sensorineural deafness (UK Cochlear Implant Study Group, 2004). Improvement in speech perception is arguably the most important single outcome of cochlear implantation and is linked to broader outcomes in linguistic, social and educational functioning. Speech perception outcomes from cochlear implantation have improved markedly with improved design of hardware and signal processing (Wilson, 1997; Zeng, 2004) and the majority of current CI users can expect significant benefit to open set speech recognition (Meyer et al., 2003). Nevertheless, even the most successful CI users are still poorer than normal hearing (NH) listeners at speech discrimination, particularly in adverse listening conditions. Moreover, there are large differences between individual CI users that are not fully understood.

One likely reason that even the best performing CI users do not achieve normal levels of speech perception, particularly in background noise, is that CI signal processing does not replicate the complex nonlinear processes involved in the normal peripheral auditory system. Instead, CI processing resembles the channel vocoder, a processing system which minimises electronic bit rate by coding the spectral envelope of a speech signal only (Dudley, 1939). Consequently, the information provided by CI processing to the auditory nervous system is somewhat impoverished compared to information provided by the normally functioning peripheral auditory system (Cohen et al., 2003; Loizou, 1999). It is therefore useful to understand the signal received by the CI user in terms of the various forms of *information loss* it has undergone compared to the equivalent signal that would be received by a NH listener. To aid in the analysis of information loss, figure 1.1 illustrates a simple communication chain describing the main stages of CI information processing.

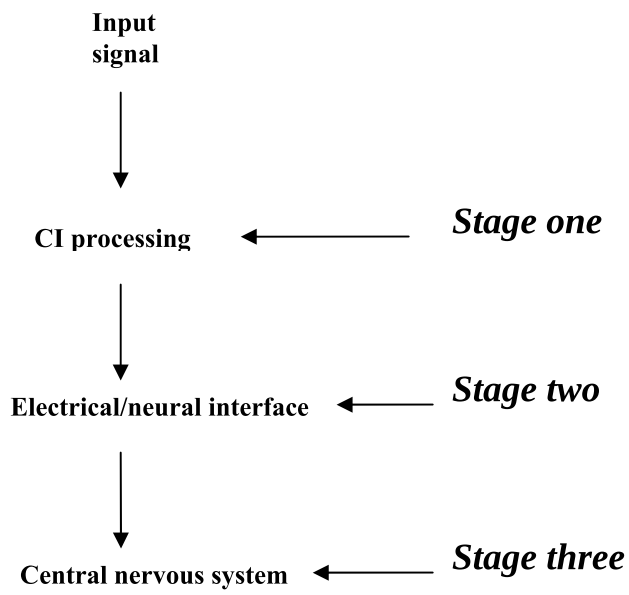

Figure 1.1 Stages of information processing in CI users

According to this figure, the first stage of the chain is the acoustic signal itself. The second stage is the processing of that signal by the CI. It is clear that the signal delivered by the CI to the electrode array is reduced in detail compared to the signal delivered by a healthy auditory system to the auditory nerve. The third stage is the interface between the CI electrode array and the auditory nervous system, referred to here as the electrical/neural interface. The fourth and final stage is the processing of the neural signal by the central nervous system. At each stage of the chain there may be loss of information necessary for accurate consonant recognition. At each stage, the degree of information loss (as opposed to simply signal loss) depends on what type of acoustic information is important for signalling a particular consonant or consonant contrast and also on the presence of any background noise or other environmental signal distortion. Thus the question at each processing stage is not simply, how does the signal differ from a signal processed by the healthy auditory system, but rather, how does the signal differ in terms of its information-bearing properties.

Although speech perception in CI users is a result of the interaction of the different domains outlined in figure 1.1, it is crucially important to understand where in the processing chain information is lost, in order to know how best to modify processing or hardware to design to optimise listener performance. An assumption in this study is that explaining deficits in CI users' consonant recognition should *start* with understanding stage 1, e.g. the effect of CI signal processing on the signal. If the explanation does not lie in this domain, e.g. stage 1 of the simple conceptual model, then stage 2, the electrical/neural interface, should be determined. If this is ruled out as the possible explanation for the perceptual deficit, only then should deficits in stage 3, central auditory function, be assumed. For most adult CI users, deafness has occurred after a lengthy period of normal hearing or at least a good level of auditory function with hearing aids, prior to the onset of severe/profound deafness. It seems reasonable to assume that, in most cases, the potential for normal or near-normal central auditory processing abilities remains (assuming that adequate re-acclimatisation to the CI signal has taken place).

In this thesis information loss is considered exclusively in the specific context of consonant recognition. There are a number of reasons for focusing on this one method of assessing speech perception. First, most information in speech is conveyed by consonants rather than vowels (Owens et al., 1968). Second, analysis of consonant recognition can be linked to underlying psychoacoustic abilities such as frequency or temporal resolution. This is because consonant recognition can be unpacked into perception of a number of features, each of which has acoustic, and therefore psychoacoustic, correlates. There is an existing framework for understanding consonant information transmission through the analysis of consonant confusion matrices and feature-specific information transmission. Third, such an approach can make use of the strong evidence base from NH listeners and the scientific disciplines of acoustic phonetics and phonology to understand the factors affecting consonant recognition. This approach is not meant to imply that understanding consonant (feature) recognition can provide a complete account of speech perception. There are a number of perceptual tasks involved in ongoing speech perception, including phonemic segmentation, whole-word recognition and the use of non-auditory cues (Liberman et al., 1967). Nevertheless, the analysis of transmission of specific consonant features provides a useful means of analysing efficiency of information transmission through a CI system.

This thesis describes a study whose main aims were, first, to investigate the factors affecting consonant recognition in CI users and, second, to compare different acoustic models (AM) of CI processing in terms of their ability to predict consonant recognition performance. The overriding question motivating the research was: "to what extent can deficits in consonant recognition by CI users be explained by information loss in CI signal processing as opposed to information loss at the electrical/neural interface?" The work described in the thesis contributes to the existing literature in a number of ways. It adds to the knowledge base on AM methodology by demonstrating that noise band carriers provide a better model of consonant recognition than sine wave carriers, and by showing that useful perceptual data can be gathered from NH subjects listening to AM stimuli within a time-efficient approach to testing, even where the AM stimuli include large spectral distortions. Moreover, the work shows the usefulness of using an AM which is based in detail on the processing of a specific device, and where a comparison is made directly between AM and CI data that are truly "equivalent" in processing terms.

The work also adds to the literature by showing that deficits in consonant recognition in (at least better performing) CI users can be attributed mainly to information loss associated with CI processing The argument for this is supported at various points in the text by acoustic analyses which demonstrate the limitations imposed on consonant information by CI processing and on the temporal response of the CI system in particular. This includes original measurements of temporal modulation transfer functions in order to describe the temporal response of the Nucleus 24 CI processor. The work adds to the existing literature on the effects of processing parameters on speech perception. The work also contributes to the understanding of speech perception in noise by CI users by showing the pattern of consonant feature recognition deficits in background noise and suggesting some reasons for the pattern of noise effects. Findings from the study also suggest some possible explanations for inter-user variation in speech perception, and in particular support the idea that channel interaction may not be the main reason for variation in performance between CI users. Two papers based on the original work in this thesis are currently being prepared for publication with further papers also likely.

Chapter 2 provides an overview of evidence and arguments relevant to the question "to what extent can deficits in consonant recognition by CI users be explained by information loss in CI signal processing as opposed to information loss at the electrical/neural interface?" This includes an overview of consonant recognition in CI users (2.1), CI signal processing (2.2), evidence regarding effects of signal processing (2.3), effects of electrical/neural interface signal distortions (2.4), use of AMs in CI research (2.5) and the likely relationship between CI processing in particular and transmission of particular consonant features (2.6). Chapters 3 to 5 are concerned with describing the original experimental work. Chapter 3 provides an overview of experimental methodology. Chapter 4 describes experimental work concerned with determining the most appropriate parametric choices for AM studies, and attempts to validate a particular AM of the Nucleus 24 device. Chapter 5 goes on to describe a "matched pair" of AM and CI experiments which form the main experimental work in the study. Chapter 6 provides an overview of results across experiments with reference to transmission of specific consonant features, while chapter 7 provides a more general discussion of results and their scientific and clinical implications. Chapter 8 briefly summarises the main conclusions of the study.

Chapter 2. Background

2.1. Overview of consonant recognition in CI users

2.1.1 Theoretical background to consonant recognition analysis

The ability to recognise speech relies on a number of underlying perceptual abilities. One of these is the ability to determine which phoneme has been uttered, out of the possible phoneme inventory of a particular language. Disregarding issues of context and semantics, the listener must make a decision based on available acoustic evidence, by decoding the acoustic patterns, or cues, which distinguish one phoneme from another. One way to understand this process is to analyse the errors made by a listener when attempting to determine which phoneme s/he has heard. The errors shed light on the perception of different consonant features. The categorisation of speech features into *a priori* categories is motivated by knowledge of the important variations in speech production, which have reasonably well-understood consequences in terms of acoustics and therefore perception.

The main theory underlying our basic understanding of speech production and its implications for speech acoustics is the "source-filter theory of speech production" which was first described by Fant (1970). The theory posits that the link between speech production and the resulting speech waveform can be described as the sum of two independent processes. First, the source of speech energy is generated via the vibration of the vocal folds (voicing) or, if the vocal folds remain open, via turbulence or friction generated by partial or complete occlusion in the upper vocal tract. Vocal fold vibration generates a quasi-periodic signal which can be characterised in the frequency domain as consisting of a fundamental frequency with multiple harmonics that decrease in amplitude as a function of frequency, as in figure 2.1. The source energy for unvoiced speech is aperiodic and is therefore associated with a wider and diffuse, spectrum, which may have more energy in higher frequencies. Second, the upper vocal tract acts as a dynamic filter which transforms the source spectrum into a more complex and varying waveform. Depending on the location within the upper vocal tract where maximum constriction occurs (place of articulation) and how the vocal tract is occluded (manner of articulation), different transfer functions will result.

The resulting acoustic waveform can be characterised as a convolution of the source spectrum and the filter effect of the upper vocal tract as in figure 2.1. Where source harmonics coincide with filter maxima, greater energy is produced than where these do not coincide. A NH listener can resolve both source harmonics and formant structure and therefore determine information about speaker voice characteristics, which are determined by characteristics of the underlying source, and also determine segmental contrasts, e.g. phoneme differences, which rely primarily on the differences in formant structure and changes to formant structure over time. It should be noted that the frequency resolution required by a listener to resolve the harmonics of the source spectrum is considerably greater than that required to resolve the formants introduced by the filter. (It is shown in 2.6.1 that Nucleus 24 processing restricts F0 information.)

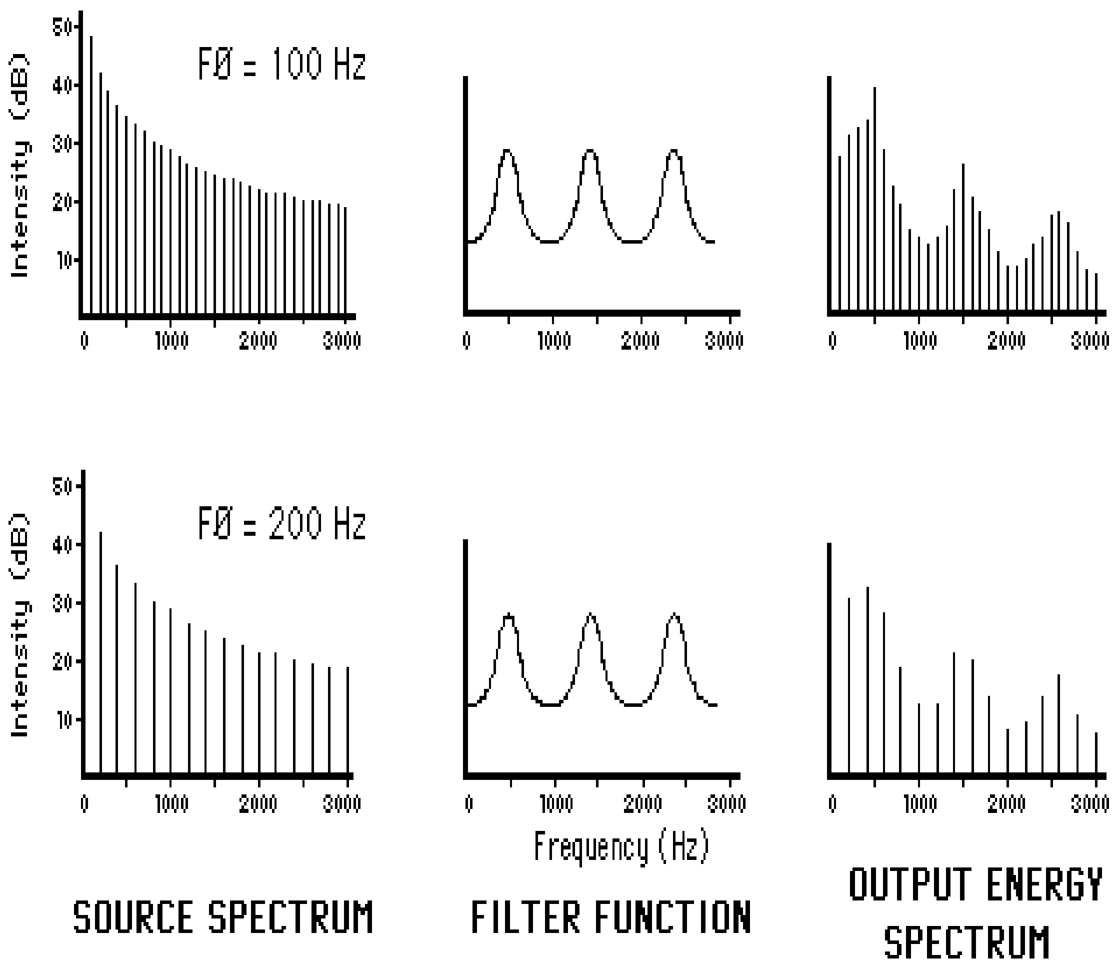

Figure 2.1. Interaction between source spectrum and filter function to produce speech waveform. Adapted from Lieberman and Blumstein, 1988.

Although the assumption of independence or source and filter has been questioned by some authors (Titze, 2004; Childers and Wong, 1994), the basic principles of the source-filter theory provide the underpinning for the classification of speech sounds. The classification of English consonants in terms of phonological features has developed with changes in phonetic science. In practice, there is no single accepted classification scheme. Chomsky and Halle (1968) described a range of binary features to describe English consonants. However, studies of CI users' consonant and vowel recognition has tended to use the tripartite distinction of voicing, place and manner, at least for non-tone languages such as English.

Voicing refers to the presence or absence of vocal fold vibration during production of a particular speech sound. Consequently, it is a binary category, at least in English. Manner of articulation refers to the way in which the vocal tract is occluded and for English consonants a convenient categorisation recognises four main manner categories: nasal, stop, fricative and approximant (the latter category can further be broken into liquids and semivowels/glides). Place of articulation refers to the locus of maximum occlusion within the vocal tract. For English consonants place classification can vary in terms of number of categories depending on how specific an analysis is required. Figure 2.2 shows the International Phonetic Association's detailed classification of consonants in terms of voicing, place and manner.

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC) © 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Figure 2.2. International phonetic alphabet classification of consonants by voice, place and manner. Voiceless cognates are indicated on left, voiced on right. Manner categories are on the y-axis and place categories are on the x-axis. Reproduced with permission from the International Phonetic Association. Copyright 2005 by International Phonetic Association (http://www.arts.gla.ac.uk/IPA/ipa.html).

Changes in feature values produce specific acoustic consequences, depending on the different acoustic patterns or cues consequent to each feature variation. Changes to place of articulation cause spectral changes as the residual volume of the unoccluded vocal tract changes. Changes in the manner of vocal tract occlusion tend to lead to differences in temporal information; for example, a stop consonant is associated with a sudden and short duration release burst whereas a fricative is associated with turbulent energy of longer duration. Finally, distinctions between voiced and unvoiced consonants tend to reflect timing differences, although there are also spectral consequences of voicing distinctions. More detail is given on the acoustic cues signalling consonant features in section 2.6.

Measures of feature-specific information transmission are obtained by using a closed set consonant recognition task from which a consonant confusion matrix can be obtained. An example confusion matrix is shown in table 2.1. Here stimuli are on the vertical axis along the left while responses are on the horizontal axis along the top.

Responses are given as total out of 100. Deviations from the diagonal line (given in bold) represent errors. It is possible to derive an analysis of perceptual errors in terms of phonological feature and from this to infer how different acoustic speech cues are being processed. To take an example, when the phoneme /b/ is presented (seen on the y-axis), 96% of responses are correct whereas 4% of responses are incorrect, namely the phoneme /d/. This represents a place of articulation error, but not a voicing or manner error (both /d/ and /b/ are voiced stops, the only difference is that /d/ is alveolar in place whereas /b/ is bilabial). The simplest feature-specific measure that can be used is therefore percentage correct. In this case, if all other phoneme response replicated the same error pattern then the result would be 100% correct for voicing and manner but 96% correct for place.

Table 2.1. Example consonant confusion matrix with 15 consonant alternatives.

	b	d	g	w	j	ɹ	l	v	z	ʤ	m	n	p	t	k
b	**96**	4	0	0	0	0	0	0	0	0	0	0	0	0	0
d	0	**89**	4	0	0	0	0	0	4	0	0	0	0	4	0
g	7	74	**4**	0	0	0	0	0	0	0	0	0	4	7	4
w	0	0	0	**19**	0	56	19	0	0	0	7	0	0	0	0
j	0	4	0	0	**15**	0	37	11	0	7	7	0	0	0	0
ɹ	0	0	0	7	0	**70**	22	0	0	0	0	0	0	0	0
l	0	0	0	7	0	11	**56**	7	0	0	11	0	0	0	7
v	0	0	7	0	0	0	0	**67**	0	0	0	0	19	0	7
z	0	4	0	0	0	0	0	33	**33**	0	0	0	0	0	0
ʤ	0	4	48	0	4	0	0	0	0	**41**	0	0	0	4	0
m	0	0	0	0	0	4	22	0	0	0	**74**	0	0	0	0
n	0	0	0	0	0	0	19	4	0	0	70	**7**	0	0	0
p	0	0	0	0	0	0	0	0	0	0	0	0	**96**	0	4
t	0	0	4	0	0	0	0	0	0	4	0	0	0	**70**	4
k	0	4	0	0	0	0	0	0	0	0	0	0	15	33	**48**

Because different features have different chance correct scores, the use of the percentage correct by feature is problematic if the intention of the researcher is to compare perception of different features. For example, in English, consonants are either voiced or unvoiced, and therefore this feature has two levels. By contrast, different categorisation schemes for manner can yield between four and seven manner categories. Therefore, the same proportion of errors for these two features must be interpreted differently and if proportion correct for each feature is used, the interpretation of results is cumbersome. A more sophisticated approach, which is used as a standard measure in consonant recognition analysis in CI users, is information transmission by feature. This measure allows for differences in the chance level across feature and thereby facilitates a more appropriate comparison of perception across features. The approach was first proposed by Miller and Nicely (1955). In the study 16 consonants in the /aCa/ vowel environment, presented by a female speaker, were presented to NH listeners at varying signal-to-noise ratios (SNRs). They analysed the pattern of consonant confusions by listening condition using five *a priori* consonant features, namely voicing, nasality, affrication, duration and place of articulation. The authors found marked differences in information transmission across the five features as a function of SNR, with place being most susceptible to noise interference.

Miller and Nicely's approach to consonant recognition testing and consonant confusion data analysis remains highly influential: both their calculation of information transmission and the use of a VCV consonant confusion task with the /aCa/ vowel environment have been used in almost all studies of consonant feature recognition in CI users or AMs. Wang and Bilger (1973) proposed a refinement of the original method, called SINFA (sequential information transfer) analysis, a statistical technique similar to multiple regression which allows analysis of confusion patterns. The authors proposed a recursive method for partialling out the independent contributions of different phonological features. This constituted a series of "iterations". The first iteration derives the unconditional transmitted information estimated for each feature in the proposed feature system. This normalizes the features for inequalities in stimulus feature information and is equivalent to the information transmission measure proposed by Miller and Nicely (1955). Generally, this has been the approach used in CI and AM consonant feature transmission studies, although Xu et al. (2005) did make use of multiple iterations of SINFA analysis. However, interpretation of consonant confusion data analyses in this way can be difficult, as the same feature may be optimally transmitted with different numbers of iterations in different conditions.

Each of these three approaches to analysis of consonant confusion data (simple percentage correct, information transmission, or multiple-iteration SINFA analysis) has potential advantages: A simple measure of percentage correct could provide lower variability with small subject numbers, while use of a multiple-iteration SINFA approach can reduce the effect of correlation between features. However, the approach in the relevant literature has been to use the information transmission measure (e.g. a single iteration of SINFA), as defined by Miller and Nicely (1955). This approach is taken in the present work, to optimise comparison against other relevant studies, although this does not contradict the possible benefits of alternative approaches:

2.1.2 Consonant recognition in quiet

To understand consonant recognition in CI users, it is first necessary to be able to describe how it differs from NH listeners. By "how" is meant "what pattern of feature error pattern?" Hence two questions can be addressed:

- How does consonant recognition differ between CI users and normal-hearing listeners?
- What effect does noise have on consonant recognition in CI users?

Additionally, the relative importance of spectral and temporal resolution was raised in relation to CI users' consonant recognition: Hence:

- Are deficits in consonant recognition in CI users due primarily to interference with temporal processing, with spectral processing, or equally with both?

More broadly, it is not adequate to simply characterise abnormalities in CI users' consonant recognition in quiet and noise without then defining:

- What factors affect consonant recognition in CI users?

With regard to the first question, the relevant literature shows that, first of all, consonant recognition in CI users is markedly worse than in NH listeners (even in quiet and even in "better listeners") and, second, that place of articulation perception in consonants in CI users is significantly worse than manner or voicing perception, to a far greater extent than is the case with NH listeners. Every study that has evaluated CI users' consonant feature information transmission has found that perception of place of articulation is poorer than manner and voicing in quiet (Donaldson and Nelson, 2000; Dorman et al., 1991; Van Tasell et al., 1992; Dorman et al., 1990; Dorman et al., 1991; Dorman, 1995). Figure 2.3 shows performance across a number of studies for voicing, place and manner in quiet. (Studies included in the chart are restricted to those studies in which CI users' consonant recognition abilities were tested and analysed in terms of information transmission of consonant features.) The figure also includes data collated by the author for over 60 adult CI users from the South of England Cochlear Implant Centre (SOECIC). Two further details of these studies should be noted: first, all the studies used performance in quiet and, second, with the exception of Geurts and Wouters (1999), all studies undertook consonant recognition measures using VCV nonsense syllables of the form /aCa/, e.g. where the vowel /a/ precedes and follows the target consonant. Data are presented for the "best" performance conditions for those studies where comparisons of different listening or processing parameter conditions were undertaken.

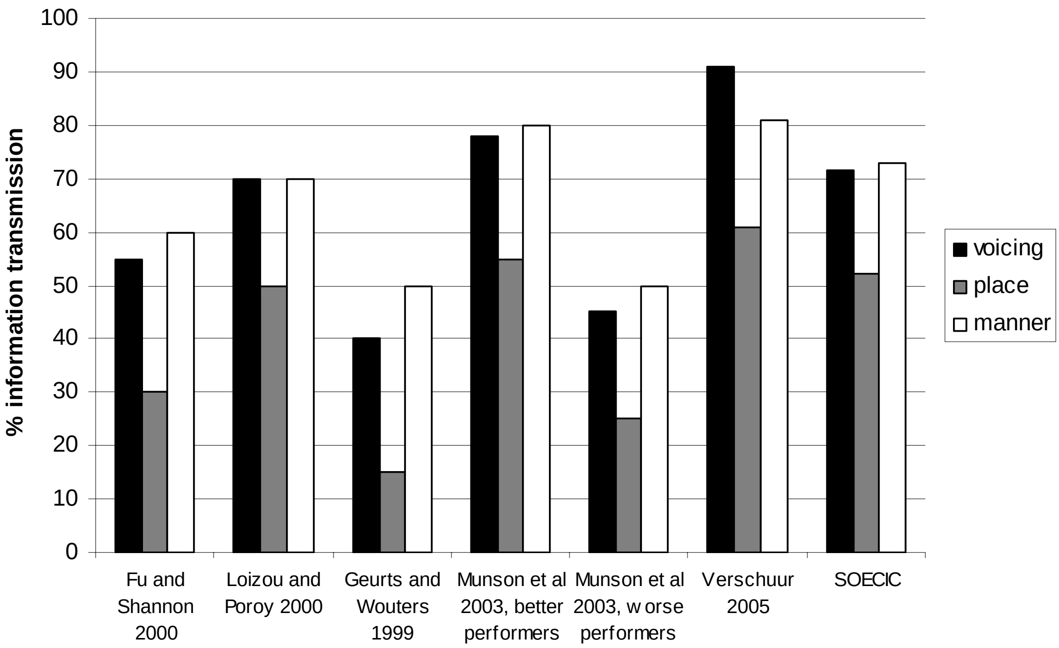

Figure 2.3. Consonant voicing, place and manner transmission from studies of CI user performance

The accompanying table 2.2 shows further details of the studies, although the table is not fully comprehensive in terms of the many variations across studies, which differed in other aspects of consonant confusion analysis, e.g. choice and number of stimuli, number of repetitions per stimulus, male vs. female vs. mixed speaker and number of tokens per speakers (although it is important to note almost all studies cited used consonant recognition in the /aCa/ vowel environment). Moreover, studies varied by subject parameters, e.g. CI devices, processing strategies and baseline speech perception abilities. Given the heterogeneity of both stimulus and subject characteristics across studies, it is interesting that the "worse place performance" pattern of results is so consistent. Although absolute levels of transmission vary between the studies, relative transmission across features is less variable. Moreover, it should be noted that NH listeners show transmission levels approximating 100% for equivalent stimuli in quiet and therefore none of the features can be said to be transmitted "normally", at least when averaged across a group of CI users.

Table 2.2. Parameters for data sets in figure 2.3. (Further details of implant types are given in 2.3)

Study	Parameters	Method	Implant
Fu and Shannon 2000	500 pps/ch x 4 CIS	16AFC,aCa, 2 tokens x 2 reps, mixed gender	N22
Loizou and Poroy 2000	2100 pps/ch x 6 CIS	20 AFC, averaged across aCa, iCi, uCu, female	MED-EL
Geurts and Wouters 1999	CIS	averaged across aCa, iCi, uCu,initial consonants, mixed gender	LAURA
Munson et al 2003, better performers	mixed	19AFC, aCa, mixed gender	N22, Clarion
Munson et al 2003, worse performers	mixed	19AFC, aCa, mixed gender	N22, Clarion
Verschuur 2005	>1500pps/ch, 12 channels	20AFC, aCa, female	MED-EL
SOECIC	mixed	20 AFC, aCa, female	N24, N22, MED-El

Several authors have suggested that this discrepancy between place and manner/voicing perception can be explained by the fact that CI users' spectral resolution is relatively poor compared to that of NH listeners (Dawson et al., 2000; Dorman and Loizou, 1997; Dorman et al., 2000; Loizou et al., 1999; Loizou et al., 2000b), whereas temporal processing is less impaired when compared to NH listeners' abilities (Busby et al.; 1993;Hescot et al.; 2000;Shannon, 1992). However, a distinction should be made between underlying psychophysical capacity as against information loss associated with CI processing. Both CI processing and electrical/neural interface factors may be implicated in poor place of articulation transmission. Spectral resolution is reduced by the way in which the CI transforms the signal into a relatively small number of envelope values (up to 22 depending on device), but spectral information could also be further affected by spread of excitation in the electrical/neural interface (these factors are discussed in 2.3.2 and 2.4.1).

Although the research literature has emphasised poor place performance, it is still worth noting that voicing and manner are still poor compared to normal performance. NH listeners obtain near to 100% in the listening conditions of the tests (e.g. in quiet at 60 dB SPL or greater). Therefore, if manner and voicing do rely on temporal envelope information then it follows that temporal envelope perception must also be impaired in CI users compared to NH listeners, whether because of information loss due to CI processing, the electrical/neural interface or the central auditory nervous system.

2.1.3 Consonant recognition in noise

For CI users, background noise has a deleterious effect on speech perception (Dorman et al., 1998a; Fetterman and Domico, 2002) although the same is true for NH listeners or hearing aid users, albeit to a lesser extent. For hearing aid users and NH listeners, upward spread of masking plays a particularly important role in reducing speech intelligibility in background noise, although other factors such as reduced frequency resolution may also play a role (Moore, 1996). CI users generally start to become worse at sentence recognition with SNRs of +10 or +15, whereas for NH listeners or even hearing aid users speech perception is robust up to negative SNRs. There is some evidence regarding difference in interference with different noise types, although these studies do not provide data about specific consonant features: Nelson et al. (2003) found that CI users' sentence perception was worse with modulated speech-like background noise compared to stationary noise, whereas the reverse is the case for NH listeners. This was hypothesised to be due to CI users' inability to use temporal modulations to achieve release from masking. Fu and Nogaki (2005) also found that CI users did not show the same release from masking with modulated noise as is shown in NH listeners.

A crucial question is the extent to which impaired spectral, or impaired temporal, processing is to blame in the deterioration of performance in noise by CI users. Spectral resolution for CI users effectively means comparison of stimulation levels between different channels (Loizou and Poroy, 2001). Ability to make these comparisons might be affected because between-channel differences would be somewhat blurred by noise. Moreover, the picture is complicated by the fact that many spectral cues to consonant recognition are dynamic, that is they represent spectral changes over time. Dorman et al. (1998b) found that a larger number of channels (12 channels) were required to obtain maximum performance on sentence recognition in noise than was required for the equivalent task in quiet (5 channels), although it should be noted that these data were obtained from NH subjects listening to an AM. Fu et al. (1998) and Fu and Nogaki (2005) suggested that noise interfered with spectral processing in CI users. However, it is also possible that within-channel temporal processing is also implicated and that the reduced temporal cues mean that reliance on spectral resolution is increased. Analysis of consonant feature transmission provides a method for determining the relative importance of temporal and spectral resolution in limiting CI users' speech perception in noise. There is almost no evidence on CI users' consonant feature recognition in background noise, despite the fact that this type of evidence could be helpful in understanding the mechanism of noise interference in CI users. In NH listeners, consonant recognition in noise is robust down to quite negative SNRs. Moreover, place of articulation perception is more easily impaired by noise interference than voicing or manner perception: For example, Parikh and Loizou (2005) found few voicing errors at –5 dB SNR with either speech-shaped or babble noise in NH listeners, although there were a number of place errors. They showed that the place errors were due largely to a perceived shift in the burst frequency of stop consonants with the addition of noise, which had the effect of masking the location of the burst. By contrast, Friesen et al. (2001) showed a similar effect for noise interference with voicing transmission compared to place or manner in a group of CI users, although this varied with channel number: at lower channel numbers place was more susceptible to noise interference while with a larger number of channels voicing appeared to be more susceptible to noise interference. This was the only study identified to look at noise interference for different consonant features in CI users, but data were not included in figure 2.3 because the authors only reported % correct rather than information transmission values (it is also worth noting that the main focus of the study was channel number). This suggests a different pattern of noise effects, and therefore a difference in the mechanism of noise interference, between NH listeners and CI users.

2.1.4 Types of information loss and consonant recognition abilities in CI users

The important question for the present study is how other authors have attempted to explain deficits to consonant recognition. Some authors have suggested that deficits in consonant recognition can be explained by electrical/neural interface information loss. Valimaa et al. (2002a and 2002b) analysed patterns of phoneme errors for vowel and consonants taken from an inventory of the Finnish language (because of language differences this study is not included in figure 2.3 and table 2.2; in any case the authors did not analyse data by information transmission because they used an open set task). They found that Finnish CI users found manner of articulation easier to perceive than place. They also found that alveolar and velar consonants were identified more accurately than bilabial consonants, and noted a tendency to confuse consonants with the closest consonant with a higher F2 transition onset frequency. A potential explanation of this might be the upward shift in perceived frequency as a result of the relatively shallow insertion depth of the electrode array (Ketten et al., 1998), although another explanation might be that electrical channel interaction shows a characteristic of creating greater unwanted spread of excitation in the basal direction (see section 2.4).

Some researchers have explicitly supported the idea that consonant recognition by better CI users can be explained by CI information loss. Summerfield et al. (2002) suggested that impairments to fricative place of articulation identification in children using the Nucleus 22 device could be explained by the reduction of formant transition information consequent to CI processing. Importantly, the authors supported this hypothesis by showing that performance (in a phoneme recognition task- discrimination of /s/ vs. /ʃ /) for the best CI users equated to the level of performance obtained with an AM. Put in the language of the conceptual model in chapter 1, the authors suggested that deficits of place of articulation perception , for fricatives at least, could be explained by CI processing rather than electrical/neural interface factors but that the latter factor (along with possibly central factors) played a role for worse-performing CI users.

Teoh et al. (2003) attempted to link performance with acoustic phonetic analysis of CI output, using the SCI-LAB programme (Lai et al., 2003). The most notable finding from this study was that CI users could not make use of formant transition information and the authors hypothesised that this was due to loss of information introduced by SPEAK speech processing. Munson et al. (2003) investigated the relationship between overall performance in a group of 30 CI users and consonant feature transmission. Of the 30 CI users, 12 were users of the Nucleus 22 device implementing the SPEAK processing strategy and 18 were users of the Clarion device implementing a range of strategies (13 used CIS, four used PPS and one used SAS- see 2.3 for a description of speech processing strategies). The authors suggested that the relative performance for different consonant features did not differ between better and worse performers (overall performance being defined by total percent correct score on the consonant recognition task), i.e. the same pattern applied to both better and worse performers with percentage information transmitted being better for voicing and manner than place. The authors suggested, on the basis of this, that it is more likely that CI processing information loss may explain the relative transmission of features, while individual differences related to absolute performance levels. However, the authors' findings do not exclude the possibility that better performers' perceptual limitations were due to both processing loss and electrical/neural interface information loss.

In order to be able to differentiate the effects of processing and the electrical/neural interface, it is useful to distinguish between the performance of "better" CI users and, second, variations in CI user performance. If a group of CI users all use the same CI signal processing but there are variations in performance, it follows that these variations must be accounted for by variations in the later stages of information processing in figure 1.1 and not in the processing itself. There is a modest amount of evidence that variations at the electrical/neural interface could explain differences in performance between individuals (these factors are outlined in section 2.5). However, for the *best* users (i.e. those obtaining the highest level of auditory-only speech perception skills) the question arises as to whether performance limitations are due entirely, or only in part, to CI signal processing, as opposed to later stages of the chain. Given that different consonant features rely on different underlying perceptual processes, it may be that the relative importance of CI processing and electrical/neural interface information loss will not be the same for each consonant feature.

2.1.5. Overview of state of knowledge and knowledge gaps

The general state of evidence about consonant feature transmission in CI users can be summarised as follows:

- CI users show worse place transmission than voicing or manner transmission in quiet when tested using the /aCa/ vowel environment. This finding is robust across a number of studies that have looked at different CI devices and test paradigms.
- However, voicing and manner transmission in quiet by CI users are still not at levels achieved by NH listeners (e.g. approaching 100%), at least as shown in the large majority of studies.
- Poorer place performance is thought to be because spectral resolution is impaired relative to temporal/envelope resolution, at least with respect to the psychoacoustic processing needed for accurate consonant recognition in quiet by CI users. However, the relative contribution of information loss from CI processing vs. electrical/neural interface is unclear.
- Studies looking at consonant feature recognition in CI users have tended to conflate users of different devices, making it difficult to derive conclusions which are specific to a particular set of processing characteristics.
- There is very little evidence as to the pattern of consonant feature transmission for features other than voicing, place and manner.
- There is very little evidence as to the pattern of noise effect across features.
- There are marked variations in CI user performance but the reasons for this are not fully understood. According to the one study evaluating variations in consonant feature perception across users, variation between users is the same for the categories voicing, place and manner in quiet (e.g. worse users are equally worse than better users across these different features). The corollary of this is that no *specific* mechanism, e.g. spectral or temporal, can be identified to explain between-user variation.

It is clear from this overview of the available evidence on relative consonant feature transmission in CI users that there is a need for further detailed evidence of transmission of consonant features, particularly in background noise also, and probably using alternative vowel environments to /aCa/ to ensure that findings are not limited to a specific vowel context. Additionally, there is the need for further research to clearly identify the extent to which CI processing, as opposed to the electrical/neural interface, can explain performance limitations, and to do in such a way that is specific for each consonant feature (given that different features can be related to somewhat different underlying perceptual processes-see 2.6). The remaining question "what factors affect consonant recognition in CI users?" is answered in sections 2.2 to 2.4.

2.2. CI signal processing

This section describes the broad principles of CI signal processing in current CI devices. Because the original experimental work in this thesis relates exclusively to a single CI device, the Nucleus 24, the description of CI function focuses primarily on the details of this device. However, where appropriate, a discussion of alternatives provided by other devices is given. Details of processing in the Nucleus 24 device are obtained from Cochlear (2004). This section is necessary as background to the subsequent section on empirical evidence about effects of signal processing on perception and to the experimental work reported in subsequent chapters.

All CIs comprise a standard set of hardware components (the description here is relevant to all CIs manufactured since 1996). The first component is the microphone, which may or may not be coupled to a speech processor. The speech processor converts microphone output to an electrical signal which can be processed in electrode array by analysing incoming signal into frequency domain and extracting the envelope of each frequency component (of which more details below). The signal from the speech processor is transmitted by a transmitting coil which converts signal into a radio frequency signal for transcutaneous transmission. The receiver-stimulator converts the incoming RF to an electrical signal for the electrode array, which in the case of the Nucleus 24 devices, comprises 22 intracochlear electrodes. Additionally, current CI devices have one or two extracochlear electrodes which act as reference electrodes. In the case of the Nucleus 24, one extracochlear is part of the receiver-stimulator while the other one provides an alternative path and is lodged in the mastoid bone. The majority of CI users have devices from one of three manufacturers: Cochlear (who manufacture the Nucleus 22 and Nucleus 24 devices), MED-EL (who manufacture the COMBI-40+ and PULSAR devices) and Advanced Bionics (who manufacture the Clarion device).

Figure 2.4 shows the signal processing stages in a multi-channel device such as the Nucleus 24. Once the signal is picked up by the microphone, the first processing stage is the input stage, or front end. The measured acoustic signal is converted to a digital signal. High-frequency emphasis may be added before or after analogue-to-digital conversion (ADC).

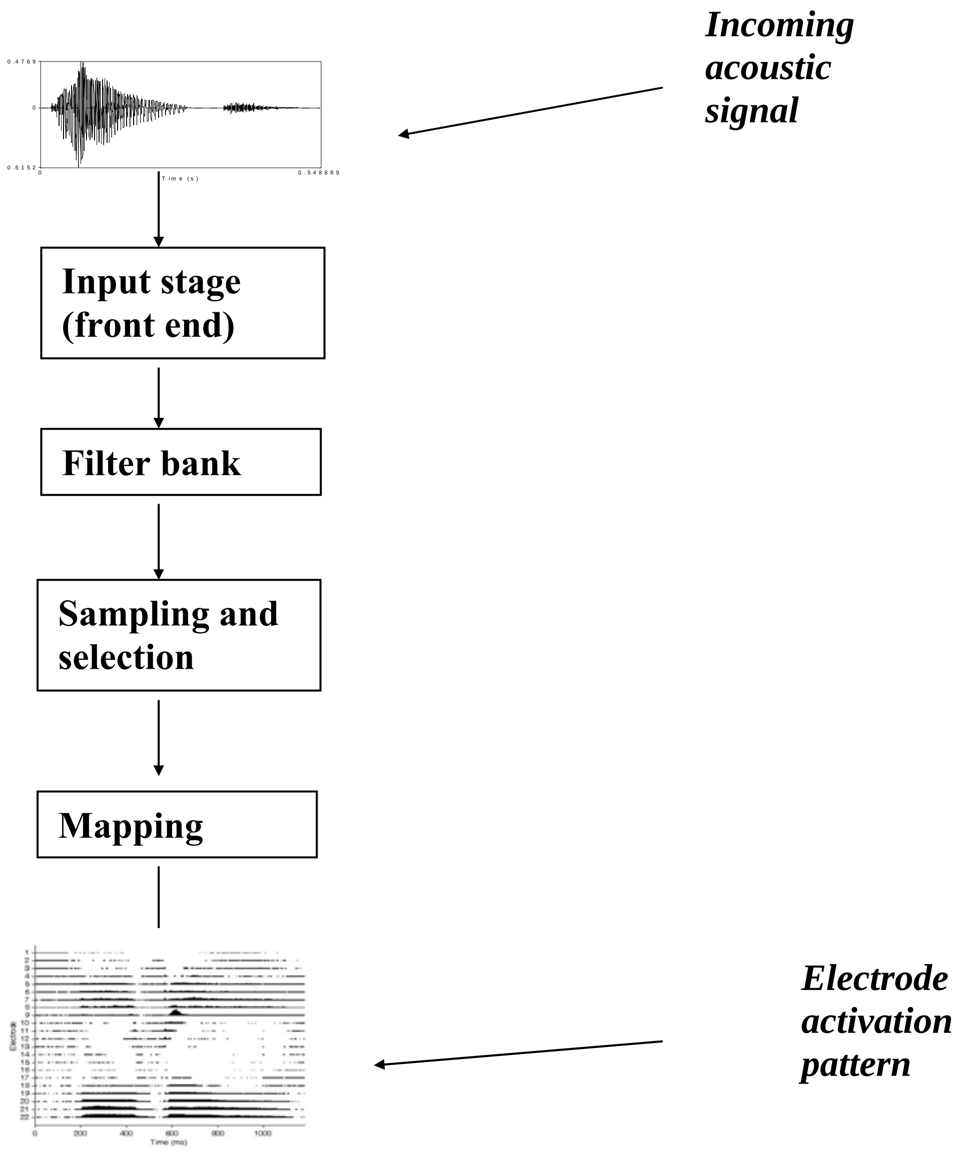

Figure 2.4 Main stages of CI processing

In the Nucleus 24 device, the incoming signal is picked up by a microphone which has a characteristic pre-emphasis frequency response (Appendix A gives a description of how the characteristics of the Nucleus 24 Sprint microphone were measured and gives the frequency response in figure A1). An anti-aliasing filter is set to 8,000 Hz and the signal is digitized at a sampling rate of 16,000 Hz with a resolution of 8 bits. The main consequences of ADC are therefore the loss of most information above 8,000 Hz and the noise introduced by quantization. The maximum possible quantization error can be calculated as 0.29* number of steps (256 for 8 bits), e.g. 1/900 the size of the signal and therefore can be ignored as it is so small.

The subsequent stage of CI processing is frequency analysis and envelope extraction. This can be done via a bank of band-pass filters (Lawson et al., 1993), in which case envelope information is extracted by rectification and smoothing of the filter bank output. Envelope information can also be derived by implementing the Hilbert transform (Anderson et al., 2002). However, the Nucleus 24M uses a Fast Fourier Transform (FFT) approach whereby envelopes are derived by recombining and weighting FFT bins in accordance with the desired number of channel outputs. The Nucleus filter bank employs a fixed 128-point FFT. This yields bin centre frequencies that are linearly spaced at multiples of 125 Hz. A Hann window is applied and gives each bin a 6 dB bandwidth of 250 Hz. Because the resulting number of bins (64) exceeds the desired number of channels (up to a maximum of 22), the bins are combined by summing powers to provide a set of frequency bands (maximum 22) as per figure 2.5. The envelope of each filter is calculated as the weighted sum of the corresponding FFT bin powers where the weights determined the frequency boundaries of the bands. Each bin appears in only one band, and the number of bins combined to form each band is determined by the total number of channels. The resulting filter bank is shown in figure 2.5. Bands are spaced linearly for low frequency/apical channels and logarithmically for high frequency/basal channels.

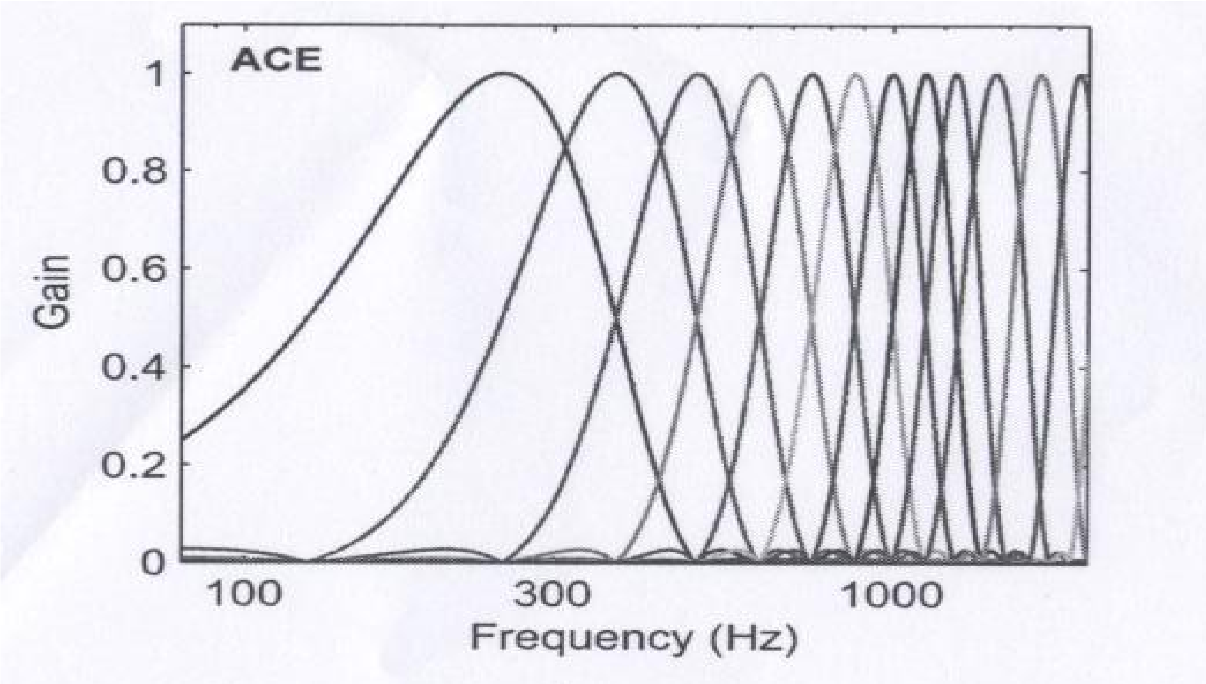

Figure 2.5. FFT filterbank used for Nucleus 24 processing with ACE or CIS speech processing strategy. Reproduced with perimission from Laneau et al. (2006)

The two main approaches to filtering, i.e. FFT vs. time-domain band-pass filter banks, differ in some respects, although both approaches yield similar outputs, e.g. a relatively small number of channels as compared with the normal hearing mechanism, in which envelope information is coded but temporal fine structure is discarded. However, the two approaches may differ in terms of their implications for temporal coding of the signal. With the IIR filterbank approach, envelopes are derived by rectifying and smoothing the outputs of each filter. This means that the temporal information coded from the incoming signal is effectively limited by the temporal response of the smoothing filter. The nominal low-pass cut-off of the smoothing filter is referred to as the “envelope cut-off frequency”. With the FFT approach, as implemented in the Nucleus 24 system, increases in stimulation rate yield increasing overlap between FFT analyses. A consequence of both filter bank approaches is that the information provided within each channel is envelope information only. Variations in level are coded via a series of pulses which are fixed in presentation rate within each channel. Consequently, different frequency components that fall within one channel cannot be accurately resolved.

Each processing channel output is coded to a corresponding electrode channel at the channel mapping stage. However, a further consideration is the approach to sampling and selection of these outputs. The general approach can be distinguished by different "speech processing strategies": with *fixed channel* strategies, all filter output samples are selected, and corresponding channels stimulated in an interleaved fashion (the interleaving is used to minimise proximity between stimulated electrodes and therefore channel interaction (*see* Boex et al. (1996)). With *peak-picking* strategies, only a subset of channels with the greatest amplitudes are selected; this means that a different subset set of channels may be stimulated with each run (Dorman et al., 2002). Each envelope is then coded to the corresponding electrode channel. The Nucleus 24 currently implements two peak-picking strategies, the Advanced Combination Encoder (ACE) and Spectral Peak (SPEAK) strategies (Skinner et al., 2002; Dillier et al.,1995). Additionally, the Nucleus 24 implements the fixed-channel Continuous Interleaved Sampling (CIS) strategy, although the specific implementation is somewhat different from equivalent implementation in the MED-EL device or as originally envisaged by Wilson et al. (1991) given the use of a different filterbank approach.

Once sampling and selection of filter outputs (envelopes) has taken place, envelope fluctuations are coded as variations in stimulus level (current * duration). Minimum and maximum permissible electrical stimulation levels are pre-determined by psychophysical measurements, in order to determine the lowest audible current level and the highest comfortable current level, for each channel. A "channel" means a particular current path, from one of the intracochlear electrodes to a reference electrode. The current path may be from the active electrode to an extracochlear electrode (monopolar), to another intracochlear electrode (bipolar) or to all other intracochlear electrodes (common ground). The dynamic range of envelope signal is compressed in order to map into the available electrical dynamic range. It should be noted that the term "MAP", is used to describe a unique set of processing parameter values used by an individual CI user, including values of minimum permissible current levels for each electrode (known as "T-levels", or electrical threshold levels) and maximum permissible current levels (known as "C-levels", or electrical maximum comfort levels. The terminology of MAPs, T-levels and C-levels adopted has tended to be used specifically by Cochlear Corporation, who manufacture the Nucleus 24 device, and is adopted here for convenience.

2.3. Effect of CI signal processing characteristics on speech perception

This sections details the likely sources of information loss associated with different aspects of CI processing and outlines the main research evidence in connection with these different areas. Again, the focus is on consonant feature recognition and the Nucleus 24 device, where possible.

2.3.1 Input stage characteristics

As noted, the main transformations that occur at the input stage of processing are the removal of higher frequencies due to the anti-aliasing filter necessary before ADC, emphasis to higher frequency components (and relative reduction in low-frequency components) due to pre-emphasis, and a reduction in amplitude information due to signal compression and limited dynamic range. The anti-aliasing filter used prior to ADC determines the absolute frequency range provided by the implant, which is limited to half the sampling frequency. There is relatively little evidence to determine whether total frequency range has a bearing on performance, although Loizou et al. (2000b) found that changes in upper frequency range from 6700 to 9900 Hz had no effect on phoneme recognition. From the point of view of the present work the important point to note is that devices vary in terms of total bandwidth provided and that AM studies (discussed in 2.5) vary widely in terms of the frequency range of the signal. It is therefore important to consider variations in other parameters in the context of a particular frequency range, although no further consideration is given to whether overall range is an important factor in itself.

Amplitude resolution and dynamic range are related factors that are related to input stage processing and could impact on performance. Some form of signal compression is needed to map the input acoustic dynamic range onto the available electrical dynamic range, which is in the order of 10-15 dB. However, this parameter is limited in part by the listener's available dynamic range (an aspect of the electrical/neural interface rather than CI processing); the larger the individual's dynamic range, the less compression that will be required. However, Nelson et al. (1995) have suggested that the main limitation to amplitude coding is to do with the number of discriminable amplitude steps rather than the absolute range. The best CI users can discriminate 40 to 50 amplitude steps (Nelson et al., 1995) whereas the number of steps defined by 8-bit quantization, as implemented in Nucleus 24 processing, is 256. This suggests that bit rate and concomitant quantization is unlikely to be a significant limiting factor in determining CI users' performance. A more important variable is likely to be the amplitude range coded by the CI and the consequent degree of audibility for quiet components in speech. The Nucleus device implements a fixed 30dB input dynamic range although this is modified by a number of more complex approaches to AGC, designed to optimize dynamic range across frequencies such as Adaptive Dynamic Range Optimization (ADRO)(Blamey, 2005). In the present study the standard fixed input dynamic range was used.

A further important aspect of input stage processing is pre-emphasis, e.g. the relative amplification of higher frequencies in the input signal. Pre-emphasis is likely to have a bearing on information transmission because of the increase in relative audibility of higher frequency spectral components. An unpublished MSc project supervised by the author of the present study did show that the addition of pre-emphasis with 6 dB per octave roll-off characteristics improved NH listeners' VCV performance using an AM. The researcher found a small but significant improvement in both place and manner transmission in the /aCa/ vowel environment with the addition of pre-emphasis. It should be noted that the study used the same 8-channel CIS AM as was used in experiments 1 and 2, reported in chapter 4.

2.3.2 Filter bank spectral characteristics

A number of variables in filter bank design and implementation have been evaluated in CI users. Total spectral bandwidth, which could be considered an aspect of filterbank as well as input stage design, has been considered in the previous section. In the same study (Loizou et al., 2000b) no effect was found no effect on consonant recognition with variations in the order of the Butterworth filters. They used a 4^{th}, 8^{th} and 10^{th} order filter (with corresponding overlaps of between –20 dB, -45 and –60 dB) in users of the Med-EL CIS strategy. No differences in word or consonant recognition were found with the different filter slopes/orders.

A related consideration is the relative allocation of different frequencies to different electrode channels. In practice, the majority of implant devices are based, loosely, on what is known about critical bands in NH listeners and therefore tend to map narrow frequency ranges to apical (low-frequency) electrodes and wider frequencies, often with logarithmic increase, to basal (high-frequency) electrodes. It is important to distinguish studies in which total spectral bandwidth is altered from those in which the relative allocation of different frequency bands is altered (within a fixed total bandwidth). An example of the latter study is Friesen et al. (1999), who found that a range of frequency allocations led to similar consonant recognition patterns in a group of Nucleus 22 users. However, Fu and Shannon (2002) altered MAPs for three Nucleus 22 users by shifting frequency allocations and found significantly reduced performance on a number of speech recognition measures, including place transmission and vowel recognition. However, it should be noted that this study altered total signal bandwidth rather than keeping this variable fixed and altering relative allocation of bands across electrodes. In more general terms, Laneau et al. (2004) suggested that current filterbank design is a limiting factor on performance and that CI user performance could be improved by alternative approaches. The authors examined the effect of filter bank design on perception of voice fundamental frequency (F0) and found that the current ACE filter bank provided very poor spectral cues to F0 discrimination but that it was possible to improve spectral representation of F0 via filterbanks with a narrower bandwidth at lower frequencies.

A critical consideration is the number of frequency channels provided by the CI processor. A number of studies into the effect of CI channel number have shown that, as channel number is increased to the maximum number available (e.g. 22 with a Nucleus 24 device).Interestingly, there is a convergence of evidence from CI user and AM studies indicating that the performance asymptote obtained with CI users, who do not generally improve on any speech perception measure beyond about 8 channels, is matched by AM studies in some cases. Evidence of the performance asymptote comes from a number of studies showing that CI users' performance does not improve beyond the level of performance obtained with between 6 and 10 active channels. (; Dorman and Loizou, 1997; Dorman and Loizou, 1998; Friesen et al., 2001; Loizou et al., 1999). For example, Friesen et al. (2001) found that, with users of the Nucleus 22 and 24 devices, with 20 and 22 active channels respectively, no significant improvements were identified beyond the range 6-10 channels. However, this effect varies with which performance measure is used. Dorman and Loizou (1997) found a higher asymptote for vowels compared to consonants, for phonemes compared to sentences and for stimuli in noise compared to in quiet. This disparity between different measures must relate to the degree of information redundancy available, e.g. with context-rich information such as sentences than any form of information reduction will have a smaller effect than on nonsense syllables and also the importance of spectral resolution. Vowel perception is reliant on spectral resolution to resolve the formant pattern that distinguishes between vowels whereas consonants rely more on temporal cues, particularly to manner distinctions. Other studies have used AM stimuli to determine whether the performance asymptote occurs dues to signal processing limitations (Dorman et al., 1997b; Dorman et al., 1998b; Faulkner et al., 2001; Dorman et al., 2000). These studies have generally found equivalence in performance between data obtained with AMs using around 6 to 8 channels and data obtained from CI users.

The results of the various studies, both with real CI users and with AMs, have been consistent across devices with rather differing characteristics. It appears that there is little benefit to increasing the number of electrode channels above about 8 for CI users. This limitation in spectral resolution achieved by CI users is thought to be due to cross-channel current spread, known as channel interaction (Throckmorton and Collins, 2002) (see 2.5.1 and 2.6.4). However, there are knowledge gaps from the literature on channel number. First, the majority of studies have used fixed-channel strategies, and there is little evidence about the performance asymptote for peak-picking strategies (a point relevant to the present study as the majority of Nucleus 24 users use peak-picking strategies). More crucially, the assumption that the performance asymptote is due to spectral channel interaction is based on a comparison between CI user performance and AM performance using varying numbers of channels. However, a larger number of channels with greater overlap between channels might not have the same perceptual consequences as a smaller number of channels without overlap. It should be possible to use an AM in which the envelope outputs are kept fixed but channel overlap is varied, to determine if this is the crucial variable determining the channel number asymptote. This issue is discussed further in the context of AMs in section 2.6.4.

2.3.3 Filter bank temporal characteristics

A number of CI processing factors come under the broad heading of “temporal”, but what they have in common is the notion of information being carried within a single channel and the associated ability of the CI to represent these changes accurately within the signals carried by individual electrodes. Until very recent innovations in CI processing, the majority of CIs have used envelope extraction. Envelope extraction strategies use a fixed rate of pulsatile stimulation in which within-channel energy changes are not coded as changes in pulse timing, but in variations in pulse level (corresponding to envelope fluctuations from the filter outputs, as described above). These strategies do not code the fine temporal structure of the band-specific signals. Recent work has attempted to utilise variations in pulse timing to represent fine temporal information (Nie at al., 2005), although one of the problems intrinsic to using variable pulse stimulation rate is the (avoidance of) simultaneous pulse presentation across channels, which is known to be associated with greater channel interaction (Boex et al., 2003). In this study only envelope extraction strategies (specifically ACE and SPEAK as implemented in the Nucleus 24) are considered.

It is important to determine whether the temporal information that is available via CI processing is adequate for speech perception and also whether temporal parameter changes, particularly stimulation rate, have an impact on speech perception in CI users. The first question is therefore, how much temporal detail is required in the signal to lead to good speech perception? Steeneken and Houtgast (1980) suggested that low modulation frequencies carry the highest information load in speech. However, Rosen (1992) argued that higher-frequency temporal information is important for various critical aspects of speech perception. According to Rosen, temporal information in speech can be divided into three separate information sources varying by modulation frequency. First, low-rate temporal information (below about 50 Hz), termed *envelope* information, conveys basic amplitude variation in speech, and is important in signalling manner of articulation, voicing, vowel identity and suprasegmental information. Second, temporal information between 50 and 500 Hz conveys *periodicity* information, e.g. information within this modulation range conveys whether the signal is aperiodic (normally unvoiced) or periodic (voiced), contributing to voicing, manner and suprasegmental information. Third, higher-frequency information (600-10,000 Hz) is termed *fine structure* by Rosen, and the main contribution to speech intelligibility is to perception of place of articulation and also vowel quality. A proviso to this account is that, in practice, NH listeners cannot code temporal information beyond about 5 kHz and therefore it is likely that information higher than this frequency must be coded as spectral rather than temporal information (e.g. must be coded via the place rather than the volley mechanism).

The question of how much temporal information CI users have access to has been addressed in some studies of temporal modulation transfer functions (TMTF) by CI users. Steeneken and Houtgast (1980) introduced the concept of the TMTF as a way of determining the temporal response of an acoustic system. The concept can be applied in both the physical and psychophysical domains. The original work by Steeneken and Houtgast (1980) defined the TMTF as a physical measure of modulation depth as a function of modulation rate, but the term is also applied to the measurement of modulation detection thresholds as a function of modulation rate as in Galvin and Fu (2005). Shannon (1992) measured TMTFs in CI users in three ways: detection of amplitude modulation, detection of low-frequency sine waves and detection of beats in two-tone complexes. For each of the three tasks the TMTF was derived. The response pattern of the TMTF was similar irrespective of which of the three tasks was used. The CI users showed TMTFs with a mean cut-off frequency of 140 Hz with a very sharper fall-off above the cut-off frequency. The TMTF varied as a function of stimulus level. With NH listeners modulation detection is independent of stimulus level across the majority of the dynamic range (Moore and Glasberg, 2001). By contrast, the subjects in Shannon's study had worse temporal modulation detection thresholds the lower the stimulus level.

The problem that should be noted in the context of the present study is that it cannot be inferred from a psychophysically measured TMTF (as with any other perceptual measure) whether restrictions in temporal information are due to CI processing information loss or electrical/neural interface information loss. The fact that there was such variability in TMTFs across CI users suggest that the electrical/neural interface may play a part in accounting for temporal information loss. A crucial question for this study is the amount of temporal information available to the CI user as a consequence of CI processing (as opposed to the subsequent information loss possibly associated with the electrical/neural interface- see 2.4.3). A particular focus of the literature has been the perceptual effect of changing stimulation rate and therefore it is important to determine extent to which temporal information changes with stimulation rate, e.g. total number of pulses provided by the CI per second. In the present study the question is addressed with specific reference to the Nucleus 24 device. Therefore, a more detailed consideration of the temporal processing of the Nucleus 24 device is needed.

Because the Nucleus 24 implements an audio sampling rate of 16 kHz and a fixed FFT length of 128 points, it undertakes 125 (=16,000/128) FFT analyses per second. The temporal response of the filter can therefore be approximated by a low-pass smoothing filter with a cut off at 125 Hz, with little information in the envelope available above this frequency (David Simpson, personal communication). However, the Nucleus 24M is able to implement channel stimulation rates ranging from 250 pulses per second per channel (pps/ch) to 1200 pps/ch (although note that the more recent device, the Nucleus Freedom, can implement channel stimulation rates up to 3,500 pps/ch). However, the extent to which increases in stimulation rate within the available range genuinely increase the temporal envelope information available is unclear, as temporal information can only be increased by increasing the degree of overlap between subsequent FFT analyses (of the same sampled signal). Stimulation rate increases are achieved by increasing the overlap between subsequent FFT analyses such that the number of (overlapping) analyses is equal to the stimulation rate (Cochlear, 2002). Let us consider the example of changing from 250 pps/ch to 500 pps/ch. For 250 pps/ch, the first stimulation frame analyses the first 128 samples, the second frame analyses points 65 to 194, and so on (e.g. there is an overlap of half the data points with each analysis). For 500 pps/ch, the second analysis uses points 33 to 160, and so on (an overlap of 3/4 the data points from each analysis). Increases in analysis rate above 125 Hz without increases in auditory sampling rate (i.e. shorter analysis windows) or a decrease in FFT length means that there is little benefit in temporal detail for the envelope. This suggests that the envelope bandwidth is effectively limited to 125 Hz, irrespective of analysis/stimulation rate, although a small amount of increased temporal information may be consequent to higher degrees of overlap between FFT analyses. In order to determine this empirically, a series of objective temporal modulation transfer functions (TMTFs) were undertaken. Sinusoidally amplitude modulated (SAM) sinusoids of 250 Hz and 2000 Hz were used as input stimuli for signal processing using the NIC-STREAM Nucleus MATLAB toolbox simulation of Nucleus 24 processing. The choice of these two frequencies was motivated by the importance of the two frequency regions for different aspects of consonant recognition. Information for voicing, nasality and fundamental frequency for higher-pitch female or children's voices occur is around 250 Hz or lower while the important second formant for most vowels occurs (and associated second formant transitions for adjacent consonants) occurs near to 2000 Hz.

The two sine waves were sinusoidally modulated at 100% modulation depth at modulation rates from 25 to 250 Hz, in 25 Hz steps. Modulation depth was measured for processed stimuli for three different stimulation rates (250 pps/ch, 900 pps/ch and 2000 pps/ch). Stimuli were processed through a single-channel CIS strategy as implemented in the Nucleus 24 CI (described in detail in 3.3.2). Figures 2.6 and 2.7 show two examples of visual representations of electrode output. The difference between the two figures is the modulation rate- in both cases, the output of a single electrode channel is given for a SAM 250Hz tone with a modulation depth of 100%. It can be clearly seen that, while for the SAM tone modulated at a rate of 25 Hz, the modulation depth approaches 100%, for the same stimulus modulated at a rate of 250 Hz, the modulation depth is markedly affected at only 9% (modulation depth for a SAM pure tone can be simply defined as the ratio of maximum to minimum signal values, expressed as a percentage).

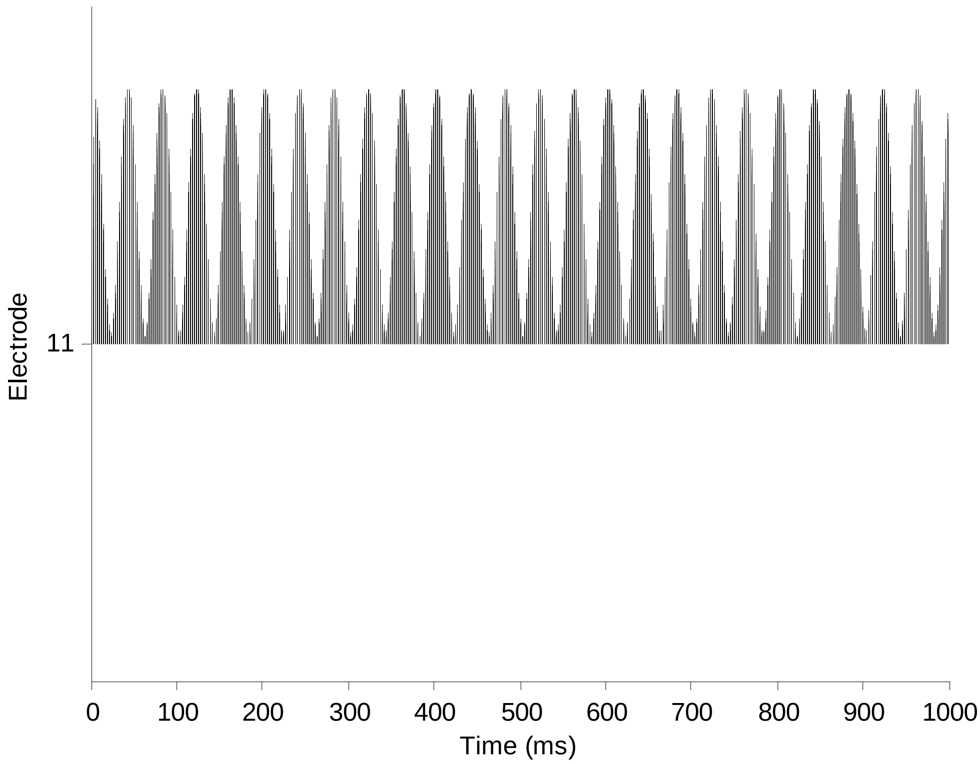

Figure 2.6. Electrode output for a pure tone modulated at 25 Hz through single channel CIS processing with a stimulation rate of 2000 pps. The input stimulus was a SAM tone with a carrier frequency of 2000 Hz, a modulation rate of 25 Hz and a modulation depth of 100%.

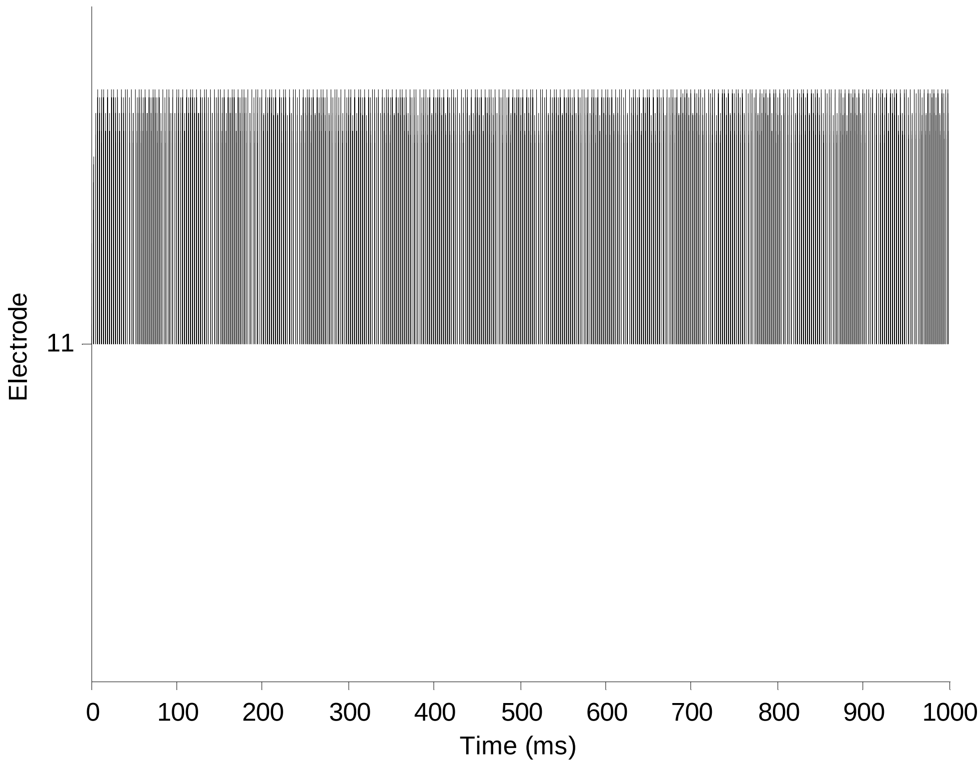

Figure 2.7. Electrode output for a pure tone modulated at 250 Hz through a single channel CIS processing with a stimulation rate of 2000 pps The input stimulus was a SAM tone with a carrier frequency of 2000 Hz, a modulation rate of 250 Hz and a modulation depth of 100%.

Figures 2.8 to 2.10 show the full range of TMTFs measured for the three stimulation rates.

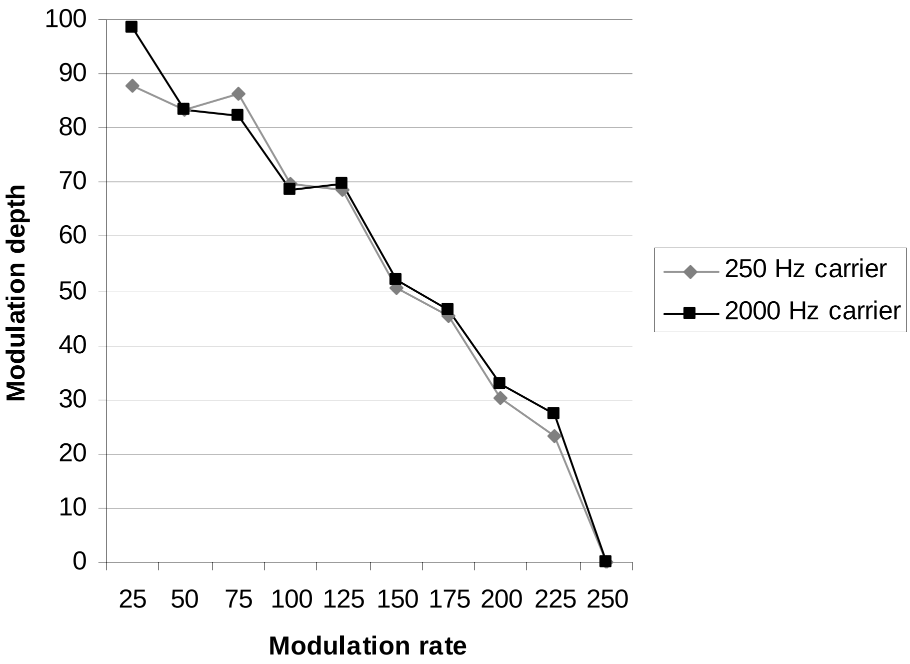

Figure 2.8. Temporal modulation transfer functions for two different carriers with single-channel CIS processing at a stimulation rate of 250 pps/ch with the Nucleus 24 processor. The original unprocessed signal was modulated at 100% modulation depth.

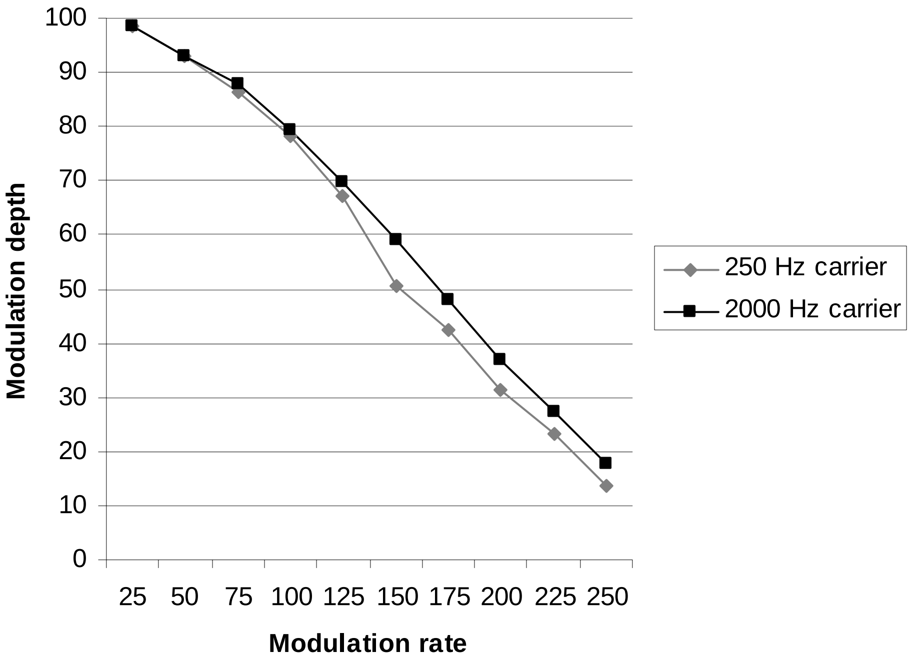

Figure 2.9. Temporal modulation transfer functions for two different carriers with single-channel CIS processing at a stimulation rate of 900 pps/ch with the Nucleus 24 processor. The original unprocessed signal was modulated at 100% modulation depth.

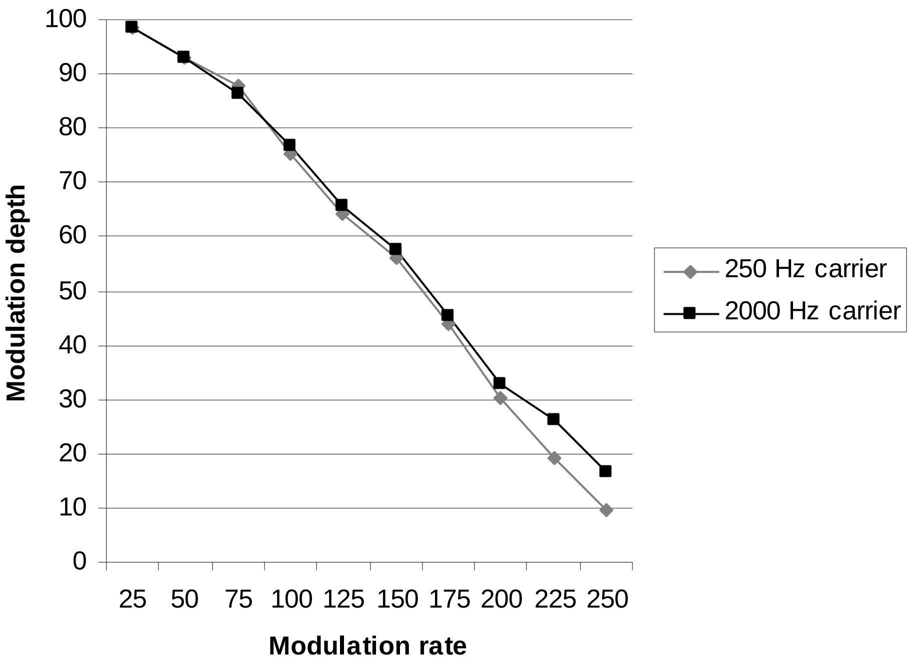

Figure 2.10. Temporal modulation transfer functions for two different carriers with single-channel CIS processing at a stimulation rate of 2000 pps/ch with the Nucleus 24 processor. The original unprocessed signal was modulated at 100% modulation depth.

It can be seen that modulation depth drops off markedly as a function of modulation rate, and that the pattern is very similar across stimulation rates and carrier frequencies. The pattern of TMTF data, showing a gradual decrease in modulation depth and a modulation depth around 70% at 125 Hz, is consistent with the hypothesis that, for a processor with a fixed FFT length and number of samples, the envelope bandwidth does not vary significantly with increased FFT overlap. For modulation rates less than 200 Hz, there appears to be a modest advantage for 900 pps/ch and 2000 pps/ch over 250 pps/ch. However, for higher modulation rates even this small advantage disappears, at least up until the modulation rate is equal to the stimulation rate as in figure 2.10.

The data, provided in figures 2.8 to 2.10 suggest that benefits to changing from lower to higher stimulation rates should be modest if present at all for the Nucleus 24 processing system. It is therefore of interest to relate this finding to empirical evidence regarding the effect of stimulation rate, particularly in users of the Nucleus 24 device. Vandali et al. (2000) evaluated sentence recognition in in a group of

Nucleus 24 CI users. In this study, six users of the Nucleus 24M CI were tested in three different stimulation rate conditions: 250, 807 and 1615 pps/ch. Users had take-home experience with the different rate conditions within a cross-over design, with order of presentation of the three rate conditions randomised across subjects. Other parameters used by the subjects were those normally used and outcome measures were tests of word and sentence recognition. The study failed to show a significant effect of stimulation rate and for some listeners even found deterioration in sentence recognition at higher rates. However, Holden et al. (2002) found that some Nucleus 24 users obtained better performance with 1800 pps/ch compared to 720 pps/ch, albeit only at 50 dB SPL but not at 60 or 70 dB SPL, and only for two of the six subjects. Interestingly, Galvin and Fu (2005) found an improvement to modulation detection at low stimulus levels when using a lower stimulation rate (250 pps/ch compared to 2000 pps/ch) in Nucleus 24 and Nucleus 22 users, although it should be noted that these differences were obtained via direct stimulation using a modulated pulse train, rather than for stimuli processed via the CI processor itself. Taken together, these findings suggest that there is very little evidence of performance benefit with higher rates in the Nucleus 22 and 24 devices and even some evidence of performance reductions. The measurements reported above suggest that the reason for this is the absence of appreciable changes to temporal envelope sampling with increases in stimulation rate in the Nucleus device, due to the inherent limitations of combining a fixed FFT length with a fixed sampling rate.

Systems other than the Nucleus CI implement IIR filterbanks followed by rectification and smoothing as with the CIS strategy in the MED-EL COMBI 40+ and CIS-PRO body-worn processor. In this case, it is possible to alter stimulation rate and envelope cut-off frequency (e.g. the low-pass cut-off of the smoothing filter) independently. It may be that the ability to increase the cut-off of the smoothing filter could lead to comparatively greater changes in temporal information transmission than is the case with devices such as the Nucleus 24 which use a fixed-size FFT approach. Recent literature suggests that both rate of pulsatile stimulation and envelope cut-off frequency may have an impact on consonant recognition, although these effects are highly variable between studies. Verschuur (2005) showed that there was little benefit to changing stimulation rate without changing envelope cut-off frequency. In that study three different stimulation rates were used (400, 800 and >1500 pps/ch) but envelope cut-off was maintained at 400 Hz. There were no differences in performance with consonant recognition measures, although there were improvements at the higher rates for sentence recognition, albeit only for 2 out of 6 subjects.

Fu and Shannon (2000) evaluated the effect of both stimulation rate and envelope cut-off frequency on consonant and vowel recognition in users of a 4-channel CIS strategy with the Nucleus 22 device. The authors used an experimental processor which implemented an IIR filterbank approach and was therefore able to separately manipulate envelope cut-off frequency and stimulation rate. The authors found improvements in performance as stimulation rate was increased from 50 to 150 pps/ch. However, they found no further significant improvement with increases in rate from 150 to 500 pps/ch, the highest rate used. They also found no improvement in consonant recognition with envelope cut-off frequencies above 20 Hz, although performance deteriorated below this frequency down to the lowest cut-off frequency used (2 Hz). This is an interesting finding, because it suggests that only very low frequency modulation rates contributed to speech perception, or at least that increasing the envelope cut-off filter above this rate did not provide more temporal information.

A final point to note is the concept of “trade-off” between stimulation rate and channel number. Brill et al. (1997) showed that different individuals performed better at higher rates and lower channel numbers while for others performance was optimal for relatively lower rates and higher channel numbers. Nie et al. (2006) found that changes in stimulation rate and channel number could be “traded off” against one another to produce similar outcomes in consonant recognition in quiet, again in a group of users of the MED-EL device. Clearly, the degree to which these two parameters can be traded off against each other must depend on the relative change in information. For the Nucleus 24 device, as indicated in 2.4.3, a doubling of stimulation rate means considerably less than doubling of temporal information. Theoretically, an increase in channel number (or number of peaks coded in a peak-picking strategy) should mean a corresponding increase in spectral detail, although this of course depends on electrical/neural interface limitations. Moreover, the trade-off would presumably be different for different consonant features, depending on the relative importance of spectral and temporal resolution for coding of the feature. The possibility of "trading off" channel number and stimulation rate was included in the design of the experimental work reported in chapter 5, although it was not anticipated that this phenomenon would be observed for users of the Nucleus 24 device given the absence of changes in temporal sampling with increased stimulation rates.

2.3.4 Sampling and selection approaches (processing strategies)

Variation in overall approach to sampling and selection is referred to as the "speech processing strategy" type. The main division in terms of CI speech processing strategies is between those strategies where information in all analysis bands is coded to the CI ("fixed-channel" strategies) and those where only certain analysis bands are coded ("peak-picking" strategies). In practice, there is little evidence to suggest that there are differences between these two classes of strategy and in any event there are a number of confounding variables affecting comparisons between strategy types (Dorman et al., 2002). They compared simulations of peak-picking and fixed-channel strategies and found no significant difference overall, and no improvement above 8 channels with a fixed-channel strategy simulation or 9-of-20 with a peak-picking strategy simulation. This suggests equivalence between the number of channels in a fixed channel strategy and the number of peaks, rather than total number of channels in a peak-picking strategy.

2.3.5 Overview of state of knowledge and knowledge gaps

- CI users achieve maximum speech perception scores with 8 to 10 channels with current approaches to processing and stimulation. Place of articulation coding requires a higher number of channels to achieve asymptote performance than manner or voicing (in quiet), presumably because of the greater reliance of place on spectral resolution. A higher asymptote is obtained with AM studies than with CI users studies, but even with AM studies using a relatively large number of channels, place transmission does not approach normal levels; taken together, these general findings suggest that both electrical/neural interface and CI processing are limiting factors on place transmission, but that place transmission is limited by CI processing even with 20 separate frequency channels.

- The performance asymptote with channel number 8 to 10 is probably associated with spectral channel interaction. However, this is inferred from equivalence between AM studies using varying numbers of channels of envelope information and not from AM studies using the actual number of channels used by CI users with varying degrees of simulated channel overlap. The issue is discussed further in 2.6.3 in the context of AM studies of channel number.
- Channel stimulation rate effects are highly variable between devices, individual users and stimuli. It is likely that variations between devices could be explained by differences in signal processing, whereas differences between individuals could be explained by differences in channel interaction (though these are speculative hypotheses). Most studies showing benefit for stimulation rates above 200 pps/ch used CI processing with IIR filterbanks rather than the approach used in the Nucleus 24. This may be due to the inability to improve temporal response by increasing overlap between fixed-length FFT analyses. This hypothesis is supported by objective TMTF data collected here and behavioural TMTFs from other authors.
- Some studies of users of the MED-EL device (which uses a linear IIR filter bank and fixed channel strategy) have suggested a possible "trade-off" between channel number and channel stimulation rate.

2.4. Electrical/neural interface factors

Although it is possible to characterise the signal produced by the CI signal processing perfectly, the same is not true of the "neural" signal which leads to the auditory percept in the CI user. This is because the way in which the signal is delivered by the electrode array to the auditory nervous system is not fully understood. The electrode array is assumed to stimulate both surviving spiral ganglion cells within the cochlea and also other peripheral elements. Researchers have identified a number of ways in which the link between the electrode array and the auditory nerve might lead to further signal distortion and therefore information loss over and above that associated with CI processing. The main areas are: interaction between electrode and neural channels; other frequency distortions, particularly the perceived upward shift in pitch experienced by CI users and the fact that only mid to high frequencies in the auditory nerve are stimulated; abnormalities in temporal coding in the auditory nerve when stimulated by a CI compared to NH.

2.4.1 Channel interaction

An important potential source of information loss associated with the electrical/neural interface is “channel interaction”. The term refers to any effect that the stimulation of one electrode channel has on the activation of a spatially separated channel (Cohen et al., 2003). An important aspect of channel interaction is that simultaneous presentation on a group of electrodes results in distorted perception because greater cross-channel electrical interaction occurs with simultaneous presentation compared to non-simultaneous presentation (Favre and Pelizzone, 1993). The majority of current CI processing strategies, including Nucleus 24 ACE or CIS, employ non-simultaneous pulse presentation to minimize channel interaction. However, it is also clear that channel interaction does occur despite the use of non-simultaneous pulse presentation as it has been measured in users of various strategies which use non-simulataneous pulse presentation.

Channel interaction has potential consequences for consonant recognition because of both spectral and temporal information. Related to this is the idea that channel interaction has a “spatial”, or spectral, aspect, in that stimulation of an individual electrode affects adjacent frequency channels and also a “temporal” aspect in that the neural response is affected for some time after stimulation (Chatterjee and Shannon, 1998; Throckmorton and Collins, 1999). The spatial aspect has been described by a space constant of exponential decay. Stimulation of different electrodes produces overlapping electrical fields and, as a consequence, the same neurons can be activated with stimulation of different electrodes. A number of studies have attempted to quantify the decay of electrical potential within the scala tympani beyond the site of the stimulation electrode as two decaying exponentials (e.g. one either side of the stimulation electrode). Wilson et al (1994) described a model of population responses of the auditory neurons by linking a description of the electrical field patterns in the cochlea with descriptions of individual neural responses derived from the large body of work on single-neurone responses to auditory stimulation. They suggested a space constant (of exponential decay of neural excitation) for monopolar stimulation of 3.6 mm. This approximate space constant was supported by a modeling study by Kral et al. (1998)

Black and Clark (1980) developed a three-dimensional discrete resistance model of the cochlea which indicated that current spread from monopolar stimulation was 1dB/mm as measured in the scala tympani. The length constant λ was defined as the inverse of the natural logarithm of the voltage 1mm from the recording site, divided by the voltage at the site.

$$\lambda = \left(\ln \frac{V_1}{V_0} \right)^{-1}$$

Equation 2.1. Current decay in the scala tympani, according to Black and Clark, 1980.

This space constant of exponential decay was used in the AM study by Laneau et al. (2006), which is discussed further in 2.6.4. One of the aims of the present study was to determine if this model could be used to explain some of the variance in consonant recognition in CI users.

Channel interaction can also be described in its temporal characteristics, which have both a “physical” and “physiological” aspect. The “physical” aspect refers to the residual charge stored in neural tissue and membrane capacitances after pulse presentation. This aspect of channel interaction is thought to be largely dealt with by use of biphasic pulses as the second phase of a stimulation pulse should remove most of the charge delivered in the first phase. However, some residual charge could still be present and therefore one recent line of work has evaluated the use of triphasic pulses (with zero net charge) to further reduce the possibility of residual charge (Bonnet et al., 2004). However, temporal channel interaction also has a more “physiological” aspect because of the refractory property of auditory neurons. Recent work has shown that much of the channel interaction, particularly the temporal aspect occurs at the neural level e.g. stimulation of one electrode does not produce as focused a neural response as might be expected given equivalent processing in the healthy cochlea (Boex et al., 2003a; Boex et al., 2003b; de Balthasar et al., 2003). However, the distribution of current with a specific electrode will depend on a number of factors, many of which are highly variable between individuals.

A number of possible methods are available to measure channel interaction. Pitch ranking, pitch scaling and electrode discrimination all provide indirect psychophysical measures of spatial channel interaction (Busby et al., 1994; Busby and Clark, 1997; Zwolan et al., 1997). Gap detection and forward masking have been used as psychophysical estimates of temporal channel interaction (Chatterjee and Shannon, 1998; Blamey and Dooley, 1993), although Throckmorton and Collins (1999) argued that these "temporal" measures also reflect spectral aspects of channel interaction as they are also affected by degree of neural population overlap. The most common method of measuring channel interaction is to measure masked thresholds in which the masker and probe electrodes vary in distance. A masking function obtained in this way will show the greatest masking effect when masker and probe coincide, but by increasing the distance between masker and probe electrodes, it is possible to determine the spread of excitation. Lim (1989) found that the spread of excitation decayed more gradually in the basal direction than the apical direction, and this finding has been supported in other studies, including Cohen et al. (2003), although the pattern, along with degree, vary quite markedly between individual CI users. The same approach to separating masker and probe electrodes has been used with the electrically evoked compound action potential; this can be measured in the Nucleus 24 system by using intracochlear electrodes as recording electrodes (Cohen et al., 2003; Cohen et al., 2004; Cohen et al. 2005). Interestingly, Cohen et al. (2004) found a good correlation between psychophysical measurements of forward masking and spatial spread of excitation as estimated using the electrically evoked compound action potential measurements. The convergence of these different types of measure suggests that the measurements of channel interaction are valid. An additional finding common to both psychophysical and electrical approaches to the masking paradigm is that channel interaction increases with current level (Abbas et al., 2004).

An important implication of recent research into channel interaction (Boex et al., 2003a; Boex et al., 2003b) is that the degree, direction, time course and spread of neural excitation may be a critical factor in explaining individual differences in CI user ability, although the evidence base for this idea is not especially strong. Zwolan et al. (1997) evaluated speech recognition for Nucleus 22 users using two different electrode configurations. In one condition, the subjects used MAPs in which only discriminable electrodes were included; in the second condition, the same users used MAPs that included all possible active electrodes. They found an overall improvement in speech perception with the first condition. Moreover, there were marked differences in electrode discriminability (presumably an indirect measure of channel interaction) across the CI users. This was given as indirect evidence that performance improves as channel interaction is reduced, although it is not in itself a direct measure of the correlation between channel interaction and speech perception. Loizou et al. (2003) found better recognition of consonants, in particular place and voicing transmission, in users of the Clarion device with users of pulsatile non-analogue strategies which were thought to produce less channel interaction, as compared with users of an analogue strategy which was thought to produce greater channel interaction. Stickney et al. (2006) measured channel interaction by measuring masked thresholds with varying probe to masker separations and then also measured vowel consonant and sentence recognition. The authors found a high degree of correlation between speech recognition and channel interaction when a simulataneous pulse presentation strategy was used, but there was no correlation between speech perception and channel interaction for users of an interleaved pulsatile strategy.

It is not wholly clear from the literature to what extent individual differences in performance are related to channel interaction and, more specifically, how consonant recognition in quiet and noise relates to channel interaction. It has been hypothesized in a number of studies looking at channel number that the reason that CI user performance does not increase beyond levels achieved with around 6-10 channels is due to spectral channel interaction (see 2.4.2). A related hypothesis is that performance in “worse” CI users can be effectively modeled by AMs with smaller numbers of channels. That is, it is hypothesized that individual variations in channel interaction place an upper limit on the number of perceptually distinct channels available to that individual CI user and that, moreover, this is an important limiting factor determining speech perception abilities. This could be tested by comparing the channel number corresponding to performance asymptote with the degree of channel interaction. Another way to approach this question, along with the more general hypothesis that variations in channel interaction determine variations in overall speech perception ability, would be to compare CI user performance with AMs that vary in terms of channel interaction characteristics. To date, no study of consonant recognition has used this approach although Laneau et al. (2006) applied the principle to measures of F0 perception and Fu and Nogaki (2005) applied this approach to measures of sentence recognition. These and other AM studies relating to channel interaction are discussion in 2.5.

2.4.2 Pitch mismatch and insertion depth

Another aspect of the electrical/neural interface is the "pitch mismatch" associated with CI use, whereby the subjective pitch sensation produced by the CI is higher than that generated by the normal auditory system. This is because existing cochlear implant systems are not inserted fully into the cochlea. Electrode arrays would be typically inserted through the round window into the scala tympani to a length no greater than 25 mm. Therefore, as the electrode array conveys a range of stimulus frequencies from the environment with a typical band pass characteristic of about 150-8000 Hz, these input frequencies are mapped onto neural elements within the vestibulocochlear nerve that, in NH listeners, would code relatively higher frequencies, e.g. above about 1000 Hz. Shannon et al. (1998) suggested that this equates to a basal basilar membrane shift of approximately 3 mm. Ketten et al. (1998) showed that variation between individuals electrode array insertions was substantial and suggested this could be measured using *in vivo* measurement methods. The majority of studies evaluating the effect of insertion depth on performance have made use of AMs and are described in 2.5.

2.4.3 Temporal coding

It is possible that the electrical/neural interface may introduce loss of temporal information as well as loss of spectral information. Section 2.4.3 implied that TMTFs were uniform across CI users; however, a number of earlier studies showed that there was considerable variation in TMTFs between individual CI users, and in one study, this was shown to be correlated with consonant recognition, suggesting that temporal aspects of electrical/neural interface information loss may be as or more important than spectral aspects in determining individual variations in consonant recognition.

Busby et al. (1993) measured perception of temporal modulations in a group of adult CI users. They found that the shape of the TMTF also approximated a low pass filter with a cut-off frequency between 50 and 100 Hz, slightly lower than was the case for the Shannon (1992) study. What is interesting in the context of a discussion of the electrical/neural interface is that Busby et al. (1993) attempted to match temporal processing characteristics with patient characteristics, in particular duration of deafness. They found that four postlingually deafened subjects were better able to perceive temporal information than three prelingually deafened subjects. It is not possible to determine whether these variations were to do with neural or central function, but they do suggest that temporal processing varies across CI users.

Other studies have also looked at the relationship between CI users' basic temporal psychophysical abilities and the level of speech perception they obtain. Cazals et al. (1991) measured perception of a silent gap in noise and interval between two clicks in five users of the Ineraid CI. They found that there was a relationship between perception of click interval at the most basal CI used and perception of consonant place of articulation. The most striking evidence of such a relationship is given by Fu (2002), who found a strong correlation between consonant recognition scores and mean modulation detection thresholds across users' electrical dynamic range. Subjects were nine users of the Nucleus 22 CI system using the SPEAK speech processing strategy. Whereas previous studies had linked speech perception abilities to TMTF performance at high input levels, Fu (2002) measured the TMTF across a range of stimulus levels and found that the mean score averaged across input levels was a significant predictor of both consonant and vowel intelligibility.

In order to convey temporal information to the CI user, the neural discharge pattern in response to CI stimulation must convey the temporal detail in the input signal. An important difference in temporal coding between acoustic and electrical hearing lies in the stochastic relationship between acoustic input and the response of the auditory nerve to stimulation. This enables high rates of temporal coding in the auditory system, up to around 4 kHz, because of the summation of neural responses across neural populations, rates which cannot be supported by individual neurons. Without stochastic resonance, phase-locking of individual nerve fibres would prevent coding of high-frequency temporal information, or temporal fine structure. Because the mechanism of stochastic resonance is thought to be cochlear in origin, it can be presumed that this does not occur with CI systems. Therefore, electrical hearing may be at some disadvantage with respect to coding of high frequency temporal information. A number of papers using physiological outcome measures have suggested that the use of very high rates and also the use of conditioning noise stimuli may improve the temporal representation and accuracy within the auditory nerve, e.g. Matsuoka et al. (2001).

There is also a question as to whether higher stimulation rates may lead to greater increased channel interaction. Brill et al. (1997) found individual variations in trade-off between channel number and stimulation rate in a group of users of the MED-EL device. It seems plausible that individual differences in this trade-off may be mediated by the degree and nature of channel interaction. McKay et al. (2005) found that sensitivity to spectral shape was less at higher rates, given a particular number of channels. Their explanation for this was that forward masking of one pulse over a successive pulse serves to blur between-channel amplitude differences. This may help to explain why there is so much individual difference in benefits with higher stimulation rate: it is possible that individual CI users who have greater channel interaction could experience increased forward masking at higher stimulation rates compared to those with lower channel interaction.

Despite these considerations, it is appears that the focus in the present study should be on information loss associated with CI processing rather than the electrical/neural interface. It appears from the evidence presented in 2.4.2 that Nucleus 24 processing preserves temporal modulations with decreasing accuracy as modulation rates increase. Moreover, it also appeared that differences in TMTF with stimulation rate were small. Consequently, it can be hypothesised that Nucleus 24 users have little access to mid-frequency modulation frequencies (those denoting periodicity according to Rosen (1992)), no access to higher frequency modulations and that stimulation rate should make only very small differences to consonant recognition.

2.4.4 Overview of state of knowledge and knowledge gaps

- Cross-channel spread of excitation has been measured in CI users using various techniques.
- Channel interaction has both a spectral and temporal aspect, although Chatterjee and Oba (2004) showed that spectral channel interaction has a stronger implication for speech perception outcomes in CI users.
- Variations in channel interaction could help to explain variations in CI user performance but the evidence base for this is limited.
- Electrode insertion is associated with an upward frequency transposition because of the alignment of the electrode array in relation to the remaining auditory elements.
- Although there is evidence that partial insertion limits performance, it is not thought that a normal insertion depth (e.g. more than 22 mm) is an important factor in limiting consonant recognition.
- There is evidence that CI users show abnormal temporal resolution, particularly at lower intensities, although the account in 2.4 suggests that some of this must be due to information loss from CI processing rather than the electrical/neural interface.
- There is also evidence of a link between temporal processing abilities and overall consonant recognition.

2.5. Acoustic models of CI function

2.5.1 Validity of acoustic models

A signal which has been processed using the same, or similar, signal processing techniques as are used in CI speech processors and which is used to generate an acoustic signal to elicit a response in NH listeners can be termed an “acoustic model “ (AM) of CI processing (Throckmorton and Collins, 2002). Because current CI signal processing techniques are very similar to channel vocoders, AMs are also sometimes referred to as “vocoded” signals (Faulkner et al., 2000; Loizou, 2006). The aim of developing an AM is to reproduce the information content of the implant output in an acoustic form, rather than necessarily reproducing the subjective auditory sensation experienced by the implant user, as the term “simulation” might imply. Therefore, the term “AM” is preferred here.

AMs have a number of potential benefits to research. The most important point for this thesis is they can help distinguish between the effects of CI processing *per se* and electrical/neural interface factors contributing to CI user performance. AMs also allow the researcher to develop and refine hypotheses so that the design of CI performance experiments maximise CI user time. It can also be argued that studies using AMs of CI processing are of intrinsic interest even without direct reference to CI research, as they provide evidence about normal speech perception under conditions of reduced acoustic information (Shannon et al., 1995).

A typical AM was described by Loizou et al. (2000a). First, the signal was processed through a pre-emphasis filter and then band passed into N frequency bands using sixth-order Butterworth filters. In order to create an AM, sine wave or narrow bands of noise with the centre frequencies of the corresponding electrode channels were generated with amplitudes equal to the RMS energy of the envelopes and frequencies equal to the centre frequencies of the band pass filter. The sine wave or noise bands were recombined to generate the final waveform. The RMS value was then adjusted to be equal to the original signal. The difference between generating a CI signal and an AM is the final output stage: in the first case, level variations within each channel are used to vary current level among corresponding electrode channels, while in the second case, they serve to vary amplitude among a set of carrier stimuli which are recombined to generate an acoustic waveform. It is also worth noting that this approach, similar to that of the majority of AM studies, is based on the multiple IIR filterbank rather than FFT analysis.

The validity of a CI AM, that is, its ability to predict and model CI user performance, is determined by a number of factors. A key question is the degree of similarity between the signal perceived by the CI user and the signal perceived by a NH listener with an equivalent AM. There are two aspects to this: first, whether or not identical signal processing methods have been used in AM listeners and equivalent CI subjects and, second, the degree to which processing in the normal auditory system transforms the signal. While the signal received by the CI user has been processed by the CI itself, the signal perceived by the NH listener has been processed not only via the AM itself but also via the external, middle and inner ear of the listener. The external ear can be characterised by a frequency response which includes both pinna and external ear canal components. For the purposes of the study here, an insert earphone was used to minimise the amplification characteristics of the pinna. The question of processing is dealt with in the present study by ensuring that the same signal processing techniques apply to both CI users and NH subjects listening to the AM stimuli (see 3.2).

Auditory acclimatisation is another factor that may impact on the validity of AMs. A CI user will normally have had a good deal of auditory experience with the CI signal when being tested, whereas a NH listener listening to an AM may have had only a few minutes acclimatisation. Faulkner et al. (2006) showed that considerable time was needed to acclimatise to the model. Their study used running speech with a conversational discourse tracking technique in which word rate was used. The authors found that many hours of acclimatisation was needed to optimise performance with pitch-shifted speech materials. However, Davis et al. (2003) suggested that initial acclimatisation to AM stimuli occurs within a few minutes, so long as the listener is given the original unaltered stimulus for comparison. It appears that there NH listeners are able to acclimatise relatively quickly to AM stimuli without significant spectral shifts, but that considerably longer time is needed to achieve optimal performance with pitch-shifted stimuli (see Rosen et al, 1999). In the current study, it was proposed to include a degree of pitch shift in the AMs which would reflect the degree of upward frequency transposition associated with a normal insertion of the Nucleus 24 electrode array. As noted in 3.3.2, this degree of pitch shift was somewhat less than that noted as causing significant acclimatisation problems in Rosen et al (1999) and Faulkner et al (2006). Therefore, In order to determine if rapid acclimatisation to this more modest degree of pitch shift was possible, a pilot study was undertaken to see if a minimal acclimatisation procedure could yield valid results (see 3.1.2.).

2.5.2. Methodological parameters of acoustic models

The majority of papers using CI AMs have used either noise band (Friesen et al., 2001; Shannon et al., 1995; Henry and Turner, 2003; Blamey et al., 1985; Qin and

Oxenham, 2003; Nelson et al., 2003) or sine wave carriers (Dorman et al., 1998b; Loizou and Poroy, 2001; Throckmorton and Collins, 2002; Loizou et al., 1999), although a few have used filtered harmonic complexes (Deeks and Carlyon, 2004) or pulse trains (Faulkner et al., 2000). Only a few have compared performance with different carrier stimuli (Faulkner et al., 2000;Dorman et al., 1997b). An important question is therefore: what, if any, are the differences in performance between simulations using different carrier stimuli, and why might these occur? Both published CI AM studies to have compared noise band with sine wave carriers found significant effects of carrier type on perception of speech information requiring good spectral resolution: Dorman et al. (1997b) found significantly better performance with multitalker vowel recognition with a sine wave AM but significantly better performance with place of articulation in consonants with the noise band model. Gonzalez and Oliver (2005) found significantly better speaker identification with a sine wave carrier compared to the noise band carrier. They also found that the noise band stimulation was more sensitive to number of channels than the sine wave simulation e.g. performance reached maximum levels with a higher number of channels with noise band simulation and ceiling effects were obtained with sine wave carriers. However, the envelope smoothing filter was higher for the sine wave model than the noise band model, undermining the validity of the comparison from their study.

Differences in results could be explained by the different physical consequences of modulating sine waves and noise bands. Sine wave simulation would provide better frequency resolution than noise band simulation. In particular, higher envelope modulation rates with a sine wave carrier would lead to much stronger periodicity cues than would be obtained with a noise band carrier. However, modulation of either type of stimulus produces spectral side bands whose spectral distance from the carrier is equal to the modulation rate (Kohlrausch et al., 2000). However, with noise band modulation, spectral side bands are masked by adjacent noise bands. Gonzalez and Oliver (2005) suggested that the additional information about modulation that would be provided by the side bands with sine wave carriers might be advantageous to some perceptual tasks. A relevant point is that NH listeners find it harder to detect amplitude modulation in signals with a noise carrier compared to a sine wave carrier. Viemeister (1979) found amplitude modulation detection in the region of 5 to 10%

with noise band carriers whereas Kohlrausch et al. (2000) found modulation detection as good as 1% with a sine wave carrier. This difference might be explained by the fact that noise band carriers have a randomly fluctuating envelope that distorts the modulations in the envelope of the incoming signal whereas sine waves have a fixed amplitude envelope (Gonzalez and Oliver, 2005). The expected corollary of this in terms of consonant recognition would be better perception of manner or voicing in consonants with a sine wave carrier, given that these contrasts rely primarily on temporal cues requiring accurate detection of amplitude variations in the envelope. This was not supported by Dorman et al. (1997b). However, this may have been because their study found ceiling effects for perception of these features.

It is not clear from the preceding discussion which of these two carrier stimulus types is likely to lead to a more a “valid” model of CI performance, and from this perspective each type of stimulus has potential advantages and disadvantages. An AM using a noise band carrier would seem to be more appropriate as a model of the effect of channel number, as indicated by Gonzalez and Oliver (2005), whereas the same paper suggested that sine wave carriers provide a more valid model of F0 perception because of the greater salience of harmonic cues. It would also seem reasonable to assume that noise band carriers would lead to better perception of consonant contrasts requiring perception of aperiodic/noisy speech components, e.g. identification of whether a sound is a fricative or a plosive.

A further consideration is choice of input stage characteristics. The initial stage of sound processing with current implant devices includes high frequency pre-emphasis and compression. The idea of emphasising higher frequencies has a theoretical benefit by boosting less audible high frequency speech cues such as those associated with voiceless consonants or sibilant sounds. Existing CI AMs studies vary as to the inclusion of pre-emphasis or input stage characteristics in the AM. For example, Loizou et al. (1999) applied a 3 dB per octave pre-emphasis whereas Dorman et al. (1997b) applied 6 dB per octave pre-emphasis, while many other studies fail to mention whether or not pre-emphasis is added to the signal prior to processing. There is no published evidence regarding the effect of manipulating pre-emphasis characteristics either with an AM or with CI users. As noted in 2.3.1., there is some unpublished data suggesting that the inclusion of pre-emphasis affects AM results. Given the possibility that pre-emphasis might affect performance, it seems logical to incorporate this into the AM (while at the same time minimising the normal pinna/external ear canals resonance in the NH listeners which itself amplifies mid to high frequencies).

2.5.3 Acoustic modelling of CI processing and electrical/neural interface variables

The main processing parameter which has received attention in AM studies is channel number. A number of the studies cited in 2.4.2 used AM results exclusively or compared AM results with equivalent CI user data. To date there have been no studies of changes in CI temporal processing characteristics using AMs. One of the possible problems with such studies is the spectral distortion caused by changes in modulation rate with sine wave carriers, as noted in 2.5.1. It is unclear whether or not AMs using noise band or sine wave carriers can provide appropriate models of changes occurring to temporal information with higher stimulation rates. One of the aims of the work reported in chapter 5 was to evaluate changes in stimulation rate using an AM in parallel with CI users, in order to determine whether the model provides an accurate representation of the changes in temporal information (or the absence of such changes) associated with stimulation rate. However, it should be made clear that such a model is only able to deal with one of two distinct aspects of stimulation rate. These two aspects are i) changes in neural response associated with increased number of pulses presented and ii) changes in temporal sampling associated with changes in stimulation rate. A model of the first aspect is beyond the scope of the present study, and would require a more sophisticated estimate of the physiological response and an understanding of what stimuli would be necessary to engender an analogous response in a NH listener. Rather, the purpose of using an AM in the context of stimulation rate is to replicate the changes in analysis filterbank output that occur with changes in stimulation rate. For CI systems which implement an IIR filterbank, analogous AM data have been obtained by looking at changes in envelope cut-off frequency which would be set at some value less than half the stimulation rate (see 2.3.3 for a brief discussion of some papers looking at variations in envelope cut-off frequency). However, with the Nucleus 24 system, the purpose of an AM is to represent the changes in envelope bandwidth that occur as a function of changing overlap between FFT analyses, as this is the parameter which varies in accordance with stimulation rate changes (as described in 2.3.3). The AMs used in the experimental work reported in chapters 4 and 5 used modulated carrier stimuli at whatever rate was determined by the output of the analysis filterbank; therefore, if the analysis filterbank were able to convey increases in envelope modulations with increased FFT overlap (the parameter associated with increased stimulation rate), this increase in envelope bandwidth would be represented accordingly in the carrier stimuli. In practice, as noted in 2.3.3, the effective envelope bandwidth of the fixed FFT length Nucleus 24 processor appears to be limited by the maximum non-overlapping FFT analysis rate (125 Hz), irrespective of stimulation rate, and this suggests that variations in envelope bandwidth across stimulation rates are minimal. However, the AM provides a faithful reflection of the changes in filter output as a consequence of increased stimulation rate. A caveat should therefore perhaps apply that the term "stimulation rate", as applied to the Nucleus 24 AM stimuli, really means "envelope bandwidth as a function of FFT analysis overlap concomitant with stimulation rate changes". This does not mean that the AM can be a good model of neural changes occurring as a consequence of stimulation rate which are independent of changes in temporal sampling (if these occur, this would be shown by increased performance in CI users with higher rates but not associated increase in performance with the AM).

It is worth noting some of the limitations of the evidence base from AM studies. The first relevant point is the majority of AM studies have used fixed-channel IIR filter processing (although Dorman et al. (2002) is an exception to this) and therefore cannot strictly be considered as appropriate models of signal processing using a peak-picking strategy such as ACE, or, in any case, of processors which implement a FFT filterbank. A more general limitation of AM studies to date is that those studies which have compared AM performance with CI user performance directly have used CI users with varied processing parameters and devices, making direct comparison with a specific set of processing parameters impossible. For example, Fu and Nogaki (2005) compared 10 CI subjects with 6 NH subjects listening to an AM. The CI users were a varied group: 4 were users of the Nucleus 22 device, one was a user of the MED-EL device, one a user of the Clarion 1 device, while 4 were users of the Clarion CII device. This meant that parameters such as total spectral bandwidth, strategy type, channel number and stimulation rate all varied across CI subjects. The AM used was a noise band model with a frequency range of 200 to 7000 Hz and 16 channels. The lack of close correspondence between processing parameters used in an AM and those used in a matched group of CI users means that the importance of the specific parameters is unclear.

As discussed in 2.4.2, one of the perceptual consequences of cochlear implantation is an effective upward shift in perceived frequency compared to NH. A few studies have attempted to incorporate these pitch shift characteristics of CI stimulation into an AM by using a mapping between analysis and carrier frequencies derived from Greenwood's work (Greenwood, 1990). Some studies have sought to compare AM performance with and without pitch-mismatch. Shannon et al. (1998) found a significant degradation in speech perception with simulated pitch shift in a four-channel AM. Dorman et al. (1997a) found that, with simulations equated to insertion depths of 22 or 23 mm, NH listeners showed reduced performance in vowel, consonant and sentence recognition. However, Rosen et al. (1999) found that the reduction in performance associated with the upward frequency transposition could be reduced by lengthy exposure to simulations. The authors found marked effect of pitch shift on AM performance in word and sentence recognition. However, the study used a four-channel implant which makes generalisation to higher number of channels used in the present study problematic.

Throckmorton and Collins (2002) described an AM of channel interaction and also other spectral anomalies that are associated with CI use, such as pitch reversals. The authors developed AMs of different aspects of electrical/neural interface signal distortions, including pitch reversals, indiscriminable electrodes and forward masking. They compared sentence and consonant recognition abilities between the different AMs to determine which might have the greatest impact on speech perception abilities. The authors found that models of spectral channel interaction had the greatest detrimental effect on consonant recognition.

Other authors have evaluated performance with different degrees of spectral smearing, which can be taken as a method of modelling channel interaction, at least in its spectral aspect. Shannon et al. (1998) used a simulation with overlap of filter skirts of the noise bands, thus creating an effective spectral smearing effect. They found that channel overlap made little difference to speech recognition. However, it is worth noting that the AM they used had only four spectral channels, which means that spectral information was highly limited even without overlap.

Two identified studies to date have attempted to compare different AMs against CI user performance directly. Fu and Nogaki (2005) compared number of channels with changes in spectral resolution using spectral smearing. The outcome measure used was release from masking as shown by sentence recognition in noise. The 10 CI subjects used a variety of CI devices. AMs were based on a fixed-channel strategy using IIR filterbanks (as usual in AM studies) and varied by channel number (16, 8 and 4) and spectral overlap between channels (24 dB/octave or 6 dB/octave slope).The authors found that release from masking in sentence recognition was modelled best by AMs in noise with broadly overlapping filters (6 dB/octave slope), although better CI users' performance was approximated with either an 8-channel or 16-channel AM and, worse users, by a 4-channel AM. However, it should be noted that the CI users were a heterogenous group from the point of view of CI processing used and, also, that the AMs used were not based on the specific processing details of a particular device.

Laneau et al. (2006) undertook a series of experiments in which spectral overlap between adjacent channels was systematically varied. The authors were interested in perception of fundamental frequency (F0) rather than consonant recognition, but the paper is of particular interest in its use of an AM based in detail on a specific device, the Nucleus 24, implementing a specific processing strategy, ACE, and where a comparison between AM and equivalent CI user performance was made. The authors used an AM with noise band carrier stimuli. They compared pitch discrimination abilities as a function of degree of carrier overlap varying from no overlap to overlap equivalent to 10mm spread of excitation. The precise pattern of filter overlap was based on the model of channel interaction of Black and Clark (1980) and assumed asymmetric spread of excitation as noted in 2.4.1. Two separate experiments showed a close match between Nucleus 24 users and AMs with 1mm spread of excitation. A further noteworthy characteristic of this study was that the AM used the same filterbank as was used in the group of Nucleus 24 users against which performance was compared (Laneau et al., 2004). This made the comparison between AM and CI data much more powerful than with other studies where a precise match between characteristics was not obtained, where hetereogenous groups of CI users were used, and where attempts to model electrical/neural interface factors did not have a specific physiological basis.

2.5.4. Overview of state of knowledge and knowledge gaps

- AMs have been found to be highly predictive of performance trends in channel number, although not absolute magnitude of performance levels.
- Choice of carrier stimulus probably does have some effect on AM results although it is unclear which carrier stimulus type would provide the best match/predictor of CI user performance.
- It is probable that a sine wave AM should over-predict frequency and periodicity resolution abilities in CI users, as compared to a noise band model.
- The majority of AM studies have sought to mimic general processing principles, rather than the fine details of processing in a specific device. Most studies have developed models based on fixed-channel processing with a IIR filter approach. This means that there is little data of direct relevance to users of the Nucleus 24 device given that this device uses an FFT filterbank and the majority of users access a peak-picking processing strategy.
- One study to date (Laneau et al., 2006) has attempted to mimic specific processing of a particular device AND aspects of the electrical/neural interface, although the study looked at F0 discrimination in vowels rather than consonant recognition. The authors found that CI user performance was well approximated by an AM in which channel overlap was equivalent to 1 mm spread of excitation.

2.6. Consonant feature transmission

This section outlines the hypothesised effects of factors identified in sections 2.3 to 2.6 on transmission of specific consonant features. Section 2.2 has detailed what is known about consonant recognition in CI users. However, as noted, there are many knowledge gaps from this literature. In order to capture information content at the electrical/neural interface stage, the ideal representation would be some sort of neural activation pattern map. However, the current state of knowledge of the electrical/neural interface precludes an accurate representation of this kind. In order to represent the information provided by CI processing, activation patterns across electrodes can be represented with an "electrodogram" in which electrode number is given on the y-axis (with apical electrodes, coding low-frequency information, at the bottom and the most basal electrode, coding the highest frequency information, at the top), time on the x-axis and, here, amplitude (representing current level) also indicated on the y-axis within each electrode channel. A number of authors have shown that a typical CI user, even if they are using a processor with a larger number of channels, only has access to around 8 perceptually distinct channels. Given this, the majority of electrodograms in the subsequent section use an 8-channel CIS representation (equating to the parameters used in experiments 1 and 2), although a 20-channel (12 maxima) ACE electrodogram is also included in one case. In all cases electrodograms represent output from the NIC-STREAM MATLAB platform (see 3.3.2) and therefore should represent precisely the information delivered by the processing of the Nucleus 24 device, although it should be highlighted these representations do not include further transformations in the electrical/neural interface.

2.6.1 Voicing

Available data show voicing transmission around 70% in CI users (in quiet with the /aCa/ vowel enviroment), compared to nearly 100% in NH listeners. It follows that there is information loss relevant to coding of voicing information, even in quiet, although this information loss would appear to be less than for place. To consider the reasons for information loss at the CI processing and electrical/neural interface stages, the acoustic cues to voicing must first be considered. The main acoustic cues to voicing are temporal. These are: voice onset time (Holden-Pitt et al., 1995), the relative onset of the voiced and voiceless components of the speech sound; relative amplitude of aspiration (Repp, 1979), silence duration and cutback of the first formant. However, the spectral cue of F1 onset frequency is also important, particularly in background noise (Stevens et al., 1992; Gonzalez and Oliver, 2005).

Figure 2.11 shows wide-band spectrograms of the stimuli /ibi/ and /idi/, along with corresponding frequency time matrices produced by Nucleus 24 processing. In this case the ACE strategy with 500 pps/ch "stimulation rate" (really FFT analysis rate, as noted in 2.5.2), was used. The best-preserved cues to voicing appear to be voice onset time and (related) closure/silence duration. The gap between onset of voiceless speech components (the "burst) and the onset of the low-frequency periodic voicing is referred to as "voice onset time", and is characteristically shorter for voiced consonants in consonant-vowel sequence. In /ibi/, the onset of the low-frequency voiced component occur at approximately the same time as mid to high-frequency activation, while for /ipi/ onset of the burst precedes voicing onset by around 100ms. In the corresponding electrodograms, the voice onset time can be seen as the difference in relative onset of activation of channel 7 as against 19 and 22 (the difference is 100 ms).

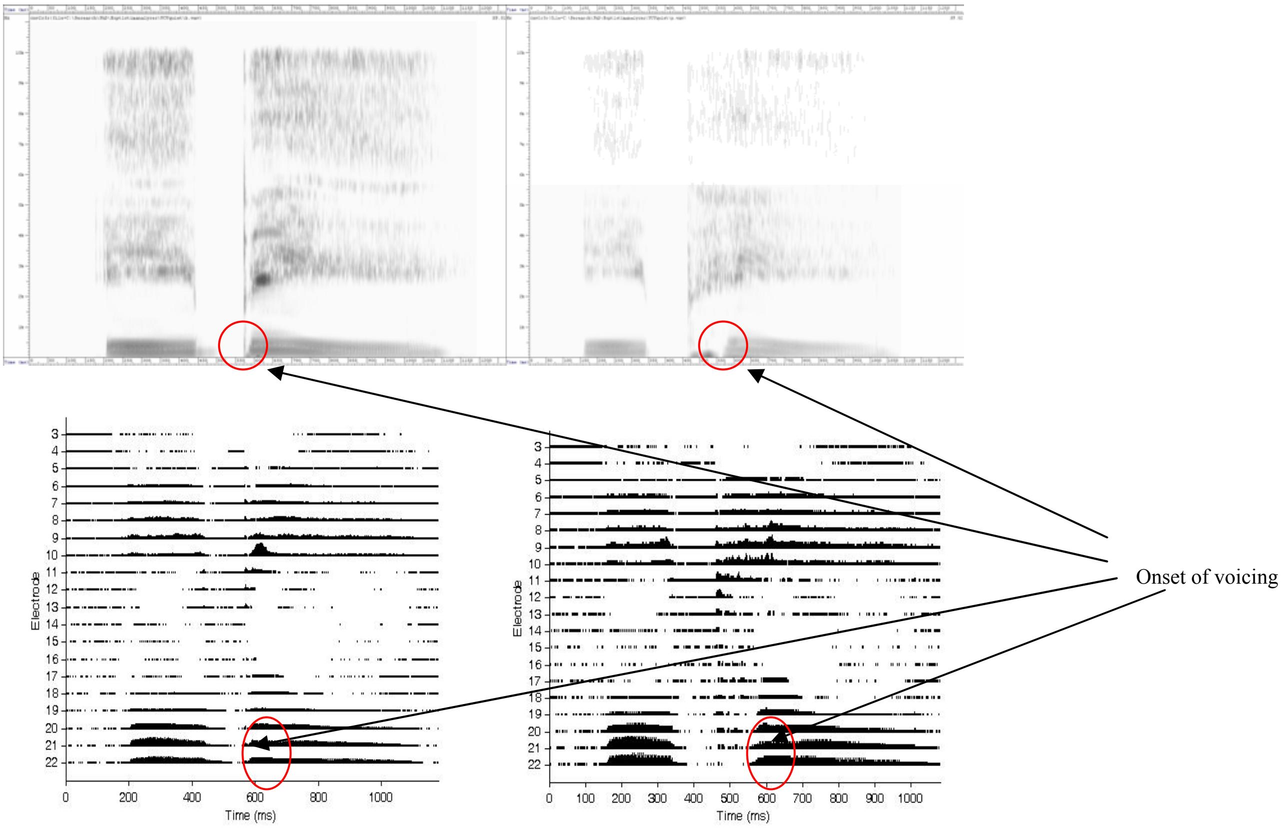

Figure 2.11./ibi/ (left) and /ipi/ (right) in quiet; unprocessed stimuli above, stimuli transformed via 12/20 ACE processing below.

However, the same two stimuli with background stationary noise added at +10 dB SNR, as in figure 2.12, show a different pattern given the introduction of noise. Here the noise interference has markedly reduced the salience of both the burst in the basal channels and the envelope fluctuations in apical channels that signal the voice bar (or its absence).

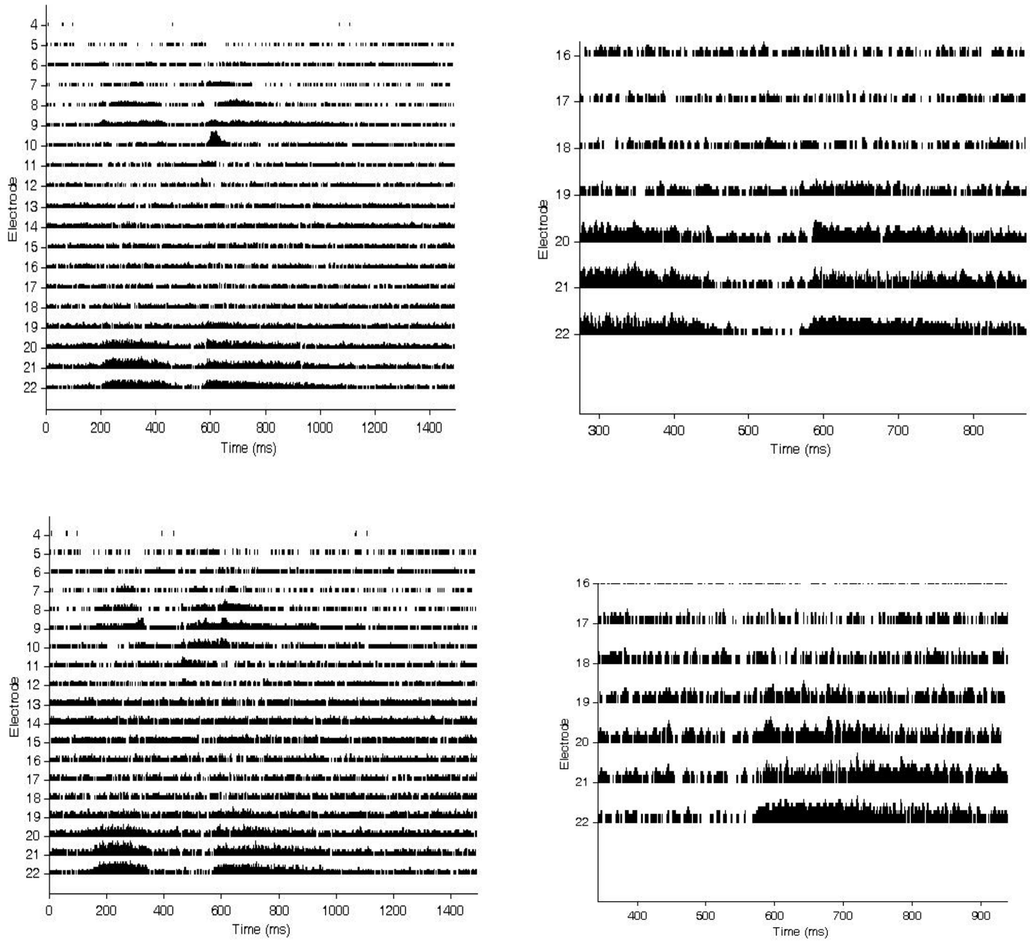

Figure 2.12. /ibi/ (above) and /ipi/ (below) in background noise at +5 dB SNR, 12/20 ACE processing; on the right are close-ups of basal channels

As noted, this representation of the electrodogram may be misleading because it overestimates the number of perceptually distinct frequency channels. Figure 2.13 shows the equivalent electrodograms, but here for an 8-channel CIS processor. Interestingly, these do not indicate a particularly different pattern of cue salience than for the 12-of-20 processor electrodograms.

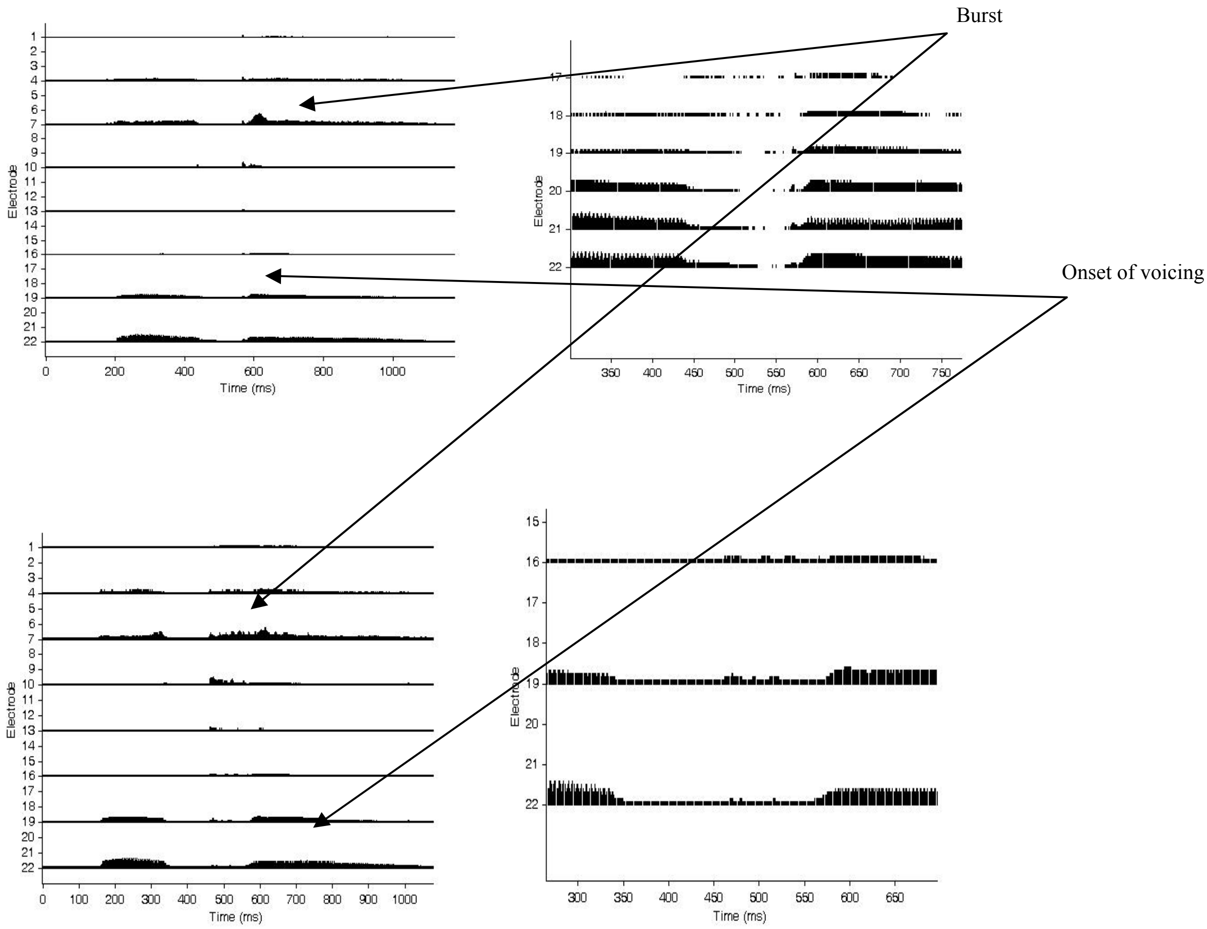

Figure 2.13. /ibi/ **(above) and** /ipi/ **(below) in quiet, 8 channel CIS processing; on the right are close-ups of basal channels**

Electrodographic analysis suggests that voice onset time is represented by the relative onset of activation of different electrodes. However, noise interference would introduce distortions to envelope fluctuations within the important low-frequency channels and therefore the voicing cues would be affected adversely. Given the possible importance of the secondary spectral cues to voicing, it might also be anticipated that parameters important to spectral coding, e.g. channel number or channel overlap/interaction, might have a greater bearing on voicing than manner (but less than place).

A further issue that is of particular relevance to voicing transmission, is the coding of fundamental frequency (F0). Figure 2.1 showed the spectrum of the source of energy for voiced speech sounds. The important characteristic of voicing, which is therefore of relevance to coding the voiced/voiceless contrast (and also some manner contrasts- see 2.6.2) is the quasi-periodic signal which has a fundamental and multiple harmonics. Given the evidence regarding limited frequency resolution in CI processing, including the Nucleus 24, it is difficult to see how this would be coded in the spectral domain, e.g. by activation across electrodes. Figure 2.14 shows a close-up of the two most apical electrodes in an electrodogram (ACE, 900 pps/ch, 12 maxima of 20 channels) during activation of the vowel /i/, as this gives a clearer picture of the representation of F0. The female speaker has a fundamental frequency of 200 Hz. It can be seen that F0 is discernible in the peak every 5 ms, although modulation depth is reduced by processing (as already shown by the TMTFs in 2.3.3.) However, figure 2.15 shows activation of more basal electrodes; here no clear periodic information is discernible.

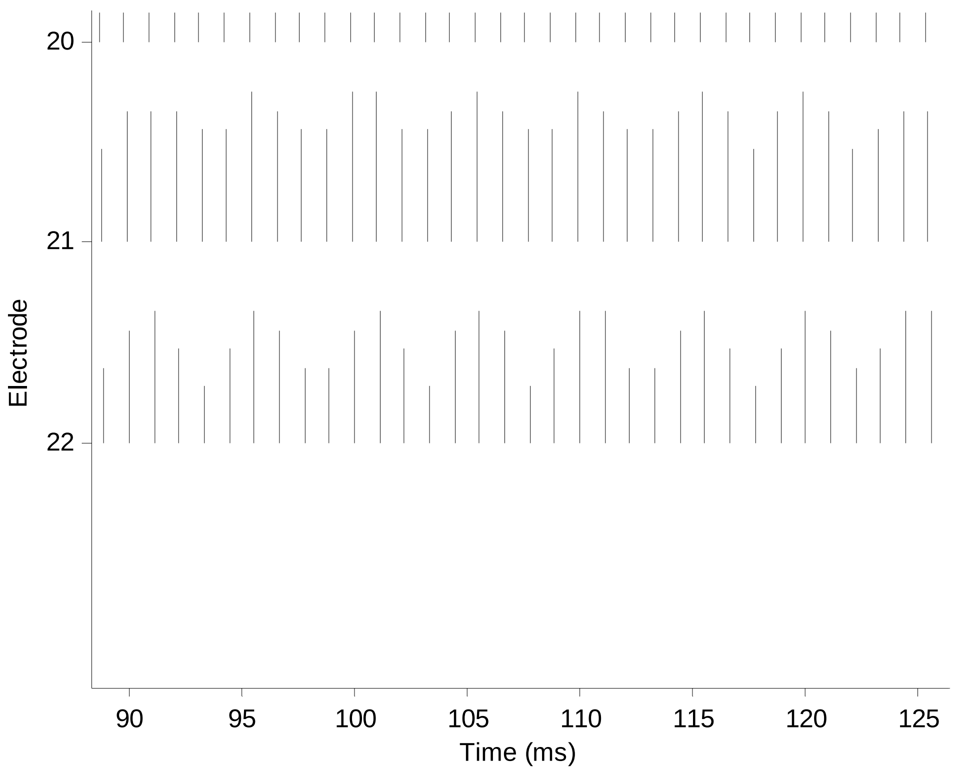

Figure 2.14. Close-up of electrodes 21 and 22. for the vowel /i/, produced by a female speaker with F0 around 200 Hz.

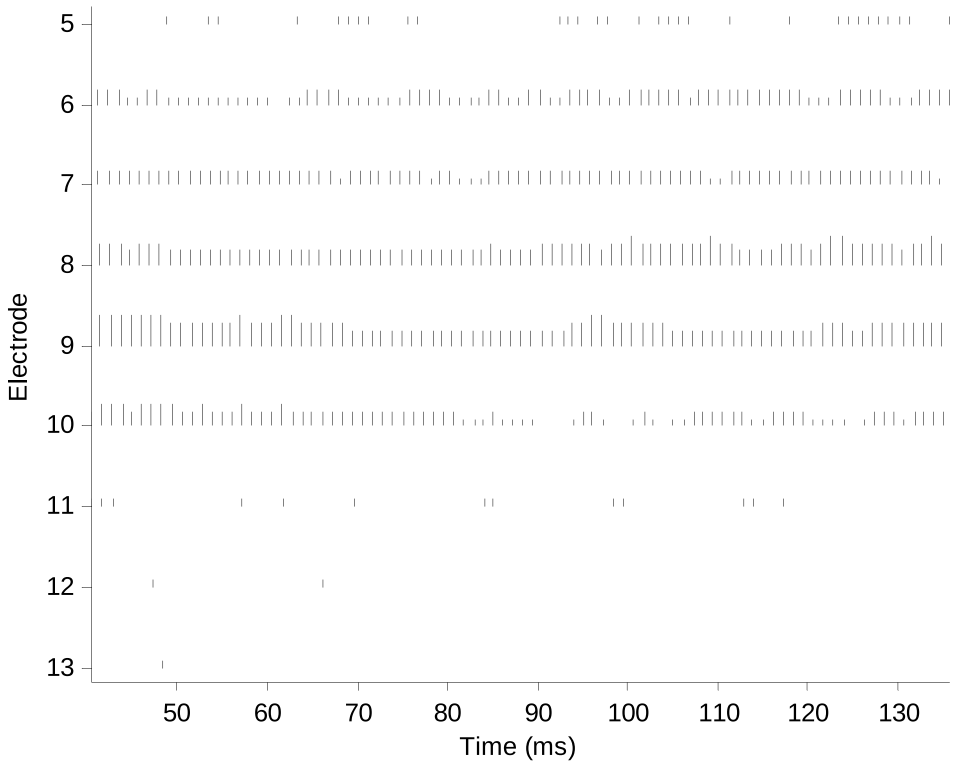

Figure 2.15. Basal electrodes during the same stimulus as figure 2.14. The stimulus is the vowel /i/, produced by a female speaker with F0 around 200 Hz.

These figures suggest that higher harmonics of F0 are poorly represented in the output of the analysis filterbank, while F0 itself is discernible in the modulation patterns of apical electrodes, albeit modulation depth is around 20-30% rather than 100%. Although the presence/absence of periodicity in itself is not the only acoustic cue to voicing, it is the main within-channel cue and therefore has an implication for voicing and other features requiring perception of periodicity in the waveform given that that F0 is present in the apical channels but there are no cues to higher harmonics; moreover, the reduction in modulation depth caused by processing means that, even in the apical channels the cue is not coded ideally through CI processing. Some authors have examined ways of making the F0 cue more salient: Green et al. (2005) noted that a form of modified CI processing, which enhanced F0, produced benefits to perception of F0, and this was thought to be, at least in part, due to improvements to modulation depth; however, the same strategy also led to reductions in perception of vowel recognition and formant frequency discrimination in the CI users accessing the modified strategy. Faulkner et al. (2000) found an improvement to voicing transmission for NH subjects listening to a variety of AMs which encoded periodicity, including one which explicitly preserved F0 information by producing a pulse sequence in time with F0.

2.6.2 Manner and manner subcategories

Manner of articulation refers to the way in which the upper vocal tract is occluded. In general, distinctions between different manner categories are considered in the literature to be signalled by temporal differences, with spectral resolution being less important than for voicing. However, this represents an over-simplification, as different manner distinctions are signalled by a variety of acoustic cues. It can be argued that the category itself is too general to be usefully linked to specific sets of acoustic characteristics, although the literature on consonant feature recognition in CI users generally uses this category in conjunction with voicing and place. While it could be argued that manner does reflect temporal/envelope information more than spectral information, it can be seen from the following acoustic analyses that the extent to which this is true depends on which specific manner distinction is being considered.

Although it is possible to subdivide consonants into manner categories in a number of different ways, here a four-way distinction is used, between stops (also known as plosives), nasals, liquids/glides and fricatives. Stops are produced by a rapid release of a complete closure of the vocal tract. The presence of a short duration (<100 ms) release burst (of aperiodic unvoiced sound) distinguishes stops from other manner categories, as does the presence of a short duration (<100 ms) formant transition (Liberman et al., 1956). Nasals are similar to stop consonants in that complete closure of the oral cavity is sustained. However, with nasals, the velum remains open, with various acoustic consequences (Malecot, 1956). There is a characteristic nasal "murmur" prior to closure release, with a characteristic low frequency prominence around 250 Hz with higher-frequency harmonics at very low amplitude. Additionally, nasals are characterised by antiformants, or zeros in the spectrum; these are unlikely to be realised by CI processing. Liquids and glides (also known as approximants) are produced with partial constriction of the vocal tract and can be distinguished by the presence of longer duration (>100 ms) formant transitions (O'Connor et al., 1957). Finally, fricatives (sometimes distinguished between high-frequency sibilants such as /s/ and broadband fricatives such as /f/) are also produced by incomplete closure of the upper vocal tract, but, unlike liquids and glides, they are associated with the generation of friction, producing a turbulent aperiodic signal of generally >100 ms duration (Raphael and Dorman, 1980).

In summary, stops can be distinguished from nasals as the latter have a much weaker formant structure and a slightly greater duration, from liquids and glides by formant transition duration and from fricatives by noise duration. Fricatives/stops can be distinguished from nasals and liquids/glides by the presence of noise in the spectrum. In order to illustrate this, figure 2.16 shows wide-band spectrograms of the unprocessed stimuli idi/, /izi/, /ini/ and /ili/.

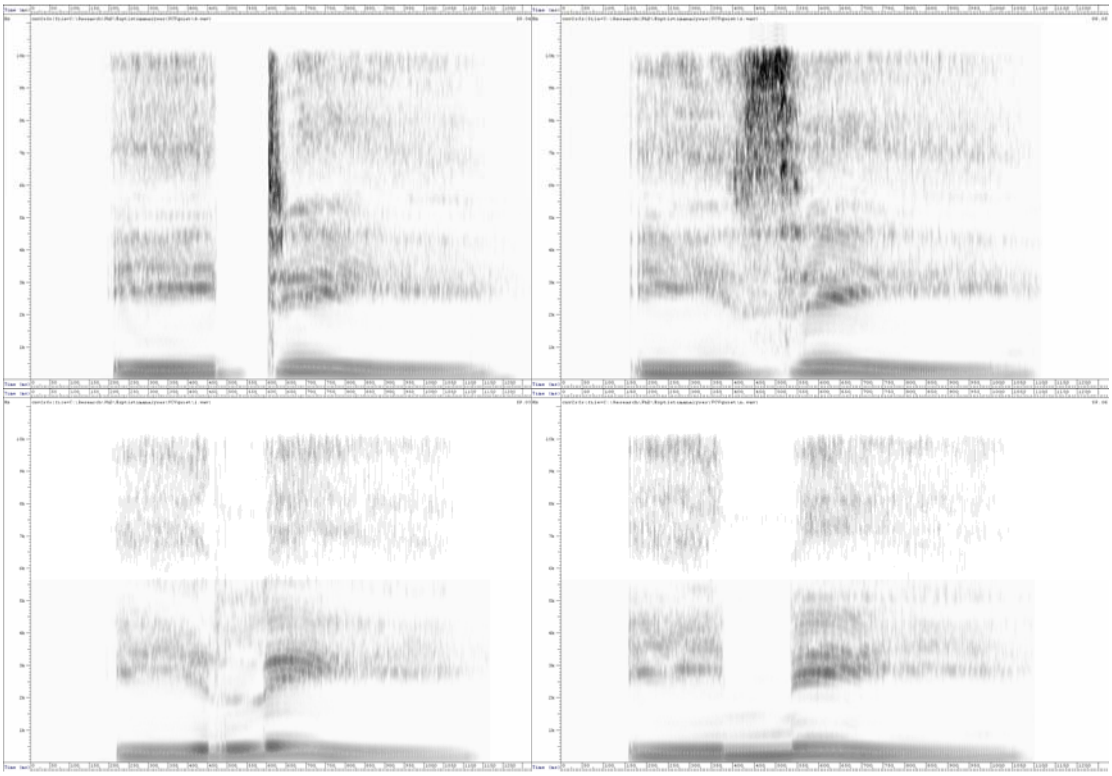

Figure 2.16. /idi/,(upper left) /izi/ (upper right) /ili/ (below left) and /ini/ (below right) in quiet. The continuant and nasal, below, can be distinguished from the stop and fricative, above, by the absence of high-frequency burst/frication energy and a period of silence or very low amplitude in low frequencies.

Figure 2.17 shows 8-channel CIS electrodograms of /idi/, /izi/, /ini/ and /ili/; these represent the four main English manner categories and all have the same voicing value and similar place values (although it should be noted that "voicing" is a confounding factor in that all nasals and approximants in English are voiced whereas this is not the case with stops or fricatives). The main acoustic cues distinguishing these four categories for NH listeners are: presence of noise (present in stops and fricatives but absent in nasals and liquids), duration of noise (distinguishing stops from fricatives), The nasal/approximant can be distinguished from the plosive/fricative via the continuous high level of activation in the most apical electrode. The difference between the nasal and the approximant lies primarily in the differences in activation in the slightly less apical electrodes. Here the difference between the stop and fricative lies primarily in the difference in activation in more basal electrodes.

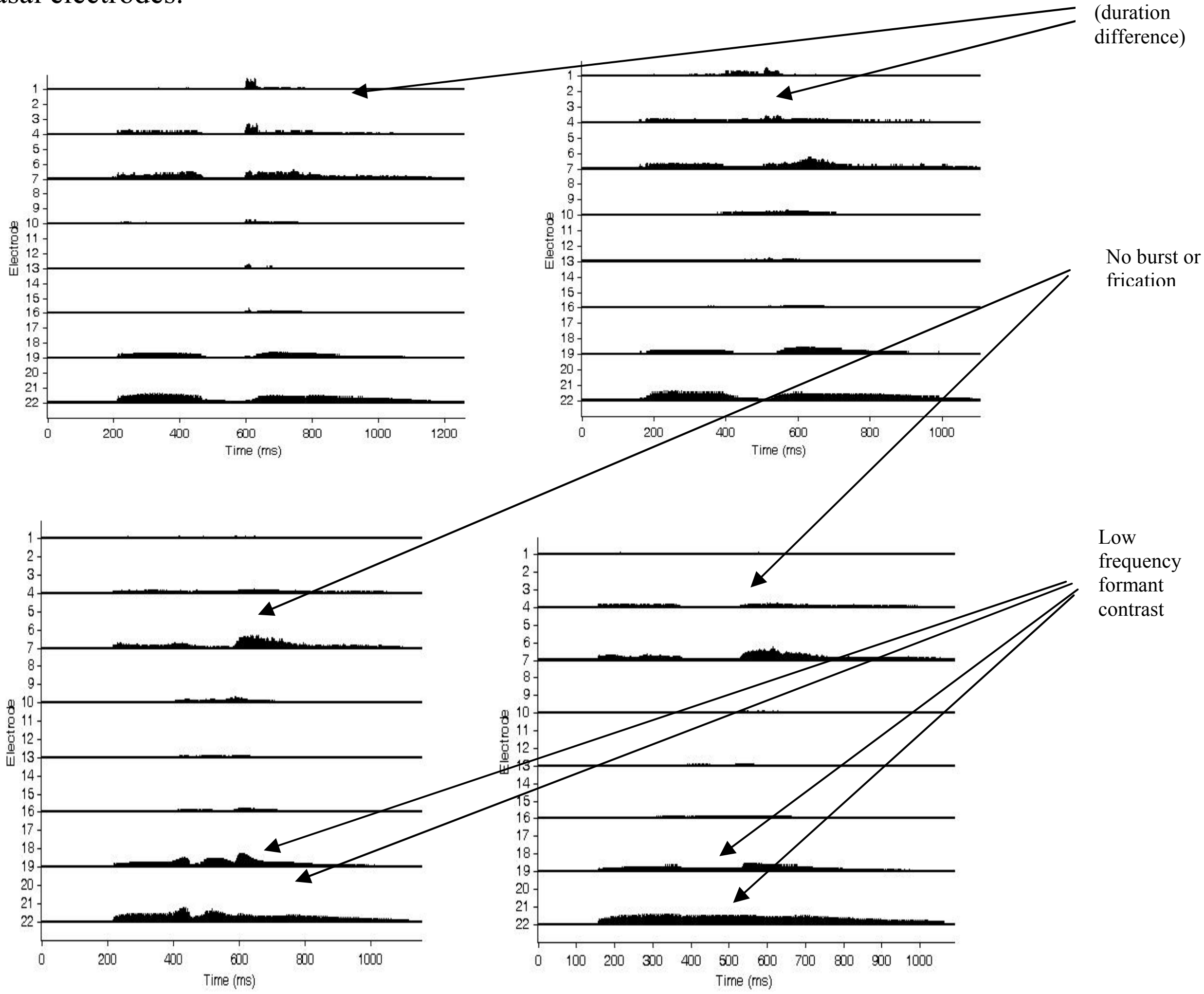

Figure 2.17. /idi/,(upper left) /izi/ (upper right) /ili/ (below left) and /ini/ (below right) in quiet, 8 channel CIS processing. The continuant and nasal, below, can be distinguished from the stop and fricative, above, by a more consistent pattern of activation in electrode 22 (e.g. there is no silence) whereas /idi/ and /izi/ have a period of low activation in electrode 22 and 19 between 400 and 600 ms.

The distinction between the stop /idi/ and the fricative /izi/ is in terms of activation between 400 and 600 ms in the basal electrodes. The nasal /ini/ can be distinguished from the three other manner tokens by the modulated pattern of activation in the apical channels- it can be seen that this could be easily masked by noise with greater energy in low frequencies. The liquid /ili/ can be distinguished from other stimuli primarily because of its consistent activation in the most apical channel. The temporal resolution of the Nucleus 24 device, as indicated in the TMTFs shown in figures 2.10 to 2.12, should allow the distinctions between the four manner categories to be coded. However, there are a number of complicating factors. First, how well does CI processing represent noise, e.g. the distinguishing characteristic of stops and fricatives, as opposed to quasi-periodic voiced components of speech, the distinguishing characteristics of voiced speech sounds in general and nasals in particular? Second, how well does CI processing represent nasals, which have a particularly weak formant structure and might be particularly susceptible to masking given the low amplitude of the component formants? The electrodograms in figure 2.17 suggest that it would be difficult to distinguish nasals from other categories, in particular from liquids/glides. Because manner categories are distinguished largely by the variance in activation pattern over time within electrodes (assuming a fairly crude high/low frequency resolution), it also seems likely that noise interference would serve to reduce the clear differentiation in level within electrodes over time.

The previous analyses suggest that the broad consensus in the literature , that manner is more reliant on temporal/envelope processing than spectral processing, can be supported and, consequently, that this feature should be transmitted better than place and possibly better than voicing. However, the addition of background noise could have a larger effect than for place because of temporal envelope fluctuations being important. However, it would be useful to assess transmission of specific manner subcategories in assessing CI user and AM performance rather than looking at "manner" as an overall category exclusively. This is because each of the four manner subcategories has distinct acoustic correlates which could be informative about the effects of CI processing and electrical/neural interface factors identified in 2.3 and 2.4.

Because of the formant structure of nasals, the potential for noise interference in perception of nasality should be greater than for voicing. How cues to nasality are represented by CI processing will depend on degree of pre-emphasis (this is critical given low-intensity cues being primarily in low frequencies) and factors affecting envelope fluctuations. Nasality should show greater susceptibility to noise than other temporal cues such as voicing or overall manner, or fricative, simply because the cues involved are of lower amplitude than equivalent cues to voicing. These relatively low amplitudes do not pose a problem for NH listeners when determining nasality in noise, but the limited amplitude resolution/dynamic range of CI users could have an impact in this regard. Also, we might expect some susceptibility to stimulation rate effects and in particular errors between nasals and approximants should be common at lower rates.

Identification of a consonant as a fricative rather than a different manner type requires identification of the presence of noise (to distinguish it from nasals or liquids) of durations greater than 100 ms (to distinguish it from plosives). Consequently, both spectral and temporal cues are available for fricative identification. The frication noise which characterises fricatives (and plosives, albeit of much shorter duration) is of greater amplitude and higher frequency than the quasi-periodic cues that distinguish nasals or approximants; consequentl y, it could be hypothesised that the fricative feature would be less susceptible to noise interference at positive SNRs. It can also be hypothesised that fricative identification is more reliant on spectral resolution than identification of nasals or liquids, hence it should be more affected by channel number of spectral channel interaction than nasality transmission, but less so than place transmission.

2.6.3 Place of articulation

The two most important cues to place of articulation, (for any of the four manner categories and for voiced or unvoiced consonants) are formant transition, particularly the second and third formant transition onset frequency, and the spectrum of the burst or frication (e.g. for stops and fricatives, respectively). In all the research literature this feature is the most poorly coded in CI users (see figure 2.4 and table 2.2). A likely reason for this is the very poor representation of formant transition information in the output of CI processing (Teoh et al., 2003). In order to illustrate this, figure 2.18 shows spectrograms of the original /ibi/ and /idi/ stimuli while figure 2.19 shows ACE 12/20 and CIS 8 channel electrodograms for the stimuli /ibi/ and /idi/ (pulse rate 500 pps/ch).

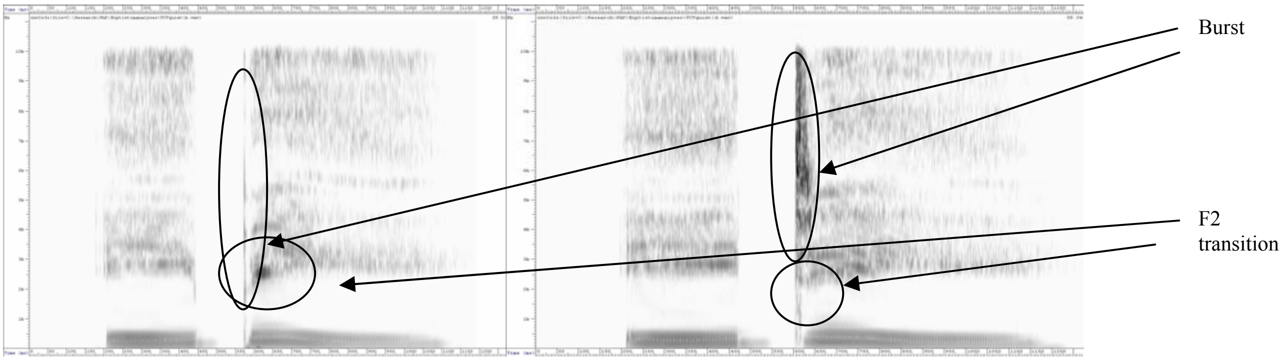

Figure 2.18. Spectrograms of /ibi/ and /idi/. The labiodental plosive /idi/ can be distinguished from the bilabial plosive /ibi/ by the presence of a higher-frequency (and longer duration) burst, in addition to a second formant frequency with higher frequency onset.

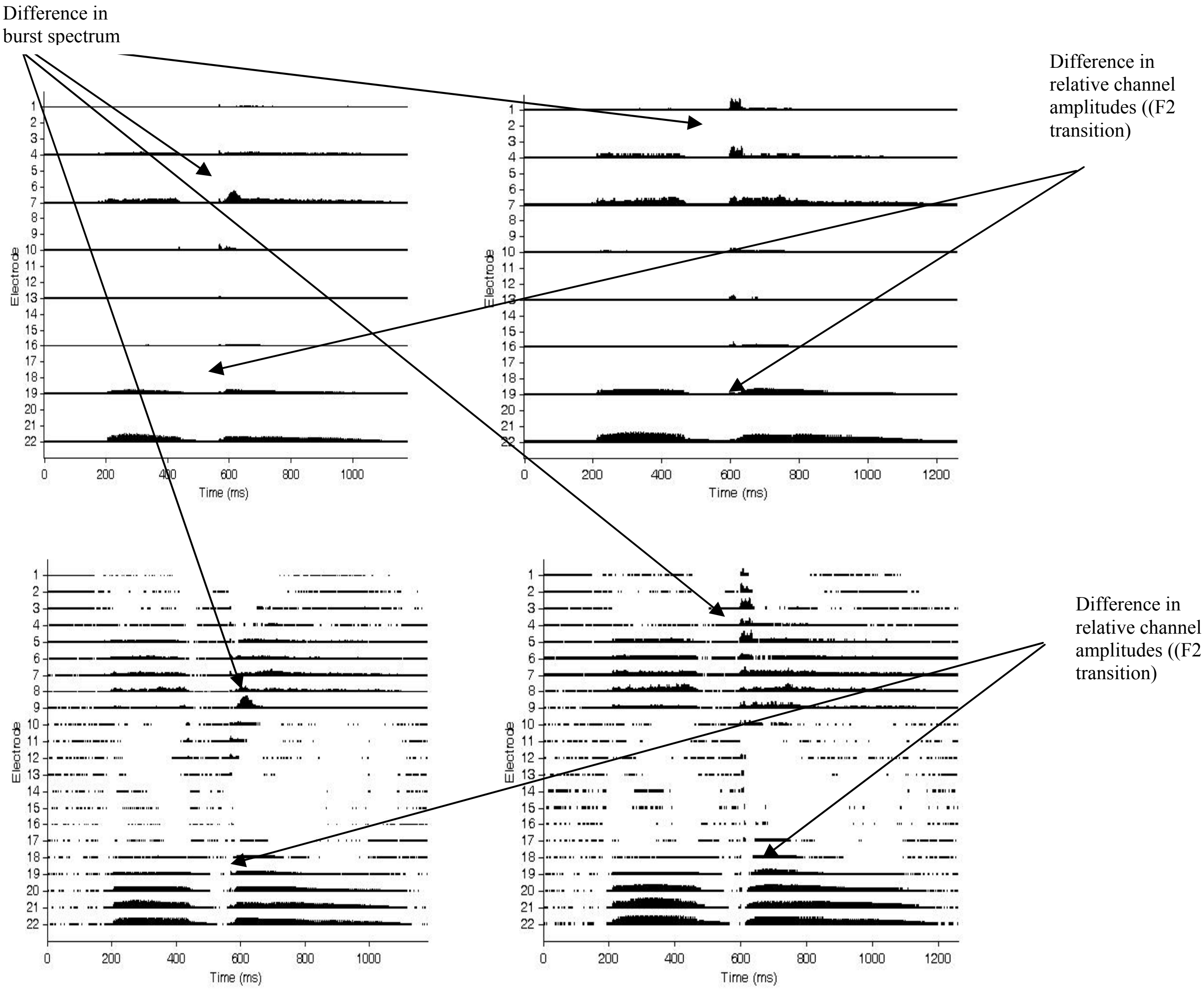

Figure 2.19. /ibi/ and /idi/; upper figures are CIS 8 x 900 , lower figures are ACE 12/20 x 900. The wider spectrum of the burst cue can be seen in /idi/ compared to /ibi/ but the difference in formant transition is not apparent, apart from a difference in degree of activation of electrode 19 from around 580 ms.

In the ACE electrodograms, the residual information coding formant transition is the relative amplitude of channel 17 compared to adjacent channels. However, for the 8 channel CIS electrodograms, even this information is lost. The richness of formant transition information is absent due to the relatively small number of channels used. The difference in second formant transition onset frequency between the /b/ in /ibi/ and the /d/ in /idi/ in the original stimulus is in the order of 400 Hz over a duration of <100 ms. Within-channel information would not be of use as formants (e.g. second formants are typically around 2000 Hz) would not be coded in the time pattern of individual channel envelope variations as these are beyond the temporal resolution of the CI system (see 2.3.3). By contrast, the figures for /ibi/ vs. /idi/' show a clear distinction in basal channels between stimuli- the high-frequency burst is coded as a local activation pattern at 600 milliseconds which extends to more basal electrodes for /idi/ than for /ibi/ The same argument applies to coding of fricative place. Figures 2.20 show spectrograms of unprocessed stimuli and 8-channel CIS electrodograms of /isi/ and /iʃi/. Here the difference is more pronounced, as might be expected given the wide bandwidth of the relevant cue.

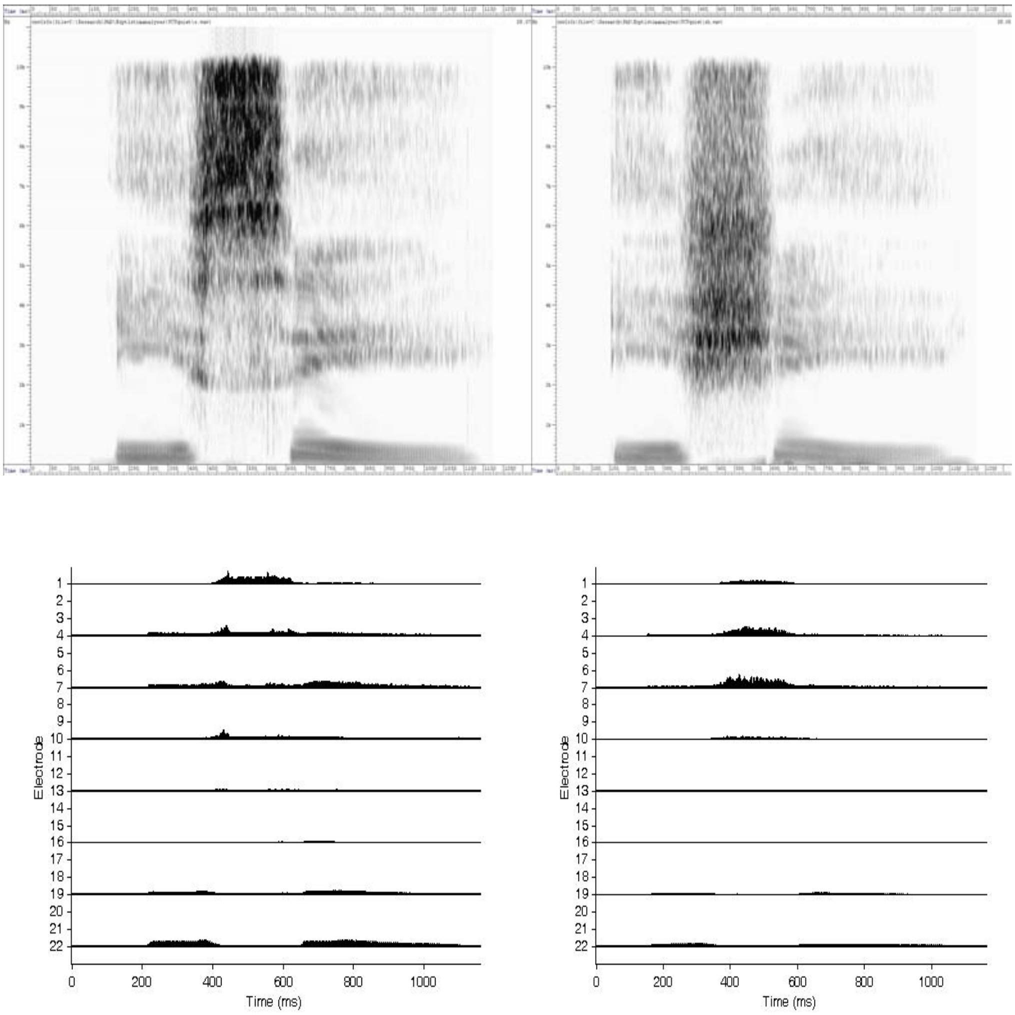

Figure 2.20. /isi/ and /iʃi/,; unprocessed stimuli above, stimuli transformed via 8 channel CIS processing below. For /isi/ there is greater activation of electrode 1 between 400 and 600 ms while for /iʃi/ there is greater activation of electrodes 4 and 7 over approximately the same time frame.

The relative salience of the burst/frication and the relative impoverishment of the formant transition cues lead to a number of hypotheses. First, place and manner coding for nasals or liquids should be more difficult for CI users than coding of fricatives or plosives as these distinctions require exclusively on formant transitions. Second, performance with place for fricatives and plosives should be sensitive to any parameters which might affect burst/frication coding, i.e. pre-emphasis, vowel environment, noise, stimulation rate and channel number.

The poor coding of formant transitions, which cue place in all manner categories, is implicated in the general finding that place of articulation is perceived less well than manner and voicing by CI users. The relative preservation of the burst must therefore become a more critical factor. The electrodograms in figure 2.19 show the high amplitude of the burst compared to activation across other channels. By the same token, the electrodograms in 2.20 show that frication energy coded in electrodes 1, 4 and 7, whose spectrum defines fricative place, is also higher in amplitude than activation in other electrodes. If place perception in quiet depends primarily on burst perception in CI users then noise interference should have only a small effect on performance up to relatively unfavourable SNRs. Another consequence of the reliance on the burst would be that place coding for nasals and liquids should be poorer than for stops or plosives, and also that place transmission in general should be better for the iCi than aCa environment (because in NH listeners the burst is more salient in iCi). Finally, the importance of the electrical/neural interface on place transmission must depend on the degree to which worse spectral resolution (e.g. associated with channel interaction) will reduce available place cues. In fact, if burst frequency is the primary cue, it is difficult to see that further channel overlap will have a worse effect on performance, as the resolution of burst spectrum can probably be achieved with 8 channels (thought to be equivalent to spectral resolution abilities in better CI users, as indicated in 2.4.2).

2.6.4 Auditory phonological categories

The previously defined categorisation scheme is based on the mechanism of speech production, albeit mechanisms which have corresponding acoustic and therefore auditory consequences. However, other further phonological categories based on purely acoustic or auditory distinctions have been used in the general speech perception literature. The category “sibilant” refers to a specific subset of fricatives with high-frequency energy loci. Another category “envelope” is of particular interest here. Blamey et al. (1985) first suggested a classification of consonants into four categories, each of which can be distinguished by gross shape when processed via a CI: unvoiced plosives, unvoiced fricatives, voiced plosives and fricatives, and nasals/liquids. Dorman et al. (1990) found a correlation between overall speech perception and transmission of the envelope feature (although note that the devices used in that study had less spectral and temporal resolution than those of more recent interest). In order to illustrate this, figures 2.21 shows electrodograms of tokens of each of the four different groups, simulated with a 1-channel CIS model and an update rate of 250 pps/ch:

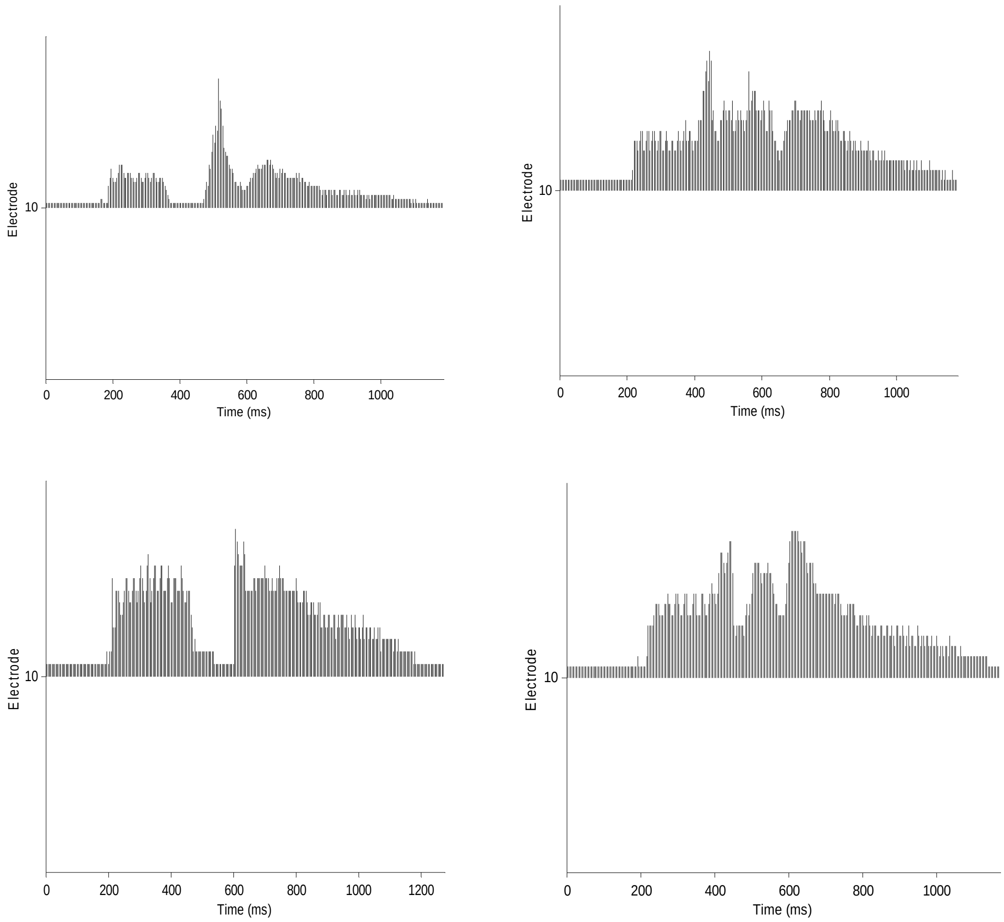

Figure 2.21. Electrodograms /iti/ and /isi/ above left and right; /idi/ and /ili/ below left and right, processed through a 1-channel CIS strategy with 250 pps/ch stimulation rate.

Here we can see that gross overall shape is distinct between different groupings. This would suggest that transmission of this feature would be the most "robust" of all the consonant features, e.g. should show the least effect for parameters such as channel number, stimulation rate and channel interaction.

2.7 Overview

The previous discussion covered a range of issues relating to the information processing stages involved in speech perception by CI users and how these processes could be evaluated using AMs. The possible link between specific stages and recognition of some consonant features has also been outlined. Figure 2.22 suggests a more detailed picture of information flow for users of the Nucleus 24 device. Input stage processing is distinguished from frequency analysis and envelope extraction as the main two stages of CI processing. The subsequent stage of the electrical/neural interface is given as a distinct stage as is processing in the central nervous system. For a set of CI users using the same signal processing characteristics, differences in individual CI user performance can be attributed to the latter two stages. At each stage of processing different sources of information loss are proposed, in line with the discussion outlined in previous sections of this chapter. For the work reported in chapter 4, the same set of processing characteristics applied to both CI users and AM listeners. Consequently, differences between AM listeners and CI users could be attributed to different processes involved in the interface with the auditory system.

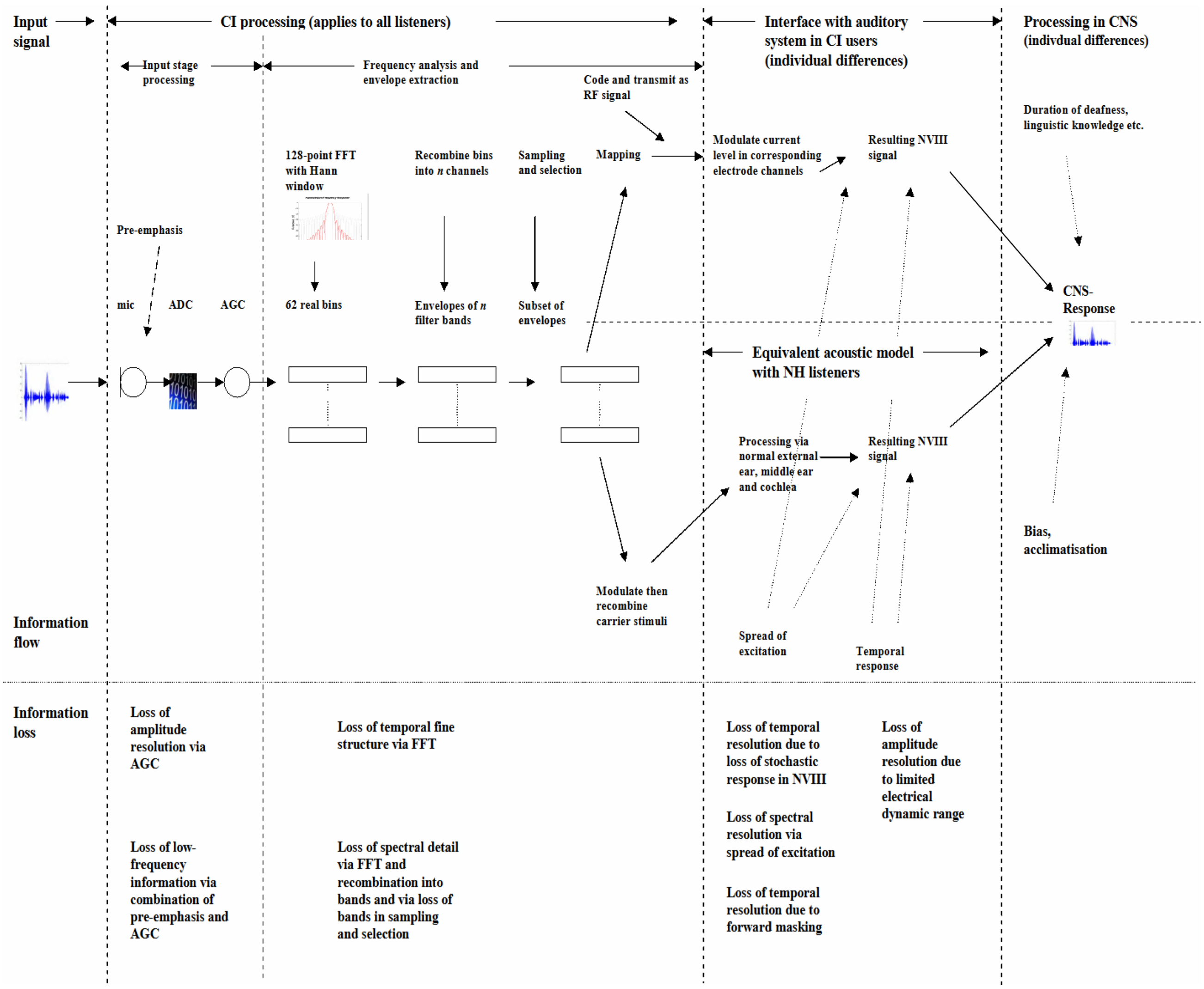

Figure 2.22. Proposed overall conceptual map of information loss and information flow in the Nucleus 24 cochlear implant, including reference to equivalent acoustic models

Chapter 3. Methods

3.1 Research questions and hypotheses

The focus in this thesis is on the relative contribution of CI signal processing information loss as against electrical/neural interface information loss in determining the pattern of consonant feature transmission in CI users. The overriding question motivating the research was: “to what extent can deficits in consonant recognition by CI users be explained by information loss in CI signal processing as opposed to information loss at the electrical/neural interface?” This question is not directly answerable but must be translated into experimentally tractable hypotheses. The main problem in determining the relative contribution of electrical/neural interface factors such as channel interaction is the difficulty in controlling variations in these factors between individual CI users. While there is good evidence that individual CI users may vary in terms of the degree of spectral and/or temporal channel interaction and in other electrical/neural interface factors, it is unclear to what extent variations in these underlying abilities contribute to individual performance (Throckmorton and Collins, 1999). Moreover, there is no consensus as to how to measure these individual differences, whether through psychophysical or objective means.

It is therefore argued that, in order to differentiate effects of CI signal processing from other factors, it is highly useful to compare results between normal (NH) subjects listening to AMs of CI processing with results obtained from CI users using equivalent signal processing. This approach has been justified by Throckmorton and Collins (2002) and Laneau et al. (2006), among others. The rationale is as follows: where CI performance and AM performance match, explanations for CI performance can be related directly to model design. More specifically, if an AM which only takes into account CI processing characteristics can predict CI performance, then it follows that CI processing information loss can explain CI user performance. If, however, the model works better if it also incorporates some aspects of the electrical/neural interface, then it follows that the information loss at the electrical/neural interface must also contribute to the CI performance.

An assumption behind much work in AM research is that a range of different AMs may account for CI user performance so long as those models have the *general* properties of CI processing that are perceptually important, e.g. the relatively small number of channels and the absence of temporal fine structure within channels. Almost all AM studies to date have used fixed-channel models and envelope extraction has been via a set of linear IIR filters followed by rectification and smoothing. However, only some CI systems implement this type of processing while others, notably the Nucleus 24 which is the focus of the present study, use an FFT filterbank and, also, most users use peak-picking strategies such as ACE rather than fixed channel strategies. There is an identifiable need to consider the extent to which the results obtained can be attributed to the *specific* set of CI processing parameters, (e.g. pre-emphasis, strategy type, FFT parameters, channel number and channel stimulation rate). Additionally, consideration must be made, specific to AMs themselves, as to the effects of specific choices of waveform output parameters, (e.g. carrier stimulus) and stimulus parameters, (vowel environment and noise type). It should be noted that each of these specific choices is evaluated in experimental work in the study by comparison with alternatives, with the exception of noise type and input stage processing.

The assessment of consonant feature information transmission provides an opportunity to determine if an AM is predictive of CI user performance. This is because transmission of different consonant features relies on different underlying psychoacoustic abilities and therefore relates to different aspects of signal acoustics. Therefore, it is useful to compare AM performance against CI user performance in a number of ways. First, the pattern of information transmission across consonant features; second, the pattern of effects of background noise across consonant features; third, the pattern of effects of CI processing parameters across features; fourth, the pattern of effects of electrical/neural interface factors across features. If an AM can predict the magnitude and/or pattern of consonant feature transmission as a function of any or all of these variables, then it can be said to have explanatory power in predicting CI performance.

A number of knowledge gaps were identified in chapter 2. This leads to a series of research questions concerning CI users' consonant recognition. Almost all the more specific research questions can be framed in the context of this more general knowledge gap, e.g. the lack of knowledge about the relative importance of processing and electrical/neural interface information loss to consonant recognition.

For CI consonant recognition, a consistent finding has been worse place of articulation perception than manner or voicing perception. However, there were identified knowledge gaps in the following areas:

1. Is the pattern of consonant feature transmission in CI users the same in quiet and noise?
2. Is the pattern of consonant feature transmission in CI users the same in vowel environments other than /aCa/?
3. What is the pattern of consonant feature transmission in users of the Nucleus 24 device?

The remaining questions relate to the ability of an AM to predict consonant recognition abilities in CI users:

4. Can an AM accurately predict the pattern of relative consonant feature transmission (in quiet or noise)?
5. Can CI consonant recognition be predicted better by an AM with or without the characteristic shift in perceived pitch associated with CI insertion (referred to as "pitch mismatch")?
6. Can CI consonant recognition be predicted better by a model incorporating channel interaction and, if so, how much channel interaction is required to optimally match CI user performance?
7. Can variations in channel interaction model variations in CI user performance?
8. Which version of an AM can best predict changes to CI user performance with changes in channel number?
9. Which version of an AM can best predict changes to CI user performance with changes in stimulation rate?
10. Does choice of AM carrier stimulus have a bearing on the prediction of CI user performance?

Although Laneau et al. (2006) compared performance with a specific device with performance using an AM which incorporated electrical/neural interface features (channel interaction/e.g. spectral channel interaction), the authors assessed different aspects of speech perception than those addressed here. The authors found a correlation in performance between AM and CI user findings with spectral channel interaction equivalent to 1 mm in the model. It can therefore be hypothesised that consonant recognition in Nucleus 24 users will be best approximated with a model in which channel interaction is equivalent to 1 mm spectral spread.

Experimental hypotheses can either be couched as overall hypotheses or as feature-specific hypotheses. Section 3.3.6 gives a justification for choosing six specific consonant features voicing, place, manner, nasality, fricative and envelope. Ideally, each of the consonant features would have a corresponding hypothesis for each variable in each experiment. A number of feature-specific hypotheses have been put forward in 2.6. More specific hypotheses, including those relevant to processing parameter variables and to specific features, are stated within the context of each experiment in chapters 3 and 4.

3.2 Aims

There are a number of questions and aims in 3.1 that are specific to AMs and to test methodology as opposed to the relationship between AMs and CI user performance. Therefore the initial experimental work, reported in chapter 4, was concerned with these areas. Because the potential complexity of further planned experiments, it was important to determine two more purely methodological questions in experiment 1. These two methodological questions were motivated by the need to keep the number of distinct variables as low as possible for further experimental work in order to minimise subject fatigue effects and provide a practical experiment. First, would a relatively small number of repetitions of each consonant give “valid” results? The second was, what was likely to be an optimally sensitive SNR for use with further experiments of consonant recognition in noise?

The experimental work had two distinguishable sets of aims, the first relating to AMs specifically, and the second relating to the ability of AMs to predict CI user performance. The first set of aims were addressed in the experimental work reported in chapter 3 and can be summarised as follows:

1. Develop an AM of a specific CI device in order to achieve the following aims:
2. Determine the relative transmission of different consonant features.
3. Determine the relative effect of noise at different SNRs on consonant features.
4. Determine the effect of carrier stimulus on consonant feature transmission.
5. Determine the effect of including pitch shift on consonant feature transmission.
6. Determine the effect of vowel environment on consonant feature transmission.
7. Decide on the "optimal" combination of model and stimulus parameters for an AM to compare directly with equivalent CI user data.

The second set of aims were addressed in experimental work reported in chapter 4 and related to the comparison between CI user and AM performance:

1. Ensure that the processing and stimulus variables in the model and CI users were, as far as possible, equivalent.
2. Determine the effects of changing channel/maxima number on consonant feature transmission in the model and in the CI users.
3. Determine the effects of changing channel stimulation rate on consonant feature transmission in the model and in the CI users.
4. Determine the effect of altering carrier stimulus overlap as undertaken in Laneau et al. (2006) (as a means of mimicking spectral channel interaction) on feature transmission in the model.
5. Determine whether the inclusion of channel interaction improved the fit between model and CI user data.
6. Determine whether variance among CI users could be modelled by variations in channel interaction in the model.

3.3 Methodology

3.3.1 Overall approach to test methodology

The approach taken in this study was to evaluate two aspects of the electrical/neural interface through AMs. In experiment two, AMs were generated with and without the characteristic "pitch mismatch" associated with electrode insertion. For the third experiment, pitch mismatch was included in all listening conditions (having been found to make only modest differences to AM performance in the second experiment) but the presence and degree of channel interaction, a proxy for assumed spectral channel interaction, was systematically varied. The rationale here was to see whether variations in performance across channel interaction conditions could mimic variations in performance across individual CI users and, more generally, whether the inclusion of channel interaction improved the "fit" between AM and CI user data.

The remainder of this section describes the methodology used for experimental work in chapters 4 and 5. Where specific experiments deviated from this methodology, details are given in the relevant chapter. The following principles were adhered to across the four experiments:

(1) The same approach to consonant feature analysis, and set of six consonant features, was used throughout (see 3.3.6 for a justification for choice of features).
(2) The same experimental paradigm was used throughout (this was established as being workable during the conduct of experiment 1 and its pilot study).
(3) The same noise type and noise addition method were used throughout.
(4) All three AM experiments used the NIC-STREAM (Cochlear, 2002) and AMO MATLAB platforms (Laneau et al., 2006). These implement the same processing as the Nucleus 24 CI system. Additionally, all stimuli were filtered using a pre-emphasis filter prior to AM processing proper.
(5) For the CI user experiment, the standard programming platform for the Nucleus 24 device was used.

In all experimental work, a 20-alternative forced-choice nonsense syllable recognition task was undertaken. Nonsense syllables took the form iCi, where the vowel /i/ is followed by one of twenty English consonants and then followed by a second token of the same vowel. The rationale for choosing such a large consonant inventory is that this allows the fullest possible analysis of different consonant features. The stimulus set represents 20 out of the 24 English consonants although it excludes /h/ which can be considered a glottal vowel, and /ð/, /ŋ/ and /ʒ/ which do not have a unique spelling indicator. Moreover, the specific stimulus set has been validated in CI users as part of a large study of adult CI outcomes (UK Cochlear Implant Study Group, 2004). Although the majority of both normal hearing and CI studies evaluating consonant recognition have used the /aCa/ vowel environment, there are a number of reasons for choosing the /iCi/ vowel environment instead. In a recent study of stop consonant recognition in noise by four normal hearing listeners (Jiang et al., 2006) the /iCi/ vowel environment yielded a larger effect of background stationary noise than /aCa/. The authors showed, through acoustic analysis, that voicing perception was determined more by F1 onset frequency than voice onset time at unfavourable SNRs but that F1 onset frequency is more salient in the /aCa/ environment. Loizou et al. (2000b) showed that consonant recognition in the /iCi/ vowel environment was more sensitive to stimulation rate in CI users than with the /aCa/ vowel environment and, more generally, performance was poorer than for /aCa/ where ceiling effects were obtained in some conditions. Although these findings are not directly relevant to the study carried out here, they do suggest that consonant recognition in the /iCi/ vowel environment may be more sensitive to small parameter changes and less likely to yield ceiling effects.

In summary, the methodology used was as follows:

- Vowel environment: /iCi/ or /aCa/
- Choice of stimuli: 20 English consonants
- Total number of stimuli: 20
- Number of presentations per stimulus: 3
- Single or multiple speakers: Single
- Speaker gender: male, female or mixed: Female
- Number of iterations of SINFA analysis: 1
- Provision of feedback: none
- Amount of acclimatisation to the model: Self-directed (as describe in 3.1.2), typically 5-10 minutes per subject.

Most of the studies on consonant feature recognition in CI users have used relatively heterogeneous groups of CI users, e.g. the CI participants in the studies used varying signal processing parameters. To control for variations in signal processing, all the work reported in this dissertation used one device, the Nucleus 24. For the main experiments (reported in chapter 5), all the CI subjects the same signal processing parameters and in most cases were most highly acclimatised to using the particular set of parameters. The corresponding AM experiment also used precisely the same parameters.

3.3.2 Stimulus processing

All stimuli used as input to the processing were recorded nonsense syllables using a female speaker, kept as digitised Microsoft sound (.wav) files with a sampling rate of 22,050 and a resolution of 16 bits. An additional stimulus was “speech-shaped noise”- this was white noise filtered to have the same long-term average spectrum as the BKB sentences (Bench et al., 1979) spoken by an adult female speaker. A randomly extracted sample (of the appropriate length) was mixed with the speech stimuli at the appropriate SNR for noise-contaminated listening conditions. Two possibilities exist with respect to how to achieve a defined SNR for VCV nonsense syllables: either the signal RMS level could be averaged across the entire signal duration or, alternatively, the signal RMS could be computed across the duration of the nominal consonant portion of the stimulus. There are disadvantages of each method: with the first option, the effective SNR with respect to the consonant itself will vary according the consonant to vowel amplitude ratio while with the second option the overall level of the signal will vary and lack of a clear definition of start and end times of the consonant portion makes the task more subjective than is ideal. For this study the first approach was used (across all experiments). In order to this, the software package Adobe Audition was used to determine the RMS level of each stimulus. For each stimulus conditions, the average RMS level of all 20 stimuli was first determined. A randomly chosen portion was copied from the sound file containing the speech-shaped noise was adjusted so that its mean RMS was at the appropriate level for whichever SNR was to be used. This was then mixed with the target stimuli at the appropriate SNR. It should also be noted that all sound files containing the target stimuli had 1 second of silence before and after the stimulus and for noise-contaminated stimuli noise also began 1 second before stimulus onset and one second after stimulus end. A final processing stage prior to AMling was down-sampling of the sound files to 16,000 samples per second as the recordings had been made using a 22,050 sampling rate whereas the input to the NIC-STREAM/AMO processing needed to be 16,000 Hz to mimic the Nucleus 24 audio sampling rate. Stimuli were also decimated to an 8-bit rate as this is the quantization used by the Nucleus 24 processor.

The remainder of this section describes the signal processing principles used to produce the AMs (e.g. simulated stimuli for presentation to normal hearing listeners), although some further details are given to specific to each experiment. Stimuli were processed using NIC-STREAM, a MATLAB software toolbox created for processing of cochlear implant signals with the Nucleus 24 cochlear implant system, designed by Brett Swanson of Cochlear Corporation to mimic the processing of the Nucleus 24 device. The platform is much more flexible than the standard clinical programming software and is designed for research use. Its advantage for this work was the fact that it implements the same filterbank, envelope extraction and channel mapping processes as are implemented in the Nucleus 24 device and therefore allowed a valid comparison between AM and CI user data. NIC-STREAM comprises a MATLAB toolbox for generation of pulse sequences in addition to a set of functions for direct stimulation of a CI (the latter were not used in this study).

Figure 3.1 shows the conceptual stages of processing, both for the Nucleus 24 device and for the NIC-STREAM stimulus processing. For the purposes of this work, only those MATLAB functions necessary to generate a channel magnitude sequence were used. At the time of initial experimental work, the MATLAB toolbox did not implement front end processing. Consequently, this aspect of processing was dealt with separately (see below) and the input to NIC-STREAM was at the filterbank stage. Consequently, the Nucleus MATLAB toolbox was used for filterbank and sampling and selection stages of stimulus processing. Audio input to the filterbank stage generates a 2-dimensional matrix known as a “frequency-time matrix” which represents variations in output for each filter (in the case of experiments 2, the filterbank was configured as having 8 filter outputs). The subsequent stage of sampling and selection was used to generate a channel magnitude sequence for ACE processing as used in experiment 3, but for CIS the frequency-time matrix and channel-magnitude sequence were effectively identical as with the CIS strategy all filter outputs are chosen. The channel magnitude sequence was used to generate acoustic stimuli for the AM experiments and also to generate visual representations of nominal electrode output (“electrodograms”) used in chapters 2 and 6.

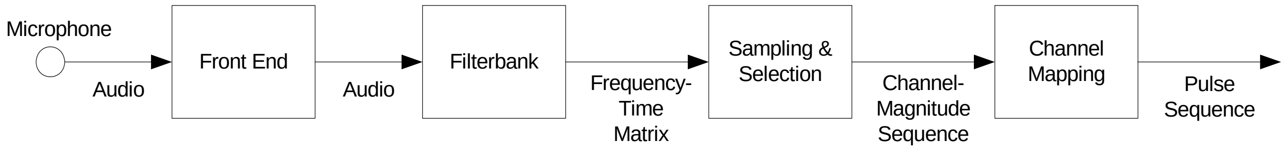

Figure 3.1. Signal flow in the ACE and CIS speech processing strategies. Reproduced with permission of Brett Swanson, Cochlear Corporation.

Additional MATLAB M-files were developed by Johan Laneau and colleagues (Laneau et al., 2006) for generation of AMs and were used for experiments 2 and 4. These additional functions allowed the inclusion of a channel interaction model that was implemented by altering the filter characteristics used to generate the noise bands used as carrier stimuli. The AM was developed and validated in a study of pitch perception (Laneau et al., 2006) and was based on the mathematical model of current spread of Black and Clark (1980), described in 2.4.1. As the unique aspects of this model were only used for generation of stimuli in experiment 4, further details are given in section 5.2. The remaining details of processing given here apply across all three AM experiments.

At the beginning of the experimental work, front end processing was not included in NIC-STREAM. Therefore, the first stage of stimulus processing was the implementation of a pre-emphasis filter to mimic the normal high frequency boost used by the Sprint and Esprit speech processors. The frequency response of the Sprint microphone was determined empirically and the measurements use to determine this are described in Appendix A. This was defined as having the following characteristics: up to 1800 Hz, 6 dB per octave was added; between 1800 and 5000 Hz there was a flat frequency response; from 5000 to 10,000 Hz a 24 dB per octave decrease was implemented. The pre-emphasis was implemented in Adobe Audition using an FFT filter with a Hamming window and an FFT size of 8192. In some cases implementation of the pre-emphasis led to clipping and therefore the filter was implemented with an overall gain reduction as necessary to reduce clipping. However, prior to subsequent processing, all stimuli were re-scaled to the same relative levels (to one another) as obtained prior to the addition of pre-emphasis. With the Nucleus 24 device, the pre-emphasis is inbuilt in the microphone and therefore the subsequent stage of processing would be ADC. However, the stimuli here had already been down-sampled to 16,000 Hz with an 8-bit resolution (e.g. the characteristics of the ADC stage within the Nucleus device) so no further processing was necessary to mimic the Nucleus device in this respect.

The next stage of processing was to band-pass filter the signal using the NIC-STREAM/Nucleus FFT filter bank. It should be noted that the same filterbank is used for both ACE and CIS processing strategies therefore this is identical across AM experiments. The input waveform was analysed at the same rate as the nominal "stimulation rate", e.g. 500 Hz for experiments 1 and 2 and 900 or 250 Hz for experiment 3. As with the Nucleus device itself, a 128-point FFT was performed. This yielded bin centre frequencies that were linearly spaced at multiples of 125 Hz and which had a 6dB bandwidth of 250 Hz. These bins were combined by summing powers to provide eight frequency bands as per figure 3.2. For experiments 1 and 2, an 8-channel CIS implementation was used: the upper and lower frequency boundaries of the 8 analysis filters are shown in figure 3.2. For experiment 4, an ACE implementation was used (in order to match the clinical parameters actually used by the CI users) and details of the corresponding analysis filters are given in chapter 5.

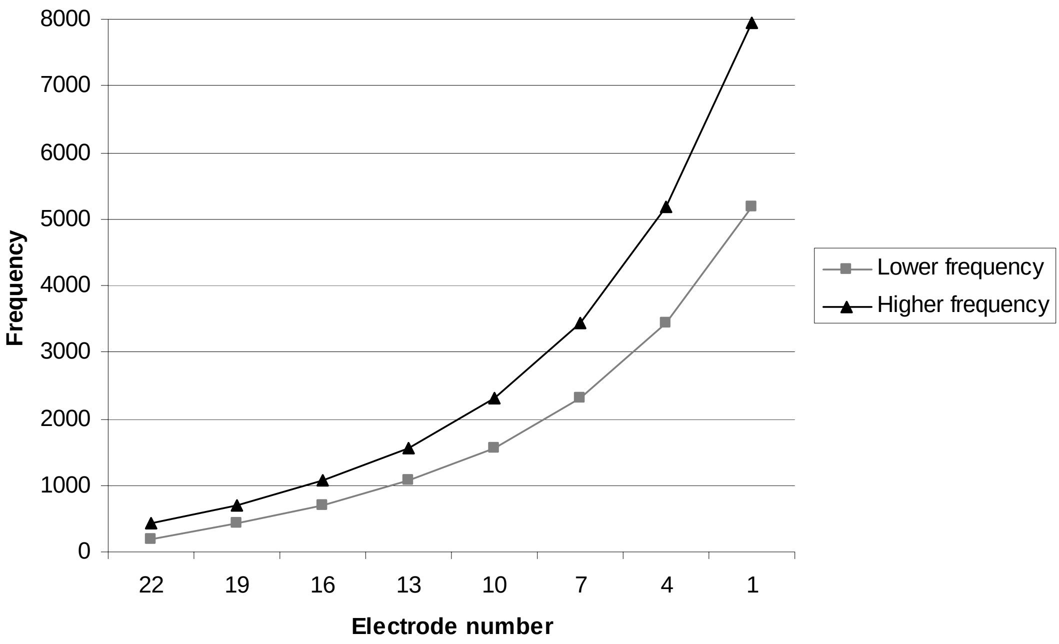

Figure 3.2. Frequency allocation for the 8-channel CIS implementation used in experiments 1 and 2.

The envelope of each filter was calculated as a weighted sum of the corresponding FFT bin powers where the weights determined the frequency boundaries of the bands. Carrier stimuli were modulated according to the fluctuations in the envelopes of the corresponding band-pass filters. The nature of the carrier stimuli varied across experiments in terms of: carrier stimulus type, choice of (centre) frequency and (in the case of noise bands for experiment 4 only) overlap between carriers.. For experiment 1, sine waves were used, whose frequencies corresponded to the centre frequencies of the 8 FFT filter outputs shown in figure 3.2. For experiment 2, noise bands and sine waves were used in different models for comparison purposes. For half of the models used in experiment 2, centre frequencies of the carriers corresponded to the centre frequencies of the FFT filter outputs as shown in figure 3.2, as in experiment 1. However, for half of the acoustic models in experiments 2, and all of the models in experiment 4, the centre frequencies of the carrier stimuli were shifted upwards in frequency so that they so that they corresponded to the assumed place of excitation along the basilar membrane (F in equation 3.1) for the corresponding intracochlear electrode (assuming the standard Nucleus 24 electrode array inserted 25 mm into the cochlea). This frequency transformation was determined according to Greenwood (1990).Consequently, the centre frequencies of the channels used in the CI processing were shifted upwards in frequency based upon the assumed frequency along the basilar membrane for an electrode array with 22 electrodes placed 25 mm into a cochlear with a length of 33 mm. To determine the appropriate frequencies, Greenwood's formula, given here as equation 3.1, was used.

$$F = A\left(10^{ax} - k\right)$$

where

F=centre frequency in Hz

A=165.4

a= 0.06

x=distance along basilar membrane in mm.

k=1

Equation 3.1. Determination of centre frequency corresponding to place along the basilar membrane according to Greenwood, 1990.

To take an example, the filter output for (virtual) electrode 13 in the 8-channel CIS model shown in figure 3.2 yielded a centre frequency of 1313Hz. The corresponding electrode along a 22 electrode array of 25 mm length along a 33mm basilar membrane was assumed to be 17.8 mm from the apex. This resulted in an assumed characteristic frequency of 1768 Hz according to Greenwood's formula. Consequently, the frequency of the carrier (sine wave frequency, or noise band centre frequency), was shifted upwards by 455 Hz. The formula, combined with information about electrode array characteristics and typical insertion depth, yielded upwards shifts in frequency which ranged from 1.2 for apical/low-frequency channels to 1.45 at basal/high frequency channels. The same shift was used to determine the frequency of the sine wave carriers (for experiment 2) and the centre frequency of the noise band carriers (in experiments 2 and 4). Because of the finding from experiment 2 that this degree of "pitch shift" had only a very modest effect on performance, the transform was applied to all models used in experiment 4. It should be noted that the filter bank frequency bands reported in figure 3.2, 5.1 and 5.2 reflect analysis filter bank characteristics (common to AM and CI processing), not necessarily AM output carrier frequencies, given that these were transformed systematically as described above.

For experiment 4, noise band carriers with centre frequencies chosen to reflect corresponding cochlear locations according to Greenwood (1990) were used as in experiment 2. Additionally, in order to model spectral channel interaction the frequency response of the filters used to generate the noise-band carriers was altered, according to the Laneau et al. (2006) model. The frequency response of the filter was designed to simulate the exponential decay of current density along the basilar membrane (Black and Clark, 1980) and is defined by:

$$F(x(f)) = \exp\left(\frac{-abs(xelectrode - x(f))}{\lambda}\right)$$

where

λ = distance along cochlear in mm (the conversion of distance on a cochlear into the frequency domain assumed the Greenwood formula)

xelectrode = the position of the simulated electrode

x(f) implements the conversion to distance along the cochlea from frequency according to Greenwood, 1990

Equation 3.2. Filter transfer function used to model spectral channel intertaction from Laneau et al., 2006

The desired frequency response was obtained by implementing a linear phase FIR filter in MATLAB. The model assumed a 35 mm cochlear length and 25 mm electrode array insertion. Laneau et al. (2006), applying the same model, found equivalent performance between Nucleus 24 users and AM listeners when a channel overlap term equivalent to 1 mm spectral spread of excitation was used. However, those papers evaluated pitch perception rather than segmental perception, e.g. consonant identification. It was therefore chosen to take three channel overlap conditions: first, no overlap between noise band carriers, second, overlap equivalent to 1 mm spectral spread, and, finally, overlap equivalent to 3.3 mm spectral spread, similar to the value suggested by Black and Clark (1980). Therefore, the three models were identical except for the definition of λ which varied across three values. Figure 4.3 shows the effect of varying λ. The figure shows wide-band spectrograms of a 2000 Hz pure tone which was sinusoidally amplitude modulated at 50 Hz with a modulation depth of 100% and processed through an AM of the ACE speech processing strategy (12 maxima out of 20 channels) and a 900 pps/ch stimulation rate. It can be seen that the spectral spread associated with the 3.3 mm channel interaction condition is very marked. It should also be noted that the effect of any given degree of channel interaction in a peak-picking strategy will be stimulus dependent, as with a wider band stimulus it is possible that the peaks will be wider apart, whereas for a narrow band stimulus the peaks will be closer together. Therefore, for a given degree of spectral spread, the consequences will differ according to the location and spacing of the peaks chosen in a particular frame. For a stimulus where peaks are selected in the same frequency region, a small amount of channel interaction (e.g. 1mm, which represents a filter bandwidth just over 1 electrode wide either side of the stimulation electrode) will cause a larger amount of channel overlap than for a stimulus which produces widely spaced peaks.

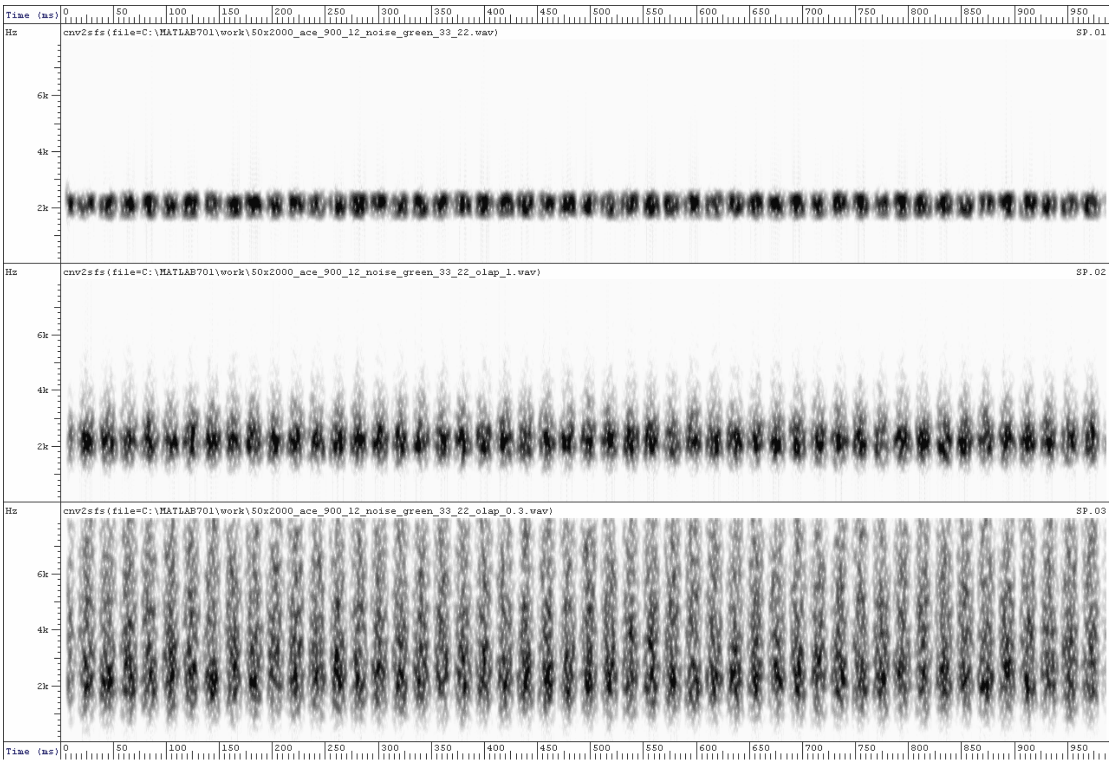

Figure 3.3. Wide-band spectrograms of AMs of a 2000 Hz pure tone modulated at 50 Hz with no channel interaction (top) with λ = 1 mm (middle) and 3.3 mm (below)

In all three AM experiments, carrier stimuli, either sine waves or noise bands, were added together and the RMS level of the resulting signal was adjusted to be equal to the original signal. Presentation level for the AM experiments was at a nominal level of 65 dB(A) as measured in a 2cc acoustic coupler, equivalent to approximately 60 dB(A) at the tympanic membrane. For the CI user experiment, stimuli were presented in the sound field at a level of 70 dB(A) as measured at the location of the subjects' speech processor microphone.

3.3.3 Stimulus presentation and calibration

For all experiments, the experimenter stayed in an observation booth while the subjects were in a sound-treated booth. For all experiments, subjects sat immediately in front of a touch screen. All equipment, e.g. PC, mixer and amplifier were in the observation room and were linked via wall plugs to the soundproof booth. The experimenter could see the subject through a one-way mirror and could also hear the subject via intercom. Stimulus generation was via a PC with a SoundBlaster sound card. The Praat speech analysis and software testing package (Boersma and Weenink, 2005) was used for stimulus presentation and response recording. For AM experiments 1,2 and 4, stimuli were routed through an INKEL MX-880E stereo mixer which delivered a mono signal to an insert earphone worn by the subjects. For the CI user experiment (no. 3), stimuli were routed through the stereo mixer and then through an INERN L140 amplifier which fed the amplified signal to a loudspeaker located in the soundproof booth.

The aim of calibration for the AM experiments was to ensure that the level of the sounds presented via the insert earphone was around 65 dB(A) at the tympanic membrane (TM) of the subjects. To do this, a real-ear-to-coupler difference (RECD) was determined for the first subject. Given the relatively small variation in level of RECD (averaged across frequency) across adults, this was used to determine the required coupler level that would give 65 dB(A) at the tympanic membrane. The first step was to create a sound file with the same mean RMS and power spectrum as the speech tokens. This was done by taking a ten second sample of the stationary speech-shaped noise and altering the level to equate to the average level of the sound files used for the experiment. This was defined as the calibration stimulus. The RECD for subject 1 was determined as follows: a probe microphone was placed in his right ear and the insert phone connected to the experimental rig was then inserted in the same ear. An AudioScan hearing aid measurement system was used as a sound level meter to measure the level at the eardrum by using the "manual" mode of operation of the test box, and setting the measurement scale to A-weighting. The output of the insert earphone was then measured in a 2 cc coupler, and output at the TM was found to be -

3.7 dB relative to the coupler output; consequently the RECD was –3.7 dB. Volume settings on the software were altered until the calibration stimulus was equal to 65 dB(A) at the TM and these settings of the software volume controls were noted. This allowed for a daily check at the beginning of each experiment to ensure that the stimulus delivered from the insert phone was at a level of 59.3 dB(A) +/- 0.5 dB in the coupler, equivalent to 65 dB(A) at the eardrum of the first subject. Given a typical head-related transfer function, the difference between the sound pressure level reaching the CI users' speech processor microphone and the sound pressure level reaching the eardrum of the NH listeners was estimated to be approximately 4 dB across frequencies 250-8000Hz, e.g. equivalent to approximately 61 dB(A) in the sound field.

Calibration for the CI user experiment was undertaken to ensure that the level of the stimuli presented in the sound field was 70 dB(A) at the microphone of the subjects' speech processors. A similar technique was applied, e.g. a proxy speech-shaped noise stimulus with the same RMS level was used for presentation in the sound field and volume controls were adjusted to ensure that the 70 dB(A) mean level for each stimulus presentation was maintained.

3.3.4 Subjects

For all three AM experiments NH listeners in the age range 18-35 years were used. Screening audiometry was undertaken to check that hearing levels were at 20 dB HL or better for all subjects. Otoscopy was also performed to check for any abnormalities of the external or middle ear. Details of CI subjects used in experiment 3 are given in 5.2.

3.3.5 Testing regime

For all experiments, subjects were seated in a double-walled soundproof booth. Stimuli in the AM experiments were presented monaurally to the subjects via an ER-3 insert earphone connected to a PC with a Sound Blaster sound card or via the sound field for the CI user experiment. Insert earphone presentation was used for AM experiments in order to minimise the effect of the pinna/outer ear transfer function on the stimuli. Monaural presentation was used as this mimics the normal listening condition for the majority of CI users. Stimuli were presented to subjects via routing of the signal into an adjacent soundproof booth. A touch screen was used for visual presentation of response options and to code subjects' responses. The ear to which the sounds were presented was alternated between subjects.

Testing was undertaken using the Praat (version 4.1) speech analysis and testing software, developed by Paul Boersma and David Weenink of the Institute of Phonetic Sciences at the University of Amsterdam (www.fon.hum.uva.nl/praat/). Stimulus presentation was controlled using scripts developed from the Praat (v 4.1) speech analysis toolkit. The software was designed to enable speech analysis but also enabled code to be generated to run and score speech perception experiments. The Praat code, or programme, randomised presentation of stimuli kept in the same source folder as the Praat script. The script also generated a graphical-user interface that was used on a touch screen to record subject responses. Each stimulus presentation required the subject to click on an icon on the screen before the next stimulus was presented. The same set-up was used in all subsequent experimental work reported in this thesis. The only difference between the three AM experiments and the CI user experiment was stimulus presentation and level: for the AM experiments, stimuli were presented by monaural insert earphone, whereas for the CI user experiment, stimuli were presented via sound field presentation as described in 5.

An important aspect of the experimental approach used was the nature of acclimatisation to stimuli. Davis (2004) noted that relatively brief familiarisation with noise-vocoded speech, e.g. of 20 minute or less, led to improved performance with sentence recognition. Some authors have noted that there is an intial "pop-out" effect of vocoded speech, e.g. when a stimulus is defined (e.g. the listener is exposed to the AM stimuli, then told what the word or speech sound is, the salience of the stimulus "pops out"). However, other authors, notably Rosen et al. (1999) have noted that considerable acclimatisation time is needed to achieve optimal performance for NH subjects listening to AMs. One of the questions for this study is whether sufficient acclimatisation would occur over a relatively short time period to yield valid results. The approach taken in the first experiment was to present all 20 stimuli on the touch-screen, each labelled (for example, /idi/ was labelled as "d") and allow the subject to listen to each stimulus as many times as s/he wished prior to testing proper. For experiments 2 and 4, there were a large number of AM conditions; consequently, it was impractical to allow this task prior to every listening condition. Therefore, the self-directed acclimatisation process took place for only one AM condition in quiet, and then one AM condition in noise, randomised across the subjects in each experiment. In practice, the self-directed acclimatisation process took between 5 and 10 minutes. A check of identification of four of the tokens was undertaken at this point to determine whether subjects had acclimatised sufficiently to AM stimuli A striking finding was that subjects stated that the stimuli were much clearer after only this short acclimatisation period. The results (see chapters 3 and 4) also indicated that this approach to AM acclimatisation yielded valid results.

For all 4 experiments, quiet listening conditions preceded noise-contaminated listening conditions, e.g. the design of the experiments was not randomised across the quiet vs. noise contrast. The rationale for adhering to this was the desire to provide further acclimatisation to the model when undertaking noise-contaminated listening conditions via testing in the quiet listening conditions (given the modest amount of acclimatisation time given in the first place). This meant that the effect of noise may have been diluted by consistent exposure to quiet AMs prior to noise-contaminated AMs and that, as a consequence, the possibility of a type II error (with respect to the noise variable) was increased. However, it also meant that the possibility of a type I error for the noise variable was minimised and, where significant effects of noise were obtained, these were more robust.

For experiment 1, stimuli were first presented in the "unaltered" condition, then "quiet AM", then "AM +10 dB SNR", then "AM+5 dB SNR" then "AM 0 dB SNR". For experiment 2, the 8 listening conditions (2 vowel environments * 2 pitch shift conditions * 2 carrier stimulus types-see 4.1.2 for further details) were randomised across subjects and in each case quiet then noise variant of the listening condition was presented. For the CI user experiment, the three MAP conditions were randomised across the 9 subjects, but again within each of these the quiet presentation was undertaken first, followed by the noise-contaminated condition . For experiment 4, testing was first undertaken in the "unaltered" condition. Following this, the three channel interaction conditions were randomised, then the three MAP conditions within each channel interaction, but, again, first the quiet then the noise-contaminated version of each listening condition were presented. Randomisation was achieved via coding of each listening condition and using a random number generator implemented in Microsoft Excel.

3.3.6 Analysis

For each subject/listening condition, a test run comprised randomised presentation of 3 instances of each of the 20 consonants. As noted below, each test run generated a series of responses which were coded as a confusion matrix. Subsequent data analysis could be divided into two main stages: first, the derivation of consonant feature transmission values and, second, descriptive and inferential statistical analyses. These stages are described below.

Before undertaking the first data analysis stage, it was necessary to decide on a set of features for information transfer analysis. As noted, consonant confusion data can be analysed with various levels of phonological detail: simple total correct values can be computed, as in the majority of studies using consonant recognition in CI users. Alternatively, a tripartite division into voicing, place and manner can be used; this is the approach that has been used in all more detailed studies, as in 2.1.2. At a greater level of detail, Chomsky and Halle (1968) described a large number of binary phonological features; it would be possible to use all of these features to analyse CI confusion data. However, such a detailed phonological analysis would be cumbersome when exploring a large number of independent variables and, moreover, it is important that data analysis methods have a clear rationale. It was clear that the three categories of voicing, place and manner needed to be included in phonological analysis, for the purposes of comparison with the existing literature and because of the fairly clear distinction in perceptual terms between these categories. However, there was also some justification for expanding on these three categories. As with some other studies, the "envelope" feature was included as this was based on perceptual abilities of CI users. It was hypothesised that this feature would be more robust than other features, and was arguably more purely "temporal" (e.g. effectively reliant on within-channel information) than the other features, e.g. even as compared with voicing (see the discussion in 2.6.4). It was also of interest to assess the perception of nasality, as this feature is similar to voicing in its reliance on low-frequency periodicity cues but distinctive in its reliance on low frequency (weak) formant structure. Finally, the fricative vs. non-fricative distinction was also used as a way of determining how well CI users can code noise information- given the long duration of the noise spectrum in fricatives the ability to resolve the noise in the time domain should not be a confounding variable. Moreover, this feature provided a larger reliance on spectral processing than other features apart from place.

Based on these choices, each confusion matrix yielded seven dependent variables: percentage total correct, and percentage information transmission for the consonant features voicing, place, manner, nasality, fricative and envelope. The following steps were taken to derive feature-specific information transmission values for all four experiments. Responses for each subject/test run generated by the Praat programme were tabulated and then converted to an Excel file. A macro transformed the data into a format usable for further analysis. Two further pieces of speech analysis software were used for consonant confusion analysis, namely FIX and SCORE, developed by the Department of Phonetics and Linguistics at University College London (www.phon.ucl.ac.uk/resource/software.html). The SCORE programme combined a defined stimulus and response data for each subject in each listening condition and generated a confusion matrix. Table 3.1 showed a typical confusion matrix in which stimuli are along the y-axis and responses indicated along the x-axis. For each confusion matrix, the FIX programme computed percent information transmission for the six features voicing, place, manner, fricative, nasality and envelope feature according to the feature transmission matrix in table 3.1 (although feature matrices are normally presented with features on the y-axis, the large number of stimuli necessitates the alternative presentation in this case). All percentage transmission values were computed from a single-iteration of SINFA analysis (see 2.1.1 for a discussion of this issue). Resulting total correct and information transmission values were entered into SPSS files. Subsequent data analysis was undertaken on the resulting feature transmission and total correct values.

	voicing	fricative	nasal	place	manner	envelope
b	yes	no	no	bil	plo	vpf
d	yes	no	no	alv	plo	vpf
g	yes	no	no	vel	plo	vpf
w	yes	no	no	bil	con	ng
j	yes	no	no	alv	con	ng
ɹ	yes	no	no	ret	con	ng
l	yes	no	no	alv	con	ng
v	yes	no	no	lad	fri	vpf
z	yes	yes	no	alv	fri	vpf
ʤ	yes	yes	no	ret	aff	vpf
m	yes	no	yes	bil	nas	ng
n	yes	no	yes	alv	nas	ng
p	no	no	no	bil	plo	vlp
t	no	no	no	alv	plo	vlp
k	no	no	no	vel	plo	vlp
f	no	yes	no	lad	fri	vlf
ɵ	no	yes	no	den	fri	vlf
s	no	yes	no	alv	fri	vlf
ʃ	no	yes	no	ret	fri	vlf
ʧ	no	yes	no	ret	aff	vlp

Table 3.1. Feature transmission matrix used for phonological feature analysis

The aim of inferential statistical analysis for each of the four experiments reported in subsequent chapters was to determine the effect of one or more independent variables, and their interactions, on the seven dependent measures derived from the confusion matrices. The independent variables were either categorical (as in experiment 2 or 3 and all variables apart from channel interaction in experiment 4) or had a small number of possible values (5 in experiment 1, 3 for channel interaction in experiment 4). Given that each listening condition generated a number of dependent variables, the appropriate statistical technique was considered to be multivariate analysis of variance (MANOVA).

For experiment 1, it was also important to test the hypothesis that voicing and manner would exceed place. This hypothesis required a direct comparison between feature transmission values, and therefore a single factor repeated measures ANOVA was used in this case, in which consonant feature was the only factor. However, this was the only use of this approach as the direct comparison between consonant features was considered of less importance than the assessment of the effect of the various independent variables on transmission of specific features. As the different feature transmission values represented multiple dependent variables, it was deemed appropriate to use MANOVA rather than a series of separate ANOVAs on each feature (it should be noted that the latter approach would have increased the possibility of type I errors). Because of the larger number of independent variables in experiments 2 and 4, the resulting number of degrees of freedom for many of the MANOVAs was low (1 or 2), which would also increase the possibility of type II errors. In general, the approach taken in various aspects of the design of the experimental work was to err on the side of minimising type I errors with a consequent increase in the possibility of increasing type II errors. This meant that the interpretation of statistically significant results could be more conclusive than might be the case otherwise.

ANOVA and MANOVA, as with other parametric statistical tests, are based on the assumption that data were normally distributed. For each experiment this was considered by applying the Kolgomornov-Smirnoff (K-S) test to each variable. The K-S test assesses the hypothesis that the distribution of a variable deviates significantly from the normal distribution. As noted in the relevant sections, the great majority of the variables in each of the 4 experiments were not found to be significant using this test, e.g. were consistent with a normal distribution. In a few cases (noted in relevant sections), distributions were skewed where the mean approached 100%, e.g. ceiling effects. However, it is considered that the F test used in MANOVA is robust to the problem of skewed distribution (Howell, 2003) (whereas it is not to outliers, a problem that did not occur) and, in any case, this only occurred for a small number of variables; consequently, it was assumed that MANOVA was appropriate from this point of view.

Chapter 4. Development of experimental methodology using fixed-channel AMs

Prior to undertaking experimental work comparing AMs and CI user data, it was important to determine whether an AM based on a set of device-specific processing characteristics could be used to determine consonant feature transmission. It was also important to determine what AM parameters were likely to affect performance and, in particular, how choice of these parameters would affect the correspondence between AM and CI user performance. It was also important to establish a time-efficient test methodology that could be applied to further experimental work involving both AMs and CI users where a larger number of variables would need to be compared. This meant that there was a need for preparatory experimental work with AMs of consonant recognition. This work, which comprised two experiments, is described in this chapter. Both experiments used 8-channel models of the Continuous Interleaved Sampling (CIS) processing strategy as implemented with the Nucleus 24 device. The first experiment applied a single model to determine patterns of consonant feature transmission with background stationary speech-shaped noise added at varying SNRs. The second experiment varied model parameters to determine their effects on consonant feature transmission.

4.1. Consonant recognition in quiet and at different SNRs with an 8-channel CIS model

4.1.1 Research questions, aims and hypotheses

The objectives of the experiment reported in this section were, first, to assess the effect of CI signal processing on consonant feature recognition by using an acoustic CI model based on a specific CI device, the Cochlear Nucleus 24M and, second, to determine the effect of noise at different SNRs on consonant feature recognition using the same model. Previous work in CI users, outlined in 2.1.2, established a consistent pattern of consonant feature identification in quiet, a pattern which suggested that temporal envelope cues necessary for speech are relatively better represented than spectral cues, at least in quiet. The little evidence that is available for consonant recognition in noise in CI users would also suggest that, conversely, noise interference may be correspondingly greater for temporal cues in speech. The experiment aimed to determine if this expected pattern of results could be replicated using an AM, although it should be noted that there is little evidence for transmission of features beyond voicing, place and manner, or for feature transmission in noise, against which to compare results.

Two specific hypotheses were tested in the experiment, although in broader terms the aim of the experiment was to enable further hypothesis formation with regard to further experimental work. First, it was hypothesised that information transmission for consonant features which rely primarily on temporal envelope resolution would be significantly better than equivalent transmission for features relying primarily on spectral cues. This hypothesis is based on the consistent observations described in 3.3 of better transmission of voicing and manner than place of articulation in CI users from a range of studies as shown in 2.1.2. On the basis of the six features discussed in 3.3.6, this should translate into the following pattern: nasality, envelope, voicing, and manner should be significantly greater than place of fricative. However, it might also be anticipated that the relative contribution of temporal envelope and spectral information within each feature will determine results. Given the likely contribution of spectral and temporal processing to the six features used, this would mean that the envelope feature would be transmitted best (as this is reliant almost entirely on temporal coding) whereas voicing and manner require some degree of spectral analysis and would be slightly less well transmitted. Moreover, fricative should be transmitted better than place given the greater importance of temporal cues to distinguishing fricatives from some other manner categories (whereas this is not the case for place transmission). It was also hypothesised that noise would have a significantly greater effect on those features relying on temporal information compared to those relying on spectral information. Friesen et al. (2001) found that noise had a greater effect on voicing than place transmission and hypothesised that the main mechanism of noise interference was the reduction in the salience of within-channel temporal fluctuations and hence an increased reliance on spectral cues.

Because the aim of this experiment was not to *compare* different processing or stimulus variables within the model, such as those identified in chapter 2, it was necessary to make a choice regarding these variables. It was decided to use a fixed channel rather than peak-picking strategy as it is simpler to interpret a fixed-channel model, e.g. in terms of number of channels information conveyed. Processing parameters were chosen to equate to the likely information transmission for a "good performer" using this device as it had been shown from the existing literature that a typical good performer's speech perception performance is approximated best by an AM of around 8 channels (see 2.3.2). Consequently, the model coded 8 channels of information in the Nucleus CIS fixed-channel processing strategy. A stimulation rate of 500 Hz was chosen (this meant that envelope bandwidth of the carrier stimuli was equivalent to that derived from the output of the FFT filterbank undertaking 500 FFT analyses per second, precisely as would occur in the real Nucleus 24 processor-see 2.3.3. for a discussion of the temporal characteristics of the Nucleus 24 filterbank). A sine wave carrier was chosen (the possible perceptual implications of choosing a sine wave carrier over a noise band carrier are discussed in 2.5.2). In this study the /i/ vowel environment for medial consonants was chosen.

The overall aims of the first experiment were therefore to:

- Use an AM of CI processing to represent the information available to a good CI performer using a specific device (Nucleus CI 24) and processing strategy (CIS, 8 channels with 500 pps/ch stimulation rate).
- Apply the model to evaluate consonant recognition in quiet and steady background noise at three different SNRs using information transmission analysis.
- Interpret the pattern of feature transmission in quiet and noise in the context of the feature-specific hypotheses formulated above.
- Determine, on the basis of results, which SNR should be used in subsequent experimental work (given the need to minimise the number of different SNR conditions with further work in which a number of other variables were to be compared)

- Assess the ability of a specific experimental approach (e.g. amount of familiarisation with stimuli, number of repetitions of stimulus) to answer research questions.

4.1.2 Pilot study and overall test methodology

Prior to the main experiment, a small pilot study with five normal hearing subjects was undertaken in order to determine, in a qualitative way, the following points: (a) the amount of acclimatisation required for the listeners to become adequately used to the stimuli, (b) the number of repetitions required to achieve stable results, (c) the amount of time taken to undertake the experiment. In addition, it enabled a determination of the likely effect size for sample size calculation. The stimulus processing and test methodology were identical to those for experiment 1 as described in 3.3. Figure 4.1 shows the mean results for the three features voicing, place and manner in the three noise conditions:

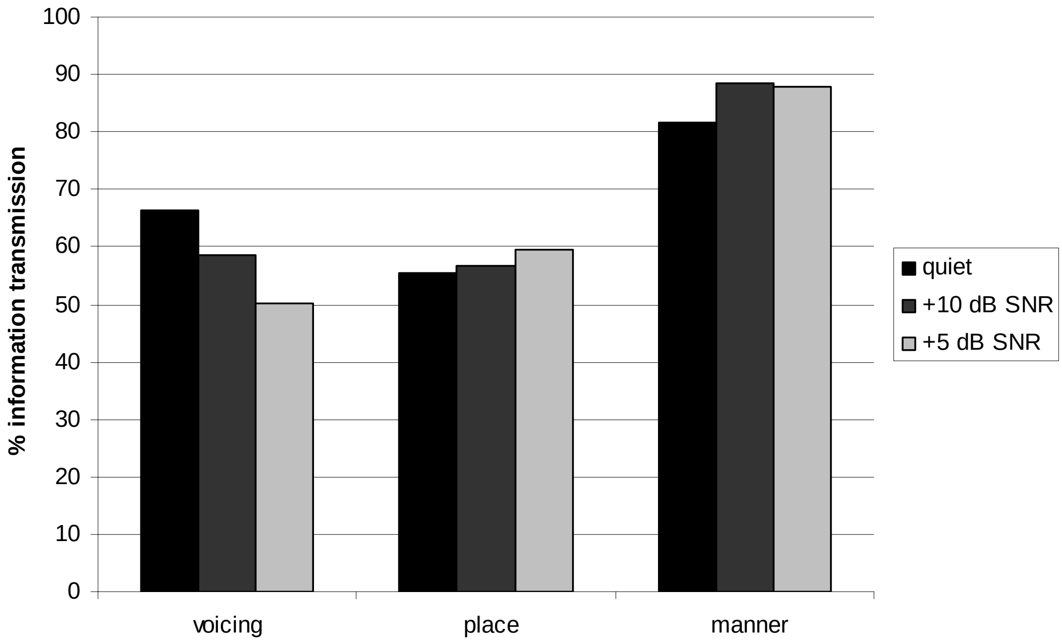

Figure 4.1. Mean feature transmission across three listening conditions from the pilot study. Data were obtained from 5 listeners using an 8-channel CIS AM

These findings suggested the following: first, the experiment should not yield floor or ceiling effects. Second, it was worthwhile to attempt three different SNRs, including one not included in the pilot study, given the absence of noise effects for place and manner down to +5 dB SNR (0, +5 and+10). Third, the approach to AM acclimatisation used in the pilot study, whereby subjects were able, prior to testing proper, to listen to each token as many times as they wished (with the identity of the token made evident) appeared to yield sensible results. Four, a testing regime with just three repetitions of each stimulus should give valid results (the need to minimise presentations was motivated by the possibility of further experimental work in which a large number of different variables, and hence listening conditions, were to compared, and fatigue effects also needed to be minimised). It was anticipated that total testing time for the main experiment would be around 80 minutes, which would be feasible without substantial fatigue effects and with some breaks provided. A further outcome of the pilot study was the availability of relevant data for sample size calculation.

For the main experiment 1, a VCV consonant recognition task was undertaken in each of five listening conditions: unaltered (e.g. original, unprocessed) stimuli and stimuli processed through an AM with four noise conditions (quiet, with background noise at +10 dB SNR, +5 dB SNR and 0 dB SNR). Further details of test methodology are described in 3.3.

4.1.3 Subjects and sample size calculation

The sample size calculation used to determine subject numbers was based on the difference in performance with voicing in quiet vs. the +10 dB SNR condition obtained from the pilot study. Based on the mean, standard deviation and cross-correlation between these two variables in the pilot study, the required sample size to obtain a power of 80% was 19. Consequently, 19 normal-hearing subjects (12 females, 7 males, mean age 25 years) were recruited to the study following local safety and ethics committee approval. Subject inclusion criteria were: age 18-35; thresholds better than 20 dB HL across the octave frequencies 250 Hz- 8000 Hz; English as native language; willingness to participate. It should be noted that the same inclusion criteria for NH subjects applied to all other AM experiments reported in this dissertation. Subjects were not paid to participate and had provide fully informed consent based on the safety and ethics application in order to participate.

4.1.4 Results

For each of the five listening conditions (unaltered, AM in quiet, AM in noise at +10 dB SNR, AM in noise at +5 dB SNR and AM in noise at 0 dB SNR) information transmission values were derived for the six consonant features outlined in 3.3 and for total percent correct. This yielded 7 measures * 5 listening conditions, therefore 35 dependent measures in total for each of the 19 subjects. The 7 measures are shown in Figure 4.2 across listening conditions (It should be noted that, in the case of the "total correct measure", the variable is simple percentage rather than percent information transmission). Where mean + standard deviation is greater than 100%, a 120% scale is used (the same applies across all subsequent figures in this and remaining chapters).

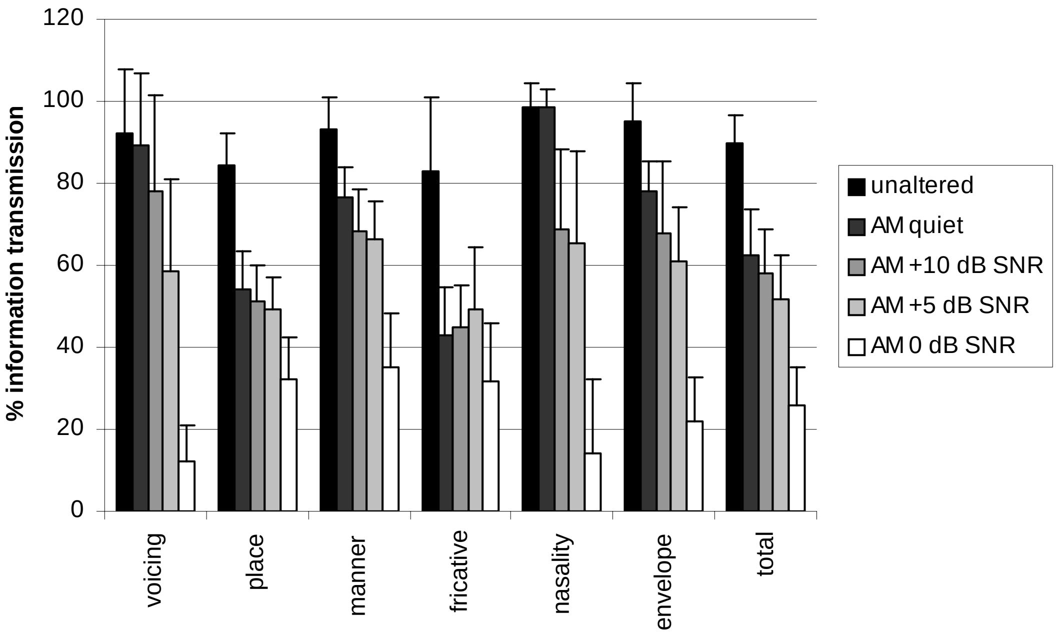

Figure 4.2. Mean (+1 SD) feature transmission, in addition to total percentage correct, as a function of listening condition.

Of the 35 resulting variables, all but two were not found to be significant at the 5% level using the Kolmogornov-Smirnoff test. Consequently, parametric statistical tests were appropriate for inferential statistical analysis. A multivariate analysis of variance (MANOVA) was undertaken in which listening condition (with five levels as noted above) was the only factor and the six feature transmission values and total percentage correct were the 7 dependent variables. The analysis showed a significant effect of listening condition on all six features and on total correct ($p<0.001$ in all cases). Further details of the MANOVA are given in Appendix B. *Post-hoc t*-tests were undertaken for each feature for each of a possible ten comparisons between the five listening conditions. Table 4.1 shows mean differences between listening conditions for each of the seven dependent variables and indicates which *t*-test comparisons were significant at the 0.005 level (0.05/10, the number of comparisons made) via bold script. For clarity of interpretation, comparisons between the unaltered condition and noise-contaminated AM conditions are not included in the table as these are not particularly meaningful in the context of the research (in any case, all of these were significant at the Bonferroni-adjusted significance level for all dependent variables). Values given in the table are mean differences for the comparisons indicated on the left of the table.

Table 4.1. Mean differences in feature transmission (or total correct) between listening conditions for the seven dependent variables. Differences reaching the Bonferroni-corrected significance level ($p<0.005$) are given in bold.

Dependent measure/ Mean difference	Total correct	Voicing transmission	Place transmission	Manner transmission	Fricative transmission	Nasality transmission	Envelope transmission
Unaltered - quiet AM	**27**	3	**30**	**17**	**40**	0	**17**
Quiet AM - AM at +10 dB SNR	**5**	**11**	3	**8**	-2	**30**	**10**
Quiet AM - AM at +5 dB SNR	**11**	**31**	5	**10**	-7	**33**	**17**
Quiet AM - AM at 0 dB SNR	**37**	**77**	**22**	**42**	**11**	**84**	**56**
AM at +10 dB SNR - AM at +5 dB SNR	6	**20**	2	2	-5	4	7
AM at +10 dB SNR - AM at 0 dB SNR	**32**	**66**	**19**	**33**	**13**	**55**	**46**
AM at +5 dB SNR - AM at 0 dB SNR	**26**	**46**	**17**	**31**	**18**	**51**	**39**

Results of *post-hoc t*-tests can be summarised briefly as follows. As noted (but not given in the table) all comparisons between unaltered and noise-contaminated AM conditions were significant. For voicing transmission, there was no significant difference between unaltered stimuli and AMs in quiet whereas all other possible comparisons were significant. However, for both place and fricative perception there was a significant difference between the unaltered and quiet AM conditions but no significant differences between AM conditions if the 0 dB SNR condition is excluded, although all comparisons between the 0 dB SNR and other conditions were significant. For manner and envelope, all comparisons were significant except for the comparison between the +5 and +10 dB SNR AM conditions. For nasality, the comparison between the unaltered and quiet AM condition was not significant, and all comparisons between AM conditions were significant except for the comparison between the +5 and +10 dB SNR AM conditions.

An additional consideration was the need to compare performance across features directly. This was needed to test the hypothesis that feature transmission in quiet would depend on the degree of importance of spectral resolution in coding the feature. In order to facilitate interpretation, the data for the six feature transmission value given in figure 4.2 are repeated in figure 4.3. It appears from this figure that nasality transmission was greatest, followed by voicing, followed by manner and envelope, which are approximately equal. Fricative transmission was lowest with place slightly higher, but both of these feature transmission values were markedly lower than the other four features. In order to determine differences statistically, a repeated measures ANOVA was undertaken in which there was a single factor of feature with six levels (the six consonant features) using the results of the quiet AM condition only. This factor was highly significant ($p<0.001$) and full ANOVA results are given in Appendix B. Table 4.2 lists the mean differences between each possible pair of features. Differences which were statistically significant (*post-hoc t*-test assuming the Bonferroni corrected significance level of 0.003) are indicated in bold.

Table 4.2. Comparison between feature transmission values in the "quiet AM" listening condition. Values given are percent transmission for features indicated on the left – percent transmission for features indicated on the top. Values in bold indicate that the post-hoc comparison was statistically significant (p<0.003).

Feature	voicing	place	manner	fricative	nasality	envelope
voicing		**22**	**13**	**46**	-9	**11**
place			**-22**	**11**	**-44**	**-24**
manner				**34**	**-21**	-2
fricative					**-56**	**-35**
nasality						**20**

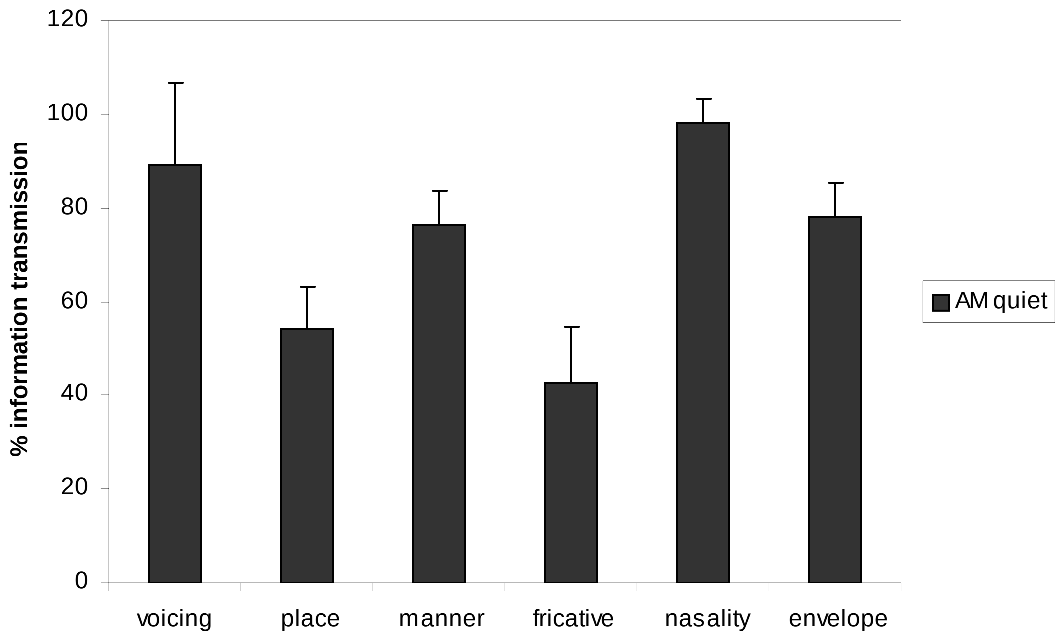

Figure 4.3. Mean (+1 SD) feature transmission by feature for the "quiet AM" listening condition only.

The pattern of *post-hoc* comparisons supports the impression from figure 4.3: for the quiet AM listening condition, nasality transmission was significantly greater than all other features except voicing. In turn, voicing transmission was greater than manner, envelope, fricative or place. Place of articulation was significantly worse than manner, nasality, envelope and voicing but significantly better than fricative transmission. Manner was significantly worse than voicing or nasality but significantly better than the remaining features. Fricative was significantly worse than all other features. Envelope was not different from manner. Envelope transmission was significantly worse than nasality but significantly better than fricative and place. In order to illustrate the error patterns directly, table 4.3 shows the confusion matrix for the quiet

AM condition. Error rates are converted to a percentage score for ease of interpretation.

Table 4.3. The consonant confusion matrix for the quiet AM condition. Data were obtained from 57 total presentations (3 presentations x 19 subjects) per stimulus but are presented as percentage responses e.g. for the stimulus/g/ the response /g/ is given 37% of the time.

	b	d	g	w	j	ɹ	l	v	z	ʤ	m	n	p	t	k	f	θ	s	ʃ	ʧ
b	91	2	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0
d	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
g	0	51	37	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	5
w	0	0	0	74	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0
j	0	0	0	4	77	2	11	5	0	0	2	0	0	0	0	0	0	0	0	0
ɹ	0	0	0	14	0	86	0	0	0	0	0	0	0	0	0	0	0	0	0	0
l	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
v	0	0	0	0	0	5	0	95	0	0	0	0	0	0	0	0	0	0	0	0
z	0	0	0	0	0	0	0	7	91	0	0	0	0	0	0	0	0	2	0	0
ʤ	0	0	28	0	0	0	0	0	0	67	0	0	0	0	0	0	0	0	0	5
m	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	68	32	0	0	0	0	0	0	0	0
p	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0
t	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	23	18	56	0	0
k	0	0	2	0	0	0	0	0	0	0	0	0	0	56	37	0	0	0	0	5
f	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	72	21	4	0	0
θ	0	0	2	0	0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0
s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	2	5	88	2
ʃ	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	4	5	89	0	0
ʧ	0	0	9	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	88

A detailed analysis of specific confusion patterns is left for the experimental work reported in chapter 4. However, a notable confusion that occurs in table 4.3 is the confusion of /ʤ/ for /g/. This confusion would have an effect on fricative, place and manner transmission. A possible methodological reason for this confusion was the confusion in orthographic representation of the phonemes concerned as the letter "g" is pronounced /ʤi/ in English. Examination of individual confusion matrices showed that a subset of the subjects consistently made this confusion (e.g. across the three repetitions in most cases) but the remaining did not make the confusion at all. Moreover, a similar error pattern was found in the unprocessed speech. These findings support the possibility that the confusion arose from not adequately instructing subjects to ensure that the sound /ʤ/ was given as the letter "j" on the screen, and not the letter "g". Instructions for all three further experiments were modified in order to reduce this possibility further.

4.1.5 Discussion

The main aim of the experiment was to assess the effect of CI signal processing on consonant feature recognition by using an acoustic CI model based on the Nucleus 24 processing. In order to determine the likely validity of the model, it was important to establish if the pattern of feature transmission in quiet was broadly in line with what is known of CI user performance, and also whether the patterns of noise effects on feature transmission were in line with experimental hypotheses. It was hypothesised that information transmission for consonant features relying more on temporal envelope resolution (manner, voicing, envelope, nasal) would be significantly better than equivalent transmission for features relying primarily on spectral cues (place and fricative), in quiet and therefore that an AM which represents the information content of CI processing accurately would yield this expected pattern of feature transmission. This was clearly supported by the findings of the AM in the quiet listening condition: place and fricative transmission were much worse than manner, nasality, envelope and voicing.

The pattern across studies of CI users' relative feature transmission, cited in 2.1.2, namely worse place transmission than manner/voicing holds very well in the present AM. However, it is worth noting that in the present experiment voicing transmission was around 90% both in the unaltered condition and with the AM in quiet. This exceeds manner transmission and appears to be inconsistent with the studies cited in 2.1.2. The likely explanation is that the sine wave carriers over-estimated the amount of voicing information available to CI users. Gonzalez and Oliver (2005b) and Dorman et al. (1998) suggested that random envelope fluctuations in the noise band carrier would lead to reduced performance compared to sine wave carriers and therefore that voicing would be one of the features where carrier stimulus might be a crucial factor. This underlines the importance of the choice of carrier stimulus, an issue addressed in the subsequent experiment reported in 4.2. Similarly, nasality transmission approached 100% in the AM in quiet and was not different from transmission in the unaltered condition. The discussion in 2.6.2 suggesting that cues to this feature could be compromised by CI processing and therefore the high (indeed normal) levels of feature transmission could be anomalous. Again, the choice of carrier stimulus may be crucial in explaining this result and needs to be evaluated in a further experiment.

As noted, there is very little evidence as to the specific effects of noise on consonant feature transmission in CI users or using AM studies. It was hypothesised that noise would have a significantly greater effect on those features which had a stronger reliance on temporal/envelope resolution rather than spectral resolution. The pattern of *t*-test results shown in table 4.2 supports this hypothesis, at least if the 0 dB SNR listening condition is excluded from the analysis. Comparisons between the quiet AM and AMs with noise at +5 or +10dB SNR show that noise had a significant effect on voicing, nasality, envelope and manner but not place or fricative. The question then arises as to what is the mechanism of the noise effect, given the pattern of results obtained above. It has been hypothesised that random fluctuations in the envelope of the noise reduces the salience of envelope fluctuations (here sine wave modulations) and consequently would have the greatest effect on temporal speech cues because of the importance of within-channel information to the latter. However, it is also worth noting that the two features with the smallest noise effects, e.g. place and fricative, achieved relatively poor performance in the quiet condition and therefore the relative lack of noise effect on these features could be interpreted as being due to floor effects (that is, if something is already bad, it is harder for it to get worse). Given the strong possibility that the absence of place and fricative noise effects was due to floor effects, further evidence would be needed to corroborate the hypothesis that noise interference relates more to within-channel processing.

Given that the main rationale for the experimental work reported in this chapter was "methodological", e.g. in preparation for the work reported in chapter 4, it is of particular interest to consider the data in order to determine what might constitute the most "sensitive" SNR or SNRs for further experimental work. The pattern of feature transmission as a function of noise shows quite clearly that performance at +5 and +10dB SNR was broadly similar, whereas performance at 0dB SNR was notably different in that almost all features showed noise effects. This would suggest that 0dB SNR yields floor effects when using the AM and therefore either +5 or +10dB SNR should be chosen for further experimental work.

It is also appropriate to consider the way in which testing was undertaken and whether it represents a valid method to assess consonant recognition. There are two particular issues of interest: first, the nature of acclimatisation used prior to testing proper; second, the number of stimulus tokens presented. The approach taken here was to allow the listener to familiarise him/herself with stimuli as much as s/he wanted prior to testing proper. However, the number of tokens of each presentation was relatively small (3 per stimulus). It seems likely, given the rich and varied pattern of phonological feature transmission which was broadly consistent with experimental hypotheses that this approach is valid and can be applied to further experimental work, where it is even more critical to trade off number of presentations with the larger number of variables and corresponding listening conditions to be tested.

In summary, the experiment showed that use of an AM based on the Nucleus 24 device with a fixed-channel strategy could be used to estimate consonant feature transmission and that this could be undertaken with a small number of repetitions per stimulus and with self-directed acclimatisation process. In general, the feature transmission patterns reflected what would be expected of CI users within the constraints of the available evidence base. However, some notable anomalies, emerged, particularly the high rate of voicing and nasality transmission. It was possible that the sine wave carrier over-estimated voicing and nasality transmission in quiet. It was also possible that the model over-estimated performance because it failed to incorporate aspects of the electrical/neural interface, or because a different vowel environment was used to the majority of other studies against which data were compared. It was therefore necessary to further explore the possible effect of specific methodological choices on AM performance before proceeding to the main experimental work reported in chapter 5.

4.2 A comparison between AM parameters

4.2.1 Research questions, aims and hypotheses

The experiment reported in the previous section showed that a model of the Nucleus 24 CIS strategy yielded, broadly, the expected pattern of feature transmission, although it appeared that voicing transmission in particular may have been over-estimated in the model. Results also indicated that the addition of stationary background noise at either +5 or +10 dB SNR could be used to determine noise effects for subsequent AM experiments. However, the experiment was based on a particular choice of simulation and stimulus parameters (sine wave carrier, no attempt to mimic any of the distortions associated with the electrical/neural interface, and choice of vowel environment) and it was not clear whether these particular methodological choices would be important in determining the validity of the model, e.g. how well the model results would predict CI user performance. Because the main experimental work planned in the thesis (reported in chapter 5) aimed to compare AM and CI user data directly, it was important to establish which of these AM or stimulus parameters would affect performance.

The experimental findings raised the possibility that the sine wave could be an inappropriate choice of carrier stimulus as the model may have over-estimated transmission of some consonant features. Moreover, the nature of the carrier stimuli meant that the model made no attempt to mimic any of the distortions associated with the electrical/neural interface discussed in 2.4. The question of channel interaction was of particular interest for subsequent experimental work, as it was hypothesised that the inclusion of spectral channel interaction in the model would improve the fit between model and CI user performance. However, before this issue was considered (in the AM experiment reported in chapter 4) it was important to determine whether the subjective pitch shift associated with CI insertion, discussed in 2.4.2, would have a bearing on consonant feature transmission. Finally, it was also of interest to determine the likely effect of vowel environment choice, as the experiment reported in section 4.1 used a different vowel environment (/iCi/) for the consonant confusion task to that used in the great majority of other studies of this nature, for reasons outlined in 3.3.1.

The aim of the experiment reported in this section was to therefore determine whether any of these variables had an effect on consonant feature transmission (in quiet and one SNR) in the AM. The research questions for experiment were:

- Does choice of carrier stimulus (noise band vs. sine wave) have a significant effect on AM results?
- Does the inclusion of "Greenwood pitch shift" have a significant effect on AM results?
- Does the choice of vowel environment have a significant effect on AM results?

Based on Gonzalez and Oliver (2005) it was hypothesised that the noise band carrier would lead to worse transmission across all features. There is some evidence about the effect of the Greenwood shift on place transmission (Rosen et al., 1997; Dorman et al., 1999); consequently, it was also hypothesised that this would adversely affect performance across features, particularly those more reliant on spectral processing, e.g. place and fricative. As for the previous experiment, it was hypothesised that noise would have a greater effect on voicing, manner, nasality and envelope than place or fricative. It was also hypothesised that feature transmission values would be less for /iCi/ stimuli than /aCa/ stimuli, particularly for place.

4.2.2 Differences in methods from experiment 1

In order to address the questions raised in 4.2.1, the second experiment was set up in such a way as to allow comparison of the noise vs. quiet, noise band vs. sine wave carrier, analysis filter vs. Greenwood filter and /aCa/ vs/ /iCi/ vowel environments for the VCV test. The same 20-alternative VCV consonant recognition test as undertaken in experiment reported in 4.1 was presented in quiet and in background speech-shaped noise, in each of the two vowel environments /iCi/ and /aCa/ at +5 dB SNR. Stimuli were processed through an AM with four different configurations: using a sine wave carrier with no pitch shift, a noise band carrier with no pitch shift, a sine wave carrier using the Greenwood pitch-mismatch formula and a noise band carrier using the Greenwood pitch-mismatch formula. Therefore, in total, each of the two versions of the VCV test (iCi or aCa vowel environment) was presented in 8 listening conditions, e.g. 2 noise conditions (quiet or background noise) * 4 AM conditions, yielding a total number of 16 listening conditions. As with other experiments, each listening condition yielded a confusion matrix from which 7 dependent variables were derived. The sample size calculation was based on data from experiment 1, specifically the difference between voicing transmission in quiet and noise at +5dB SNR. This yielded a required sample size of 5 to achieve 80% power. Consequently, 5 normal hearing subjects were recruited to the study.

4.3.3 Results

Information transmission analysis was undertaken for each listening condition as described in 2.3.6. Figures 4.4 to 4.9 show information transmission across listening conditions for each of the six features while figure 4.10 shows total correct across listening conditions.

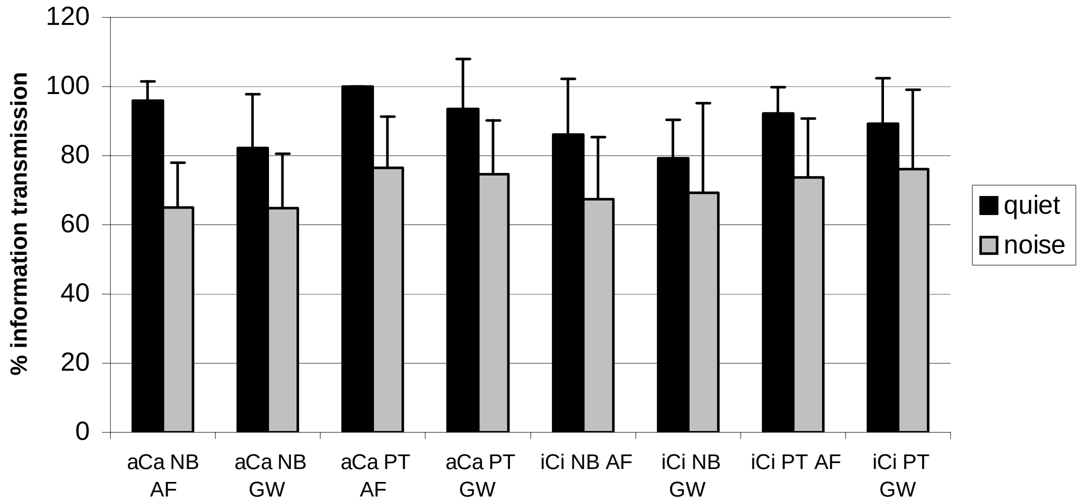

Figure 4.4. Mean (+1 SD) voicing transmission across listening conditions, in quiet and with background stationary noise at +5 dB. Data label key: aCa vs. iCi refers to vowel environment; NB = noise band carrier; PT= sine wave carrier; AF = analysis frequencies (no pitch shift); GW = Greenwood pitch shift. Error bars indicate +1 standard deviation.

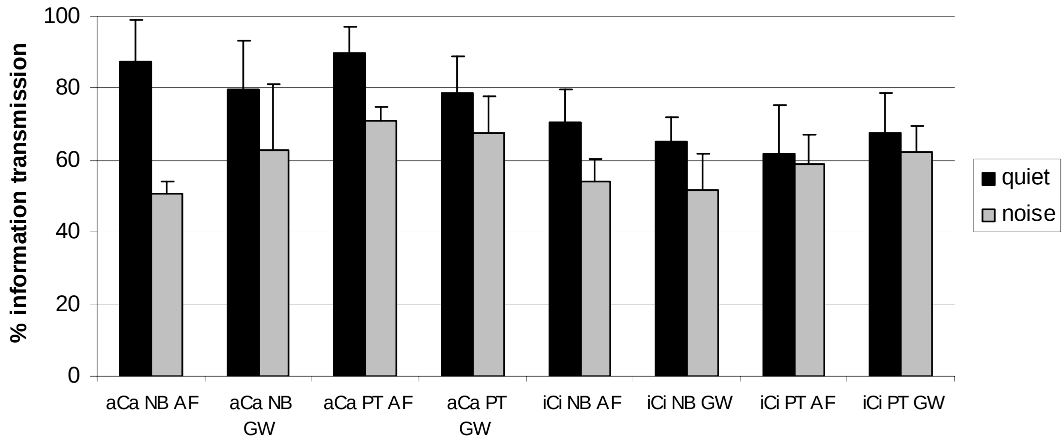

Figure 4.5. Mean (+1 SD) place transmission across listening conditions, in quiet and with background stationary noise at +5 dB. See figure 4.4 for key to data labels.

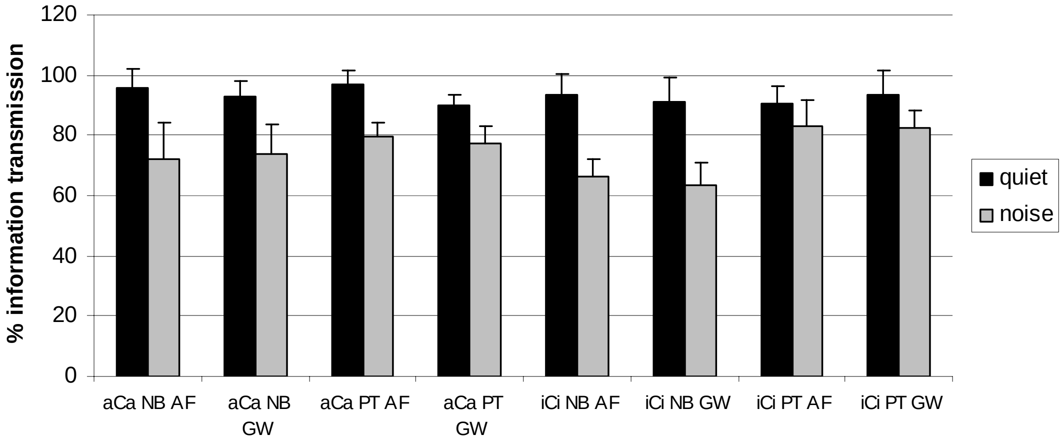

Figure 4.6. Mean (+ 1 SD) manner transmission across listening conditions, in quiet and with background stationary noise at +5 dB. See figure 4.4 for key to data labels.

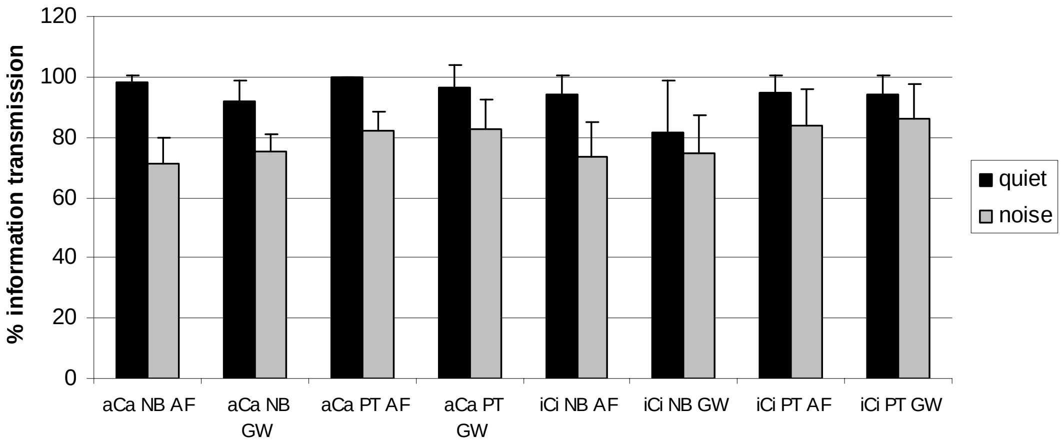

Figure 4.7. Mean (+ 1 SD) fricative transmission across listening conditions, in quiet and with background stationary noise at +5 dB. See figure 4.4 for key to data labels.

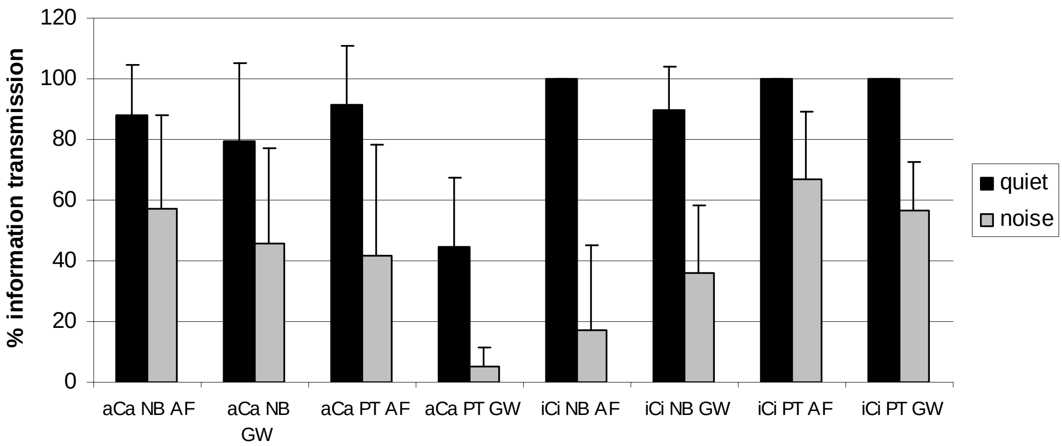

Figure 4.8. Mean (+ 1 SD) nasality transmission across listening conditions, in quiet and with background stationary noise at +5 dB. See figure 4.4 for key to data labels.

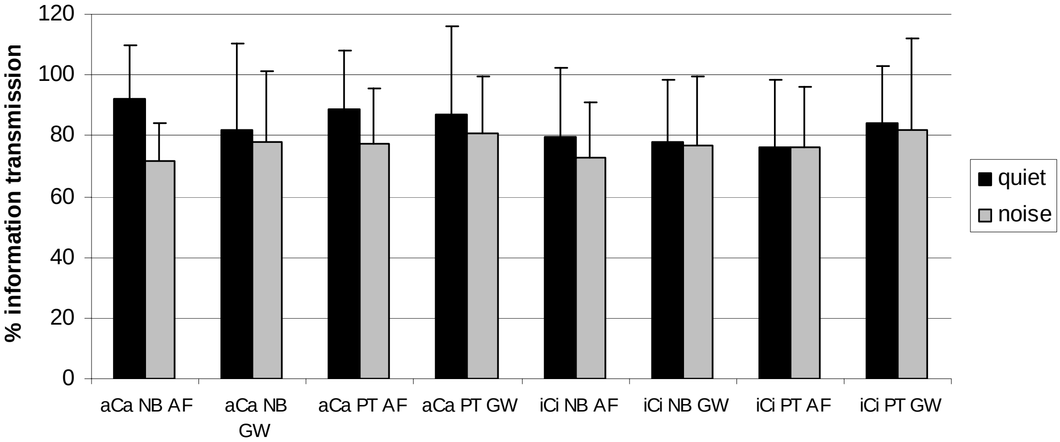

Figure 4.9. Mean (+ 1 SD) envelope transmission across listening conditions, in quiet and with background stationary noise at +5 dB. See figure 4.4 for key to data labels.

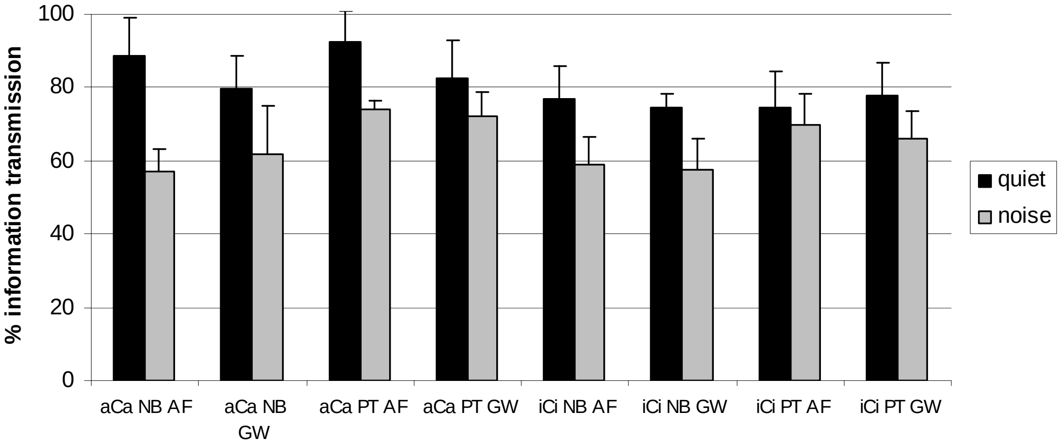

Figure 4.10. Mean (+ 1 SD) total correct across listening conditions, in quiet and with background stationary noise at +5 dB. See figure 4.4 for key to data labels.

Each of the 7*16 (112) variables was checked for normality of distribution using the Kolmogornov-Smirnoff (K-S) test. Of the 112 variables, only 5 were found to be significantly different (at 5% level) from the normal distribution. These were each cases of 100% mean scores. In each case, paired-sample K-S test comparisons with over variables were found to be non-significant using the K-S test. A MANOVA was therefore undertaken with seven dependent variables (each of the six consonant feature and total correct) and four binary categorical predictor variables: noise (presence/absence at +5 dB SNR), Greenwood pitch shift (presence/absence), carrier stimulus (narrow band noise vs. sine wave) and vowel environment (aCa vs. iCi). For each dependent variable this yielded F and significance values for each of the four factors and 11 possible interactions. For the sake of brevity, only factors and interactions which were significant at the 0.05 level are reported here (full MANOVA results are given in Appendix B).

The carrier stimulus factor had a significant effect on voicing, place, manner and fricative information transmission and on total correct. In all cases, better performance was obtained with the sine wave carrier as compared to the noise band carrier. However, the pitch shift factor only had a significant effect on nasality transmission but was not a significant predictor of variance for any other feature or total correct. Vowel environment had a significant effect on place and nasality transmission and also on total correct. The direction of this effect differed between the three dependent variables: nasality was transmitted better within the iCi vowel environment but place transmission and total correct were better with the aCa vowel environment. Noise had a significant effect on total correct and for all the consonant features except envelope The direction of the effect was as anticipated, i.e. the addition of stationary noise at +5 dB SNR was associated with lower transmission scores.

There were no significant factor interactions for voicing, envelope or fricative. For total correct there was a significant interaction between carrier stimulus and noise. For place transmission there were significant two-way interactions between carrier and noise and between vowel and noise. For manner transmission there was a significant two-way interaction between carrier and vowel. For nasality, there were two-way interactions between carrier and pitch shift, between carrier and vowel and between pitch shift and vowel. For nasality there was also a significant three-way interaction between carrier, vowel and noise.

The interaction between carrier and noise for total correct and for place of articulation transmission can be explained in the same way. Figures 4.11 and 4.12 show data averaged into four categories of 2 carrier types * 2 noise conditions, for total correct and place of articulation, respectively. For both measures there was a smaller difference between quiet and noise conditions for the sine wave carrier conditions compared to the noise band conditions. This is shown by the pattern of *post-hoc t*-tests given in tables 4.4 and 4.5: there was a significant difference between carriers in noise but not in quiet. It can also be seen that the effect of introducing noise was less for the sine wave carrier.

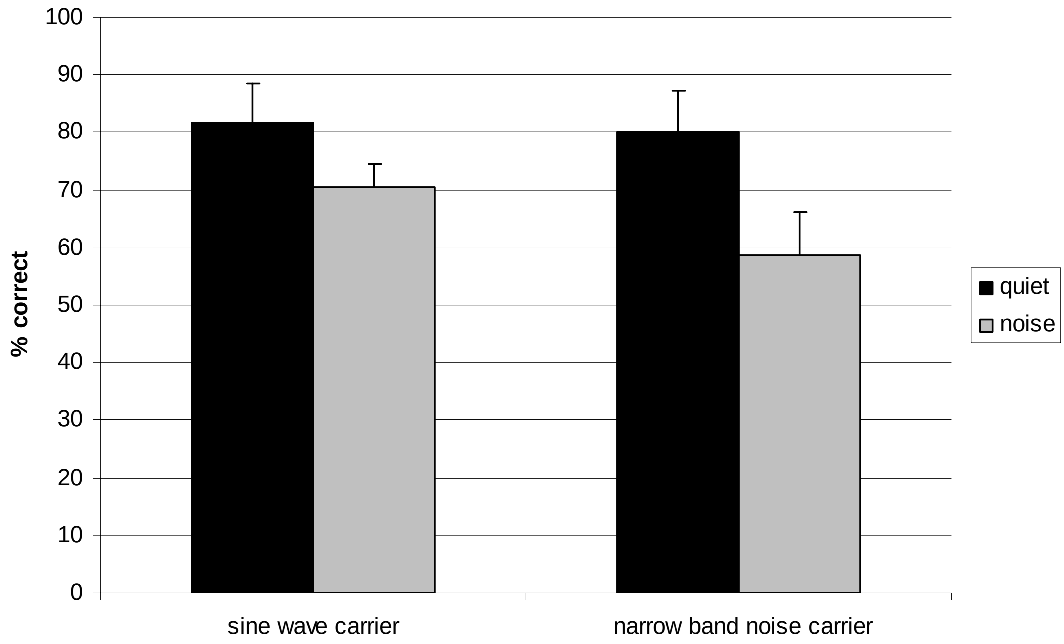

Figure 4.11. Mean (+ 1 SD) total correct by carrier type and noise condition.

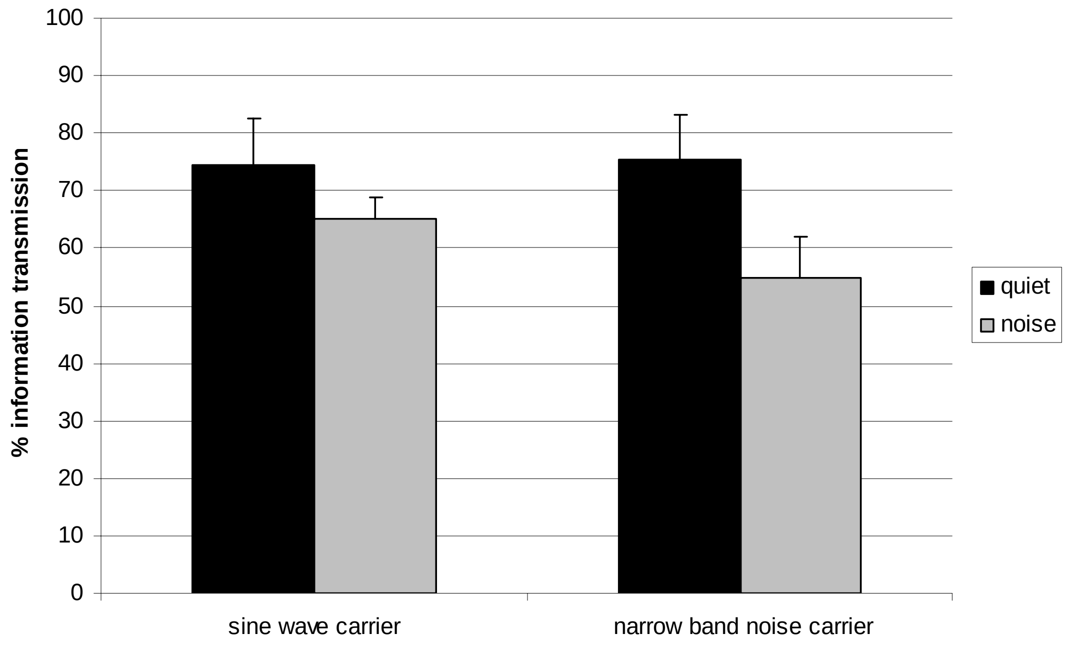

Figure 4.12. Mean (+ 1 SD) place transmission by carrier type and noise condition.

Table 4.4. ***Post-hoc*** **t-tests comparing total percentage correct in different listening conditions in order to explore the interaction between carrier stimulus and noise.**

Comparison conditions	Mean difference	T value	Significance level
Narrow band, quiet Sine wave, quiet	-1.56	-0.72	0.51
Narrow band, noise Sine wave, noise	-11.74	-6.61	0.00
Narrow band, quiet Narrow band, noise	21.19	19.25	0.00
Sine wave, quiet Sine wave, noise	11.01	5.16	0.01

Table 4.5. ***Post-hoc*** **t-tests comparing place of articulation transmission in different listening conditions in order to explore the interaction between carrier stimulus and noise.**

Comparison conditions	Mean difference	T value	Significance level
Narrow band, quiet Sine wave, quiet	-0.81	-0.27	0.80
Narrow band, noise Sine wave, noise	10.30	4.24	0.01
Narrow band, quiet Narrow band, noise	9.40	2.59	0.06
Sine wave, quiet Sine wave, noise	20.51	11.98	0.00

The interaction between noise and vowel type for place transmission is illustrated in figure 4.13 and can be explained by the larger effect of noise in the /aCa/ vowel environment compared to /iCi/ or to the smaller difference between vowel environments for noise compared to quiet. Another way of looking at the same data is to say that quiet performance was better for /aCa/ than /iCi/ whereas the difference between the vowel environments was much less for performance in noise.

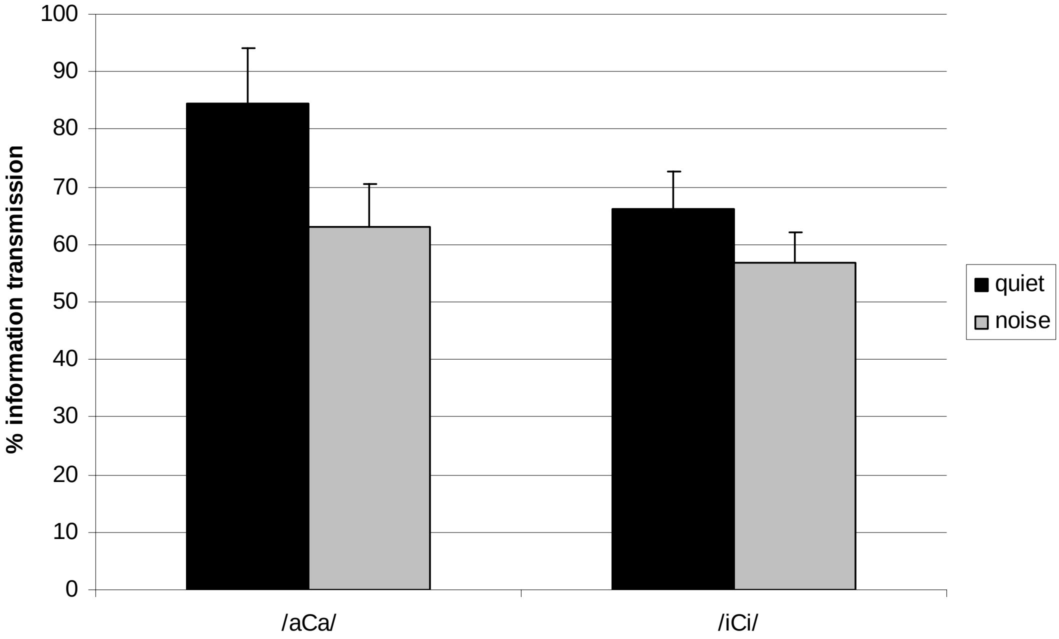

Figure 4.13. Mean (+ 1 SD) place transmission by vowel environment and noise condition.

Table 4.6. *Post-hoc* t-tests comparing place of articulation transmission in different listening conditions in order to explore the interaction between vowel environment and noise.

Comparison conditions	Mean difference	T value	Significance level
/aCa/, quiet /iCi/, quiet	18.15	4.98	0.01
/aCa/, noise /iCi/, noise	6.32	1.75	0.16
/aCa/, quiet /aCa/, noise	21.26	4.82	0.01
/iCi/, quiet /iCi/, noise	9.44	3.99	0.02

The interaction between carrier and vowel environment for manner is illustrated in figure 4.14. The difference between carriers is greater for /iCi/ than /aCa/.

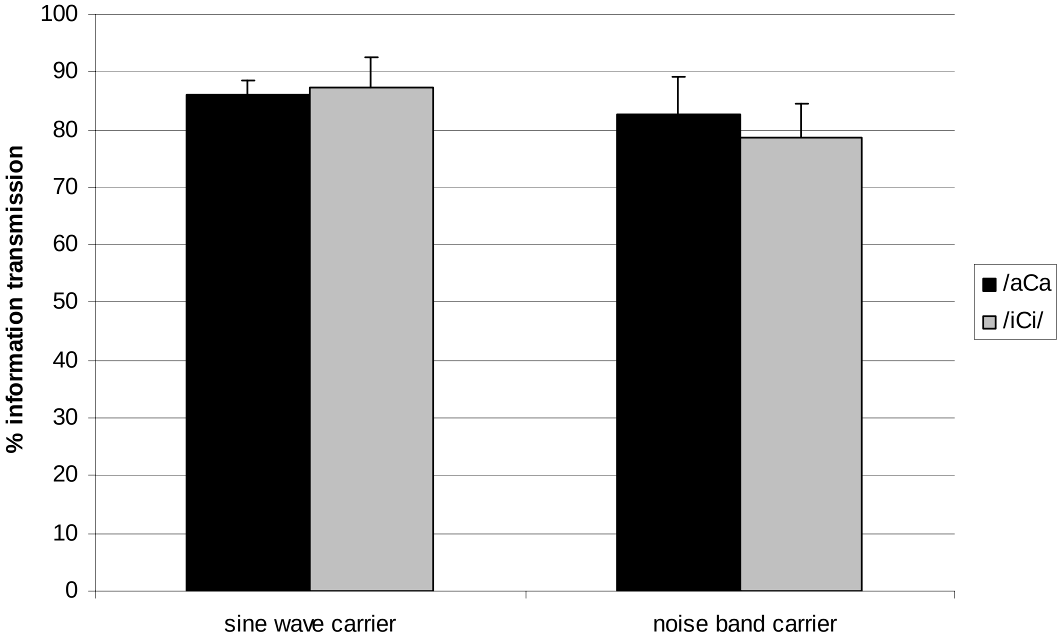

Figure 4.14. Mean (+ 1 SD) manner transmission by carrier type and vowel environment.

Table 4.7. *Post-hoc* t-tests comparing manner transmission in different listening conditions in order to explore the interaction between vowel environment and carrier.

Comparison conditions	**Mean difference**	**T value**	**Significance level**
Sine wave, /iCi/ Narrow band, /iCi/	8.78	5.57	0.01
Sine wave, /aCa/ Narrow band, /aCa/	3.11	1.11	0.33
Sine wave, /aCa/ Sine wave, /iCi/	-1.42	0.64	0.56
Narrow band, /aCa/ Narrow band, /iCi/	4.25	-1.08	0.34

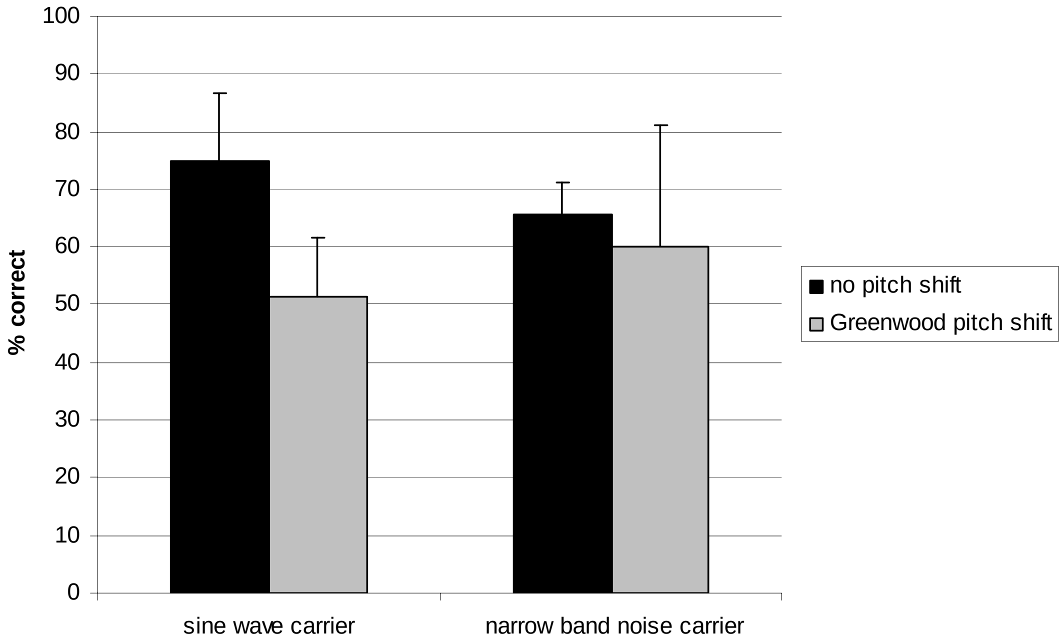

Figure 4.15. Mean (+ 1 SD) nasality transmission by carrier type and presence/absence of Greenwood pitch shift.

Table 4.8. *Post-hoc* t-tests comparing nasality transmission in different listening conditions in order to explore the interaction between carrier and inclusion of Greenwood pitch shift.

Comparison	**Mean difference**	**T value**	**Significance level (2-tail)**
Sine wave, no pitch shift Sine wave, pitch shift	23.50	4.54	0.01
Narrow band, no pitch shift Narrow band, pitch shift	5.59	0.66	0.55
Sine wave, no pitch shift Narrow band, no pitch shift	9.36	1.34	0.25
Sine wave, pitch shift Narrow band, pitch shift	-8.54	-0.98	0.38

The interaction between carrier and pitch shift for nasality is illustrated in figure 4.15. The difference in nasality transmission with the inclusion of pitch shift is much greater for the sine wave carrier than the noise band carrier.

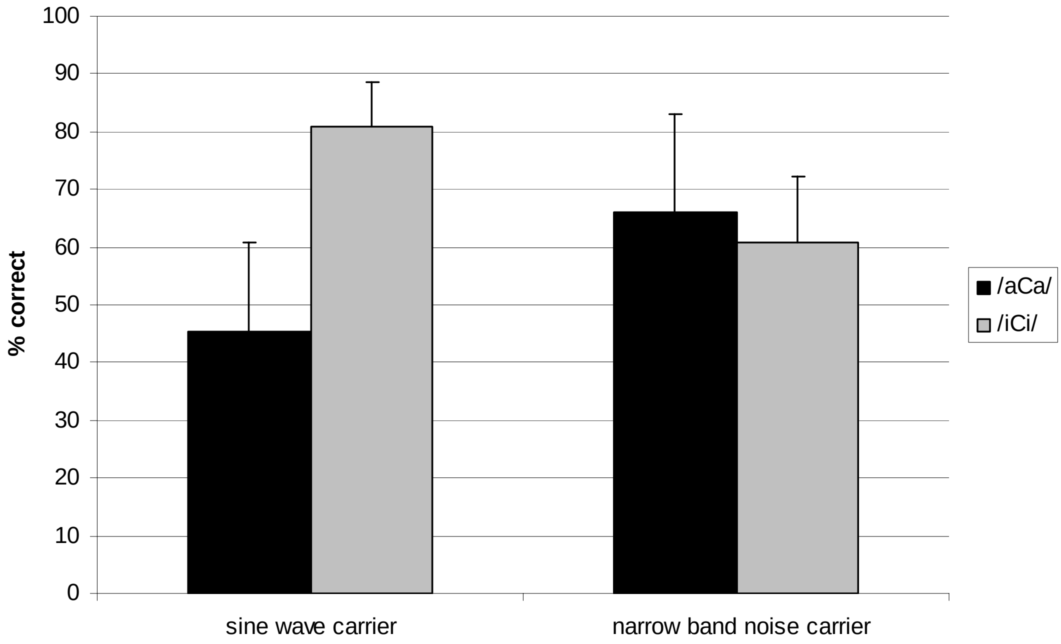

Figure 4.16. Mean (+ 1 SD) nasality transmission by carrier type and vowel environment.

Table 4.9. *Post-hoc* t-tests comparing nasality transmission in different listening conditions in order to explore the interaction between carrier and vowel environment.

Comparison	**Mean difference**	**T value**	**Significance level (2-tail)**
Sine wave, /iCi/ Narrow band, /iCi/	-35.44	-4.78	0.01
Sine wave, /aCa/ Narrow band, /aCa/	5.15	0.60	0.58
Sine wave, /aCa/ Sine wave, /iCi/	-20.43	-2.20	0.09
Narrow band, /aCa/ Narrow band, /iCi/	20.16	3.05	0.04

The interaction between carrier and vowel for nasality is illustrated in figure 4.16. Here the reason for the interaction can be seen in that there is a large difference in nasality transmission for the sine wave carrier but no the narrow band noise carrier.

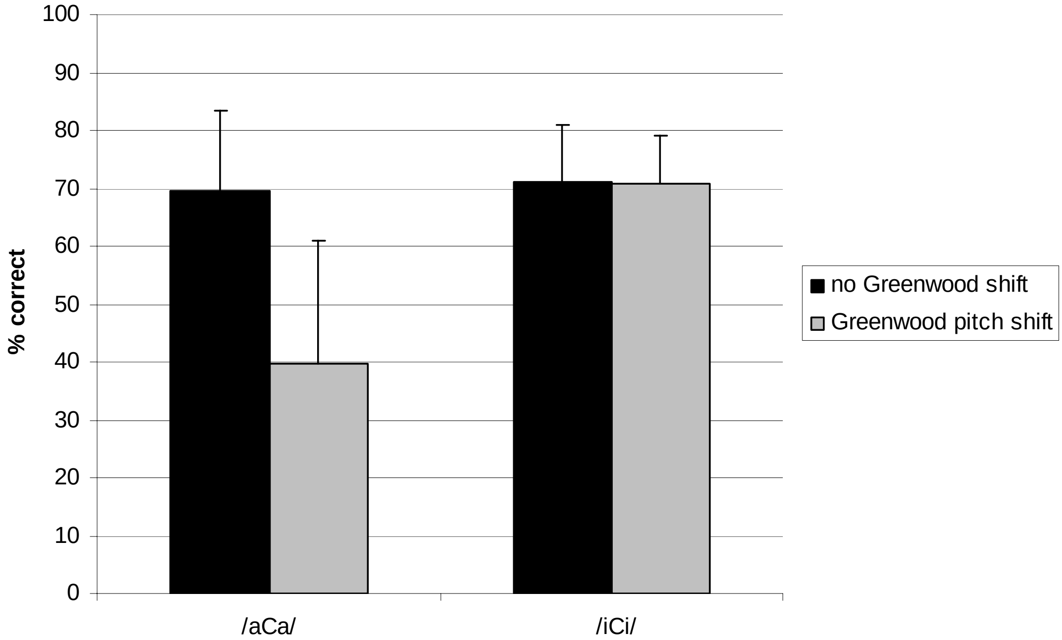

Figure 4.17. Mean (+ 1 SD) nasality transmission by vowel environment and presence/absence of Greenwood pitch shift.

Table 4.10. ***Post-hoc*** **t-tests comparing nasality transmission in different listening conditions in order to explore the interaction between vowel environment and inclusion of Greenwood pitch shift.**

Comparison	**Mean difference**	**T value**	**Significance level (2-tail)**
No pitch shift, /aCa/ No pitch shift, /iCi/	-1.55	-0.16	0.88
Pitch shift, /aCa/ Pitch shift, /iCi/	-30.87	-4.91	0.01
No pitch shift, /aCa/ Pitch shift, /aCa/	29.73	3.03	0.04
No pitch shift, /iCi/ Pitch shift, /iCi/	0.41	0.08	0.94

Figure 4.17 shows the interaction between pitch shift and vowel environment. It is evident that there is a large reduction in nasality transmission with the inclusion of the Greenwood pitch shift for the /aCa/ vowel environment but not for the /iCi/ vowel environment.

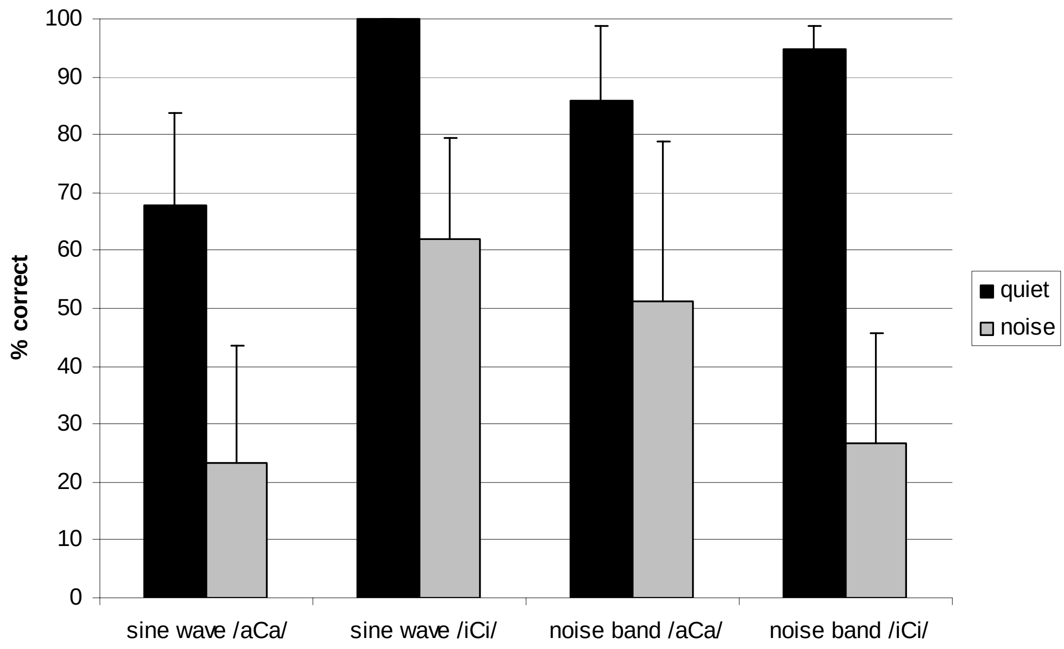

Figure 4.18. Mean (+ 1 SD) nasality transmission by vowel environment, carrier stimulus and noise condition.

Table 4.18 illustrates the three-way interaction of vowel environment, carrier stimulus and noise condition in relation to nasality transmission. Here the difference between carrier stimulus conditions depends on both noise conditions and vowel environment. For both carrier stimulus conditions in quiet there is an improvement in nasality transmission with /iCi/ compared to /aCa/, whereas in noise there is better performance with /iCi/ with the sine wave carrier but the opposite pattern with the noise band carrier.

Table 4.11. ***Post-hoc*** **t-tests illustrating three way interaction on nasality transmission (vowel environment, carrier stimulus and noise condition).**

Comparison	**Mean difference**	**T value**	**Significance level (2-tail)**
Sine wave, /aCa/, quiet Sine wave, /iCi/, quiet	-32.24	-4.51	0.01
Narrow band, /aCa/, quiet Narrow band, /iCi/, quiet	-8.95	-1.10	0.33
Sine wave, /aCa/, noise Sine wave, /iCi/, noise	-38.64	-3.51	0.02
Narrow band, /aCa/, noise Narrow band, /iCi/, noise	24.59	1.66	0.17
Sine wave, /aCa/, quiet Narrow band, /aCa/, quiet	-18.17	-1.86	0.14
Sine wave, /aCa/, noise Narrow band, /aCa/, noise	-28.03	-2.11	0.10
Sine wave, /iCi/, quiet Narrow band, /iCi/, quiet	5.12	1.63	0.18
Sine wave, /iCi/, noise Narrow band, /iCi/, noise	35.20	3.20	0.03
Sine wave, /aCa/, quiet Sine wave, /aCa/, noise	44.49	4.95	0.01
Sine wave, /iCi/, quiet Sine wave, /iCi/, noise	38.09	4.87	0.01
Narrow band, /aCa/, quiet Narrow band, /aCa/, noise	34.63	2.32	0.08
Narrow band, /iCi/, quiet Narrow band, /iCi/, noise	68.17	8.69	0.00

4.3.4 Discussion

The experiment reported in this section was motivated by the need to determine the best choice of parameters for the main AM experiment 4, the aim of which was to reproduce the information content of CI processing as accurately as possible to facilitate direct comparison with CI user data. Four independent variables were considered: Greenwood pitch shift, vowel environment, noise, and carrier stimulus type. The main issue of interest was the extent to which combination of variables led to model performance being close to what has been observed, or could be anticipated, in CI users, and also which versions of the AM were likely to be most sensitive to differences in processing or electrical/neural interface variables. Each of the variables, and their interactions, are considered in turn. Because of the need to focus on methodological preliminaries to the experimental work reported in chapter 4, a detailed discussion of some of the more complex effects for specific feature transmission, particularly nasality, are left for chapter 5 in which patterns of transmission for each feature are considered in more detail.

The first, and arguably most important, consideration for this experiment was choice of carrier stimulus. The sine wave carrier was significantly associated with better transmission of the features voicing, place, manner and fricative and total correct as compared to the noise band carrier. Although these differences were not very large (generally less than 10% difference in transmission), they were statistically significant and it can therefore be concluded that choice of carrier stimulus was a significant factor in determining AM performance. Given the tendency of the model to over-estimate absolute transmission values, it seems reasonable to assume, on purely empirical grounds, that a noise band model which leads to lower transmission values is more appropriate than a sine wave model for representing the information for consonant recognition available to CI users. However, this does chime with theoretical arguments, set out in 2.5.2, that both frequency resolution and within-channel periodicity information are coded better with the sine wave carrier and that this over-estimates the information available to CI users. The results suggest that the very high levels of voicing transmission in quiet obtained in experiment 1 were probably due, at least in part, to the choice of carrier stimulus. This provides indirect support for the notion that a noise band carrier is a more appropriate model as it does not provide periodicity information nor can frequency resolution be enhanced through the use of spectral side bands.

It is also of interest to note the interaction between choice of carrier stimulus and inclusion of noise, as indicated in figures 4.11 and 4.12. For both total correct and place transmission, there was a larger noise effect for the noise band carrier than the sine wave carrier. Although there is little data available on specific feature transmission values in noise in CI users or CI AMs, the effect on total correct with the addition of +5 dB SNR can be compared with available data. This again suggests that the noise band carrier is more appropriate as a model of consonant information. In general, the finding that noise band carrier AMs yielded poorer results for many features, but at a level that is nearer to levels of transmission to be expected with CI users, and that some of the sine wave carrier stimuli were nearer to achieving ceiling effects, suggests that noise bands are a more appropriate carrier stimulus for consonant models.

A further consideration is the noise variable. One of the aims of the two AM experiments reported in this chapter was to determine the most sensitive SNR to show *differences* in patterns of feature transmission, given the anticipated need to choose a single SNR for comparison with quiet to avoid multiplication of conditions given the inclusion of a number of other variables in the experimental work reported in chapter 4. The AM experiment in 4.1 gave reasonably similar results at +5 and +10dB SNR and experiment 2 only +5dB SNR was used. However, the findings of experiment 2 showed that reductions in performance with the addition of stationary background noise +5dB SNR were large across most features, for some features/listening conditions in excess of 30% reduction. Moreover, statistical analyses showed that the addition of noise at +5dB SNR had a significant effect on all consonant features except for envelope. This would suggest that an SNR of +10dB SNR might be more appropriate to tease out differences in transmission across different features and also to cater for the possibility that CI users would perform worse in background noise than AM listeners.

A further model variable was included, namely the alteration of carrier frequencies to mimic the assumed subjective pitch shift (using data from Greenwood (1990). Interestingly, the presence vs. absence of Greenwood pitch shift in the model had no effect on transmission of any feature except for nasality. The important question here was whether subjects could adequately acclimatise to pitch-shifted stimuli. Previous work on pitch-shifted AMs suggests that considerable acclimatisation time is necessary for optimal performance; however, this depends on the degree of shift. The shift undertaken in experiment 2 (described in 3.3.2) represents a mean upward transposition of between 1.2 and 1.45 depending on frequency channel, which is markedly less than values obtained from some previous studies e.g. Dorman et al (1997a), and moreover using AMs with a larger number of channels which might offset the effect of any pitch shift. The likely absence of a strong pitch shift effect can be assumed to relate to the relatively small shift used. Given that the inclusion of this increases AM validity, in the sense that an additional variable is included which reflects one aspect of the electrical/neural interface, and given that the generally small or non-existent effects of this variable, it seemed appropriate to include it in models which attempt to go further in mimicking electrical/neural interface factors e.g. channel interaction.

A final consideration related to a potentially important variable in test methodology, namely vowel environment. As noted in chapter 2, almost all work on consonant confusion analysis in CI users has used the /aCa/ vowel environment. In this experiment vowel environment had an effect on features place and nasality and on total correct, albeit the effects had differing directions. Place of articulation transmission was better for the /aCa/ vowel environment than /iCi/, though the interaction illustrated in figure 4.13 shows this to be due specifically to better transmission in quiet with the /aCa/ vowel environment. It was suggested by Loizou et al. (2000b) that place is coded more via the burst than the formant transition in a front vowel (/a/) context but more by formant transition in a back vowel context (/i/).. It was observed in 1.7.3 that formant transition is coded very poorly through CI processing whereas the burst spectrum is better preserved. The finding of better place transmission with the /aCa/ vowel environment therefore supports this hypothesis. However, it is notable that the absolute levels of place transmission for /aCa/ from this experiment somewhat exceed those obtained in the literature and do not fit with the pattern of worse place transmission compared to manner and voicing transmission as indicated in 1.4. Given that the performance in the model for /iCi/ where the burst is thought to be less prominent is nearer to observed transmission levels than for /aCa/ where the burst is more prominent, findings suggest that the models used in experiment 2 may over-estimate the representation of the burst for some reason. This issue is addressed further in the specific discussion of place transmission in Chapter 5. Most of the interactions involving a nasality transmission also related to choice of vowel environment. The specific reasons for the interactions shown with nasality are discussed in 5.6.

The two anomalous results from experiment 1 have been addressed by experiment 2. In this experiment, much more stringent efforts were made to avoid the frequent /idʒi/-/igi/ confusion due to the orthographic representation of the two sounds in English, Probably as a consequence, transmission of fricative was at a much higher level in experiment 2 compared to experiment 1. Also, voicing transmission values were around 80% in experiment 2, (although only if Greenwood-shifted noise bands were used as carriers). This contrasts with voicing transmission values around 90% in experiment 1.

The overall significance of the two experiments taken together can be stated as follows. First, it is possible to obtain meaningful results using a consonant confusion measure with normal hearing subjects listening to an AM, without a large number of repetitions or extended acclimatisation time being necessary. Second, it is likely that important differences across feature transmission as a function of noise can be captured by using only two listening conditions: quiet, and one of either +5 or +10 dB SNR stationary noise. For reasons stated above, a choice of +10 dB SNR was deemed preferable and was therefore used for the AM experiment reported in the next chapter. This meant that it was possible to construct an AM experiment for comparison with equivalent CI user performance in a way that allowed a number of variables to be compared over a single test session without strong fatigue effects. Third, noise bands should be used as carrier stimuli rather than sine waves as it seems likely that the sine wave carrier may over-estimate spectral and periodicity information available to CI users. Fourth, similar results are obtained if noise band centre frequencies are aligned to assumed pitch-shifted values rather than analysis frequencies, despite evidence to the contrary reported in 1.6.4. Given the desire to mimic aspects of the electrical/neural interface as accurately as possible, it therefore seemed appropriate to include the Greenwood pitch shift in any model used for comparison with CI users. Fifth, the vowel environment /iCi/ led to slightly worse performance, particularly for place transmission, than /aCa/, as hypothesised. Given the rationale, proposed by Loizou et al. (2000b), that this vowel environment might be more sensitive to parametric variations such as stimulation rate than the more commonly used /aCa/, it was therefore deemed appropriate for use in the experimental work in chapter 4.

Chapter 5. A comparison between AMs and CI users

5.1 Research questions, aims and hypotheses (both experiments)

This chapter describes two parallel experiments, one with Nucleus 24 users and one with normal hearing subjects listening to three different versions of an AM of the Nucleus 24 device. The two experiments were matched in terms of CI processing characteristics and test methodology. The primary aim was to determine whether a carefully matched AM could predict consonant feature recognition in a group of CI users and whether the inclusion (and degree) of spectral channel interaction had a bearing on the model's predictive power. However, the aim was achieved not merely by determining consonant feature recognition in a single listening condition but by looking at variations in feature transmission patterns with the addition of noise and with changes to the processing parameters of channel number and stimulation rate. This meant that the experimental work also had a set of secondary aims, namely to determine i) whether there is a trade-off between channel number and stimulation rate for consonant recognition, ii) whether these parameters have effects on particular consonant features and iii) whether they interact with noise. It should be noted that all of these questions are framed within the context of a specific CI device and processing strategy.

The most important research questions were therefore:

- Is the pattern of consonant feature recognition between the AM and CI users the same?
- Does this correspondence depend on the inclusion of spectral channel interaction in the AM?
- Does this correspondence vary across individual CI users or between subgroups of CI users, in particular is the correspondence greater for better-performing or worse-performing CI users?

- Are variations in consonant recognition among CI users the same as differences between AMs with and without channel interaction?

Further questions can be addressed to specific feature categories:

- Is the effect of noise the same between the AM and CI users?
- Is the effect of channel number the same between the AM and CI users?
- Is the effect of changing envelope information as a consequence of changes to stimulation rate the same between the AM and CI users?
- Is the interaction between any of these factors the same between the AM and CI users?

Some additional questions can be framed specifically for the CI user experiment, in connection with the question of variance between users. If, as Munson (2004) has suggested, variation between better and worse performers is quantitative not qualitative, e.g. is generic across psychophysical abilities and not specific to a specific subset of abilities, such as spectral resolution, then the pattern of phoneme errors should be similar in better and worse CI users. If, however, there are differences in the relative transmission of spectral features, then it can be hypothesised that differences between better and worse performing CI users would be mirrored by the difference between higher and lower levels of channel interaction in the AM. Alternatively, if individual differences are more to do with differences in temporal/amplitude coding at the electrical/neural interface, then spectral channel interaction might not show equivalent variations, but it would be anticipated that there would be larger between-individual variation in temporal features.

5.2 Methods

The general methods have been described in chapter 2. However, a number of more detailed considerations are outlined in this section. First, the choice of CI users and the associated choice of processing parameters is considered. The clinical population available to the author at the South of England Cochlear Implant Centre were users of the Nucleus implant, and work reported in chapter 3 used an AM of the Nucleus 24 implant with the CIS processing strategy. However, the great majority of Nucleus 24 users use the ACE or SPEAK processing strategy. Therefore, to continue to use a fixed channel (CIS) AM would introduce an additional confounding variable when comparing AM and CI user results. Therefore, it was decided it to undertake the AM experiment using the Nucleus 24 ACE speech processing strategy. In this way, all processing variables could be matched between the AM and CI user experiments.

A further consideration was the choice of channel number and channel stimulation rate values. It was decided to reduce channel number and stimulation rate selectively from "typically used" values. In order to match the two experiments as closely as possible, the default parameters were those most commonly used by the clinical population accessed for the CI user experiment. The most common parameter settings for adult users of the Nucleus 24 implant in the available clinical population (adult Nucleus 24 users in the South of England Cochlear Implant Centre) were: 900 pps/ch stimulation rate, 12 maxima out of 20 channels with the Advanced Combination Encoder (ACE) speech processing strategy. This was therefore chosen as the default "high rate high channel number" condition for both experiments. The low stimulation rate condition was chosen as 250 pps/ch as this was the minimum permissible stimulation rate allowed by clinical software and also corresponds to the stimulation rate used with the SPEAK speech processing strategy. However, the TMTF measurements shown in 1.4.3 showed that differences in temporal envelope coding between these two rates are modest at low modulation rates and non-existent at higher modulation rates. Moreover, the differences in envelope coding, e.g. effective envelope bandwidth, varied in the noise band AM to the same extent as in the CI processor. Consequently, if the temporal information provided by the implant is the key factor in determining perceptual abilities (as opposed to some physiological mechanism associated with higher pulse rates), little if any change between rates would be anticipated. Nevertheless, the design of the consonant recognition task, e.g. using the /iCi/ vowel environment which (suggested by Loizou et al. (2000b)) should be more sensitive to rate changes and the inclusion of a background noise condition should be such that any perceptual effects would be evident. The majority of the CI users had ACE maps with a channel stimulation rate of 900 pps/ch. For all conditions which did not use the CI subject’s standard stimulation rate (e.g. all 250 pps/ch conditions), re-mapping was undertaken by globally adjusting T-levels and C-levels along all 20 active electrode channels. Re-mapping was undertaken in order to account for the change in loudness (which is a function of stimulation rate, so long as pulse duration remains unchanged). Only small changes in overall electrical dynamic range were observed in the altered low-rate MAPs, of the order of 2-3% reduction in dynamic range overall compared to the 900 pps/ch MAPs.

The decision about reducing channel number was less straightforward, given the choice of the peak-picking strategy ACE. The question was raised previously as to whether channel number is perceptually equivalent to the number of spectral peaks selected, or equivalent to the number of channels available, or to something in between the two. Dorman et al. (2002) found that performance was equivalent between fixed channel and peak-picking models where peak number in a peak-picking strategy was around the same as channel number in a fixed-channel strategy. In the present study the decision was taken to reduce both of these correspondingly to a level where channel number effects have been determined in previous work- thus the normal 12/20 condition was changed to 4/7, e.g. both channel number and peak (maxima) number were altered threefold. The 900*4/7 condition used channels 3,6,9,12,15,18 and 21. Figure 5.1 shows the frequency weighting and boundaries for the 20-channel MAPs while figure 5.2 shows frequency boundaries for the 7-channel MAP. Table 5.1 summarises the parameter values for the three MAP conditions. (The term "MAP", coined by Cochlear Corporation, is used here to describe the particular set of parameter configurations, and their implementation, used for a particular CI user in a particular listening condition.)

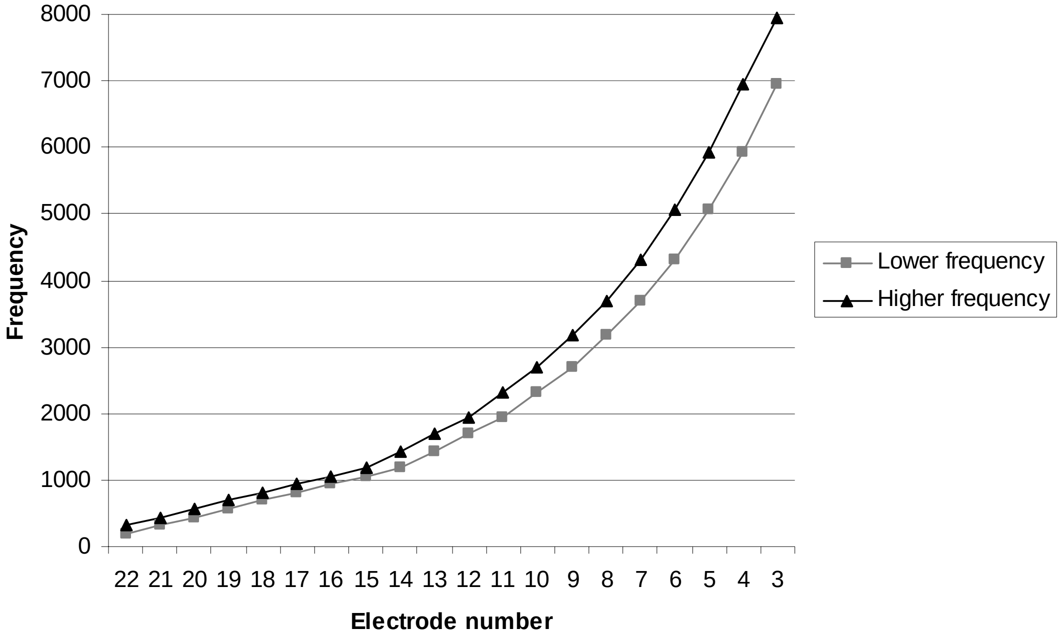

Figure 5.1. Upper and lower frequency boundaries for 12*20 channel MAPs.

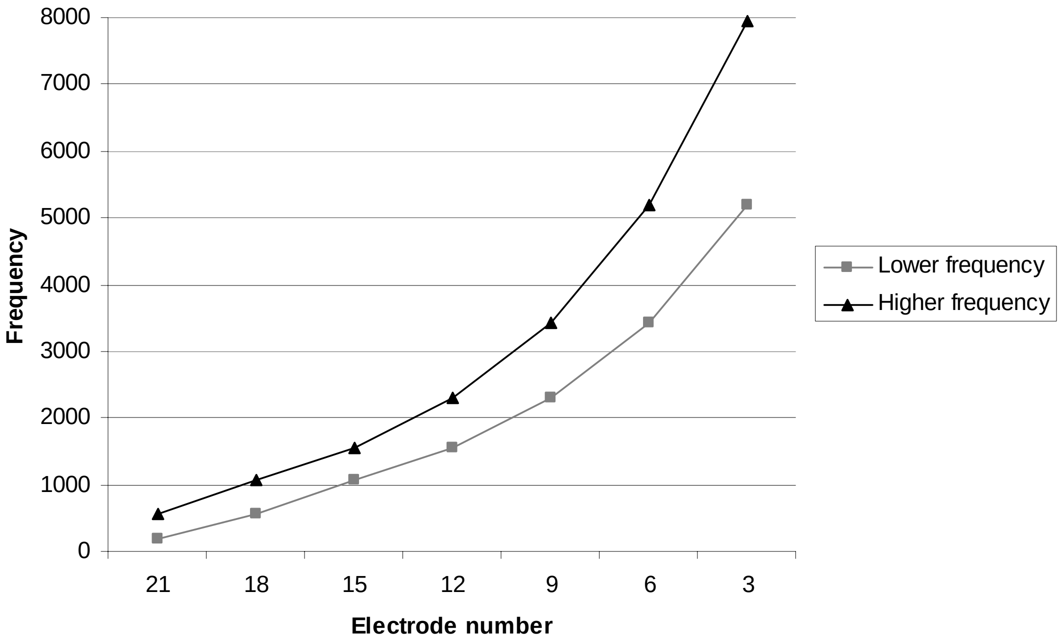

Figure 5.2. Upper and lower frequency boundaries for 4*7 channel MAP.

Table 5.1. Summary of three MAP conditions.

Map	Stimulation rate	Channel/maxima number
Default	900 pps/ch	12 maxima/20 channels
Reduced stimulation rate	250 pps/ch	12 maxima/20 channels
Reduced channel number	900 pps/ch	4 maxima/7 channels

An additional advantage of this design was that equally "unfamiliar" MAP conditions could be compared in order to reduce the effect of familiarity. Where the default MAP was the one normally used, deterioration in performance in the other MAPs could be taken as a results of inadequate acclimatisation to the MAP, given that exposure was relatively limited. However, the possibility of comparing two equally unfamiliar MAPs, one with a reduced rate and one with a reduced channel number, would allow at least one comparison that was unaffected by the familiarity/acclimatisation issue and would also allow a direct comparison between lowering rate and lowering channel number from the normally used MAP condition. Finally, it was the possibility that there was a linear equivalence in performance when trading off channel number and stimulation rate was raised in 2.3.4. It should be noted that there is no theoretical basis as such for assuming a direct comparison between linear reductions in channel number and stimulation rate for the Nucleus 24 device. Nevertheless, it was of interest to explore the possibility of "trade-off" between stimulation rate and channel number and therefore that the two "reduced" MAP conditions were roughly equivalent e.g. stimulation rate was reduced to slightly less than a third, as was channel resolution.

It was considered that, for the purposes of sample size calculation, the most important effect was the effect of noise on feature transmission values. Friesen et al. (2001) found that, for Nucleus users, consonant recognition was reduced by 10% with the addition of stationary background noise at +10 dB SNR, the same noise type and SNR used in this experiment. In the Friesen study standard deviation was also around 10. A sample size calculation based on these values (even assuming a highly conservative value for correlation in scores of over 50%) yielded a desired sample size of 11. Eleven NH subjects were recruited to experiment 4. In the event recruitment problems in the study meant that data were collected on 9 CI users. It should be noted that the sample size calculation was appropriate for within-subject comparisons but not for between-group comparisons; it was considered that the more important objective of the research was to determine differences in feature transmission as a function of processing and other variables within subjects, while the comparison between better and worse CI users, which is presented in 5.3, was of secondary importance to the overall design of the study.

The aim of the corresponding CI user experiment was to duplicate the AM processing and stimulus parameters just described with a group of adult users of the Nucleus 24 CI. Subjects were recruited from the South of England Cochlear Implant Centre and were all experienced users of the Nucleus CI24M or Nucleus CI24R Contour. Subject criteria were:

- Post-lingually deafened adult cochlear implant users aged 18 or over.
- Users of the Nucleus 24 device.
- Normally users of the ACE strategy
- Implant users for at least nine months.
- Score of at least 60% in the BKB sentences test in quiet at last review session.
- English as their first language.

No formal attempt was made to choose “better” and “worse” performers *a priori*. Due to difficulties in recruiting an adequate sample, the inclusion criteria were expanded to include two subjects who normally used the SPEAK processing strategy in the bilateral condition. Both of these subjects had had experience using the ACE strategy since receiving their implants. Subject 5 had achieved a score of only 57% in the Bamford-Kowal-Bench, known as BKB sentence test (Bench et al.,1979) at his most recent review but had achieved scores above 60% on all previous occasions. Three of the subjects had bilateral implants but performed the tests using only the implant which they had had the longest. Subjects who used a hearing aid on their non-implanted ear used their implant on its own for these tests. The subjects’ ages ranged from 25 to 85 with a mean age of 61. There were six males and three females. Subject details are given in table 5.2. It should be noted that the *post hoc* separation into “worse” and “better” CI users, described in 5.3, did not co-vary with distinctions between those who normally used the ACE 900 pps/ch strategy vs. those who did not, nor did it co-vary with those who normally used bilateral CIs vs. those who did not. Consequently, it was thought that the relaxation of the inclusion criteria did not adversely affect results.

Table 5.2. Subject details for CI user experiment.

Subject number	Sex	Age	BKB score	Duration of implant use	Implant type	Normal strategy	Other info
1	M	25	81	1yr 5m	CI24R	ACE 900 pps 12 of 20	
2	M	70	92	2 ½ yr	CI24R	ACE 900 pps 12 of 20	
3	M	65	90	1 yr	CI24R	ACE 900 pps 12 of 20	
4	F	73	94	L – 6yr R – 4 yr	CI24M	SPEAK 250 pps 8 of 20	Bilateral implants
5	M	85	57	R – 7 yr L – 3 ½ yr	CI24M	ACE 720 pps 8 of 20	Bilateral implants
6	F	62	80	2 yr	CI24R	ACE 900 pps 12 of 20	
7	M	49	98	2 yr	CI24R	ACE 900 pps 12 of 19	
8	M	72	94	L – 6 yr R – 3 yr	CI24M	SPEAK 250 pps 8 of 20	Bilateral implants
9	F	48	100	1 yr	CI24R	ACE 900 pps 12 of 20	

Each subject was tested using three different MAPs, e.g. 900*12/20 ACE, 250*12/20 ACE and 900*4/7 ACE. Mapping was undertaken using the Cochlear Custom Sound programming software by the researcher. Order of MAP condition was randomised and testing was conducted first in quiet then in noise for each MAP condition. A spare Esprit 3G processor was used to provide alternative MAPs. For most listeners, they could use their normal MAP as this was already 900*12/20 ACE. For the reduced channel condition a new MAP was created using the same seven channels as in experiment 3 (see figure 5.2). For the lower stimulation rate MAP, it was necessary to adjust T-levels and C-levels (minimal audible and maximum comfortable current levels) because of the change in loudness associated with changes in stimulation rate. Subjects were given as much time as needed to acclimatise to the new MAPs; in practice, this was not more than 15 minutes.

In summary, two parallel experiments were undertaken, one with 11 normal hearing listeners listening to an AM and the other with 9 users of the Nucleus 24 CI device. In each experiment the VCV test (as described in chapter 2) was undertaken in two noise conditions (with and without stationary background noise at +10 dB SNR) * 3 MAP conditions. Additionally, for the AM experiment, all testing was undertaken with three different AMs, varying by the degree of the term λ from equation 3.2 from 0 to 1 to 3.3. This meant that there were a total of 2*3 listening conditions for CI users and 2*3*3 listening conditions for AM subjects. The results of the CI user experiment are reported in 5.3 while results of the AM experiment are reported in 5.4. In section 5.5 the two sets of data are considered together. All results are reported separately by transmission of six consonant features and also by total correct scores.

5.3 Results of CI user experiment

Data analysis methods are described in 2.3.6. For convenience, total correct values are also included in graphical presentation along with feature transmission values although it should be noted that this represents the absolute number with correct responses rather than averaged information transmission. Figure 5.4 shows performance across features and total correct averaged across MAP conditions and figures 5.5 to 5.7 show equivalent data separately for each of the three MAP conditions.

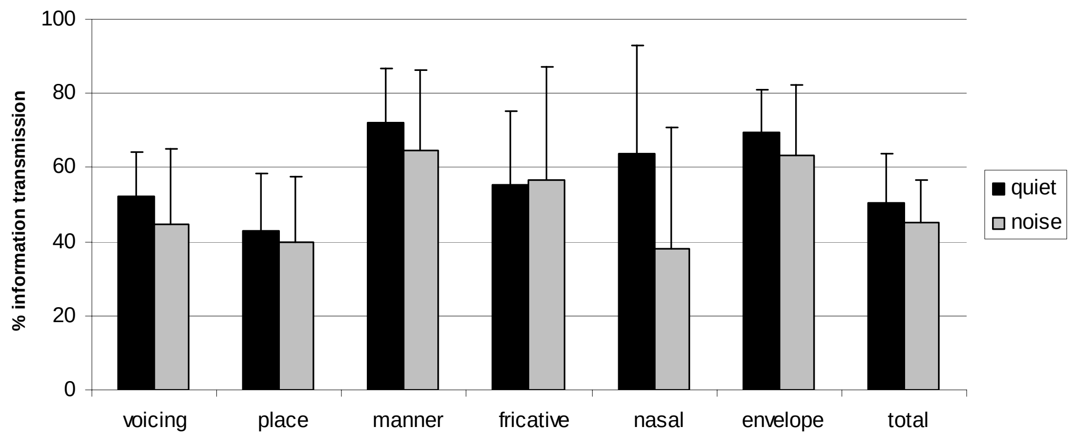

Figure 5.4 Mean (+ 1 SD) feature transmission values and total correct (across MAP conditions) in 9 users of Nucleus 24 CI with ACE processing.

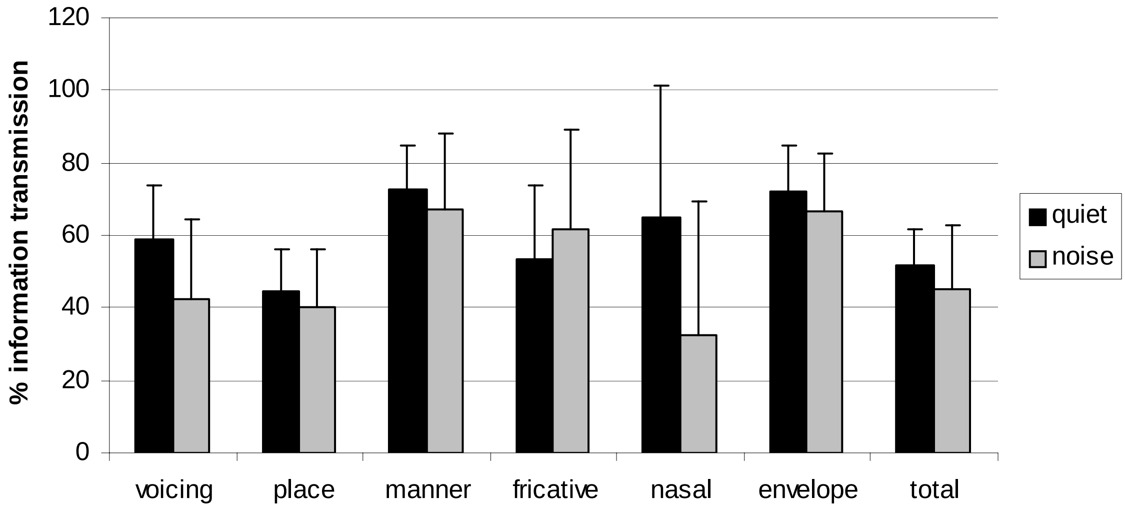

Figure 5.5. Mean (+ 1 SD) feature transmission values and total correct in 9 users of Nucleus 24 CI with ACE MAPs with 12 maxima out of 20 channels and channel stimulation rate of 900 pps/ch.

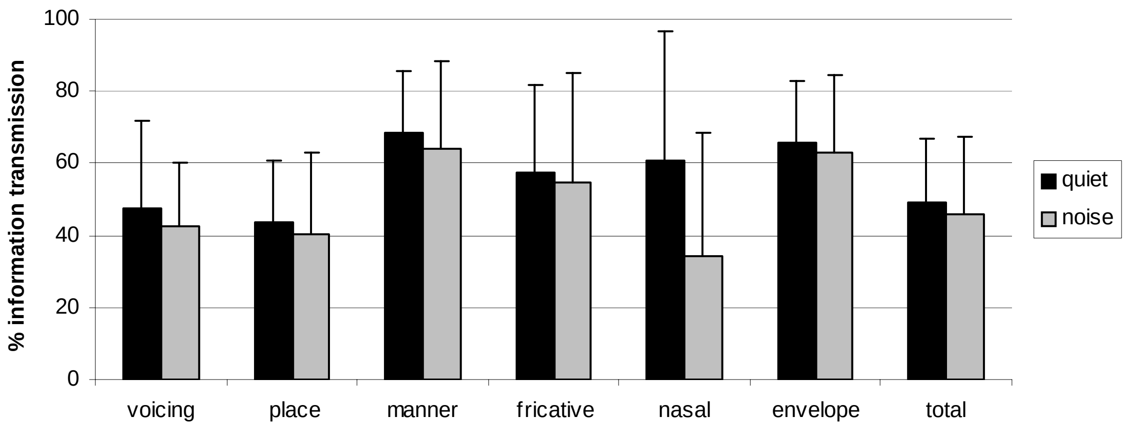

Figure 5.6. Mean (+ 1 SD) feature transmission values and total correct in 9 users of Nucleus 24 CI with ACE MAPs with 4 maxima out of 7 channels and channel stimulation rate of 900 pps/ch.

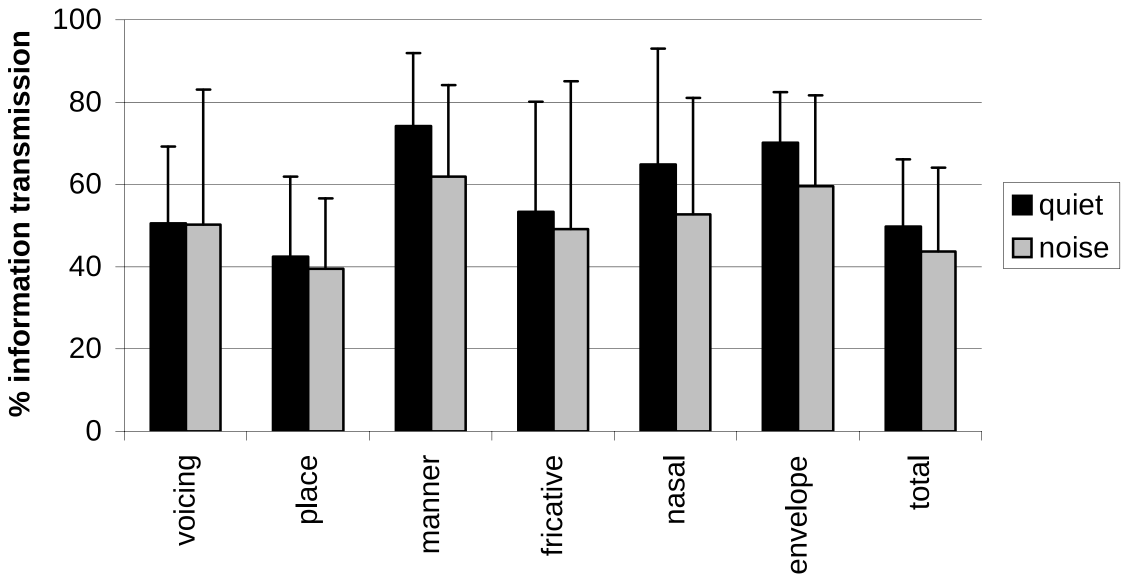

Figure 5.7. Mean (+ 1 SD) feature transmission values and total correct in 9 users of Nucleus 24 CI with ACE MAPs with 12 maxima out of 20 channels and channel stimulation rate of 250 pps/ch.

Figures 5.5 to 5.7 indicate only small differences across MAP conditions. Noise effects also appear to be small, with the exception of nasality, which shows a more marked noise effect. As might be anticipated in any group of CI users, variance appears large for most measures. A MANOVA was performed with seven dependent variables (percentage information transmission for six consonant features percentage total correct) and three predictor variables (channel number, channel stimulation rate and noise). As with previous analyses, only significant factors or interactions are reported here and full details are given in Appendix B. The noise factor had a significant effect on nasality ($p<0.05$) but no other dependent variable. Neither channel number nor stimulation rate had a significant effect on any dependent measure. There were no significant factor interactions for any variable.

It was also of interest to consider differences between individual CI users. Although the 9 CI users had been chosen on the basis of overall BKB sentence score in quiet >60%, one of the aims of the experiment was to determine, *post hoc*, the degree and nature of variance in performance. In practice, variance in most-recent BKB score was low (see table 5.2) and the consonant recognition measure itself showed much higher variance. Therefore, the definition of "better" vs. "worse" performer was taken as consonant recognition total correct in quiet during the test sessions, using the "high channel number, high stimulation rate" MAP that represented the default MAP for most of the listeners. On the basis of this, of the 9 subjects, 5 had baseline consonant recognition scores in quiet (e.g. with the normal high rate/high channel number MAP) of 50% or more while the other 4 had scores of less than 50%. Therefore, separate analyses of these two subgroups were undertaken. Figures 4.8 and 5.9 show performance across features (averaged across MAP conditions) in quiet and noise for the two groups.

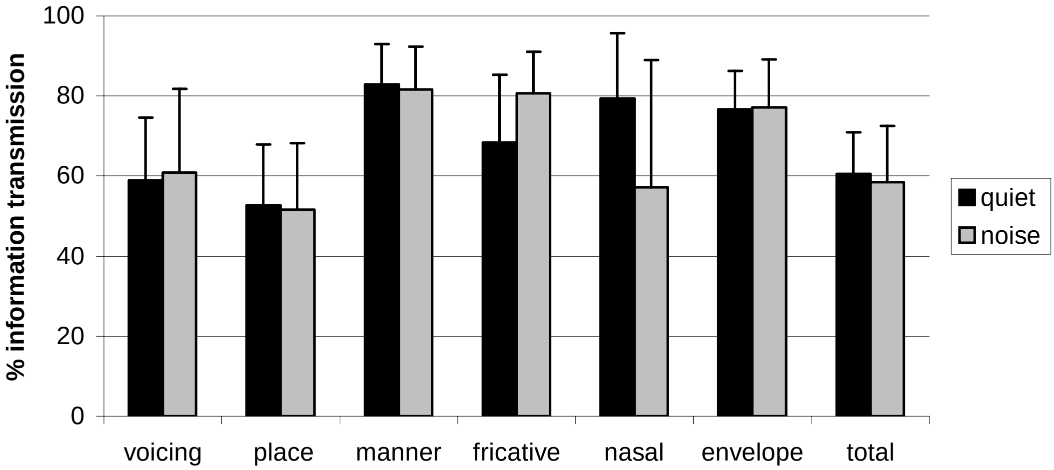

Figure 5.8. Mean (+ 1 SD) feature transmission (across MAP conditions) in 5 users of Nucleus 24 CI with ACE processing with baseline consonant recognition scores of 50% or better.

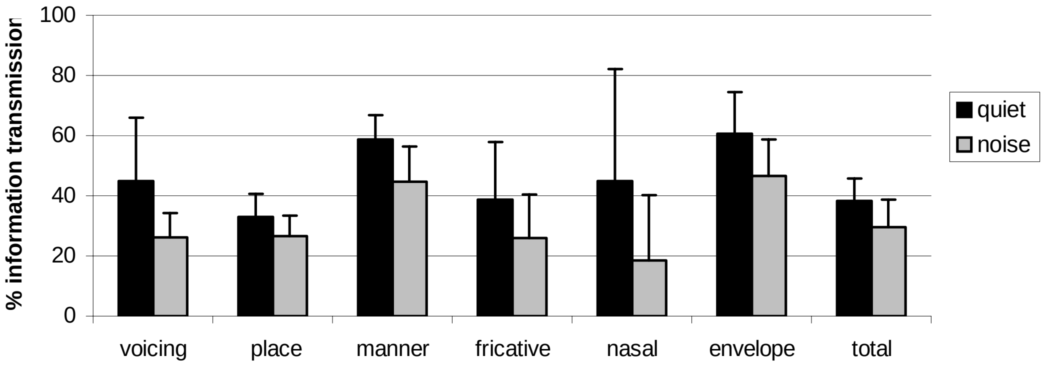

Figure 5.9. Mean (+ 1 SD) feature transmission (across MAP conditions) in 4 users of Nucleus 24 CI with ACE processing with baseline consonant recognition scores of less than 50%.

It is also of interest to represent the magnitude of differences between better and worse users. Therefore, figure 5.10 shows difference in transmission of different features between better and worse CI users in quiet and noise, averaged across MAPs. This clearly shows that differences between the two subgroups were much greater for noise-contaminated stimuli.

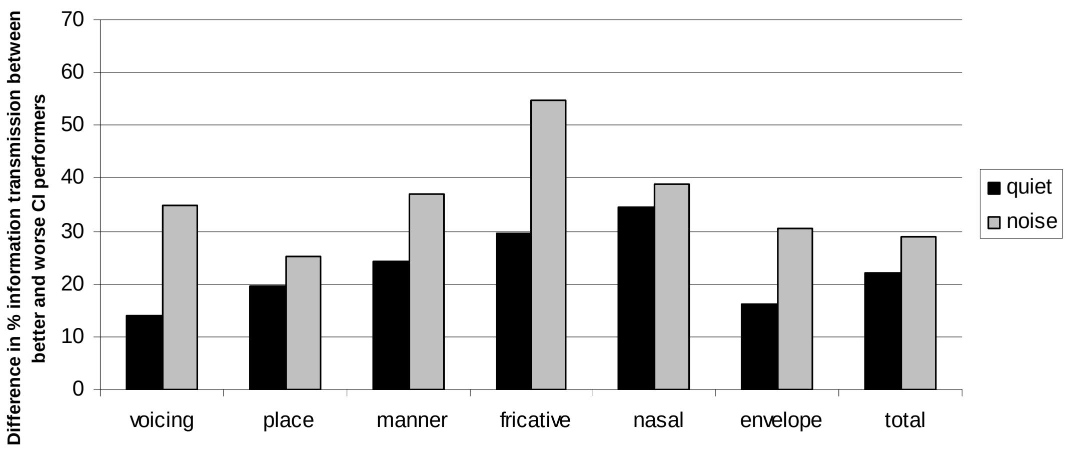

Figure 5.10. Difference between better and worse CI performers across MAPs.

In summary, there appeared to be marked differences in overall performance levels and particularly in degree of noise effect between the two subgroups of CI users. In order to explore this statistically, MANOVA analyses were undertaken on the same basis as for the overall group results, but here split into the two groups of subjects. Full details of these two sets of analyses are in Appendix B. For the "better user" group (N=5), no factor or interaction had an effect on any variable, although the effect of noise on nasality just failed to reach significance (p=0.065). For the "worse listener" group (N=4), the noise factor had a significant effect on transmission of voicing, manner and envelope and also on total correct scores. As with other analyses, neither processing factor had a significant effect on any measure; there were also no factor interactions.

5.4 Results of AM experiment

For the AM experiment results, data were first combined into confusion matrices for each listening condition/individual. As with previous experiments, total correct values and information transmission values for six features were computed for each confusion matrix and the data were then used for further analysis. In this experiment, testing was also undertaken in the “unaltered” listening condition, e.g. with no AM applied. These data are shown in figure 5.11.

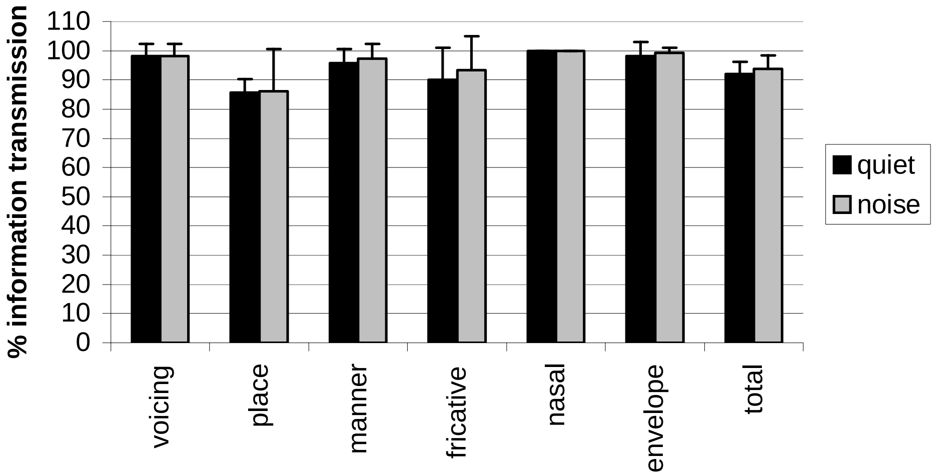

Figure 5.11. Mean (+ 1 SD) feature transmission in the “unaltered” condition.

All feature transmission values exceeded 90% except for place of articulation and all except place and fricative exceeded 95%. None of the features showed worse performance in the “noise” conditions. These values were not included in subsequent data analysis but are presented here as baseline data for comparison. The only likely effect on interpretation of AM data was the relatively low place transmission values in the unaltered condition. However, all AM conditions yielded place transmission values of 60% or less; therefore, the fact that place transmission in the unaltered condition was at 87% and 86% (for quiet and noise) and was, therefore, not considered to have any impact on interpretation of model findings.

Subsequent data analysis and illustration relates to the three AMs. In order to illustrate the overall pattern of feature performance in quiet and noise, figures 5.12 to 5.14 show performance across features for the three different “channel interaction” conditions, averaged across MAP conditions.

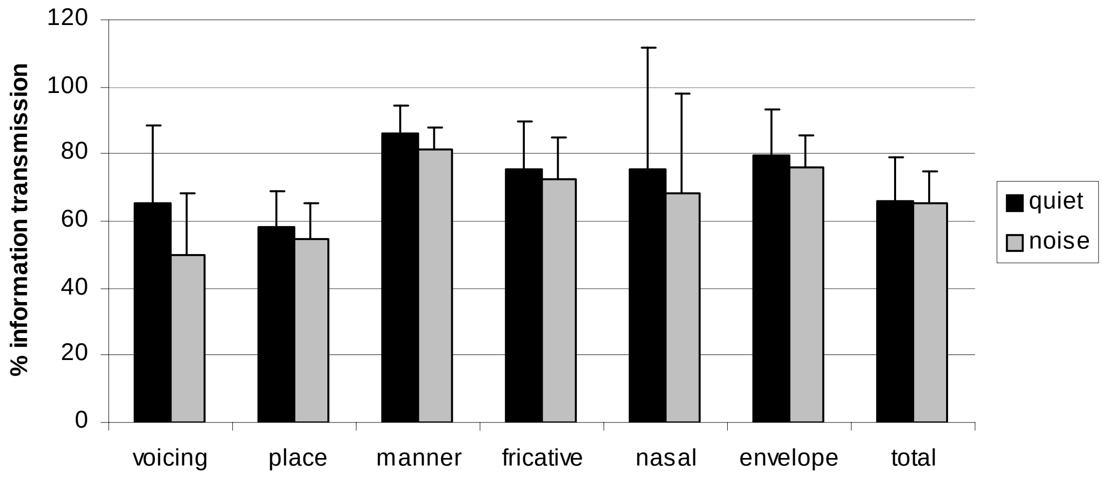

Figure 5.12. Mean (+ 1 SD) feature transmission (across MAP conditions) with and without stationary background noise at +10 dB SNR with an AM with no "channel interaction".

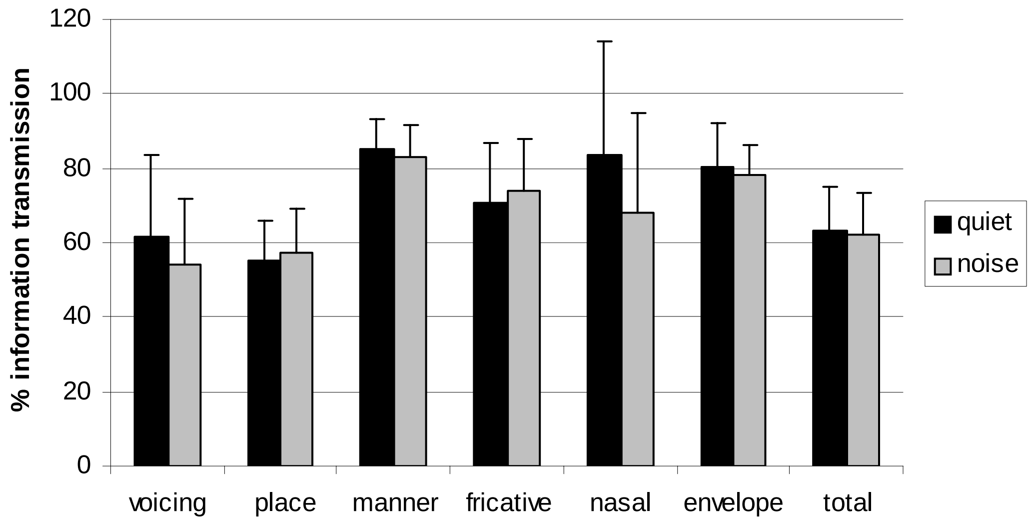

Figure 5.13. Mean (+ 1 SD) feature transmission (across MAP conditions) with and without stationary background noise at +10 dB SNR, AM with 1 mm "channel interaction".

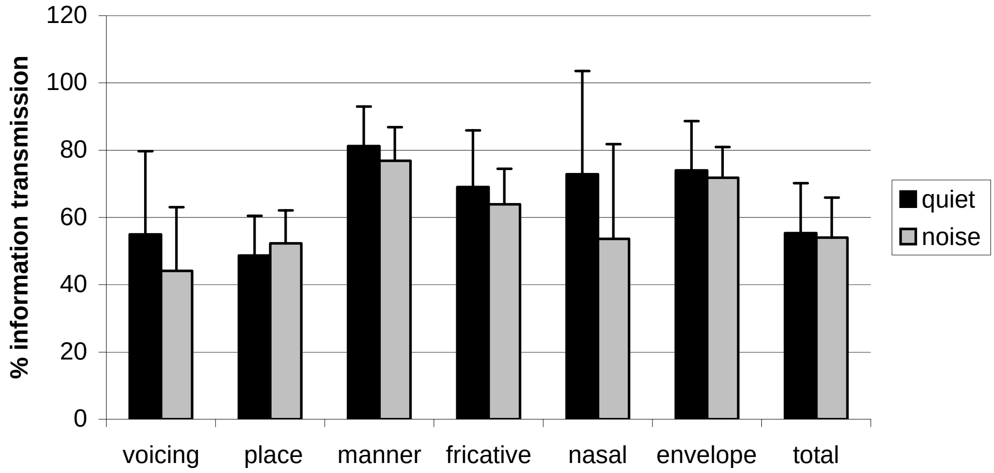

Figure 5.14 Mean (+ 1 SD) feature transmission (across MAP conditions) with and without stationary background noise at +10 dB SNR, AM with 3.3 mm "channel interaction".

In order to make a clearer comparison between AM conditions, figures 5.15 and 5.16 show performance for each of the three channel interaction conditions for quiet and noise separately, again averaged across MAP conditions.

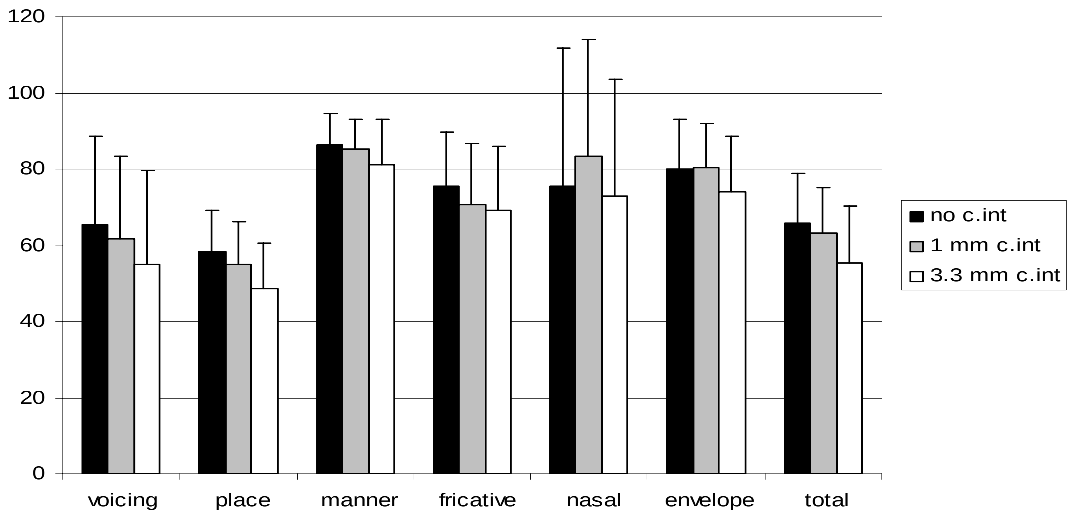

Figure 5.15. Mean (+ 1 SD) feature transmission (across MAP conditions) in quiet with varying degrees of "channel interaction".

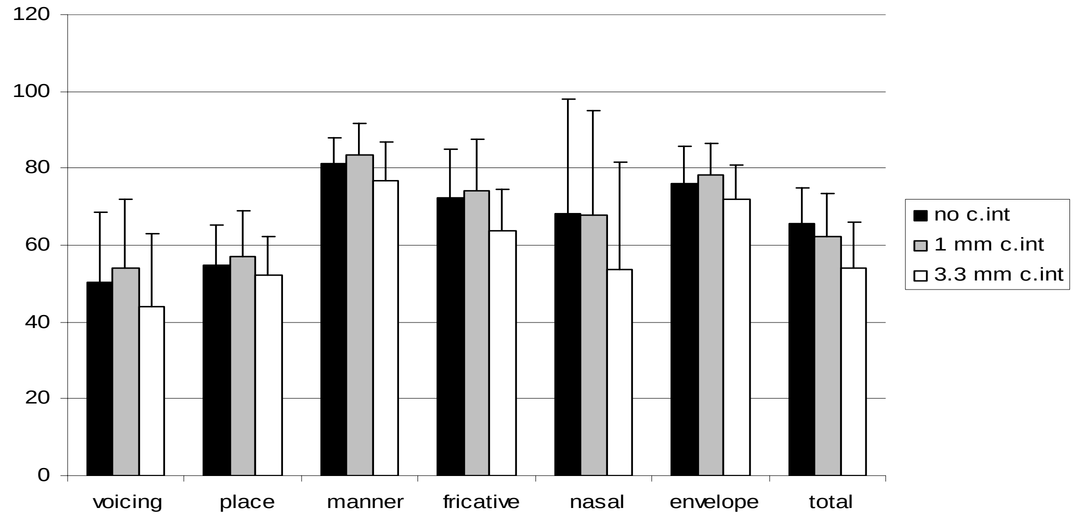

Figure 5.16. Mean (+ 1 SD) feature transmission (across MAP conditions) in stationary backgound noise at +10 dB SNR using AMs with varying degrees of "channel interaction".

These figures suggest a trend for worse performance for most features in the 3.3 mm channel interaction condition compared to the other two conditions, little difference between the "no channel interaction" and "1 mm channel interaction" conditions and worse performance in noise compared to quiet for some features, particularly nasality and voicing. In terms of differences between features, there appeared to be a trend for manner and envelope transmission being greatest and voicing or place worst.

Figures 5.17 to 5.25 show equivalent feature transmission values as shown in figures 5.12 to 5.14, but here data are presented for each of the specific MAP conditions, e.g. the high channel number high stimulation rate condition (900 pp/ch x 12/20), the low channel number high stimulation rate condition (900 pps/ch x 4/7) and the high channel number low stimulation rate condition (250 pps/ch x 12/20). Figures 5.17 to 5.19 are for the "no channel interaction" model, figures 5.20 to 5.22 are for the "1 mm channel interaction" model and figures 5.23 to 5.25 are for the "3.3 mm channel interaction" model. Only mean data are presented.

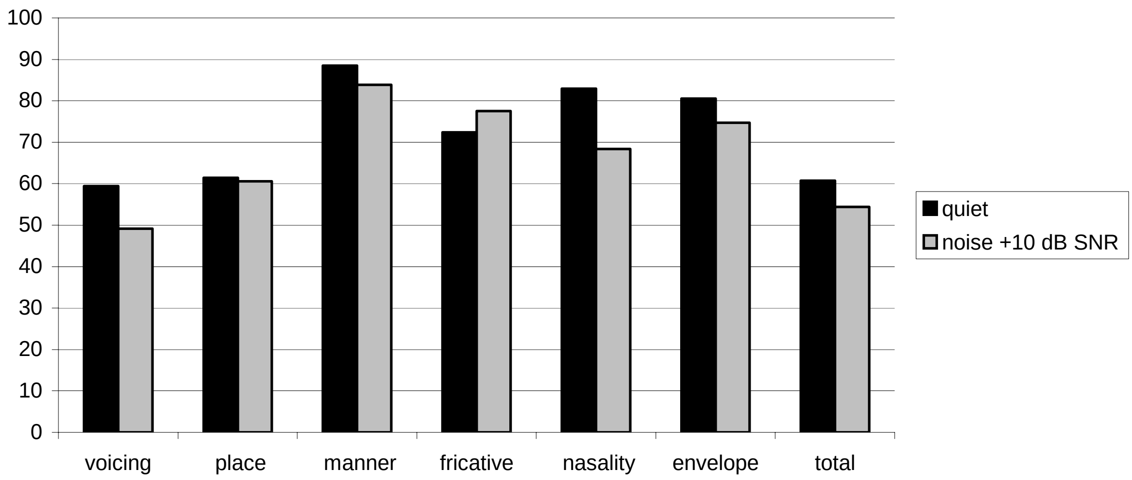

Figure 5.17. Mean feature transmission for the 900 pps/ch x 12/20 MAP condition with and without stationary background noise at +10 dB SNR with an AM with no "channel interaction"

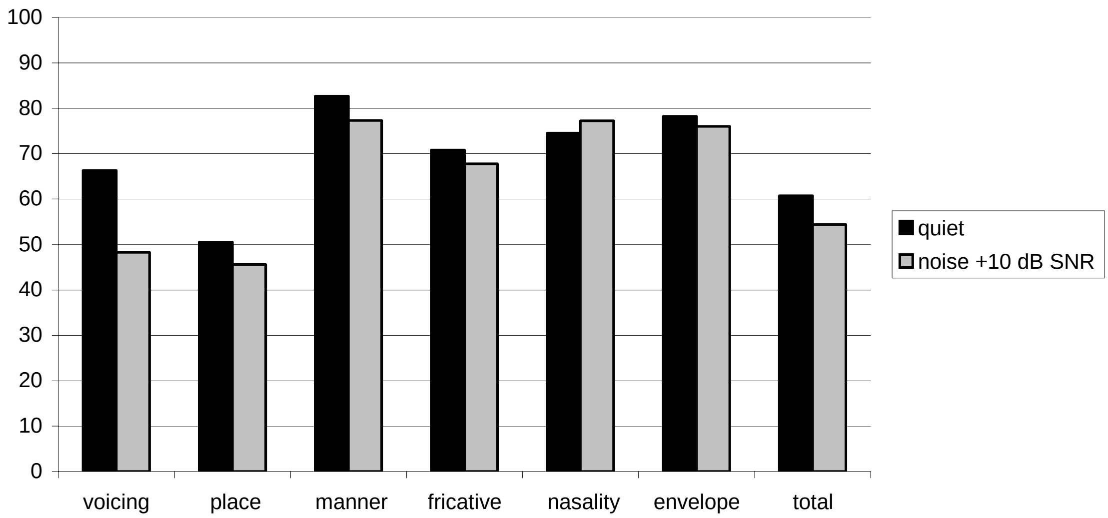

Figure 5.18. Mean feature transmission for the 900 pps/ch x 4/7 MAP condition with and without stationary background noise at +10 dB SNR with an AM with no "channel interaction".

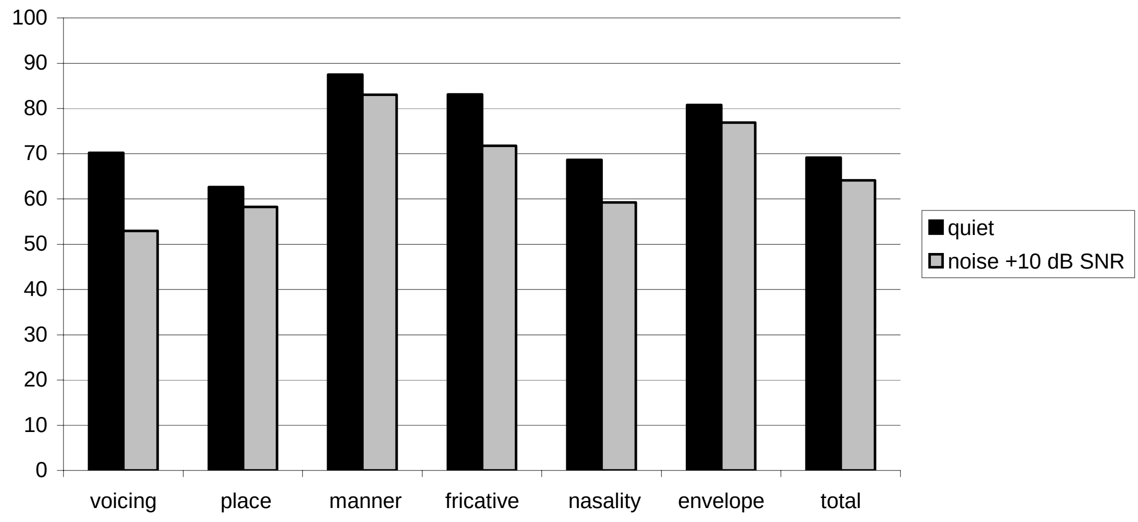

Figure 5.19. Mean feature transmission for the 250 pps/ch x 12/20 MAP condition with and without stationary background noise at +10 dB SNR with an AM with no "channel interaction".

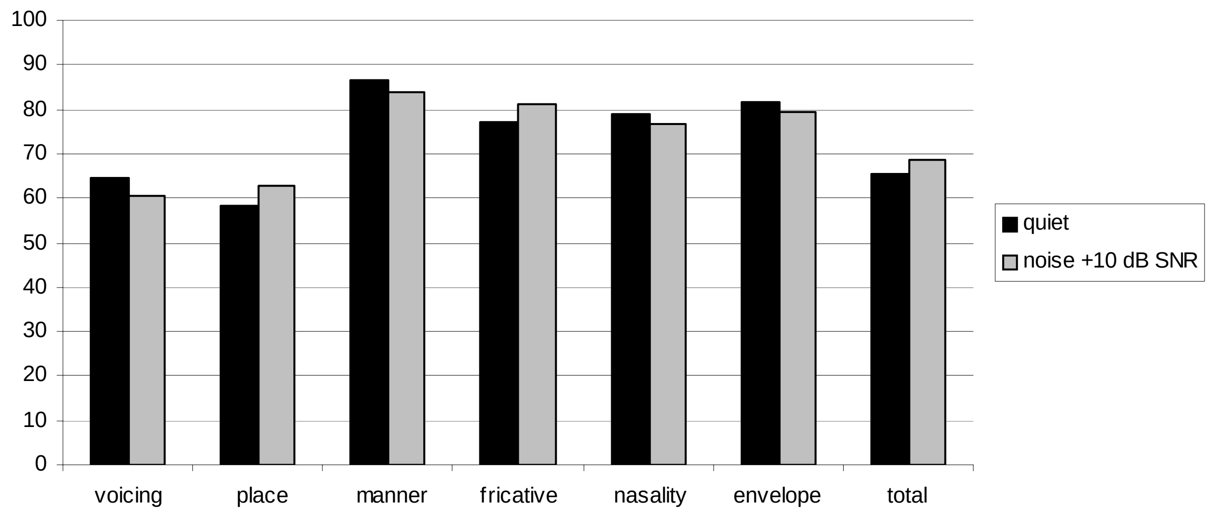

Figure 5.20. Mean feature transmission for the 900 pps/ch x 12/20 MAP condition with and without stationary background noise at +10 dB SNR with an AM with 1 mm "channel interaction"

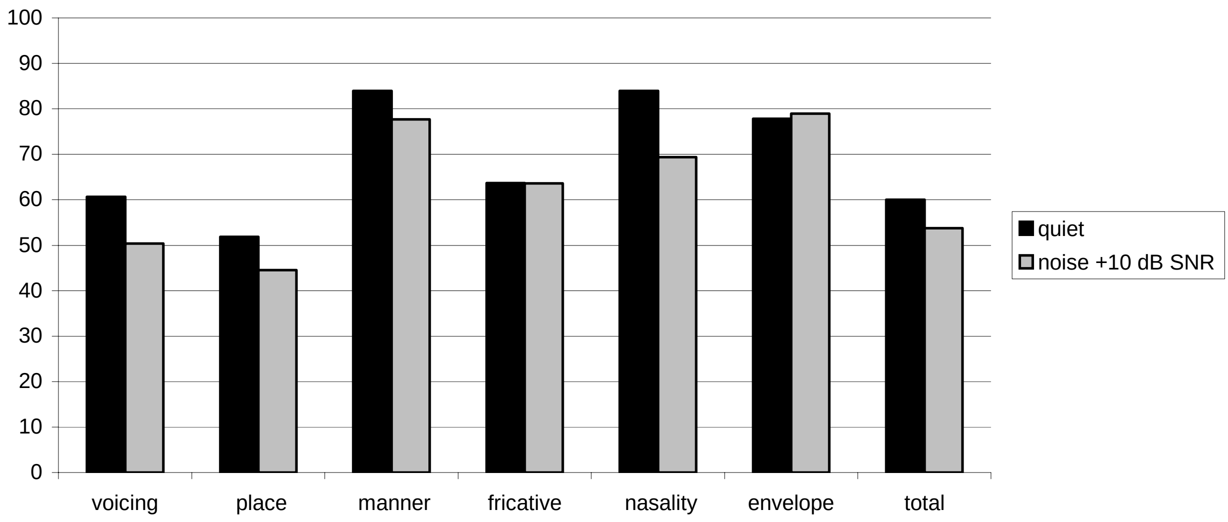

Figure 5.21. Mean feature transmission for the 900 pps/ch x 4/7 MAP condition with and without stationary background noise at +10 dB SNR with an AM with 1 mm“channel interaction”.

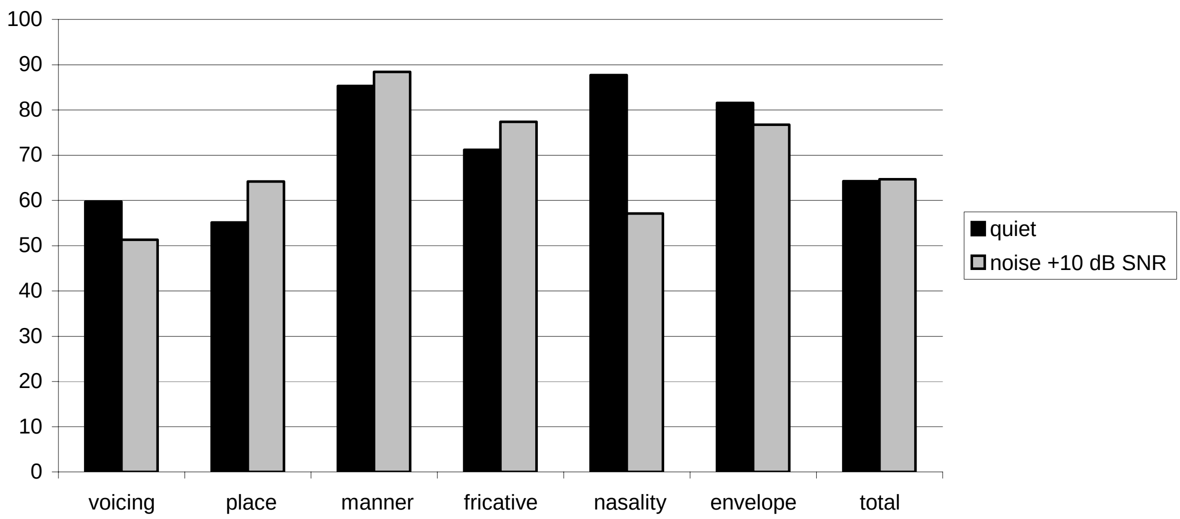

Figure 5.22. Mean feature transmission for the 250 pps/ch x 12/20 MAP condition with and without stationary background noise at +10 dB SNR with an AM with 1 mm “channel interaction”.

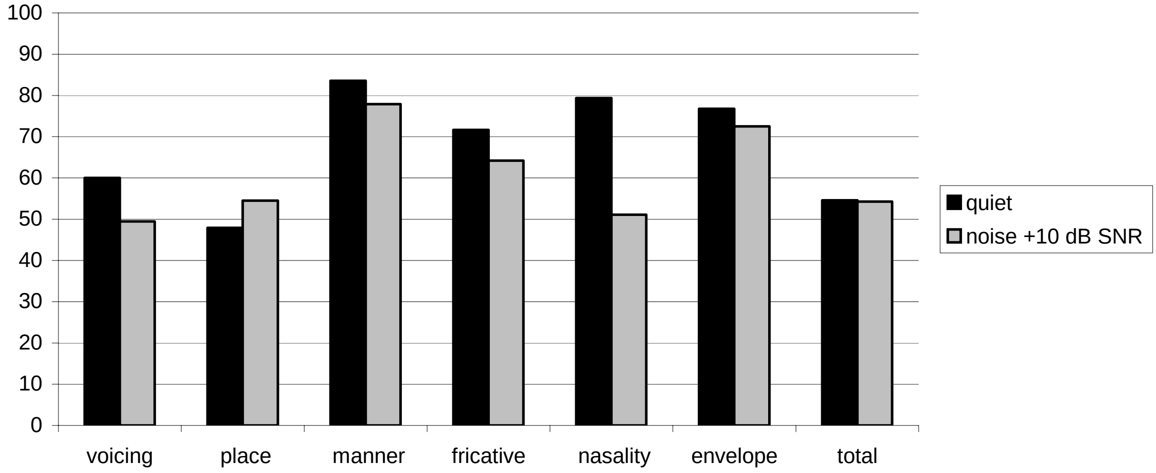

Figure 5.23. Mean feature transmission for the 900 pps/ch x 12/20 MAP condition with and without stationary background noise at +10 dB SNR with an AM with 3.3 mm "channel interaction"

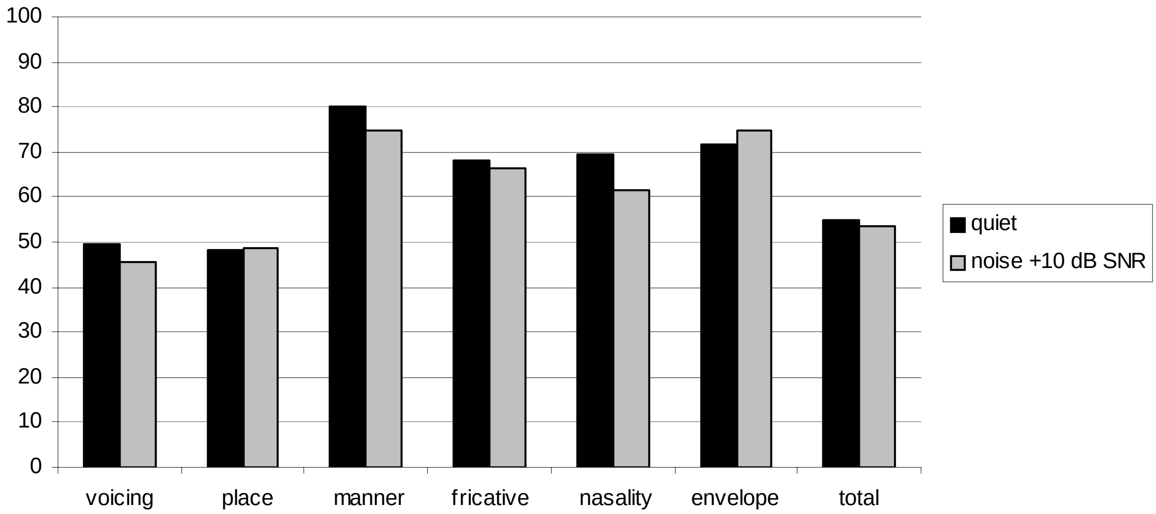

Figure 5.24. Mean feature transmission for the 900 pps/ch x 4/7 MAP condition with and without stationary background noise at +10 dB SNR with an AM with 3.3 mm"channel interaction".

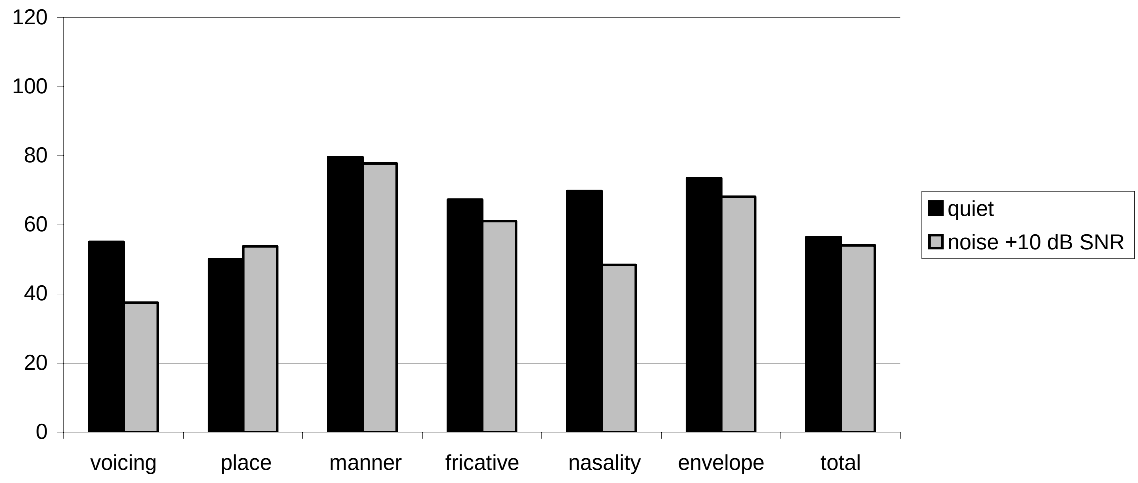

Figure 5.25. Mean feature transmission for the 250 pps/ch x 12/20 MAP condition with and without stationary background noise at +10 dB SNR with an AM with 3.3 mm "channel interaction".

It appears from these figures that differences between processing, or MAP, conditions are small. It is notable that place transmission appears to be less for the 4-channel MAP compared to the high-channel conditions. In order to analyse results, MANOVA was undertaken with the same approach as for analysis of CI user data: again, with seven dependent variables (information transmission for each of the six features and total percent correct). For the main analysis there were four factors: noise, channel number and channel stimulation rate and the additional factor for the AM experiment, channel interaction (with three levels). Again, only significant factors and interactions are mentioned in the text; the full MANOVA report is included in Appendix B.

The noise factor was found to have a significant effect on the features voicing, nasality and manner. In each case the effect was in the expected direction, e.g. worse transmission of those features with the inclusion of background noise at +10 dB SNR. Stimulation rate had no effect on any variable. Channel number had a significant effect on total correct, place, manner and fricative, again in the expected direction, e.g. worse transmission with 4 channels compared to 12 channels.

Channel interaction had a significant effect on total correct, voicing, place, manner, fricative and envelope (all $p<0.05$). As this factor had three levels, it was necessary to use *post-hoc* t-tests to determine where the differences lay. The results are indicated in table 5.3, which shows the magnitude of the difference between each of the three possible comparisons between channel interaction conditions for the six dependent variables which showed a significant channel interaction effect (numbers in bold indicate that the comparison was significant).

Table 5.3.Comparisons between three different channel interaction conditions. Values given are mean differences (rounded to nearest 1%) for the variables indicated on the left. Differences highlighted in bold were statistically significant.

Dependent variable/ comparison	**No channel interaction – 1 mm channel interaction**	**No channel interaction – 3.3mm channel interaction**	**1mm channel interaction – 3.3 mm channel interaction**
Total correct	1	**9**	**8**
Voicing	0	**8**	**8**
Place	0	**6**	**6**
Manner	-1	**5**	**5**
Fricative	2	**7**	**6**
Envelope	-1	**5**	**6**

These tests can be summarised as follows. For the six dependent measures which showed a channel interaction effect, *post-hoc* comparisons between either the "no channel interaction" condition or the "1 mm channel interaction" condition on the one hand and the 3.3 mm channel interaction condition on the other were significant, whereas the comparison between the "no channel interaction" and "1 mm channel interaction" condition was not significant for any variable. Effectively, the "no channel interaction" and "1 mm channel interaction" conditions were equivalent, whereas there were significant differences of between 5 and 9% information transmission for either of these conditions and the 3.3 mm channel interaction condition. There was also a two-way interaction between noise and channel number on place, due to a significant difference between place transmission in quiet vs. noise within the 4-channel condition but not the 12-channel condition. There were two-way interactions between channel interaction and channel number on place and fricative. Because the other interactions involved channel interaction (an inescapably inelegant repetition of the word "interaction"), and because of the importance of determining differences between the three AMs, three further MANOVAs were performed for each of the three channel interaction conditions. In each case the seven dependent variable were analysed in terms of the combined effects of the three factors channel number, stimulation rate and noise but only considering one channel interaction condition in each case. Given the number of possible effects, the results here are tabulated below for clarity.

The differences across the three channel interaction models given in table 5.4 can be summarised as follows. The pattern of noise effects differed in that nasality was the only dependent variable effect for the 1 mm and 3.3 mm channel interaction conditions, whereas voicing and manner were affected in the no channel interaction condition. There were also differences in the pattern of channel number across models. In the condition without channel overlap or with 1 mm channel overlap there were channel number effects for a number of measures while there were no channel number effects in the 3.3 mm channel overlap condition. It is worth noting that the pattern of effects for the 3.3 mm channel overlap condition was the same as for the CI user group, e.g. no effects of any processing parameter condition and noise having an effect on nasality only. The following section explores the relationship between the data sets from the two experiments quantitatively.

Table 5.4. Summary of significant factors/interactions from MANOVAs undertaken separately for three different channel interaction conditions.

Factor (interaction)/ Channel interaction condition	No channel interaction	1 mm channel interaction	3 mm channel interaction
Noise	Voicing, manner	Nasality	Nasality
Channel number	Total, place, manner	Total, place, manner, fricative	-
Stimulation rate	-	-	-
Noise x channel number	-	Place	-
Noise x stimulation rate	-	-	-
Stimulation rate x channel number	-	-	-

5.5 Combined data analysis of AM and CI user experiments

This section aims to quantify the degree of match between the two data sets, in order to address the core questions noted in 5.1. In order to illustrate the similarities or differences in patterns of feature recognition in the two groups, figures 5.26 and 5.27 show performance for CI users and for the three versions of the AM, averaged across MAP conditions, for quiet and noise respectively. The variance has already been shown for most variables and was not substantially different between CI and AM subjects across features. Therefore, in order to facilitate visual comparison between means, figures in this section, other than those concerned with statistical interactions, do not include standard deviation values.

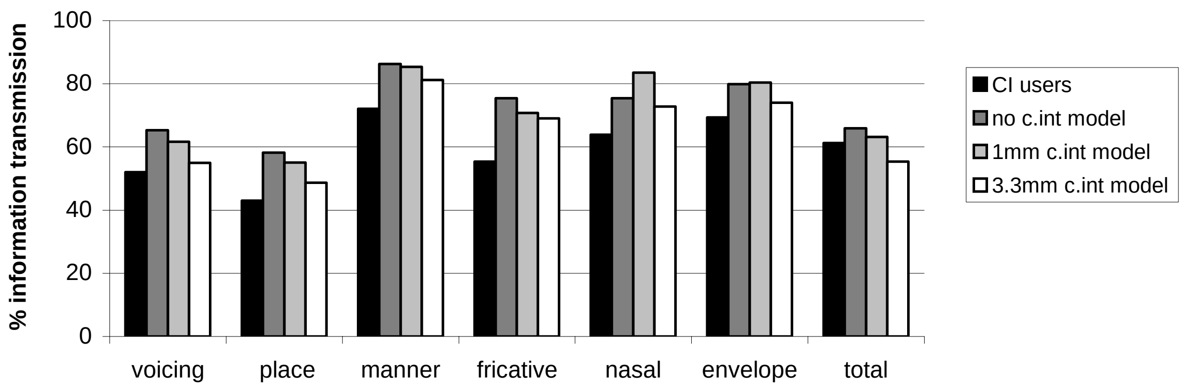

Figure 5.26. Mean feature transmission for CI users and listeners to three different AMs, in quiet.

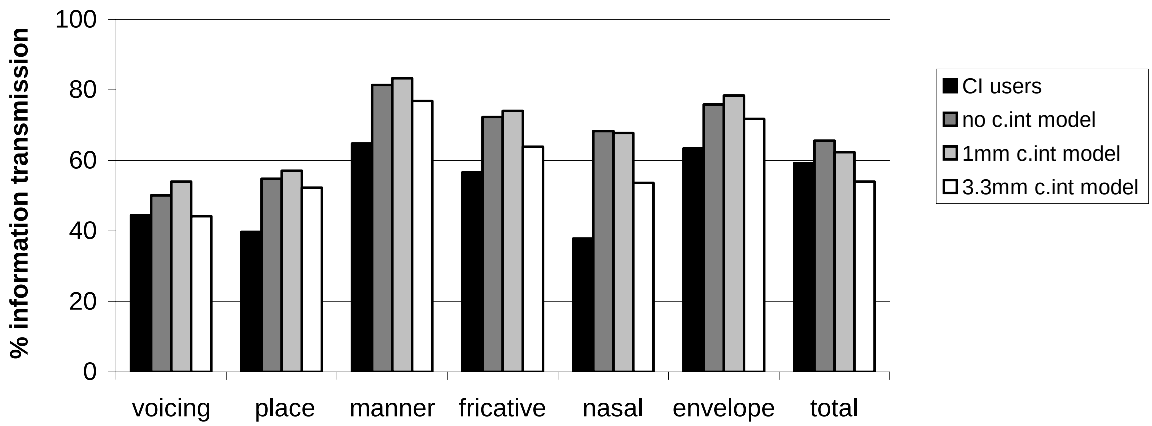

Figure 5.27. Mean feature transmission for CI users and listeners to three different AMs, in stationary background noise at +10 dB SNR.

These figures suggest that the models over-predict absolute transmission values, although the overall pattern between features is similar between model and CI user data. It also appears that the model with greatest channel interaction approximates absolute feature transmission values of CI users most closely, particularly in quiet listening conditions. Given the differences between the models (but particularly between the 3.3 mm channel interaction and other two models), it was appropriate to consider the relationship between CI user data and each separate model, rather than averaging across all the AMs. In any case, this was also necessary in order to address the question of whether inclusion or degree of channel interaction affected the predictive power of the model. Therefore, a series of MANOVAs were undertaken; in each case, a "group" factor was included. This factor distinguished between CI users and AM subjects and therefore had two levels. If this factor was found to be significant, it could be assumed that the model was not a good predictor of performance, e.g. the analysis showed a significant difference between AM and CI user results. On the other hand, if the factor was not significant, the model would be shown to have predictive power (e.g. indicating no difference between AM and CI user results). The other factors, namely noise, channel interaction and channel stimulation rate were not of particular interest in themselves as any difference in the significance of these factors between these analyses and analyses with either CI user data or AM data only would not be very meaningful. However, both the "group" factor and any interactions between "group" and the other factors were of interest. Where there were interactions involving the group factor, this would be indicative that the model was predictive of CI user results in one level of the second factor but not the other. For example, if there were an interaction between "group" and "noise" for any particular analysis, this would indicate that the AM in question had predictive power for either the quiet or the noise-contaminated condition, but not both.

The first set of MANOVA analyses were undertaken across all CI users and each of the three AMs. Significant factors are summarised in table 5.5 (as there were no significant interactions, these are not included in the table). Full details of all MANOVA analyses are given in Appendix B. With subsequent analyses comparing groups, particularly where further subgroup analyses are reported in tables 5.6 and 5.7, the possible deficit in statistical power, noted in 5.2, should be borne in mind.

The relatively small sample size, particularly with regard to subgroup analyses, does increase the possibility that a type II error. Nevertheless, it can also be argued that the group numbers still equal, or exceed, those obtained in numerous reported CI and AM studies.

Table 5.5. Summary of significant factors/interactions from MANOVAs undertaken separately for three different channel interaction conditions. Here data from all 9 subjects in the CI user experiment are also included in the analysis. The "group" factor has two levels (CI users vs. AM).

Factor (interaction)/ Channel interaction condition	No channel interaction	1 mm channel interaction	3.3 mm channel interaction
Group	All six features and total correct	All six features and total correct	Total, place, manner, fricative, envelope
Noise	Voicing, nasality, manner	Nasality	Nasality and manner
Channel number	Place	Place	-
Stimulation rate	-	-	-

This analysis shows that the model was poor in predicting CI performance in all cases except for voicing and nasality in the 3.3 mm channel interaction model. This corresponds well with the impression given by figure 5.26 that the model, even with 3.3 mm channel interaction over-predicts feature transmission values across almost all features. However, it was also important to determine the predictive power of the models separately for better- and worse-performing CI users, as defined in 5.3. Figures 5.28 to 5.31 shows feature transmission values for the two CI user subgroups separately for quiet and noise-contaminated conditions.

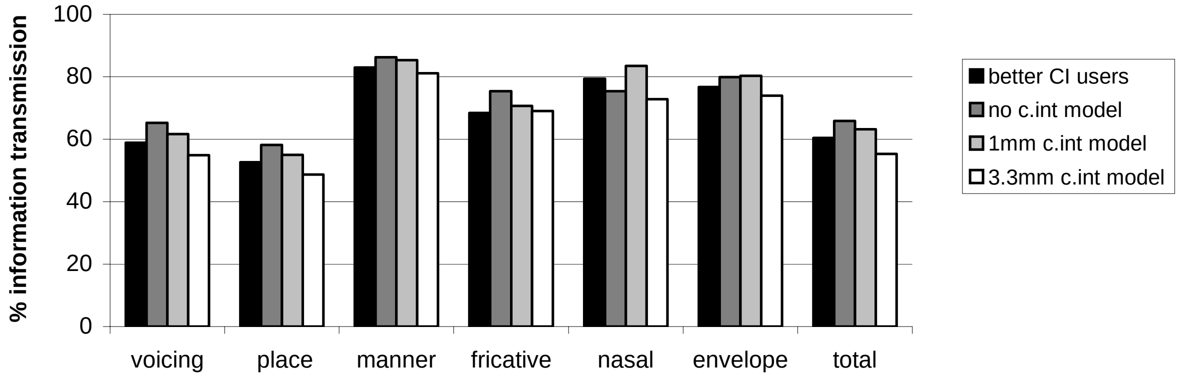

Figure 5.28. Mean feature transmission for CI users with baseline consonant recognition of 50% or more and listeners to three different AMs, in quiet.

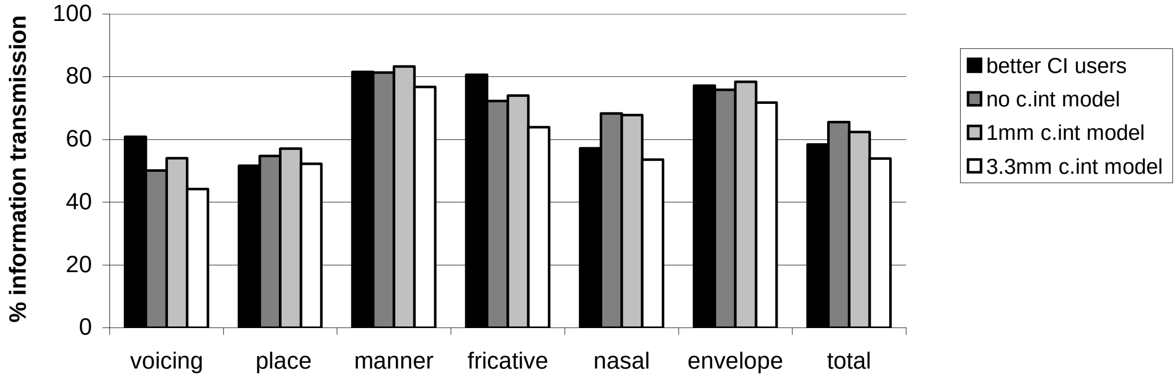

Figure 5.29. Mean feature transmission for CI users with baseline consonant recognition of 50% or more and listeners to three different AMs, in background stationary noise at +10 dB SNR.

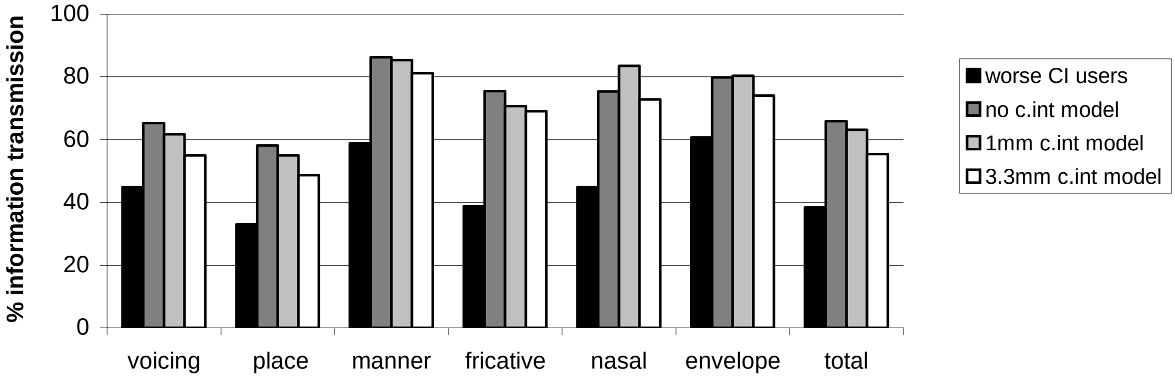

Figure 5.30. Mean feature transmission for CI users with baseline consonant recognition of less than 50% and listeners to three different AMs, in quiet.

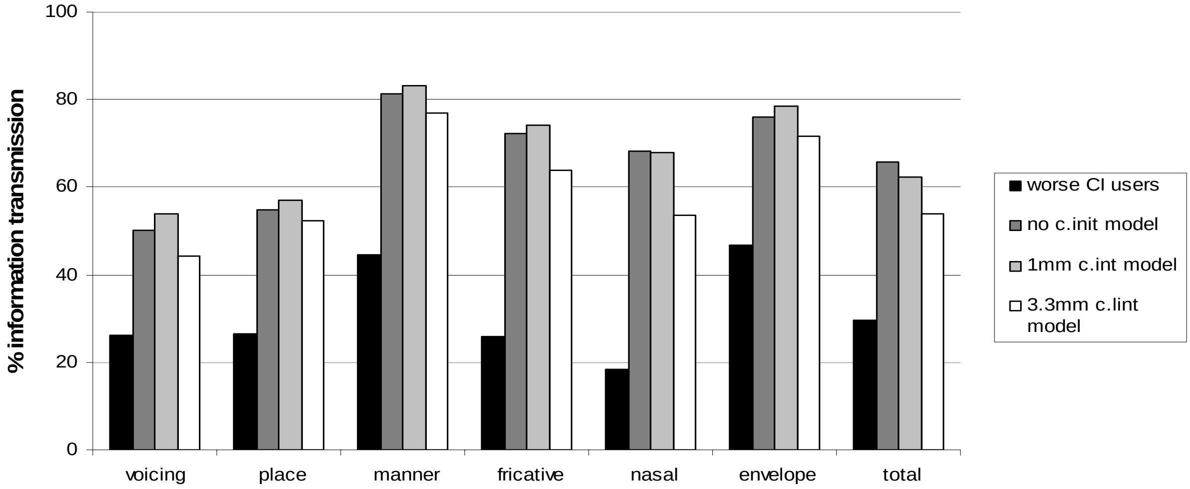

Figure 5.31. Mean feature transmission for CI users with baseline consonant recognition of less than and listeners to three different AMs, in background stationary noise at +10 dB SNR.

The figures suggest that the AM is predictive of performance of "better" CI users whereas it markedly over-estimates performance for "worse" CI users. In order to determine this quantitatively, a further two sets of MANOVAs were undertaken in the same way as the analysis summarised in table 5.6, but in this case including either the "better user" and "worse user" subgroup only. These analyses are reported in tables 5.6 and 5.7, respectively; full MANOVA details are given in Appendix B.

Table 5.6. Summary of significant factors/interactions from MANOVAs undertaken separately for three different channel interaction conditions. Here only data from the better CI users (N=5) in the CI user experiment are included in the analysis-see 5.3 for definition of the subgroup. The "group" factor had two levels (CI users vs. AM).

Factor (interaction)/ Channel interaction condition	No channel interaction	1 mm channel interaction	3 mm channel interaction
Group	-	-	Voicing, fricative
Noise	-	Nasality, fricative	Nasality
Channel number	-	-	-
Stimulation rate	-	-	-
Noise x channel number	-	-	-
Noise x stimulation rate	-	-	-
Noise x group	Fricative, voicing		Fricative
Channel number x group	Place	Place	-
Stimulation rate x channel number	-	-	-

The striking finding from table 5.6 is the lack of effects for the "group" factor. This indicates that the model works well for the subgroup of better users. Interestingly, the model appears to be better (e.g. for a larger number of features) for the conditions with no channel interaction or 1 mm channel interaction.

There are some interactions involving the group factor that need to be considered. First, there was an interaction between group and channel number with place transmission, due to the pattern of effects of channel number on place: There was a significant effect of channel number on place with AMs with no channel interaction ($p < 0.001$) or 1 mm channel interaction ($p < 0.001$). However, for CI users, and for the 3.3 mm channel interaction condition, there was no significant difference in place transmission between 12 and 4 channels. The interaction is shown in figure 5.32.

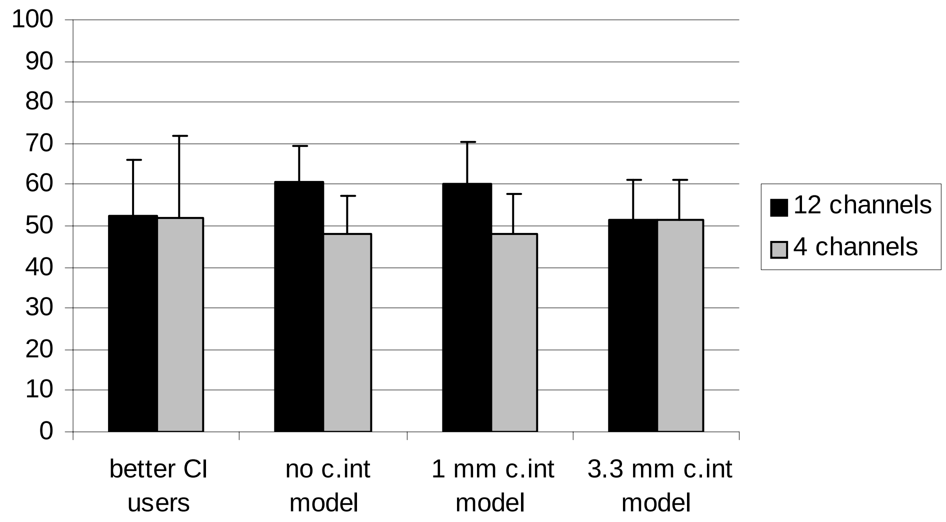

Figure 5.32. Mean (+ 1 SD) place transmission in 12-channel and 4-channel listening conditions, across better CI users and the three AMs.

A further interaction between noise and group on fricative occurred for both the no- and 3.3 mm channel interaction conditions (the effect falls just short of significance in the 1 mm channel interaction condition). Therefore, it appears that the amount of channel interaction is not the only issue, but rather the estimation of the noise effect on fricative transmission in particular. As it happens that the 3.3 mm model falls down with respect to transmission of voicing and fricative, it is convenient to illustrate both the interactions between group and noise and this effect via figures focusing on these two features. Figure 5.33 shows fricative transmission in quiet and noise across better CI users and the three AM. Figure 5.34 shows equivalent data for voicing. In figure 5.33 it can be seen that fricative transmission actually gets better in background noise whereas the differences between quiet and noise conditions are smaller for the AM conditions. This is supported by *post-hoc* t-tests: The difference with and without noise was significant for better CI users ($p<0.05$) but not for no channel interaction, ($p= 0.37$), 1 mm channel interaction ($p= 0.37$) or 3.3 mm channel interaction ($p= 0.14$).

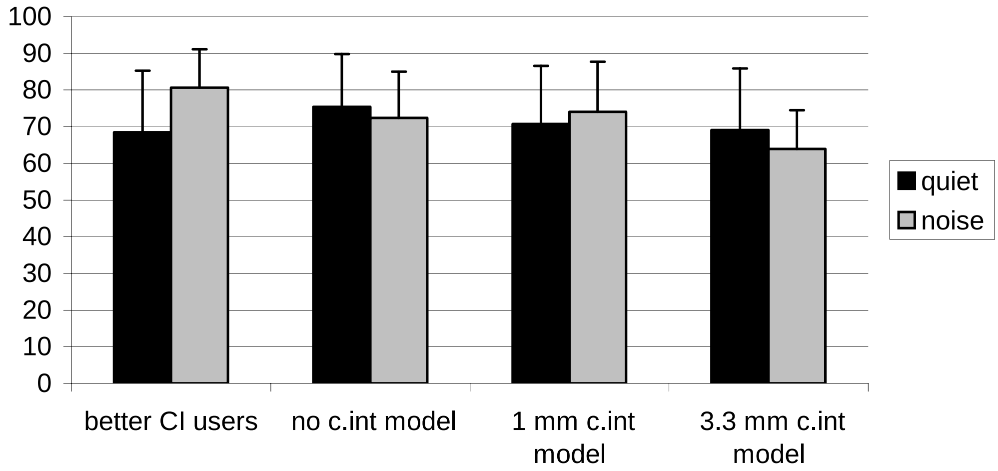

Figure 5.33. Mean (+ 1 SD) fricative transmission in quiet and noise listening conditions, across better CI users and the three AMs.

For voicing transmission, the interaction between noise and group appears to be for rather different reasons. Figure 5.34 suggests no difference with the addition of noise for voicing with better CI users, but there is a clear difference with noise across AMs, although the difference is less for 1 mm channel interaction. *Post-hoc* t-tests were non-significant for better CI users ($p=0.78$), significant for the no channel interaction model ($p<0.01$), not significant for the 1 mm channel interaction model ($p=0.12$) and just short of reaching significance for the 3.3 mm channel interaction model ($p=0.051$).

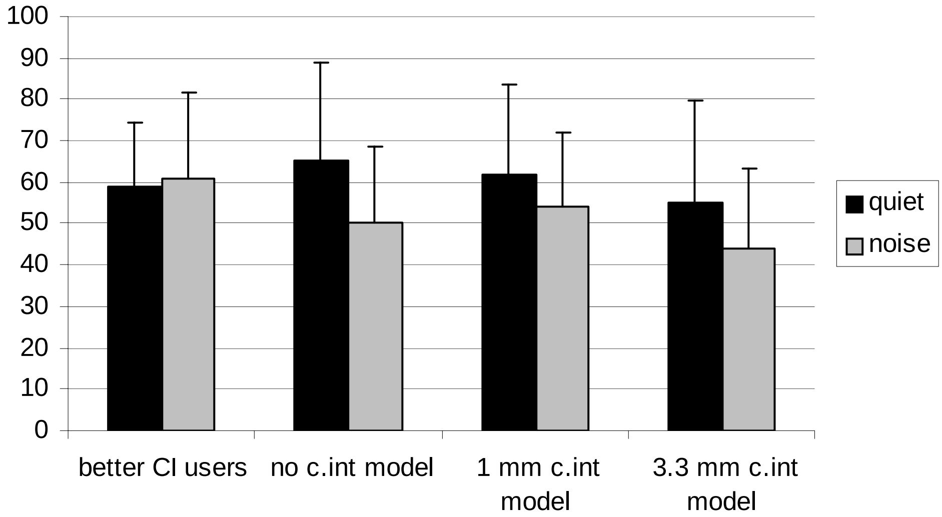

Figure 5.34. Mean (+ 1 SD) voicing transmission in quiet and noise listening conditions, across better CI users and the three AMs.

As noted, a final MANOVA comparing worse CI users with AMs was undertaken and significant factors and interactions are tabulated below.

Table 5.7. Summary of significant factors/interactions from MANOVAs undertaken separately for three different channel interaction conditions. Here only data from the worse CI users (N=4) in the CI user experiment are included in the analysis-see 5.3 for definition of the subgroup. The "group" factor has two levels (CI users vs. AM).

Factor (interaction)/ Channel interaction condition	No channel interaction	1 mm channel interaction	3.3 mm channel interaction
Group	All features and total correct	All features and total correct	All features and total correct
Noise	Total and all features except nasality	Voicing, nasality, manner, envelope	All features except place
Channel number	Total, place,manner	Total, place, manner, fricative	
Stimulation rate			
Noise*channel number			
Noise*stimulation rate	-	-	-
Noise*group	Manner	Manner, fricative, envelope	
Stimulation rate*channel number	-	-	-

These analyses clearly show that the model is very poor when considering those with baseline consonant recognition scores <50%, given that the "group" factor was significant for all dependent variables and for all AM conditions. Some two-way interactions between noise and group were found for manner transmission. Figure 5.35 shows manner transmission for the two AMs in which there was a significant interaction. It appears that there was a large noise effect for the worse users but not for the AM conditions. *Post-hoc t*-tests showed a significant difference with and without noise for worse CI users ($p<0.005$) and the no channel interaction AM ($p<0.05$) but not the 1 mm channel interaction model ($p=0.31$) or the 3.3 mm model ($p=0.11$).

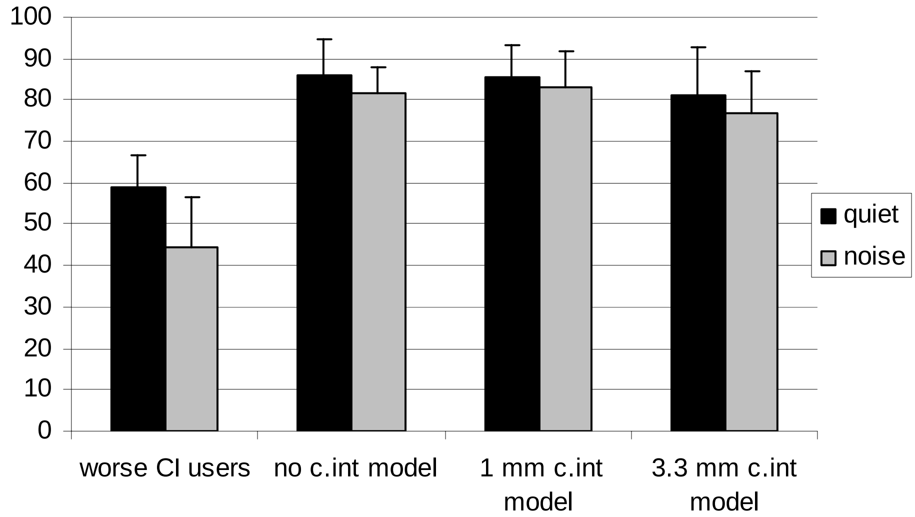

Figure 5.35. Mean (+ 1 SD) manner transmission in quiet and noise listening conditions, across worse CI users and two of the AMs.

The final two interactions noted were for the 1 mm channel interaction model for fricative and envelope. Figures 5.36 and 5.37 show transmission of those two features in quiet and noise within this model and for worse CI users. It appears that, in both cases, the interaction is produced by the larger noise effect in worse CI users than in the AM. For fricative transmission, *post-hoc* t-tests showed no difference with noise for either worse CI users ($p=0.077$) or the AM ($p=0.36$), although the difference in CI users just failed to reach significance. For envelope transmission, the noise effect for worse CI users was significant ($p<0.05$) but was not for the AM ($p=0.16$).

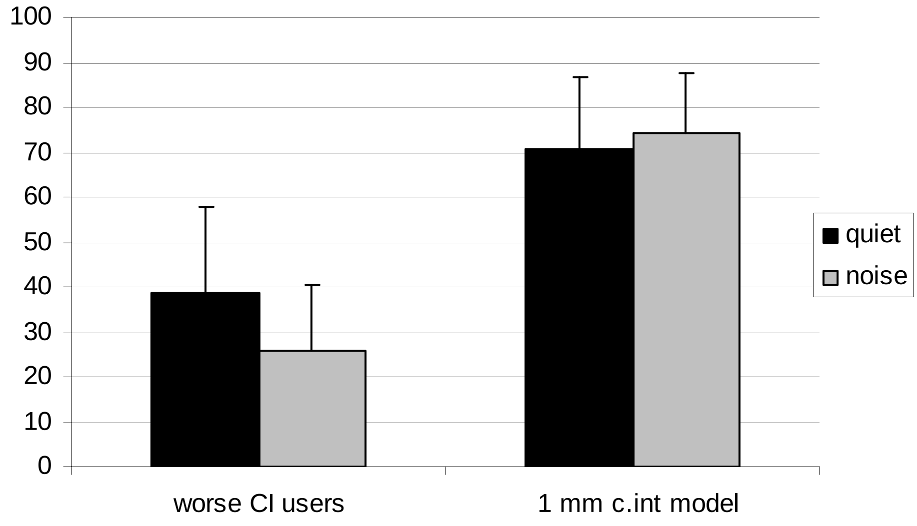

Figure 5.36. Fricative transmission in quiet and noise listening conditions, across worse CI users and 1 mm channel interaction model.

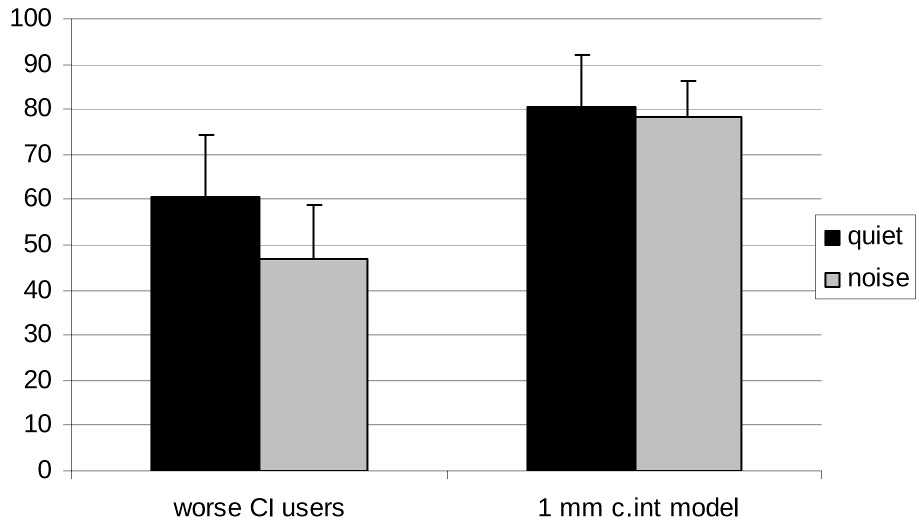

Figure 5.37. Envelope transmission in quiet and noise listening conditions, across worse CI users and 1 mm channel interaction model.

A final issue in presentation of the data is the question of whether differences between channel interaction can mimic differences between better and worse users. In order to illustrate whether this might the case, differences between "no channel interaction" and the "worst" "3.3 mm channel interaction" AM conditions are presented in figure

5.38 while differences between “better” and “worse” CI users are presented immediately below in figure 5.39. (data are averaged across MAP conditions in both cases).

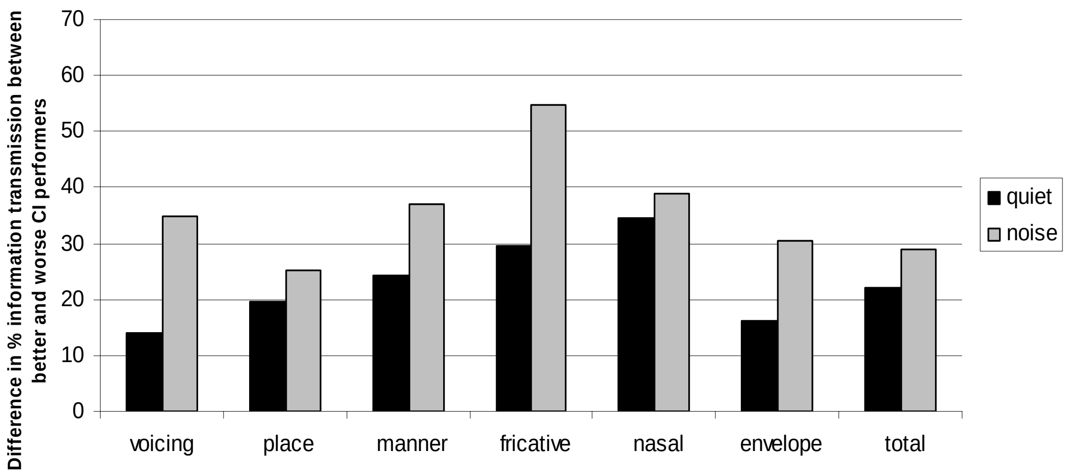

Figure 5.38. Difference between better and worse CI performers across MAPs.

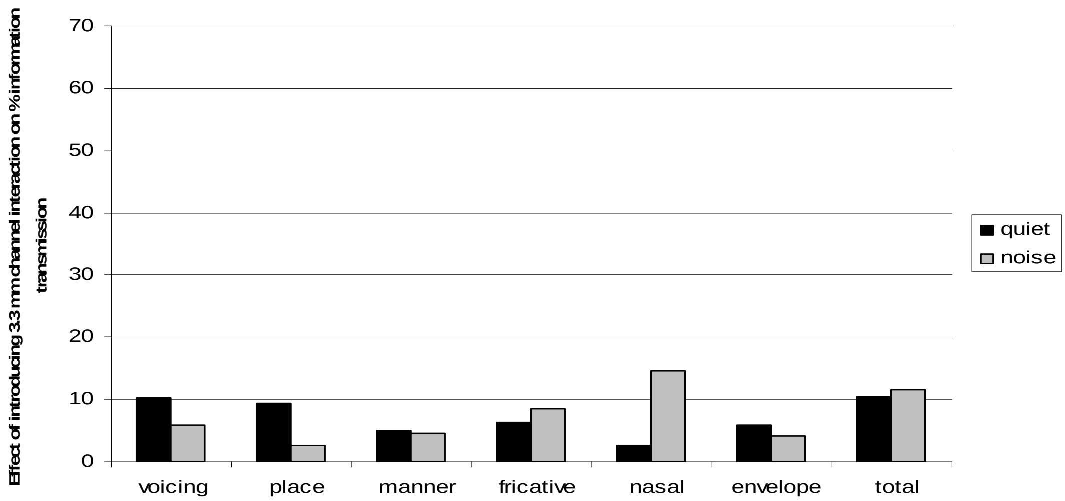

Figure 5.39. Difference between AM with no channel interaction and AM with 3.3 mm channel interaction, across MAPs. Values are no channel interaction conditions – equivalent 3.3 mm channel interaction conditions.

The clear impression is that the two patterns do not match. The absolute magnitude of differences between better and worse CI users far exceeds that for the difference between no channel interaction and 3.3 mm channel interaction. Moreover, the differences between better and worse CI users are greater for the noise-contaminated conditions whereas this is not generally the case for the AM differences.

A further useful way of looking at the data is in terms of information loss between conditions. Figures 5.40 and 5.41 illustrate this by showing three comparisons: first, between the unaltered condition and the AM with no channel overlap (referred to as "processing effect", second, the difference between the unaltered condition and the AM with 3.3 mm overlap (referred to as "processing +channel interaction effect") and, third, the difference between NH listeners in the unaltered condition and better CI users ("CI user effect").

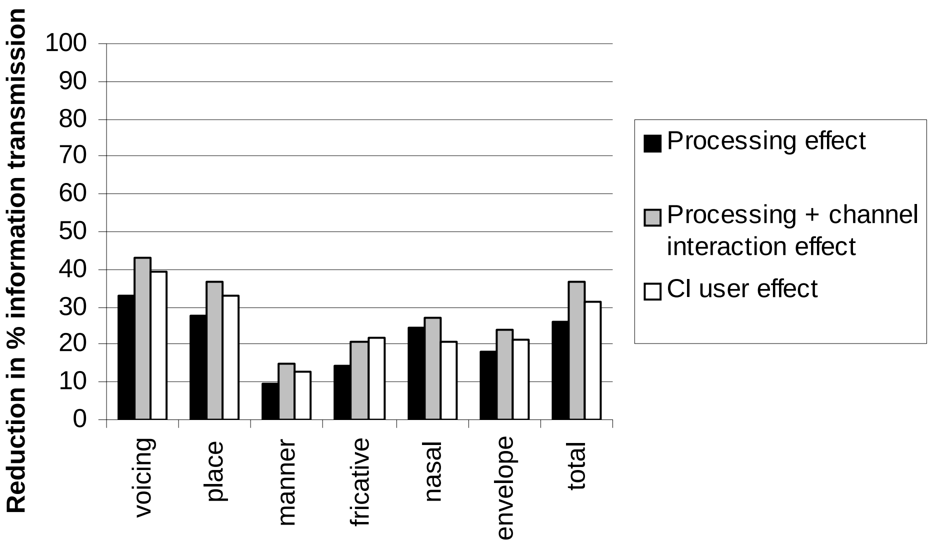

Figure 5.40. Reduction in mean feature transmission across listener and channel interaction conditions in quiet. The first comparison is between NH listeners with the unaltered stimuli vs. stimuli with AM with no channel overlap. The second comparison is between NH listeners with the unaltered stimuli vs. stimuli with AM with 3.3 mm channel overlap. The third comparison is between NH listeners with unaltered stimuli and better CI users. Data are averaged across MAP conditions in quiet.

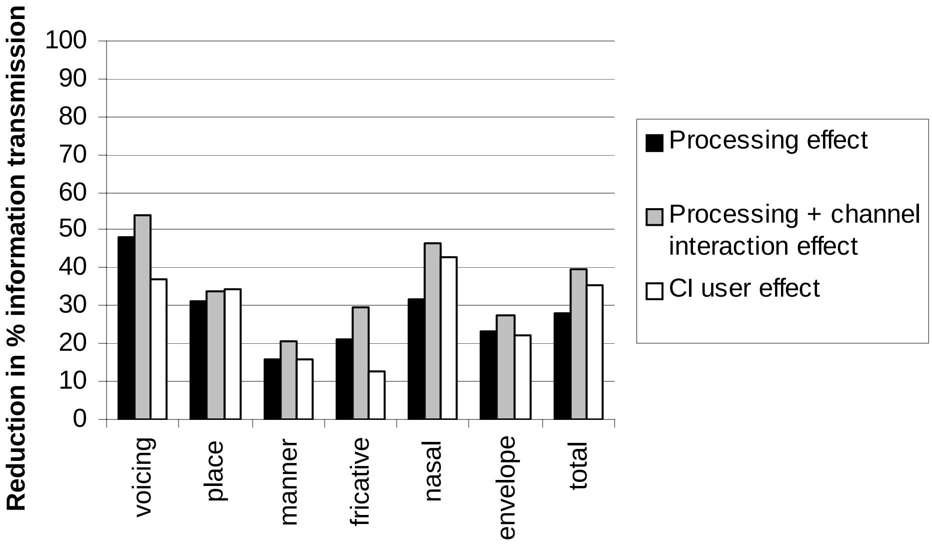

Figure 5.41. Reduction in mean feature transmission across listener and channel interaction conditions in background noise. The first comparison is between NH listeners with the unaltered stimuli vs. stimuli with AM with no channel overlap. The second comparison is between NH listeners with the unaltered stimuli vs. stimuli with AM with 3.3 mm channel overlap. The third comparison is between NH listeners with unaltered stimuli and better CI users. Data are averaged across MAP conditions in background noise at +10 dB SNR.

The noteworthy aspect of the data presented in this way is the equivalence in magnitude between the "CI user effect" and "processing + channel interaction effect", at least in quiet listening conditions. Interestingly, the effect of adding channel interaction does not reduce information transmission anything like as much as the "processing effect", e.g. the effect of the AM without channel interaction. The reduction in performance in the better CI users is modelled well by using an AM with 3.3 mm channel interaction, but the contribution of channel interaction is a smaller part of the overall effect than the effect of processing *per se* (albeit with pitch shift included also). This important finding is expanded on in chapter 7.

5.6 Consonant confusion matrices

Analysis of information transmission measures was used to determine the correspondence between CI and AM data in quantitative terms. However, it is also important to note, qualitatively, what types of phoneme error were made, given the possibility that two different error patterns could lead to the same information transmission values. Evaluation of every consonant confusion table for each listening condition for both experiments would be impractical. Instead, just a handful of important confusion matrices are considered here. Given the findings of analyses in 5.5 that indicated a strong convergence between AMs and CI user data, the most important comparison was between confusion matrices for "better" CI users and AMs. Therefore, tables 5.8 to 5.12 show confusion matrices for better CI users, no channel interaction AM and 3.3 mm channel interaction AM, respectively. Given the minimal effects of processing parameters across experimental conditions, data are averaged across the three MAPs. The matrices are briefly discussed in this section and referred to at various points in the discussions in chapters 6 and 7.

Table 5.8. Confusion matrix for better CI users in quiet, averaged across MAP conditions.

	b	d	g	w	j	ɹ	l	v	z	ʤ	m	n	p	t	k	f	ɵ	s	ʃ	ʧ
b	100																			
d		95							2					2						
g		64	14								2		14	2			2			
w				50		43					7									
j	2				17		36	12	2	2	21						7			
ɹ				24		76														
l				14		17	33	10			19	5				2				
v								88			2			5	5					
z								14	50							5	7	17	7	
ʤ		5	7		2					76				7						2
m							2				9	7								
n							2				86	12								
p													100							
t										2				79	2					17
k		2	2							12			17	14	50					2
f								12								71	5	12		
ɵ									10					2		7		76	5	
s								2	7								2	81	5	2
ʃ																5	5	5	86	
ʧ		5	7							38				7					2	40

Table 5.9. Confusion matrix for no channel interaction AM, averaged across MAP conditions.

	b	d	g	w	j	ɹ	l	v	z	ʤ	m	n	p	t	k	f	ɵ	s	ʃ	ʧ
b	66							1		1			31							1
d		8								1				15	3					1
g		19	46				1			2			2		29					
w				13		86												1		
j	1		1	1	69	1	16	1	2		1		2				4	1		
ɹ						100														
l						3	96					1								
v			1			25	2	61	9		1							1		
z				1		2		1	84								2	9	1	
ʤ			11							73				1		1				14
m						23	1				71	5								
n				1		14					84	1								
p			1						1	4			88	1	5					
t		2	2							6				57			2			31
k	1		1						1	7			9	12	66					3
f								2	2							62	17	15	1	1
ɵ									5							5	15	74		1
s									2								2	93	3	
ʃ																2	4	6	88	
ʧ		1	5							15										79

Table 5.10. Confusion matrix for 3.3 mm channel interaction AM, averaged across MAP conditions.

	b	d	g	w	j	ɹ	l	v	z	ʤ	m	n	p	t	k	f	ɵ	s	ʃ	ʧ
b	45	5						2		1			43		1	2				
d	2	62	2							4			11	14	2		1			2
g	6	2	22							3		1	25	1	20					1
w				38		38	1	12			7	3								
j	2	1	1	4	45	8	17	1	1		2	3	2				1	1		1
ɹ				24	1	56	8	5	1		2	2						1		
l				5	2	11	56	12			7	5			1		1			
v				1		2		8	6					1		6	2	2		
z								8	75							7		7	3	
ʤ			11							74										15
m				1		5	2				86	6								
n		1		1		2	2			1	82	10	1							
p	1		1			1				2			74	2	18					1
t			1							10				40	1	1	1			45
k		1	2	1		1		1		2	1		18	26	38		2			6
f		2							1	1						79	11	4	2	
ɵ								3	5							48	21	20	2	
s								1	7							6		66	20	
ʃ									1	1						1	3	4	81	
ʧ		1	4							11										84

The most striking impression from the confusion matrices is the high degree of correspondence for error patterns for most, but not all, consonants. Both models corresponded well with better CI users with respect to the perception of nasals (in particular the misperception of /n/ as /m/), the perception of voiceless fricatives, including the misperception of /θ/ as /s/. The 3.3 mm channel interaction corresponded better in terms of liquid confusion patterns, e.g the high error rates for /ɹ/ and /w/. However, there were a number of areas where the better CI user confusion patterns diverged from the model data: in particular, the perception of bilabial stops (/b/ and /p/) was notably better with better CI users than with AM listeners.

Table 5.11. Confusion matrix for CI users in quiet, averaged across MAP conditions.

	b	d	g	w	j	ɹ	l	v	z	ʤ	m	n	p	t	k	f	θ	s	ʃ	ʧ
b	91	5						1						1	1					
d	1	92	1						1				1	3						
g	4	64	9								1		12	8	1		1			
w				27	1	44	19				8			1						
j	1	3			13	1	33	13	1	5	17	5				1	5	1		
ɹ				13		62	19	1			4							1		
l				9		13	51	6			13	4			3	1				
v	1		9		1	1		63			1		15	3	3	1	1			
z		6	1					19	40							4	4	13	6	6
ʤ		6	40		1					41				9						3
m						4	18	1			69	8								
n						1	14	1			63	19						1		
p	3												87		10					
t		1	6							3			1	74	3		3			9
k		1	4			1				8			19	28	37					1
f		1						14		1				1		56	6	8	5	6
θ			3		3		1	4	8					1		5	3	51	5	17
s		1	1					1	6				1				4	6	12	13
ʃ			3							3				1		3	4	3	79	5
ʧ		4	19							22				19			3		3	31

First it is important to note whether two specific hypotheses, which were put forward in 1.7.3, are supported by the data. The first stated that place coding for nasals and liquids should be more difficult for CI users than fricative or plosive place. This is supported by the confusion matrices for the better CI users.

Table 5.12. Confusion matrix for worse CI users in quiet, averaged across MAP conditions.

	b	d	g	w	j	ɹ	l	v	z	ʤ	m	n	p	t	k	f	ɵ	s	ʃ	ʧ
b	81	11						3						3	3					
d	3	89	3										3	3						
g	8	64	3										8	14	3					
w					3	44	42				8			3						
j		6			8	3	31	14		8	11	11				3	3	3		
ɹ						44	42	3			8							3		
l				3		8	72	3			6	3			6					
v	3		19		3	3		33					33			3	3			
z		14	3					25	28							3		8	6	14
ʤ		8	78											11						3
m						8	36	3			44	8								
n						3	28	3			36	28						3		
p	6												72		22					
t		3	14							3			3	69	3		6			
k			6			3				3			22	44	22					
f		3						17		3				3		39	8	3	11	14
ɵ			6		6		3	8	6							3	6	22	6	36
s		3	3						6				3				6	36	19	25
ʃ			6							6				3			3		72	11
ʧ		3	33							3				33			6		3	19

The most striking difference in "worse CI user" error patterns was the worse nasality transmission. Specifically, these subjects consistently confused the nasals for the liquid /l/, suggesting a reduced ability to determine the pattern of envelope modulations within apical channels.

5.7 Overview of experiments 3 and 4

Two experiments were undertaken, the aims of which were to determine the relative contributions of CI processing and electrical/neural interface factors on consonant recognition. It was found that the results of the AM experiment matched the results obtained with a subset of the CI users. It was also found that the magnitude of deterioration in consonant recognition as a consequence of CI processing was markedly greater than the effect of channel interaction. The results are discussed from two perspectives. The first perspective is the question of information transmission of specific features. This is dealt with in the subsequent chapter, while chapter 7 provides a more general discussion of findings.

Chapter 6. Analysis of consonant feature transmission

In this chapter the transmission of specific consonant features is considered. In 6.1 differences between findings in experiments 3 and 4 and previously available findings are considered. Sections 6.2 to 6.7 considers transmission of the six main features separately.

6.1 Comparison with available data

Section 2.1.2 outlined available evidence on voicing, place and manner transmission in quiet for CI users from recent studies. Figure 2.3 is repeated here as figure 6.1, for convenience, in order to compare against performance obtained in experiments 3 and 4. Although the data from figure 6.1 were collected from users of a variety of different CI devices and with a variety of different approaches to consonant confusion testing (see table 2.2 and the discussion in 2.1.2 for an overview), it is of interest to note whether the present experimental work replicated the same pattern of performance. For comparison purposes, figure 6.2 shows percentage information transmission for voicing, place and manner in quiet across experiments 3 and 4, averaged across MAP conditions.

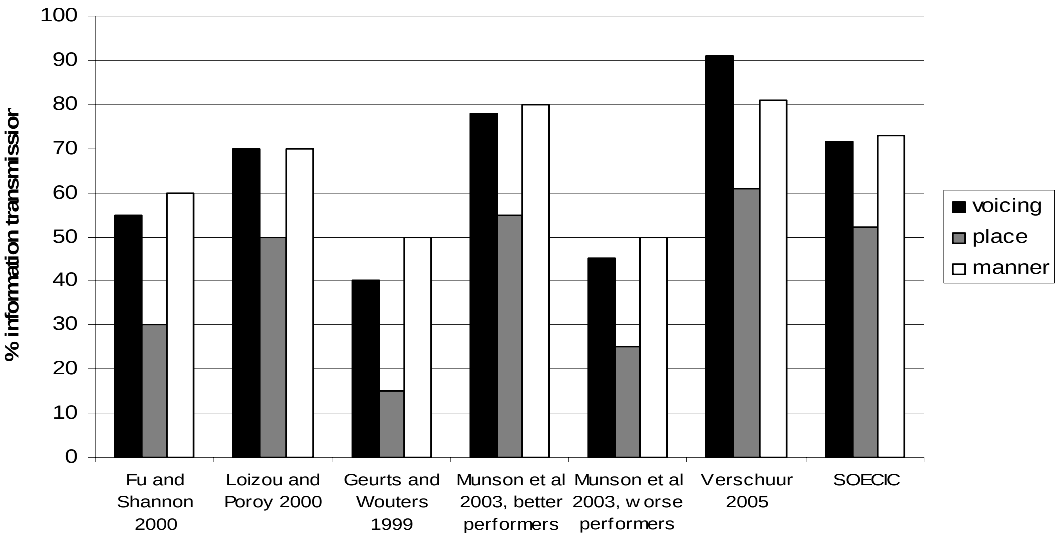

Figure 6.1. Mean consonant voicing, place and manner transmission from various studies of CI user performance (repeats figure 2.3).

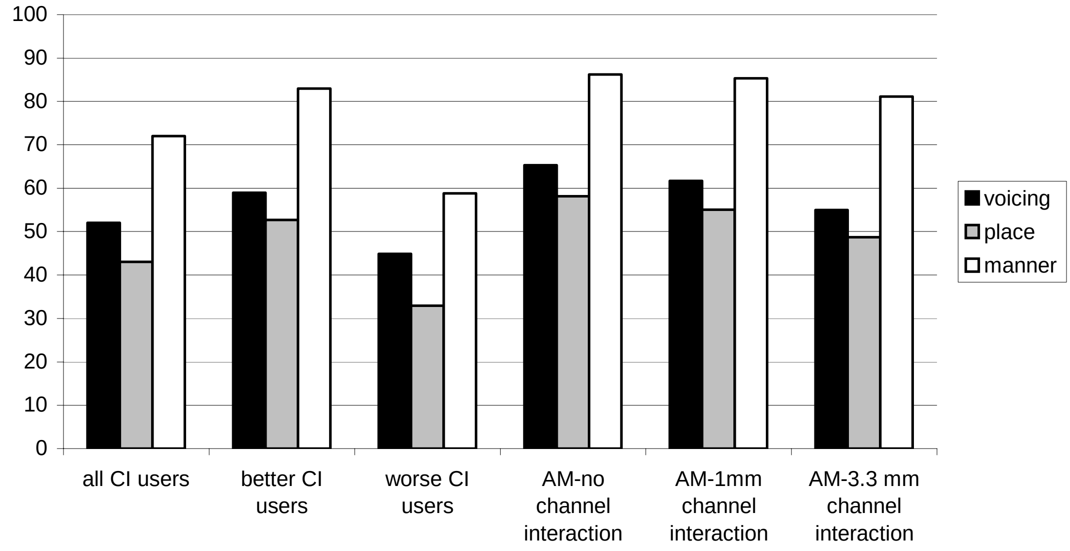

Figure 6.2. Mean voicing, place and manner in quiet across MAP conditions from experiments 3 and 4.

Both present and previous studies show that place transmission is worst and that manner transmission is much better than place transmission. However, the difference between the present and previous studies is the relatively worse transmission of voicing compared to place and (particularly) manner, which contrasts with studies in figure 6.1 showing broad equivalence between voicing and manner transmission. This disparity might be due to a number of stimulus or processing differences between the work discussed in this thesis and previous studies cited in figure/table. The most obvious difference is the choice of vowel environment for the VCV stimuli: in all the studies cited in figure 6.1 (with the exception of Geurt and Wouters, 1999-but here data were averaged across vowel environments) subjects were tested with aCa whereas in this study iCi was used, for reasons outlined in 3.3.1.

In experiment 2, feature transmission was compared between /iCi/ and /aCa/ vowel environments and significantly worse overall performance was shown for /iCi/. Overall, performance for voicing, as for other features, was worse with the /iCi/ vowel environment, although, if data from the analogous AM conditions are compared (noise band with Greenwood shift), there was relatively little difference. However, a number of parameters differed between experiment 2 and the latter experiments, in particular strategy type (ACE vs. CIS), channel stimulation rate and channel number, making this inference problematic. A more useful comparison can be made between experiment 3 results and data from SOECIC in figure 6.1. The latter data were collected using the same test room and in all respects an identical testing regime (albeit data are from a mixture of Nucleus 22, Nucleus 24 and MED-EL users), with the exception that testing was undertaken with the /aCa/ vowel environment. It can be seen that the “SOECIC” data also show the relatively better voicing transmission, comparable with manner transmission. Given this, it seems likely that the relatively poorer voicing transmission may be due at least in part to use of a different vowel environment.

It is now appropriate to consider transmission of each of the six specific features in more depth. In each case, the main findings from experiments 3 and 4 are illustrated and the statistical findings are summarised. Because of the general lack of effects of processing parameter conditions (e.g. channel number and stimulation rate), data were averaged across the three MAPs used and discussed in this context. Where appropriate, additional consideration is given to results from the experiments reported in chapter 3. The most important consideration is what can be concluded about the key question, the relative contribution of CI processing vs. the electrical/neural interface, for each of the features.

6.2 Voicing

Figure 6.3 shows voicing transmission across experiments 3 and 4, averaged across MAP conditions. As with all other features, channel interaction had a small but significant effect on voicing transmission. Voicing was not significantly different between better CI users and AM results with no/1 mm channel interaction but was significantly better than AM performance with 3.3 mm channel interaction. Voicing was significantly reduced by the inclusion of noise in most listening conditions, except for better CI users. The comparison between voicing in the no channel interaction AM and better CI users was significant for transmission in noise but not in quiet (e.g. the AM under-predicted transmission for better CI users-there was no effect for the latter group but there was for AM conditions). In short, the 1 mm channel interaction AM was the best predictor of voicing transmission but the general problem with the AM prediction of voicing transmission was the absence of a noise effect for voicing transmission in the better users, an effect that was obtained in AM conditions, albeit least for 1mm channel interaction. In this respect, AM data were better at predicting worse CI user performance, although the absolute magnitude of voicing transmission and degree of noise effect were over-estimated even by the 3.3 mm channel interaction model.

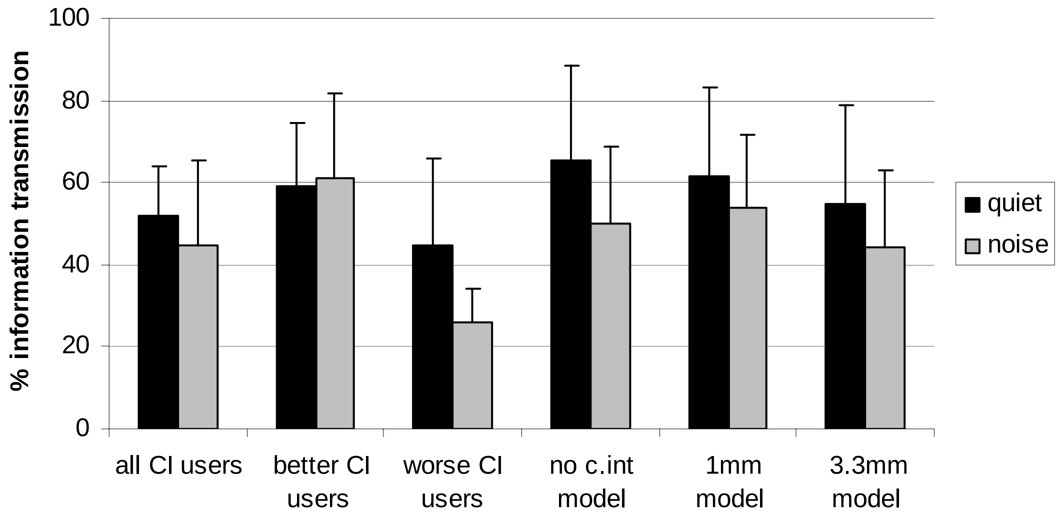

Figure 6.3. Mean (+ 1 SD) voicing transmission for CI users and AM listeners across MAP conditions.

As discussed in 2.6.1, voicing is signalled by a variety of acoustic cues, most of which rely on temporal resolution. However, there is also a spectral contribution to voicing because of the first formant transition onset frequency cue (thought to be more important in noise) and also because the voice onset time distinction requires a comparison of different frequency components, albeit the spectral differences of the two components, e.g. voice bar and burst, are large (in CI processing terms many electrodes apart). Stickney (2001) found a correlation between voicing transmission and channel interaction in a group of Clarion implant users, although the relationship was less strong than for place of articulation, while there was no correlation between manner transmission and channel interaction. This could be seen as supporting evidence for a role of spectral resolution in voicing transmission in CI users, although it should also be noted that the spectral contribution to voicing could be less in the /iCi/ vowel environment

Across CI users and AM listeners, voicing transmission in quiet was around 60% or less, and transmission in noise was worse (except for better CI users). In NH listeners there is no effect of noise on voicing transmission at positive SNRs (in studies of NH subjects such as Miller and Nicely, 1957, or Jiang et al, 2006, noise only affects voicing transmission at negative SNRs). Also, figure 5.11. shows voicing transmission both in quiet and noise around 100%. Consequently, two related questions emerge: first, why was voicing transmission markedly reduced by the addition of noise across AMs and worse CI users, but not better CI users? Second, why was voicing transmission so much worse than for NH listeners even in quiet?

Given the absence of effects at +10 dB SNR in NH listeners (unaltered condition) the susceptibility to noise must be related, at least in part, to information loss with CI processing. It has been proposed that noise reduces accuracy of coding of within-channel envelope fluctuations and that therefore the listener becomes more reliant on spectral information, which is of course impoverished by CI processing. However, it also appeared that the AM data over-predicted the noise effect for better CI users. A useful comparison here is between the noise effects with voicing and the noise effects with nasality-for the latter, even better CI users showed a very strong noise effect. In many respects, cues to voicing are similar to cues to nasality; however, voicing cues are higher in amplitude than cues to nasality. It therefore seems likely that the difference in noise susceptibility across CI users relates to some aspect of audibility or dynamic range in the apical electrodes and, moreover, that better CI users have a better access to within-channel information. The simplest possibility might be electrical dynamic range. This possibility could be tested simply by checking the MAPs of the better and worse CI users with respect to the electrical dynamic range of the relevant apical electrodes. However, a comparison of average apical electrode (15 and above) electrical dynamic range showed very similar means between the better and worse CI users (around 40 current units difference between T- and C-levels for both subgroups), so this simple explanation cannot be used.

There is also the further question of why voicing transmission is so poor in quiet in the first place. A further contributing factor may be the fact that combining pre-emphasis with a relatively small dynamic range means that the audibility of the low-frequency components is relatively reduced compared to NH. A frequent voicing error was for /g/-/k/ and also /s/-/z/. After pre-emphasis and compression, F1 (or the voice bar, or, for nasals, the characteristic ultra-low nasal murmur) is of relatively lower amplitude than in the unprocessed signal. In this context, it is interesting to note that Goedegebure et al. (2002) found that a form of compression used to reduced upward spread of masking had a negative effect on transmission of nasality and voicing in hearing aid users. This supports the idea that pre-emphasis combined with reduced dynamic range could be a factor which reduces both nasality and voicing transmission.

It is interesting that Verschuur (2005) found voicing transmission around 90% in quiet for a group of users of the MED-EL device with the CIS processing strategy. This difference is suggestive of a relationship between CI temporal processing and voicing transmission. In particular, the MED-EL device uses a set of IIR fiters (in the Verschuur, 2005 study, all with envelope cut-off frequencies at 400 Hz) rather than the FFT approach described in 2.2. As noted, it seems possible that this approach might improve the representation of higher modulation frequencies as compared to the Nucleus device (particularly at those frequencies said by Rosen, 1992 to relate to periodicity), although this needs to be supported by TMTF measurements undertaken for an IIR based processor. It is likely that that voicing transmission in quiet is limited in part by inadequacies in temporal coding in the Nucleus device, and perhaps in part by the modest spectral contribution to voicing. Voicing transmission in noise may be further limited by individual differences in electrical dynamic range in apical electrodes.

6.3 Place of articulation

Figure 6.4 shows place of articulation transmission across experiments 3 and 4, for all MAP conditions averaged together. Place transmission was not affected by noise, except for worse CI users (even here the effect was smaller than for other features) and there was a good prediction of transmission by AM data. One particularly interesting aspect of place transmission was the better prediction with the 3.3 mm channel interaction mode of better users, in particular the absence of a channel number effect for the 3.3 mm channel interaction model and CI users but the presence of such an effect of the other AM conditions. This indicated that better CI users are limited in benefiting from the higher channel number for place transmission by an increase of around 10% transmission because of channel interaction. In this one domain, the inclusion of channel interaction in the model made an important difference in the predictive power of the model.

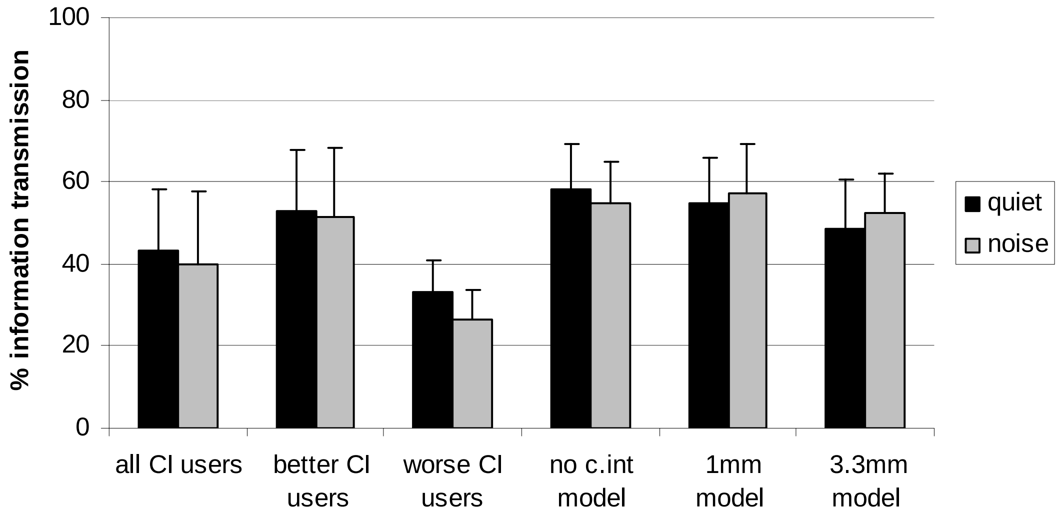

Figure 6.4. Mean (+ 1 SD) place of articulation transmission for CI users and AM listeners across MAP conditions.

It can be seen from figure 6.4 that the magnitude of place transmission was best predicted by the 3.3 mm channel interaction model , although this was not specifically supported by the MANOVAs summarised in table 5.6. It is interesting to compare the findings with place to those obtained for F0 discrimination in Laneau et al. (2006). In that study, which used the same AM as the present study, the authors found the best match between CI and AM performance with a channel interaction equivalent to 1 mm. Moreover, in the present study, the effect of channel interaction was rather more modest than for F0 discrimination in Laneau et al. (2006). This would suggest that the type of cues to F0 discrimination differ to the types of cue to place transmission, although in both cases the cues are considered “spectral”.

Previous studies (as in figure 6.1) have shown, at best, around 50 to 60% place transmission, and the findings of the present study are consistent with this. It was also interesting that performance was less variable across AM variations. This suggests that AMs are broadly equivalent in their (accurate) prediction of the magnitude of place transmission and the lack of a noise effect. Place of articulation perception in English consonants relies on a number of spectral cues, particularly the spectrum of the burst and the onset frequency of the formant transitions into following vowels, particularly F2 and F3. The electrodographic analyses shown in 2.6.3 suggested that the burst spectrum should be better preserved by Nucleus 24 than the onset frequency of formant transitions. CIs and AMs across studies have some characteristics in common, in particular the relatively small number of channels and the use of pre-emphasis. It seems likely that the combination of these two characteristics leads to a more generalisable result, namely poor representation of formant transition onset frequency and a much better representation of burst or frication frequency. The consequence of these processing characteristics means poor but not terrible place transmission and robustness to noise.

It was hypothesised in 2.6.3 that place transmission should be worse for nasals and liquids than for stops and fricatives, because with the latter the burst/frication should be well represented by CI processing, whereas for nasals and liquids/glides place transmission relies almost entirely on formant transition information. Although this hypothesis was not explicitly addressed in the statistical analysis, the confusion matrices shown in 5.6 do appear to support the hypothesis: Better CI users showed a high proportion of place errors for liquids and also showed the consistent /n/ for /m/ error. By contrast, for the majority of fricatives and plosives, place errors were relatively few. The notable exceptions were velar stops-these were consistently mistaken for alveolar or bilabial stops. The latter finding agrees with Valimaa et al. (2002a) who showed a consistent trend (albeit in Finnish CI users) for place of articulation errors in the direction of higher frequency place cues. The question here is whether the trend for errors was due to the coding of the formant transition onset frequency or the coding of the burst. The AMs converged with the CI user data in that they showed errors for /g/, but they diverged markedly in that the error patterns were less consistent than for CI users (who generally misperceived /g/ as /d/). It is also worth noting that, although the models under-estimated performance with the other stops, this was primarily because the models over-estimated stop voicing errors, not place or manner errors.

In summary, place of articulation transmission is unaffected by most variables, even noise (at the +10 dB SNR in any case) and its transmission remains poor, compared to other features, across listening conditions, but also does not appear to get worse with noise or channel interaction. It is therefore not a particularly good measure to use for variations across existing processing parameters within a limited-channel envelope extraction processing scheme. It could, however, be a very effective method for demonstrating genuine improvements in spectral processing, e.g. as might be introduced by improved within-channel frequency coding that improves the representation of formant transition information.

6.4 Manner of articulation

Figure 6.5 shows manner of articulation transmission across experiments 3 and 4, for all MAP conditions averaged together. Manner is a notably more “robust” feature than the previous categories and was around 80% transmission across all AMs and better CI users. The category was unaffected by most variables, the only exception being an effect for channel number in the no channel interaction model condition. Otherwise, the overall pattern was very similar to the “envelope” feature. As noted in 2.6.2, it may be that the “manner” category is too broad to provide a sensitive handle on perceptual differences- rather, specific manner subcategories such as nasality and fricative (some other possibilities not included in this study are continuant and plosive, sibilant and affricate) may be more useful measures and may provide more explanatory power. As suggested in 2.6.2, the contrasts between manner categories are signalled by different modulation patterns across a small range of electrodes. However, it should also be noted that manner was more affected by noise at +5 dB SNR than envelope (in experiment 2).

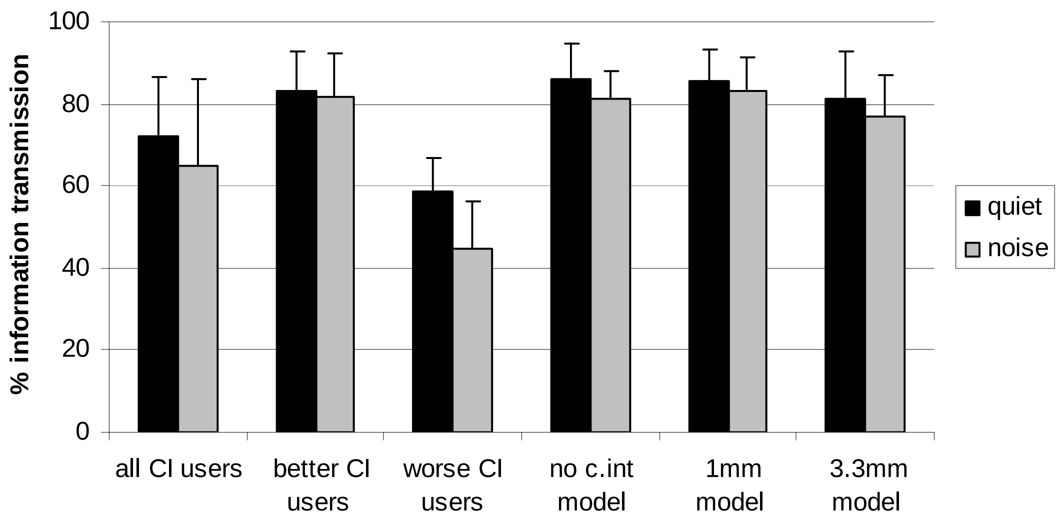

Figure 6.5. Mean (+ 1 SD) manner transmission for CI users and AM listeners across MAP conditions.

6.5 Nasality

Figure 6.6 shows nasality transmission across experiments 3 and 4, for all MAP conditions averaged together. The most striking aspect of nasality transmission was its high susceptibility to noise across all experimental conditions, including better CI users whose performance was not significantly affected by the addition of +10 dB SNR background noise for any other feature.

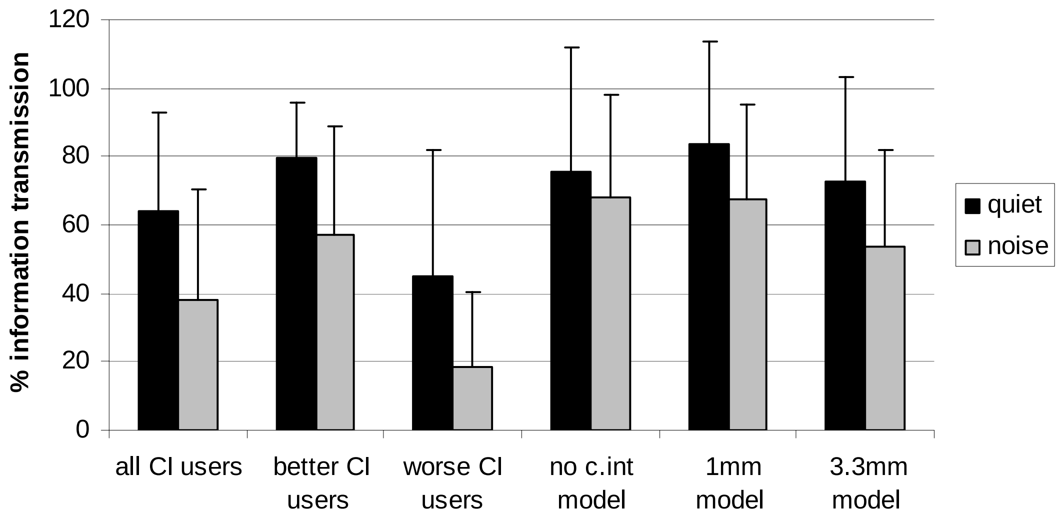

Figure 6.6. Mean (+ 1 SD) nasality transmission for CI users and AM listeners across MAP conditions.

The ability to distinguish nasal from non-nasal consonants depends on low-frequency audibility and amplitude resolution, specifically the identification of the characteristic low frequency nasal murmur, low-amplitude formants and spectral zeroes. As noted, this means that nasality can be distinguished from voicing in the low amplitude of the important cues to this feature. Nasality perception was unaffected by noise at positive SNRs in NH listeners, as in figure 5.11. As noted in 2.6.2, acoustic cues to nasality are converted into differences in envelope modulation pattern in apical electrodes by CI processing. The fact that noise interference had a significant effect across CI users and across AMs, and that this effect was at SNRs where NH listeners do not experience difficulties in nasality perception, suggests that the low-frequency audibility and amplitude resolution is compromised by CI processing, and that the degree of compromise is such that even better CI users are unable to cope with noise interference. It is also possible that the low amplitude of these cues relative to cues to other features, in particular the presence of spectral zeroes, or antiformants, in the spectrum of a nasal consonant, made them more susceptible to noise interference.

In the same vein, it was also notable that variance between better and worse CI users, and across users overall, was greater for nasality than any other feature category except fricative. The confusion matrix shown in table 5.12 showed a frequent error pattern for worse CI users, namely misperception of both nasals /n/ and /m/ as /l/. This error pattern was absent from better CI users or from AM results. As indicated in figure 2.16, the distinction between nasals and the liquid /l/ is the slowly varying modulation pattern of the most apical electrodes. One possible reason for this inter-user variation could therefore relate the listeners' ability to make use of within-channel envelope variations in the apical channels. Taken together, these findings suggest that this feature is affected by inadequate CI processing of lower-amplitude cues (and consequent susceptibility to noise masking) and to electrical/neural interface variations relating to amplitude resolution or electrical dynamic range.

There are some further issues relating to nasality transmission which relate exclusively to AM characteristics. Nasality transmission was sensitive both to choice of carrier stimulus and the inclusion of pitch shift, as indicated by the significant interactions in experiment 2. It was the only feature to show a significant statistical association with the inclusion of the Greenwood pitch shift, albeit this association was stronger for the /aCa/ vowel environment that was not used in subsequent experimental work. The difference between nasals in the two vowel environments is due to the marked difference in F2- much higher for iCi (typically 2500 Hz +) than aCa (typically 1100 Hz). Presumably the perceived shift was greater for a lower-frequency formant transition (or whatever was left of the formant transition after CI processing, e.g. locus of relative amplitude shift across channels) than for the higher frequency formant transition.

6.6 Fricative

Figure 6.7 shows fricative transmission across experiments 3 and 4, for all MAP conditions averaged together.

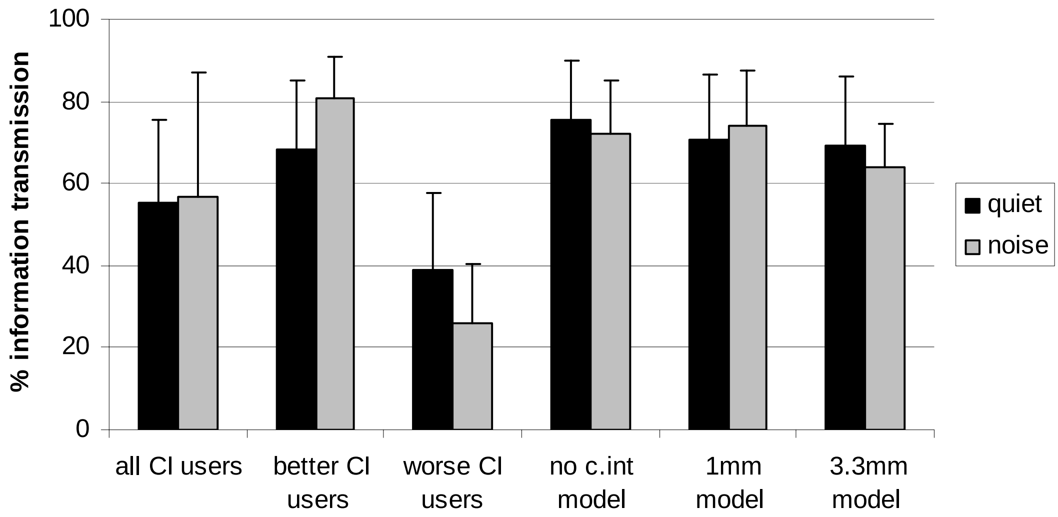

Figure 6.7. Mean (+ 1 SD) fricative transmission for CI users and AM listeners across MAP conditions.

Fricative transmission was similar to nasality transmission in that there were marked differences between better and worse CI users. However, unlike nasality, fricative transmission was not sensitive to noise effects for better users or for AMs. The identification of both the presence and duration of the characteristic frication noise that is required to distinguish fricatives from other consonant types appeared to be relatively well represented by CI processing according to figures 2.16 and 2.18. For better CI users and AM subjects, the feature was robust to noise interference, presumably because of the salience of the high amplitude of the frication cue (as for the burst cue, noted in 6.3), due in part to pre-emphasis.

At the same time, fricative transmission in noise showed the greatest difference between better and worse CI users than any other feature (this can be seen in figure 5.38.). The difficulty in determining the reason for high between-user variation for this feature is the presence of both within-channel (envelope/temporal) and cross-channel (spectral) cues. One temporal cue is duration- e.g. the duration of the noise in the basal channels signals that a consonant is a fricative and not a stop. The difference in duration of the frication/burst cue between fricatives and stops is of the order 80 ms. The TMTFs shown in 2.3.3 showed that modulation depth for a 2000 Hz carrier at a modulation frequency of 25 (which is more than adequate to code a distinction of 80 ms) was 99%, regardless of stimulation rate. A further distinguishing feature of fricatives is the presence of aperiodic (corresponding to voiceless e.g.burst, frication) rather than quasi-periodic (e.g. voiced) energy in the basal channels. However, in practice, it is likely that this distinction is unimportant given that there is a marked place distinction anyway-presumably it would become more important if the CI user had access to only one or two channels. Therefore, it seems likely that the spectral cues are more important. It could be that worse users have more channel interaction than the 3.3 mm channel interaction model, rendering the spectral cues to fricative transmission vulnerable, or it could be that audibility or amplitude resolution is the important factor. One way to distinguish these possibilities is to consider a specific fricative vs. non-fricative confusion in which spectral cues are relatively unimportant. The distinction between the fricative /ʃ/ and the affricative /ʧ/ provides such a contrast. The confusion matrix for better users show that all the errors made for the /ʃ/ stimulus were for place, whereas for worse users (table 5.12) /ʃ/ is mistaken for the affricates /ʧ/ and /ʤ/. The error patterns for the AMs show the same patterns as for better CI users, irrespective of channel interaction. This suggests that within-channel envelope processing, rather than spectral information (and therefore channel interaction) is implicated in the inter-user variation in fricative transmission.

A final interesting aspect of fricative transmission was that, in experiment 2, the sine wave carrier was associated with better fricative transmission than the noise bands. If, as has been suggested, the advantage from sine wave carriers is in the spectral side bands, this lends further support to the idea that noise bands are a more appropriate carrier stimulus because they are less likely to over-estimate the amount of spectral information available to the CI user. However, a caveat to this is that the sine wave carriers might also carry more information about within-channel envelope fluctuations (e.g. the important periodic vs. non-periodic distinction), although this explanation seems less likely given that within-channel cues are probably less important for distinguishing fricatives from non-fricatives, as suggested above.

6.7 Envelope

Figure 6.8 shows envelope transmission across experiments 3 and 4, for all MAP conditions averaged together.

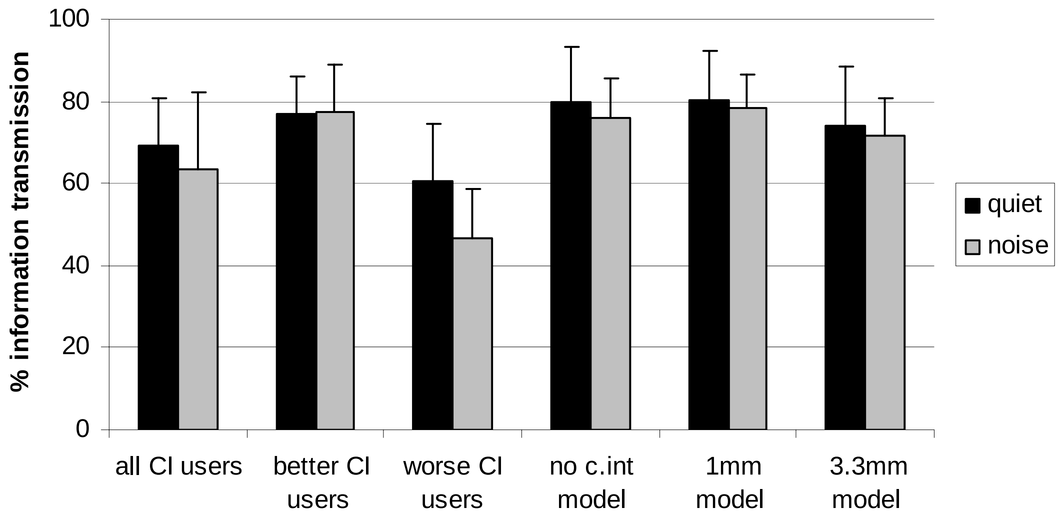

Figure 6.8. Mean (+ 1 SD) envelope transmission for CI users and AM listeners across MAP conditions.

Performance with the "envelope" feature was expected to be more robust, i.e. less affected by noise or channel overlap, and with a higher level of transmission than other features. This was borne out by the findings, although it should be noted that manner transmission followed almost the same pattern across conditions. As with manner, performance was relatively unaffected by differences in AM parameters, and performance was largely unaffected by background noise across conditions (with the exception of "worse CI users"); even with AM listeners in experiment 2, envelope was relatively unaffected by +5 dB SNR. However, it is still worth noting that envelope transmission at best is around 80%, contrasting to 100% with NH listeners (unaltered condition- see figure 5.11) in both quiet and noise. Assuming that this feature truly does reflect temporal envelope differences exclusively, this is further support for the idea that the Nucleus 24 processing does not code temporal envelope information optimally (at any stimulation rate), for reasons discussed in 2.3.3.

It is of interest to determine why envelope transmission did not approach 100%. Tables 5.8 to 5.10 showed a large number of misperception of the voiceless plosive /t/ as the voiceless affricate considered as "fricative" within the feature categorisation scheme used /ʧ /. In this classification system, the distinction between affricates and fricative counted as an "envelope" distinction, whereas it can be seen from figure 6.9 that the difference between these two consonants is in fact a spectral rather than temporal/envelope distinction. It is likely that this particular error pattern diluted the effectiveness of the “envelope” feature as a true measure of envelope perception.

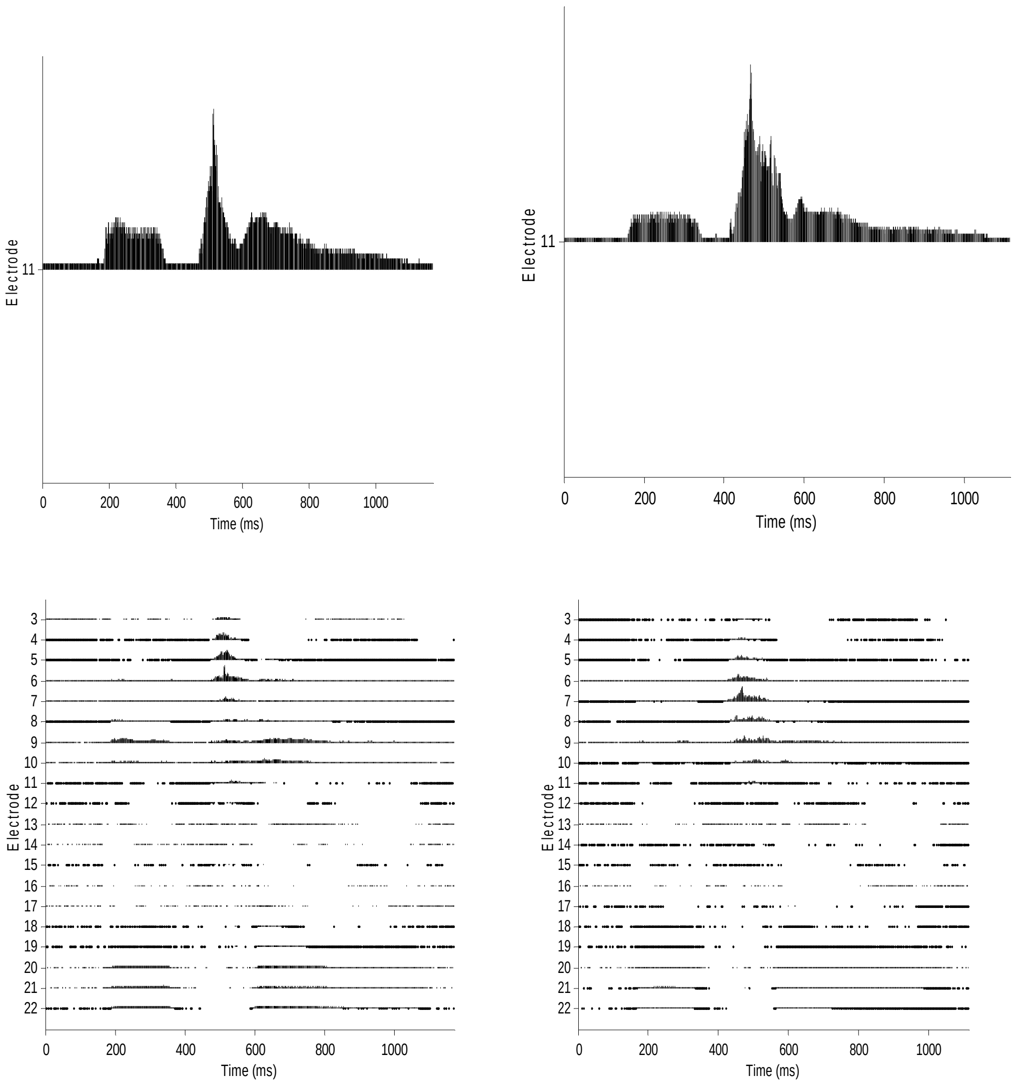

Figure 6.9. /iti/ on left and /itʃi/ on right, single channel CIS above, 900*12/20 ACE below.

Chapter 7. General discussion

7.1 Overview of design and aims of the study

The key question to be addressed in this thesis was: “to what extent can deficits in consonant recognition by CI users be explained by information loss in CI signal processing as opposed to information loss at the electrical/neural interface?” It was argued that a comparison between CI user performance and equivalent AM performance might help to answer this question. An initial review of literature suggested a number of factors likely to affect consonant recognition in CI users. Two CI signal processing parameters were identified as being particularly important: number of channels and channel stimulation rate. Also, two electrical/neural interface factors were deemed to be particularly important: pitch shift and channel interaction. Background noise was also identified as a factor likely to impact on feature transmission. The broad aim of the work was therefore to determine if it was possible to model changes to consonant feature recognition as a function of these processing and stimulus variables using a carefully matched AM which incorporated some spectral distortions associated with the electrical/neural interface. An important assumption was that transmission of different consonant features would be affected by different processing and electrical/neural interface factors in different ways, according to the relative importance of temporal or spectral information to coding each specific feature.

All experimental work related to a specific device implementing one of two processing strategies, and great care was taken to ensure that device characteristics were simulated as precisely as possible in the AMs. To achieve this, the Nucleus NIC-STREAM (Cochlear, 2002) and AMO/CISIM (Laneau et al., 2006) MATLAB toolboxes were used to generate AMs that were identical in processing details to the speech processing implemented for the Nucleus 24 device. A further advantage of using the toolboxes to generate AMs was the fact that this has been validated as a means of simulating spectral channel interaction in Laneau et al. (2006).

Each of the four experiments employed the same consonant recognition task comprising a forced choice between 20 possible English consonants in an intervocalic position between preceded and followed by the vowel /i/ (or /a/ for part of experiment 2). A single iteration of sequential information transfer analysis (Wang and Bilger, 1973), equivalent to the information transmission measure of Miller and Nicely (1955), was used to determine relative information transmission rates for the same six phonological features, across all experiments.

7.1.1 Overview of preliminary experimental work

Two preliminary experiments were undertaken to investigate the effects of AM characteristics on consonant recognition in quiet and noise by using an AM implementing the fixed-channel CIS strategy with 8 channels and a channel stimulation rate of 500 pps/ch. It was evident that voicing transmission in quiet probably over-estimated CI users' likely perceptual abilities- this was thought likely to be due to the choice of carrier stimulus (sine wave) and possibly the absence of any electrical/neural interface factors. It was also found that fricative transmission was very poor- a possible methodological shortcoming to do with subject instruction was identified as a contributing factor to this. However, in other respects, the experimental findings suggested that the model would be appropriate, albeit that various other model characteristics, particularly choice of carrier stimulus, needed to be considered in a further experiment. Results showed a varied pattern of noise effects across different consonant features: transmission of manner, voicing, nasality and envelope were markedly affected by the addition of background stationary noise at SNRs of +10 and worse whereas place and fricative were not. This disparity was consistent with the hypothesis that noise interference disrupts within-channel information, which would have a disproportionate effect on consonant features that relied more on temporal/amplitude resolution than spectral resolution.

The second experiment compared two AM parameters, namely carrier stimulus and inclusion of Greenwood pitch mismatch, and two more general stimulus parameters, noise and vowel environment. It was found that choice of carrier stimulus and vowel environment had a greater effect on performance than pitch mismatch. It was also found that the choice of +5 dB SNR would probably be less than sensitive to differences across features than +10 dB SNR (the two SNRs having been found to be broadly equivalent in experiment 1). Based on the combined results of experiments 1 and 2, parameters were chosen for a further AM experiment that was intended for direct comparison with CI user data. The model parameter choices were: noise band carrier implementing Greenwood pitch shift. Stimulus parameters chosen were: the /iCi/ vowel environment and stationary noise at +10 dB SNR.

7.1.2 Overview of main experiments

A further two experiments were undertaken and were designed to form a “matched pair”, one with an AM and the other with CI users. Because a review of literature had identified the two processing parameters of channel number and stimulation rate as being particularly important in determining general speech perception performance, these two processing parameters were varied, with equivalent variations in both AM and CI user experiments. This meant that all testing was undertaken in three processing parameter, or MAP, conditions: one with a higher stimulation rate and higher channel/peak number, the second with a higher stimulation rate and lower channel/peak number and a third with a lower stimulation rate and a higher channel/peak number. Testing was undertaken in quiet and with the addition of background speech-shaped stationary noise at +10 dB. A model of spectral channel interaction, based on Laneau et al. (2006), varying from no interaction, 1 mm interaction and 3.3 mm interaction, was included as an additional variable in the AM by changing the filter used to generate the noise bands serving as carrier stimuli.

The design of the experimental work had a number of advantages over previous work and therefore provided an opportunity to add to the existing knowledge base. Stimulus and processing conditions were matched as precisely as possible between the AM and CI user experiments, in a way that has not been achieved in other AM studies with the exception of Laneau et al. (2006) and in no other study of consonant recognition. This close matching between AM and CI user experiments allowed stronger inferences about the likely contributions of different factors to performance. All the experiments evaluated consonant feature transmission in noise as well as quiet, both in CI users and AM listeners, and therefore helped to address the lack of knowledge about the effect of noise on transmission of different consonant features and the possible mechanism of noise interference for CI users. Six features with contrasting acoustic attributes were included in the analysis of consonant recognition data in order to provide a greater level of detail to the acoustic-phonetic analysis of results than has previously been obtained. Spectral overlap between channels was included as a variable in the AM experiment in order to determine whether the inclusion of this term would better predict performance in the CI users; the values of this parameter and the method of simulation were based on Laneau et al. (2006). Although there has been previous work on spectral smearing in relation to consonant recognition (Shannon et al., 1995), there has been no previous attempt to combine this with a model of specific device characteristics, or in relation to changes in processing parameters, or in relation to performance in noise as well as quiet. It is also worth noting that, in the second set of experiments reported here, a peak-picking strategy, ACE, was used. The majority of studies of channel number, channel stimulation rate and consonant feature recognition have used other strategies, either CIS or the earlier lower-rate peak-picking strategy SPEAK. Moreover, almost all previous AM studies have used AMs based on fixed-channel devices and moreover based on the implemention of a band of linear IIR filters followed by smoothing, rather than the FFT approach used in the Nucleus device. Therefore, both the AM and CI user data collected in this study has further direct clinical relevance to users of the Nucleus 24 device, as well as broader relevance to other peak-picking strategies, although the close matching between CI and AM characteristics was the most important aspect of the study.

7.2 Overview of methodology and methodological limitations

The most important point about the test and analysis methodology used was that it was identical across experiments. Nonetheless, it is important to consider what impact the particular choice of methodology might have had, and what this means in terms of comparison with other studies. One limiting factor was the need to devise a consonant recognition task that was time-efficient, given the number of different listening conditions used in the experimental work and the fact that the task is inherently tiring or boring and therefore can lead to fatigue effects. Because a large number of consonants were used (20) in order to represent the majority of consonants in the English language, there was little scope for using a large number of repetitions. The pilot study in experiment 1 suggested, albeit qualitatively, that 3 repetitions of each stimulus would be adequate to obtain meaningful and repeatable results. Work in the literature has varied from 2 to 7 repetitions per stimulus. It should be argued that all subsequent experimental work in this study supports the argument that 3 repetitions was adequate, given the reasonably low variance of most features, the meaningful pattern of results across experiments and the finding of various significant effects between listening conditions and features.

Another methodological issue is that of acclimatisation to AMs, and, in particular, the acclimatisation to pitch-shifted stimuli. In the work here a specific (and quick) approach to acclimatisation was used: stimuli were presented visually and the listener took as much time as they wished to learn which stimuli corresponded to which consonants. In practice, this self-directed acclimatisation process never took more than 10 minutes. Despite this, and despite the fact that stimuli were not only processed using an AM, but also (for most experiments) shifted in pitch and in some conditions having large channel overlap, performance levels were remarkably high. This contrasts with other work, such as Rosen et al. (1999), who showed that performance to basally-shifted AM stimuli was very poor without substantial acclimatisation. . However, there a number of important differences between this work and the Rosen et al. paper, in particular the fact that those authors employed a four-channel AM. It appears that, given a richer spectral representation (in the AMs in the present study), and possibly because of a more constrained stimulus set, it was possible for NH listeners to acclimatise rapidly and effectively to pitch-shifted stimuli. Faulkner (2006) also found that listeners needed at least some hours to adapt to stimuli that were spectrally warped.

Another important methodological issue was the choice of vowel environment. Because of the large number of variables in experiments 3 and 4, it was not possible to add vowel environment as an additional variable and therefore the work relied on results of experiment 2, which showed that transmission of some features tended to be above 90% for the /aCa/ vowel environment but less with /iCi/. However, a weakness of using a consonant recognition task with /iCi/ is the difficult of comparing current findings with those from the research literature in which almost all data have been collected with using consonant confusion tasks with the /aCa/ vowel environment. Previous work has identified the possibility that the /aCa/ vowel environment is more likely to lead to ceiling effects in CI user performance and that the emphasis on the burst produced by the /iCi/ stimuli might mean greater sensitivity to rate effects.

Although the latter proved to be inconclusive (e.g. there were no rate effects), it seems that the meaningful pattern of results obtained across experiments 3 and 4, and in particular the absence of floor or ceiling effects in any listening condition, vindicates the choice of the /iCi/ vowel environment in this study. It should also be emphasised that the most important comparison was between the results of experiment 3 and 4 rather than comparison with other evidence in the literature. (Also, it is very difficult to compare across studies very closely in any case as there are very large differences in test methodology and CI user/processing characteristics, as discussed in 2.1.2).

A further methodological limitation may have been the choice of a single SNR of +10 dB for the third and fourth experiments. Although the worse performing CI users showed marked noise effects at this SNR, other CI users did not, although here the effect on nasality perception was apparent. The choice of a single SNR was motivated by the desire to minimise the number of listening conditions and the particular choice of SNR was motivated by findings of experiments 1 and 2 taken together. However, the interpretation of the correspondence between AM and CI findings may have been strengthened if noise effects overall had been stronger. A related weakness in methodology may have been the fixed order of noise conditions used in experiments 3 and 4, e.g. within each listening condition quiet stimuli were followed by nois-contaminated stimuli. This approach may have diluted and under-estimated the true effect of noise as, in each case, subjects had extra time to acclimatise to stimuli in quiet prior to exposure to the same stimuli in noise. This can be seen in some individuals (from the "better performing" CI group) who showed better transmission of the fricative feature (see figure 6.7) and for some AM conditions, e.g. for place, as shown in figure 6.4. In retrospect, the order of quiet and noise-contaminated conditions should probably also have been randomised as the fixed order may have diluted noise effects. Nevertheless, the presence of significant reductions with background noise for some features despite this makes conclusions about noise effects, where these obtain, even stronger.

A more fundamental question concerning methodology is whether a consonant confusion task is able to truly distinguish different perceptual processes. In the present study, as in various other studies, transmission of "temporal" consonant cues was not normal. It is interesting to ask whether this is an artefact of the methodology. Neither

CI users nor AM listeners obtained 100% information transmission for manner, nasality or even for the “envelope” category. Figure 6.5, 6.6 and 6.8 shows around 80% transmission for these features, even for better CI users; this contrasts with performance around 100% for normal hearing listeners as shown in figure 5.11. It was noted in 6. 7 that the “envelope” category used in the feature analysis in the present study may have been affected by the inclusion of the affricative/plosive distinction which, in reality represents a spectral contrast. Certainly, voicing and manner are assumed to be temporal in the CI perception literature whereas it is clear that from the literature that voicing has a spectral component (albeit reduced in the /iCi/ vowel environment) and even manner requires resolution of different frequency components. One possible way to prove more definitively that temporal envelope information is impaired in CI users would be to use a categorical perception task in which exclusively envelope cues are varied. Although this is difficult to achieve with natural-sounding speech synthesis, it should be possible to construct, for example, a continuum of stimuli varying from fricative to affricate in which only envelope information is varied (Faulkner et al., 1995).

7.3 AM findings

A striking finding was the close match between AM performance and performance by “better” CI users, across a range of features in different listening conditions. These results were markedly different from results obtained with NH listeners (e.g. in the “unaltered” listening condition in the present study). Equally striking was the fact that inclusion of a channel overlap in the model had only a modest effect on performance. The important comparison here is between the effect of processing and the effect of channel interaction as shown in figures 5.40 and 5.41. That is, the processing of the signal with a specific set of Nucleus 24 processing characteristics led to marked reductions in performance across consonant features, and both the magnitude and pattern of these deficits were mirrored by better CI users. By contrast, the difference in performance between AMs with and without channel overlap was relatively modest. This was an unexpected finding, given that the degree of spectral overlap implied by 3.3 mm spectral spread is considerable in the context of a total electrode length of 25 mm (see figure 5.3). Moreover, Laneau et al. (2006) and Laneau et al. (2004) found equivalent performance in pitch perception between the Nucleus 24 user performance and performance with an AM using a channel overlap equivalent to 1 mm spectral spread. Although pitch discrimination is more implicated in frequency resolution than overall consonant recognition, at least some consonant feature (particularly place) are reliant on spectral resolution, and therefore it seemed a reasonable hypothesis that the channel overlap model would also explain some variance in consonant recognition performance. Fu and Nogaki (2005) found that sentence recognition by the best CI users is approximated by an AM with 8 to 16 channels with some channel overlap, although direct equivalence between sentence and consonant recognition data cannot be assumed, given the difference in acoustic and contextual cues accessible in sentence materials compared to those accessible from nonsense syllables.

The implication of these findings is that, for better CI users, deficits to consonant recognition are due more to CI processing information loss than channel interaction. This conclusion is tempered by the fact that Greenwood pitch shift was included in all the models used in experiment 4. However, the finding of experiment 2 that pitch shift had very little effect on consonant recognition, combined with experiment 3 findings, lends weight to the argument that CI processing is more important than electrical/neural interface factors in determining better CI users' consonant recognition abilities. It is therefore important to determine which aspects of signal processing are likely to have had an effect on performance.

The work also provides some further information about the design of AM experiments. One of the important issues identified in 2.5 was the choice of carrier stimulus. The findings generally support the hypothesis that the noise band carrier is more appropriate when modelling consonant recognition. The presence of sidelobes produced by modulation of sine waves appears to be of benefit to many aspects of consonant recognition, to an extent that this over-estimates CI user abilities. This does not mean that a noise band model is in any sense a perfect model of all aspects of speech perception in CI users, but the data from this study suggest that it is a more than adequate model of consonant feature recognition. Other stimuli, notably pulse trains (Carlyon et al., 2002; Carlyon and Deeks, 2002) have been used, although these cannot be used with higher rate stimulation as for higher pulse rates harmonics are resolved.

7.4 CI filterbank characteristics

For any CI, information loss from processing must be determined in part by the choice of filterbank characteristics, which in effect means the choice of which type of information to reduce at the expense of other types of information. It is clear from the evidence and analyses in the present study that the Nucleus 24 filterbank imposes limitations in both spectral and temporal information. The present work has highlighted some shortcomings of the particular approach taken in the Nucleus 24, although some of these limitations may apply to other filterbank approaches also.

The finding that there was no change in consonant recognition when changing from 250 pps/ch to 900 pps/ch is broadly consistent with previous work using the Nucleus 24 device as noted in 2.3.3, e.g. Vandali et al. (2000) who found no increase in speech recognition beyond 250 pps/ch. What is novel in the present work is the clear demonstration that this must relate to an absence of significant increases in temporal sampling with increasing stimulation rate. The TMTF analyses in 2.3.3 showed that temporal information provided by the Nucleus 24 processing decreases as modulation rate increases from 25 to 250. Moreover, this effect was obtained for different carrier stimuli and at differing stimulation rates. It was therefore hypothesised that changes to stimulation rate should have little effect, assuming that the benefit to higher rates was in the improved temporal representation of the signal rather than some other benefit to neural coding, or an indirect benefit due to increased dynamic range. However, the present study showed that changes in channel stimulation rate from 250 to 900 pps/ch had no effect on any consonant recognition measure, either for CI users for AM listeners, and irrespective of feature or noise condition. This supports the hypothesis that changes in stimulation rate have little or no effect on perception because they have little or no effect on the temporal information available through the CI. It also fits with the evidence noted in 2.3.3 that the majority of studies showing benefit to changing rate above about 200 pps/ch have been in users of the MED-EL device which implements a bank of IIR filters with variable envelope cut-off frequency, whereas studies of the Nucleus system have not shown consistent benefit.

Another way to determine if the chief limiting factor on temporal information is the fixed FFT length is to compare with devices in which envelope variations are definitely coded at higher frequencies. Verschuur (2005) showed manner transmission around 90% in six MED-EL users where envelope cut-off frequency was at 400 Hz. This supports the idea that the fixed FFT approach does restrict temporal envelope information and therefore limits the upper range of performance with consonant features that are more reliant on temporal envelope processing. In order to determine whether this hypothesis can be supported, a similar exercise in temporal analysis as undertaken in 2.3.3 (use of an objective TMTF measurement across stimulation rates and for phonologically relevant carrier frequencies) would need to be undertaken for other CI processing strategies and other devices.

A further question is whether the filterbank (and also sampling and selection) used in the Nucleus 24 is an optimal approach to spectral analysis. It is therefore important to consider how the current work has contributed to the literature on channel number. The first point to make is that relatively little work has looked at channel number in the context of peak-picking strategies. Dorman et al. (2002) has stated that there is equivalence (in terms of perceptual effects) between channel number in a fixed-channel strategy and number of peaks in a peak-picking strategy. If this is the case, then the comparison made in experiments 3 and 4 between 4-of-7 and 12-of-20 ACE strategies is a comparison between 4 and 12 channels. This is quite a marked difference in spectral resolution, in theory, and it is therefore perhaps surprising that CI user showed absolutely no effect for changes in channel number. This is despite the fact that listeners had much less experience of the 4-of-7 condition and therefore a confounding effect of acclimatisation in itself might have been expected to yield worse performance. Why should this be? In order to understand this, it is important to consider the AM data regarding place transmission. Here channel number did have an effect on transmission of place (the consonant feature which is most reliant on spectral information), but only with AMs with no channel interaction or 1 mm channel interaction, as shown in 5.5. By contrast, the 3.3 mm channel interaction model showed no effect for channel number for any feature, and the same finding was obtained in the CI users. The corollary of this is that channel interaction is implicated in the lack of improvement in place transmission (of around 10%) when changing from 4/7 to 12/20 MAP condition for the CI users. It is also worth noting that transmission in the best AM or CI user condition was still worse than NH listeners' performance by some 30%. By contrast, the difference between the 12/20 MAP conditions and 4/7 MAP condition for the no channel interaction model was around

10%. By the same token, for most features, the difference between the best AM and NH was still far greater than the difference between AMs with and without channel interaction. This underlines the finding that CI processing is a more dominant factor in determining performance than channel interaction, whether for more “spectral” or more “temporal” features.

The findings underline the need to look at better ways of filtering the incoming signal as it seems clear that the FFT filterbank approach used in the Nucleus 24 is not an optimal approach. It may be that improvements in filterbank processing would mean that channel interaction becomes a bigger problem for better CI users, but it is first necessary to establish those improvements before this can be determined. In this context, it is worth noting that the particular signal processing approaches used in CIs are largely based on previously available techniques rather than based on data of direct relevance to auditory or speech processing. Other approaches to filterbank processing have been suggested in the recent literature, e.g. wavelet analysis {Yao and Zhang, 2002). It can be anticipated that newer techniques should provide better frequency and temporal resolution than those in current devices. It is essential that a filterbank used for CI processing should provide temporal information with better accuracy up to higher modulation rates than was indicated by the TMTFs measured for the Nucleus 24. This recommendation adds to the more well-established finding that CI processing limits spectral resolution with all currently available CI devices. Moreover, the finding that place and fricative transmission were so poor even in the best AMs with least channel interaction, implementing a 12/20 ACE strategy, shows that the filterbank also provides inadequate spectral information and this imposes limitations on even the best performers with (presumably) the least electrical/neural interface information loss.

7.5 Electrical/neural interface and variations between users

A number of questions and hypotheses arose concerning the role of the electrical/neural interface in determining consonant recognition. It was hypothesised that spectral channel interaction determines differences between individual CI users. The simple assumption here would be that place of articulation perception should show the greatest variance between users, as this feature relies on spectral resolution to a greater extent than other features. This was clearly not the case here or in the study by Munson et al. (2003). This in itself does not disprove a role for spectral channel interaction, as the lack of further deterioration with place with channel overlap in the AM could be explained by the fact that formant transition information is essentially removed by CI processing, and further spectral channel interaction cannot therefore worsen performance for this feature further, at least until very high values of overlap are reached.

It is essential to understand the source of variations between CI users, and the data obtained from experiment 3 can help to illuminate this area and to follow on from the work of Munson et al. (2003). The comparison between better and worse CI users in this study does not support the argument of Munson et al. (2003) that there is no quantitative difference, e.g. pattern of feature transmission, between better and worse users. In the present study the smallest differences between better and worse users were found for voicing and envelope (around 15%) while the largest differences were found for nasality (35%) and fricative (30%). By contrast, Munson et al. (2003) found a uniform difference of around 30% in feature transmission in quiet for voicing, place and manner tested in the /aCa/ vowel environment, for a group of 30 users of either the Nucleus 22 or Clarion (version 1.2) device. The study differs from the current work in a number of ways: larger subject number, different (and varied) devices and processing parameters, vowel environment and number of stimuli used in the consonant confusion task and the number of features used in analysis and inclusion of a noise condition in this study. In the present study, the greatest difference between the two subgroups was in nasality and fricative transmission. Nasality is the feature which most relies on low-amplitude formant cues and therefore low-frequency audibility (and frequency resolution). By contrast, voicing and envelope show the smallest difference between better and worse CI users. The difference in the cues signalling these features as compared with nasality is the relative amplitude (less for nasality and greater for envelope and voicing). Taken together, these findings suggest a role for low-frequency temporal/amplitude resolution in determining performance differences. This fits well with the finding of Fu (2002) that temporal resolution (measured by TMTFs and averaged across sensation levels) was a good predictor of consonant recognition.

Another possibility is that forward masking is more marked with worse performing CI users. This would suggest that the difference between better and worse CI users would look similar to the difference in performance with and without noise, as in both cases the same mechanism would apply, e.g. a reduction in salience of within-channel amplitude fluctuations. To some extent this is supported by the data, as nasality is highly affected by both. However, these possibilities must remain as untested hypotheses. (Suggestions for how these might be tested are given in 7.8.) The more general point is that, for worse CI users, performance was considerably worse than predicted with an AM with even with 3.3 mm spread of excitation. The arguments in this section suggest that these further variations are unlikely to be solely or mainly due to spectral channel interaction.

7.6. Effects of background noise

A further important question is what implications the present study has for the understanding of the effects of background noise on speech perception in CI users. Noise had the expected effect on temporal envelope cues, both in the AM(s) and in the CI user study. This effect applied across model configurations, MAP conditions and individual users. This suggests that this is a robust finding and supports the very limited data available from the literature, e.g. Friesen, 2001.

The question is: how to reduce the distortions in the envelope fluctuations introduced by noise. It should be noted that the effect will be even greater at less favourable SNRs than those used in the present study and for more "realistic" noise types with non-random envelope fluctuations, e.g. babble noise. The greatest noise effect was with nasality transmission, which relies on the resolution of low-intensity low-frequency spectral components. NH listeners do not experience any difficulty in determining nasality for positive SNRs (see figure 5.11) and yet a marked reduction in nasality transmission was found across CI users and different types of AM. This suggests that audibility and dynamic range are likely to be important factors in explaining susceptibility to noise interference in CI users and, for better users at reasonably favourable SNRs, this may be the dominant factor determining performance. One aspect of CI processing that has been somewhat ignored in this work is the amplitude range and also the quantization introduced by the CI processing at the mapping stage (not to be confused with quantization in the stricter sense of the term at the point of ADC). Inadequate amplitude quantization would impose an upper limit on CI performance as would inadequacies in the residual neural capacity to code envelope level fluctuations. In these AMs no explicit attempt to match amplitude resolution characteristics was undertaken, and it may therefore be fortuitous that these perceptual abilities were mapped so well between AM and CI conditions.

The worse performing CI users were far more susceptible to noise interference across feature types. This is likely to mean that those users were at the less favourable end of their individual SNR functions and that a more favourable SNR (e.g. +15 dB SNR) would be needed for those individuals to tease out differences between features. One possible interpretation is that the worse the transmission of any feature the worse the noise interference for that feature. However, that is clearly not borne out by the findings. First, place transmission was the worst feature in quiet but there was no effect of noise for better CI users or for AM subjects and the magnitude of the noise effect was less for this feature than for other features for CI users. Individual CI user data also failed to support this possible explanation. The reverse also did not apply (as it did in experiment 1), e.g. there did not appear to be a correlation between better-transmitted features and worse noise effects.

7.7. Overall conceptual map

It is appropriate to consider the findings of the present study in the context of the overall conceptual model of information transmission/loss associated with CI processing, with particular reference to the Nucleus 24 CI system, shown in figure 2.22.

At the input stage low-pass filtering in the analogue filter determines the maximum frequency available in the signal. ADC could introduce quantization noise (though in practice this is likely to be low even with 8-bit resolution) and the limited input dynamic range of the Nucleus 24 system could reduce amplitude resolution and therefore reduce the salience of envelope fluctuations. Pre-emphasis increases high-frequency audibility at the expense of low-frequency audibility. The findings related to voicing and nasality transmission suggest that the combination of pre-emphasis and reduced amplitude information in apical channels could be a limiting factor affecting transmission of voicing and nasality, particularly in noise. This would need to be tested by comparing performance with and without pre-emphasis and with and without increased dynamic range (e.g. such as provided by Adaptive Dynamic Range Optimisation). However, the focus of the present study was not input stage characteristics (and these parameters were not varied in the experiments). Consequently, these conclusions must be taken as suggestive rather than definitive.

In terms of frequency analysis and envelope extraction, it was anticipated that there is loss of temporal information through envelope extraction and loss of spectral detail associated with FFT analysis and recombination into a small number of channels. What is of particular interest from the present study is the finding that even temporal envelope (and periodicity) information is lost via the filterbank used in the Nucleus 24 system, and that this information loss is implicated in the fact that no consonant features were transmitted with 100% accuracy. Moreover, increases in stimulation rate had little or no bearing on coding of temporal information. The spectral limitations of current CI processing strategies have been noted in a number of studies and are not unique to the present study, nor are they unique to the specific filterbank approach used in the Nucleus 24. However, the present study has strengthened the argument that limitations in spectral information in consonant, e.g. place transmission in particular, can be explained by information loss due to CI processing rather than channel interaction.

At the electrical-neural interface, it was suggested that there should be loss of temporal information because of abnormal temporal coding in the excited auditory nerve and loss of both temporal and (particularly) spectral information due to channel interaction. The striking finding from the present study was that differences between better and worse CI users were related primarily to differences in temporal envelope processing, not spectral processing, or at least this is what the differences in feature transmission and error patterns strongly suggested. This finding contrasted with Munson et al. (2003), probably because different feature and also noise effects were considered in the present study, and they support the idea that there is a stronger relationship between consonant recognition and individual variations in electrical/neural interface temporal envelope resolution abilities than spectral resolution abilities.

It is proposed that the conceptual model suggested in figure 2.22 could provide the basis of a genuine "model", in the broader sense of a mathematical and physiologically realistic model of CI neural stimulation. Inclusion of a simple channel overlap term is a small step in this direction. What is needed is a means of modelling both temporal and spectral aspects of the electrical/neural interface and more central processing.

7.8. Recommendations for further research and development

The present study showed that it was possible to predict the magnitude and pattern of consonant feature transmission in CI users by using a "closely matched" acoustic model. It was shown that, for better CI users, the magnitude of the deficit to consonant recognition due to loss of information with CI processing was much greater than the loss of information due to spectral channel interaction. Moreover, for worse CI users, the pattern of consonant recognition suggested a more important role for deficits in the processing of within-channel temporal information, rather than increased amounts of spectral channel interaction, in explaining why these users were worse at consonant recognition than better users. The findings also showed that there was no benefit to changing stimulation rate even more than threefold, and that this lack of benefit was due to CI processing limitations in the device used, and corresponding AMs. Acoustic measurements showed the marked reductions in temporal information available at higher modulation rates through the Nucleus 24 implant. It also showed that there was no benefit from increasing channel/maxima number approximately threefold, although this was due in part to channel interaction as well as to CI processing. The loss of spectral information was also apparent from acoustic analyses, although this is better-established in the existing literature for a range of CI devices.

Recommendations can be divided into those concerning: filterbank design; electrical/neural interface factors and methodology. Concerning temporal aspects of filterbank design, there is a need to determine whether other currently available processing approaches provide more temporal information than provided by the Nucleus 24 FFT filterbank. This can be achieved by undertaking objective TMTF measurements, along the lines of those reported in 2.3.3, for other approaches. It was suggested that a set of IIR band-pass filters could provide an advantage in this respect and therefore it would be of interest to measure TMTFs for this type of filterbank at varying envelope cut-off frequencies as well as varying stimulation rates, to determine whether there are advantages over the FFT approach in temporal processing. TMTFs could also be used to compare standard envelope extraction with Hilbert envelope extraction, as it has been claimed that the Hilbert transform provides a better representation of the envelope. The TMTF measurements undertaken in the present study should also be undertaken at different intensity levels (those undertaken in 1.4.3 all used stimuli which were near to saturation, e.g. at upper levels within the electrical dynamic range of the system), as it has been claimed that benefits to stimulation rate obtained in the Nucleus 24 system have been obtained at lower intensity levels (Holden, 2002). Also, Fu (2002) showed that behavioural TMTFs in CI users showed cut-offs at lower frequencies (e.g. worse temporal resolution) with decreasing intensity. Finally, within the specific context of the Nucleus FFT filterbank approach, might a shorter FFT be useful? Although frequency resolution would be reduced with, say, a 64-point FFT (e.g. 32 real bins), this might benefit temporal processing and the trade-off might be worthwhile.

Related to this is the need to improve audibility/dynamic range in low frequencies, particularly for nasality and voicing transmission in noise, also in quiet and perhaps for some manner distinctions and envelope. As noted, it would be of interest to co-vary pre-emphasis and input dynamic range (e.g. via ADRO or other forms of dynamic range optimisation) to determine if consonant recognition, particularly in noise, could be improved.

Concerning spectral aspects of filterbank design, it was clear from the current work, as in other studies, that the main limitation on consonant recognition is transmission of place (though here fricative was also implicated, albeit to a lesser extent). Place transmission is almost certainly restricted primarily by loss of formant transition information- this is the case with only 20 channels even if no channel interaction is assumed, as is shown by AM with no channel interaction (and by the uniformity of <60% place transmission across different types of AM and varying CI studies). It would be of interest to determine what channel number is required to adequately code formant transitions which are important to place transmission. This could be tested empirically using an AM with varying channel number to values well above numbers currently available in CI devices or by looking at place transmission in the very best (e.g. lowest channel interaction) CI users. However, AM work is needed first. Place transmission would be a useful outcome measure to determine any processing modification whose aim is to improve access to formant and formant transition information, or otherwise to provide an increase in effective channel number. Voicing in noise is probably also implicated as F1 transition could be a useful cue.

As noted in 7.5, there is a clear need to further understand the electrical/neural interface factors determining performance in worse CI users. The likely factors are:

- A greater degree of spectral channel interaction than that modelled by the 3.3 mm spread with the Laneau et al. (2006) model
- A more sophisticated model of spectral channel interaction and other spectral anomalies (Throckmorton and Collins, 2002)
- Temporal channel interaction
- Amplitude resolution or dynamic range

For reasons outlined previously, it seems possible that within-channel temporal/amplitude resolution is more important than spectral channel interaction. This is supported by Fu and Shannon (2000b) who found a strong relationship between within-channel temporal resolution and consonant perception.

Additionally, there are some more purely "methodological" issues raised. A direct comparison between performance with /iCi/ and /aCa/ vowel environments in CI users would be useful to definitively resolve the question of consonant recognition as a function of vowel environment. An alternative approach to consonant recognition for this type of research is also implied by the findings. This would be to use a specific subset of English consonants. One logical approach would be to use one manner category only and to assess voicing and place errors within that category. Further work has been undertaken by an MSc student supervised by the author. In that work two "reduced" forms of the consonant confusion measure were used in CI users, one with fricatives only and the other with stops only.

Chapter 8. Conclusions

Consonant feature transmission in CI users can be modelled with a high degree of accuracy using a carefully matched acoustic model in which great care is taken to match processing parameters to those used by the CI subjects, and where more general acoustic model parameters, such as carrier stimulus, are carefully chosen. However, the accuracy of the model is very good for better performing CI users and very poor for worse performing CI users. Deficits in consonant recognition in better performing CI users can be attributed primarily to information loss in CI processing, with channel interaction playing a markedly smaller role. Consonant recognition in worse-performing CI users is worse even than that predicted by a model with quite a high degree of channel interaction. The pattern of consonant feature transmission suggested that deficits in temporal and amplitude resolution may be more important than channel interaction in explaining performance variations.

It is possible to obtain useful CI acoustic model data with a relatively small amount of pre-experiment acclimatisation time, even with considerable spectral distortions in the acoustic model stimuli (introduced to mimic aspects of information loss due to the electrical/neural interface). By comparing CI user data with equivalent acoustic model data in which an identical set of processing parameters are implemented, it is possible to make strong inferences about the relative contribution of different factors in determining deficits in speech perception abilities experienced by cochlear implant users. The use of a detailed phonological analysis of consonant confusions can also reveal perceptual abnormalities that cannot be analysed using more generic or global speech perception measures.

The combination of acoustic and behavioural measurements undertaken in the study show that the filterbank used in the Nucleus 24 processor reduces both temporal and spectral information in speech, and this defines the ceiling of performance. The loss of spectral resolution because of CI processing explains the poor transmission of the consonant place feature, and this occurs even with a relatively large number of channels and no channel interaction. However, the Nucleus 24 also shows a poor temporal response at higher modulation rates, which in turn limits transmission of consonant features that are more reliant on temporal information. The absence of benefit with increases stimulation rate shown in the present study, and in most other studies of the Nucleus device, can be predicted by the fact that the temporal response of the processor changes very little across stimulation rates. The absence of benefit associated with increasing channel number can be attributed to loss of spectral detail associated with CI processing although spectral channel interaction also plays a modest role in restricting benefit with a higher channel number. The loss of audibility and dynamic range in lower frequencies is also implicated in some of the deficits in consonant recognition shown by CI users, although further work is clearly needed to understand reasons for within-user variation.

References

Abbas PJ, Hughes ML, Brown CJ, Miller CA, South H (2004) Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential. *Audiol Neurootol* **9**: 203-213

Anderson I, Weichbold V, D'Haese P (2002) Recent results with the MED-EL COMBI 40+ cochlear implant and TEMPO+ behind-the-ear processor. *Ear Nose Throat J* **81**: 229-233

Bench J, Kowal A, Bamford J (1979) The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol* **13**: 108-112

Black RC, Clark GM (1980) Differential electrical excitation of the auditory nerve. *J Acoust Soc Am* **67**: 868-874

Blamey PJ (2005) Adaptive dynamic range optimization (ADRO): a digital amplification strategy for hearing aids and cochlear implants. *Trends Amplif* **9**: 77-98

Blamey PJ, Dooley GJ (1993) Pattern recognition and masking in cochlear implant patients. *Prog Brain Res* **97**: 271-278

Blamey PJ, Martin LF, Clark GM (1985) A comparison of three speech coding strategies using an acoustic model of a cochlear implant. *J Acoust Soc Am* **77**: 209-217

Boersma P, Weenink D (2005):*Praat: doing phonetics by computer (Version 4.4)* [Computer program].http://www.praat.org/

Boex C, de Balthasar C, Kos MI, Pelizzone M (2003a) Electrical field interactions in different cochlear implant systems. *J Acoust Soc Am* **114**: 2049-2057

Boex C, Kos MI, Pelizzone M (2003b) Forward masking in different cochlear implant systems. *J Acoust Soc Am* **114**: 2058-2065

Boex C, Pelizzone M, Montandon P (1996) Speech recognition with a CIS strategy for the ineraid multichannel cochlear implant. *Am J Otol* **17**: 61-68

Bonnet RM, Frijns JH, Peeters S, Briaire JJ (2004) Speech recognition with a cochlear implant using triphasic charge-balanced pulses. *Acta Otolaryngol* **124**: 371-375

Brill SM, Gstottner W, Helms J, von Ilberg C, Baumgartner W, Muller J, Kiefer J (1997) Optimization of channel number and stimulation rate for the fast continuous interleaved sampling strategy in the COMBI 40+. *Am J Otol* **18**: S104-S106

Busby PA, Clark GM (1997) Pitch and loudness estimation for single and multiple pulse per period electric pulse rates by cochlear implant patients. *J Acoust Soc Am* **101**: 1687-1695

Busby PA, Tong YC, Clark GM (1993) The perception of temporal modulations by cochlear implant patients. *J Acoust Soc Am* **94**: 124-131

Busby PA, Whitford LA, Blamey PJ, Richardson LM, Clark GM (1994) Pitch perception for different modes of stimulation using the cochlear multiple-electrode prosthesis. *J Acoust Soc Am* **95**: 2658-2669

Carlyon RP, Deeks JM (2002) Limitations on rate discrimination. *J Acoust Soc Am* **112**: 1009-1025

Carlyon RP, van Wieringen A, Long CJ, Deeks JM, Wouters J (2002) Temporal pitch mechanisms in acoustic and electric hearing. *J Acoust Soc Am* **112**: 621-633

Cazals Y, Pelizzone M, Kasper A, Montandon P (1991) Indication of a relation between speech perception and temporal resolution for cochlear implantees. *Ann Otol Rhinol Laryngol* **100**: 893-895

Chatterjee M, Oba SI (2004) Across- and within-channel envelope interactions in cochlear implant listeners. *J Assoc Res Otolaryngol* **5**: 360-375

Chatterjee M, Shannon RV (1998) Forward masked excitation patterns in multielectrode electrical stimulation. *J Acoust Soc Am* **103**: 2565-2572

Childers DG, Wong CF (1994) Measuring and modeling vocal source-tract interaction. *IEEE Trans Biomed Eng* **41**: 663-671

Chomsky N and Halle M (1968) *The sound patterns of English.* New York: Harper and Row

Cochlear Ltd. (2002) Nucleus Implant Communicatior (NIC) system overview. N95291 Iss.1, Cochlear Ltd.

Cohen LT, Lenarz T, Battmer RD, Bender vS, Busby PA, Cowan RS (2005) A psychophysical forward masking comparison of longitudinal spread of neural excitation in the Contour and straight Nucleus electrode arrays. *Int J Audiol* **44**: 559-566

Cohen LT, Richardson LM, Saunders E, Cowan RS (2003) Spatial spread of neural excitation in cochlear implant recipients: comparison of improved ECAP method and psychophysical forward masking. *Hear Res* **179**: 72-87

Cohen LT, Saunders E, Clark GM (2001) Psychophysics of a prototype peri-modiolar cochlear implant electrode array. *Hear Res* **155**: 63-81

Cohen LT, Saunders E, Richardson LM (2004) Spatial spread of neural excitation: comparison of compound action potential and forward-masking data in cochlear implant recipients. *Int J Audiol* **43**: 346-355

Davis M, Taylor K, Johnsrude I, Carlyon B (2004) How do cochlear implant users learng to understand speech? Transfer of learning between different carriers with vocoded speech. *Poster presented at British Society of Audiology Experimental Short Papers Meeting*, London, September 2004

Dawson PW, McKay CM, Busby PA, Grayden DB, Clark GM (2000) Electrode discrimination and speech perception in young children using cochlear implants. *Ear Hear* **21**: 597-607

de Balthasar C, Boex C, Cosendai G, Valentini G, Sigrist A, Pelizzone M (2003) Channel interactions with high-rate biphasic electrical stimulation in cochlear implant subjects. *Hear Res* **182**: 77-87

Deeks JM, Carlyon RP (2004) Simulations of cochlear implant hearing using filtered harmonic complexes: implications for concurrent sound segregation. *J Acoust Soc Am* **115**: 1736-1746

Diller N, Battmer RD, Doring WH, Muller-Deile J (1995) Multicentric field evaluation of a new speech coding strategy for cochlear implants. *Audiology* **34**:145-149

Donaldson GS, Nelson DA (2000) Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies. *J Acoust Soc Am* **107**: 1645-1658

Dorman MF (1995) Test of the use of amplitude envelope cues to consonant identity by patients who use the Ineraid cochlear implant. *Ann Otol Rhinol Laryngol Suppl* **166**: 223-224

Dorman MF, Dankowski K, McCandless G, Parkin JL, Smith L (1991) Vowel and consonant recognition with the aid of a multichannel cochlear implant. *Q J Exp Psychol A* **43**: 585-601

Dorman MF, Loizou PC (1997) Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants. *Am J Otol* **18**: S113-S114

Dorman MF, Loizou PC (1998) The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels. *Ear Hear* **19**: 162-166

Dorman MF, Loizou PC, Fitzke J (1998a) The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors. *Ear Hear* **19**: 481-484

Dorman MF, Loizou PC, Fitzke J, Tu Z (1998b) The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *J Acoust Soc Am* **104**: 3583-3585

Dorman MF, Loizou PC, Kemp LL, Kirk KI (2000) Word recognition by children listening to speech processed into a small number of channels: data from normal-hearing children and children with cochlear implants. *Ear Hear* **21**: 590-596

Dorman MF, Loizou PC, Rainey D (1997a) Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J Acoust Soc Am* **102**: 2993-2996

Dorman MF, Loizou PC, Rainey D (1997b) Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J Acoust Soc Am* **102**: 2403-2411

Dorman MF, Loizou PC, Spahr AJ, Maloff E (2002) A comparison of the speech understanding provided by acoustic models of fixed-channel and channel-picking signal processors for cochlear implants. *J Speech Lang Hear Res* **45**: 783-788

Dorman MF, Soli S, Dankowski K, Smith LM, McCandless G, Parkin J (1990) Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant. *J Acoust Soc Am* **88**: 2074-2079

Dudley, H (1939) The vocoder. *Bell Labs Record* **17**:122-26

Fant G. (1970) Acoustic Theory of Speech Production. Mouton, The Hague

Faulkner A (2006) Adaptation to distorted frequency-to-place maps: implications of simulations in normal listeners for cochlear implants and electroacoustic stimulation. *Audiol Neurootol* **11 Suppl 1**: 21-26

Faulkner A, Darling, AM, Rosen S, Huckvale M (1995) Modelling cue interaction in the perception of the voiceless fricative/affricate contrast. *Lang Cog Proc* **10**: 369-375

Faulkner A, Rosen S, Smith C (2000) Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: implications for cochlear implants. *J Acoust Soc Am* **108**: 1877-1887

Faulkner A, Rosen S, Stanton D (2003) Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth. *J Acoust Soc Am* **113**: 1073-1080

Faulkner A, Rosen S, Wilkinson L (2001) Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. *Ear Hear* **22**: 431-438

Favre E, Pelizzone M (1993) Channel interactions in patients using the Ineraid multichannel cochlear implant. *Hear Res* **66**: 150-156

Fetterman BL, Domico EH (2002) Speech recognition in background noise of cochlear implant patients. *Otolaryngol Head Neck Surg* **126**: 257-263

Friesen LM, Shannon RV, Baskent D, Wang X (2001) Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am* **110**: 1150-1163

Friesen LM, Shannon RV, Slattery WH, III (1999) The effect of frequency allocation on phoneme recognition with the nucleus 22 cochlear implant. *Am J Otol* **20**: 729-734

Fu QJ (2002) Temporal processing and speech recognition in cochlear implants. *Neuroreport* **13**: 1635-1639

Fu QJ, Nogaki G (2005) Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J Assoc Res Otolaryngol* **6**: 19-27

Fu QJ, Shannon RV (2000a) Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners. *J Acoust Soc Am* **107**: 589-597

Fu QJ, Shannon RV (2000b) Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users. *Ear Hear* **21**: 227-235

Fu QJ, Shannon RV, Wang X (1998) Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. *J Acoust Soc Am* **104**: 3586-3596

Galvin JJ, III, Fu QJ (2005) Effects of stimulation rate, mode and level on modulation detection by cochlear implant users. *J Assoc Res Otolaryngol* **6**: 269-279

Geurts L, Wouters J (1999) Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants. *J Acoust Soc Am* **105**: 2476-2484

Goedegebure A, Goedegebure-Hulshof M, Verschuure H, Dreschler WA (2002) The effects of phonemic compression and anti-upward-spread-of-masking (anti-USOM) on the perception of articulatory features in hearing-impaired listeners. *Int J Audiol* **41**: 414-428

Gonzalez J, Oliver JC (2005) Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *J Acoust Soc Am* **118**: 461-470

Green T, Faulkner A, Rosen S (2004) Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J Acoust Soc Am* **116**: 2298-2310

Green T, Faulkner A, Rosen S, Macherey O (2005) Enhancement of temporal periodicity cues in cochlear implants: effects on prosodic perception and vowel identification. *J Acoust Soc Am* 118: 375-385

Greenwood DD (1990) A cochlear frequency-position function for several species--29 years later. *J Acoust Soc Am* **87**: 2592-2605

Henry BA, Turner CW (2003) The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *J Acoust Soc Am* **113**: 2861-2873

Hescot F, Lorenzi C, Debruille X, Camus JF (2000) Measurement of the temporal-modulation transfer function for a single listener with cochlear hearing loss and left-hemisphere damage. *Br J Audiol* **34**: 341-351

Holden LK, Skinner MW, Holden TA, Demorest ME (2002) Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy. *Ear Hear* **23**: 463-476

Holden-Pitt LD, Hazan V, Revoile SG, Edward DM, Droge J (1995) Temporal and spectral cue use for initial plosive voicing perception by hearing-impaired children and normal-hearing children and adults. *Eur J Disord Commun* **30**: 417-434

Howell D (2003) Fundamental Statistics for the Behavioral Sciences. Brooks Cole

Jiang J, Chen M, Alwan A (2006) On the perception of voicing in syllable-initial plosives in noise. *J Acoust Soc Am* **119**: 1092-1105

Ketten DR, Skinner MW, Wang G, Vannier MW, Gates GA, Neely JG (1998) In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays. *Ann Otol Rhinol Laryngol Suppl* **175**: 1-16

Kohlrausch A, Fassel R, Dau T (2000) The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *J Acoust Soc Am* **108**: 723-734

Kral A, Hartmann R, Mortazavi D, Klinke R (1998) Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents. *Hear Res* **121**: 11-28

Lai WK, Bogli H, Dillier N (2003) A software tool for analyzing multichannel cochlear implant signals. *Ear Hear* **24**: 380-391

Laneau J, Moonen M, Wouters J (2006) Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. *J Acoust Soc Am* **119**: 491-506

Laneau J, Wouters J, Moonen M (2004) Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *J Acoust Soc Am* **116**: 3606-3619

Lawson DT, Wilson BS, Finley CC (1993) New processing strategies for multichannel cochlear prostheses. *Prog Brain Res* **97**: 313-321

Lieberman P and Blumstein SE (1988) *Speech physiology, speech perception and acoustic phonetics.* Cambridge: Cambridge University Press

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* **74**: 431-461

Liberman AM, Delattre PC, Gerstman LJ, Cooper FS (1956) Tempo of frequency changes as a cue for distinguishing classes of speech sounds. *J Exp Psychol* **52**: 127-137

Lim HH, Tong YC, Clark GM (1989) Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea. *J Acoust Soc* **86**: 971-980

Loizou PC (1999) Signal-processing techniques for cochlear implants. *IEEE Eng Med Biol Mag* **18**: 34-46

Loizou PC, Dorman M, Tu Z (1999) On the number of channels needed to understand speech. *J Acoust Soc Am* **106**: 2097-2103

Loizou PC, Dorman MF, Tu Z, Fitzke J (2000a) Recognition of sentences in noise by normal-hearing listeners using simulations of speak-type cochlear implant signal processors. *Ann Otol Rhinol Laryngol Suppl* **185**: 67-68

Loizou PC, Poroy O (2001) Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners. *J Acoust Soc Am* **110**: 1619-1627

Loizou PC, Poroy O, Dorman M (2000b) The effect of parametric variations of cochlear implant processors on speech understanding. *J Acoust Soc Am* **108**: 790-802

Loizou PC, Stickney G, Mishra L, Assmann P (2003) Comparison of speech processing strategies used in the Clarion implant processor. *Ear Hear* **24**: 12-19

Loizou PC (2006) Speech processing in vocoder-centric cochlear implants. *Adv Otorhinolaryngol* **64**: 109-143

Malecot A Acoustic cues for nasal consonants. *Lang* **32**: 274-278

Matsuoka AJ, Rubinstein JT, Abbas PJ, Miller CA (2001) The effects of interpulse interval on stochastic properties of electrical stimulation: models and measurements. *IEEE Trans Biomed Eng* **48**: 416-424

McKay CM, Henshall KR, Hull AE (2005) The effect of rate of stimulation on perception of spectral shape by cochlear implantees. *J Acoust Soc Am* **118**: 386-392

Meyer TA, Frisch SA, Pisoni DB, Miyamoto RT, Svirsky MA (2003) Modeling open-set spoken word recognition in postlingually deafened adults after cochlear implantation: some preliminary results with the neighborhood activation model. *Otol Neurotol* **24**: 612-620

Miller GA and Nicely PE (1955) An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am* 27: 338-352

Moore BC (1996) Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear Hear* **17**: 133-161

Moore BC, Glasberg BR (2001) Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners. *J Acoust Soc Am* **110**: 1067-1073

Munson B, Donaldson GS, Allen SL, Collison EA, Nelson DA (2003) Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability. *J Acoust Soc Am* **113**: 925-935

Nelson DA, Van Tasell DJ, Schroder AC, Soli S, Levine S (1995) Electrode ranking of "place pitch" and speech recognition in electrical hearing. *J Acoust Soc Am* **98**: 1987-1999

Nelson PB, Jin SH, Carney AE, Nelson DA (2003) Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *J Acoust Soc Am* **113**: 961-968

Nie K, Barco A, Zeng FG (2006) Spectral and temporal cues in cochlear implant speech perception. *Ear Hear* **27**: 208-217

Nie K, Stickney G, Zeng FG (2005) Encoding frequency modulations to improve cochlear implant performance in noise. *IEEE Trans Biomed Eng* **52**: 64-73

O'Connor JD, Gerstman LF, Liberman AM, Delattre PC, Cooper FS Acoustic cues for the perception of initial /w,j,r,l/ in English. *Word* **13**: 22-43

Owens E, Talbott CB, Schubert ED (1968) Vowel discrimination of hearing-impaired listeners. *J Speech Hear Res* **11**: 648-655

Qin MK, Oxenham AJ (2003) Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *J Acoust Soc Am* **114**: 446-454

Raphael, LJ and Dorman MF (1980) Acoustic cues for a fricative-affricate contrast in word-final position. J Phonetics **8**:397-405

Repp BH (1979) Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Lang Speech* **22**: 173-189

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci* **336**: 367-373

Rosen S, Faulkner A, Wilkinson L (1999) Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *J Acoust Soc Am* **106**: 3629-3636

Shannon RV (1992) Temporal modulation transfer functions in patients with cochlear implants. *J Acoust Soc Am* **91**: 2156-2164

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* **270**: 303-304

Shannon RV, Zeng FG, Wygonski J (1998) Speech recognition with altered spectral distribution of envelope cues. *J Acoust Soc Am* **104**: 2467-2476

Skinner MW, Arndt PL, Staller SJ (2002) Nucleus 24 advanced encoder conversion strategy: performance versus preference. *Ear Hear* **23(S1)**: 2S-17S

Steeneken HJ, Houtgast T (1980) A physical method for measuring speech-transmission quality. *J Acoust Soc Am* **67**: 318-326

Stevens KN, Blumstein SE, Glicksman L, Burton M, Kurowski K (1992) Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *J Acoust Soc Am* **91**: 2979-3000

Stickney GS (2001) *Analyses of speech processing strategies for cochlear implants and the effects of electrode channel interaction.* PhD dissertation, University of Dallas

Stickney GS, Loizou PC, Mishra LN, Assmann PF, Shannon RV, Opie JM (2006) Effects of electrode design and configuration on channel interactions. *Hear Res* **211**: 33-45

Summerfield AQ, Nakisa MJ, Mccormick B, Archbold S, Gibbin KP, Odonoghue GM (2002) Use of vocalic information in the identification of /s/ and /sh/ by children with cochlear implants. *Ear Hear* **23**: 58-77

Teoh SW, Neuburger HS, Svirsky MA (2003) Acoustic and electrical pattern analysis of consonant perceptual cues used by cochlear implant users. *Audiol Neurootol* **8**: 269-285

Throckmorton CS, Collins LM (1999) Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects. *J Acoust Soc Am* **105**: 861-873

Throckmorton CS, Collins LM (2002) The effect of channel interactions on speech recognition in cochlear implant subjects: predictions from an acoustic model. *J Acoust Soc Am* **112**: 285-296

Titze IR (2004) A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice. *J Voice* **18**: 292-298

UK Cochlear Implant Study Group (2004) Criteria of candidacy for unilateral cochlear implantation in postlingually deafened adults II: cost-effectiveness analysis. *Ear Hear* **25**: 336-360

Valimaa TT, Maatta TK, Lopponen HJ, Sorri MJ (2002a) Phoneme recognition and confusions with multichannel cochlear implants: consonants. *J Speech Lang Hear Res* **45**: 1055-1069

Valimaa TT, Maatta TK, Lopponen HJ, Sorri MJ (2002b) Phoneme recognition and confusions with multichannel cochlear implants: vowels. *J Speech Lang Hear Res* **45**: 1039-1054

Van Tasell DJ, Greenfield DG, Logemann JJ, Nelson DA (1992) Temporal cues for consonant recognition: training, talker generalization, and use in evaluation of cochlear implants. *J Acoust Soc Am* **92**: 1247-1257

Verschuur CA (2005) Effect of stimulation rate on speech perception in adult users of the Med-El CIS speech processing strategy. *Int J Audiol* **44**: 58-63

Viemeister NF (1979) Temporal modulation transfer functions based upon modulation thresholds. *J Acoust Soc Am* **66**: 1364-1380

Wang MD, Bilger RC (1973) Consonant confusions in noise: a study of perceptual features. *J Acoust Soc Am* **54**: 1248-1266

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) Better speech recognition with cochlear implants. *Nature* **352**: 236-238

Wilson BS (1997) The future of cochlear implants. *Br J Audiol* **31**: 205-225

Xu L, Thompson CS, Pfingst BE (2005) Relative contributions of spectral and temporal cues for phoneme recognition. *J Acoust Soc Am* **117**: 3255-3267

Yao J and Zhang Y (2002) The application of bionic wavelet transform to speech signal processing in cochlear implants using neural network simulations. *IEEE Trans Biomed Eng* 49: 1299-1309

Zeng FG (2004) Trends in cochlear implants. *Trends Amplif* **8**: 1-34

Zwolan TA, Collins LM, Wakefield GH (1997) Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects. *J Acoust Soc Am* **102**: 3673-3685

Appendix A. Investigation of the Front End of the Nucleus Sprint Speech Processor and Headset

M L Grasmeder, C A Verschuur

Introduction

The NIC-STREAM software includes simulations of the processing of the Nucleus 24 cochlear implant but did not (at the time of initial experimental work) currently include a simulation of the Nucleus 'front-end'. This includes the effect of the headset microphone, subsequent amplifier, anti-alias filter and AGC. In order to produce a simulation of these aspects of the processing, some investigations were made.

Method

At the calibrated spot in a soundproof room (approximating a free field environment), a Sprint microphone was placed on an artificial pinna and this was attached to a Sprint processor, set to sensitivity 10. A pair of modified monitor earphones was attached to the audio output socket on the Sprint and this was fed through to a line-in socket on a laptop computer. A sound sample of pink noise was played using standard clinic loudspeakers. Recordings of the processed sound were made using CoolEdit (now Adobe Audition) software, using a sampling rate 44100 Hz. The pink noise was played at different levels between 40 and 70 dB (A). Spectra were derived for the recorded samples and original sample. However, there was some difficulty in measuring the effect of the AGC, as the effect of the line-in input on the sound level could not be found independently. Hence no attempt was made to simulate the effect of the AGC, but instead a more simple comparison was made between the original signal before and after processing through the implant. Figure A1 shows the difference in energy between the original and processed signal as a function of frequency.

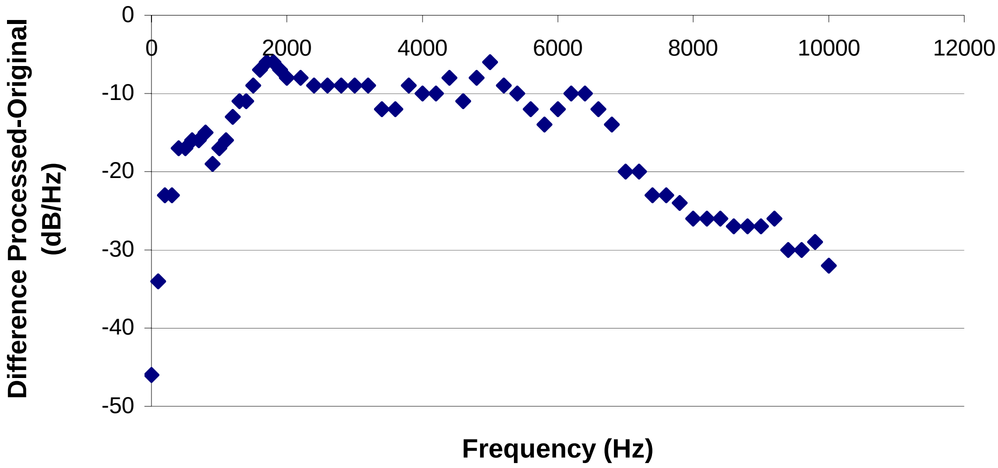

Figure A1. Effect of Nucleus input stage processing on frequency response of incoming sound

In order to design a filter that would mimic these results, the effect of the front end (as shown above) was separated into 3 frequency regions. The low frequencies showed an increasing output as a function of frequency. The mid frequencies showed a near flat response and the high frequencies show a decreasing output with frequency. Linear regression was used to fit the low and high frequency areas with a straight line graph. The mid frequency area was assumed to be a flat line. In summary, there is approximately a +6dB/octave slope for the low frequencies up to about 1700 Hz (5.4 dB/octave was measured). The frequency response is approximately flat until 5000 Hz, after which there is a –24 dB/octave slope to 10 kHz (measured value = -25.6 dB/octave).These data were used to produce a filter, which can be used prior to processing a sound sample through the NIC software, having the following characteristicis:

- Up to 1800 Hz, +6 dB per octave
- 1800 – 5000 Hz, flat
- 5000 – 10000 Hz (or above), -24 dB per octave

Summary

The Nucleus input stage processing includes a microphone, amplifier, anti-alias filter and AGC. In this investigation, the effect of these aspects of processing were analysed with respect to their effect on the amplitude spectrum of pink noise, although this did not include a characterisation of AGC effects. A filter was derived based on the measurements, and this was used prior to processing samples through the NIC software.

Appendix B. Summary of MANOVA and ANOVA details for experiments 1 to 4.

Table B1. Summary of MANOVA from Experiment 1, as described in 4.1.4. Seven dependent variables (six feature transmission values and total percentage correct) were entered. The factor was "listening conditions", which had five levels (unaltered, quiet AM, AM+10 dB SNR, AM +5 dB SNR and AM 0 dB SNR). Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
Listening condition	total	43514.822	4.000	10878.706	161.086	**<0.001**
	voicing	77254.933	4.000	19313.733	56.630	**<0.001**
	place	29064.156	4.000	7266.039	88.955	**<0.001**
	manner	43934.733	4.000	10983.683	213.232	**<0.001**
	fricative	51918.489	4.000	12979.622	134.337	**<0.001**
	nasality	89890.511	4.000	22472.628	97.013	**<0.001**
	envelope	55303.142	4.000	13825.785	96.458	**<0.001**
Error	total	5740.3	85.0	67.5		
	voicing	28989.2	85.0	341.0		
	place	6943.0	85.0	81.7		
	manner	4378.4	85.0	51.5		
	fricative	8212.7	85.0	96.6		
	nasality	19689.9	85.0	231.6		
	envelope	12183.5	85.0	143.3		

Table B2. Summary of results of ANOVA on from Experiment 1, as described in 4.1.4. The factor "feature" had six levels, corresponding to the six consonant feature transmission values. Only quiet acoustic model conditions were included in the analysis.

Source	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
Feature	40089.583	4	10022.396	103.661	<0.001
Error (feature)	8701.606	90	96.685		<0.001

Table B3. Summary of MANOVA from Experiment 2, as described in 4.3.3. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors, each with two levels: carrier, shift, vowel and noise. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
carrier	total	921.4848	1	921.4848	12.62628	**0.001**
	voicing	1330.578	1	1330.578	5.623042	**0.021**
	place	412.8483	1	412.8483	4.106917	**0.047**
	manner	583.57	1	583.57	11.41062	**0.001**

	nasality	18.08704	1	18.08704	0.039574	0.843
	envelope	152.4639	1	152.4639	0.319315	0.574
	fricative	1132.196	1	1132.196	13.02732	**0.001**
shift	total	124.6202	1	124.6202	1.707558	0.196
	voicing	229.6561	1	229.6561	0.97053	0.328
	place	24.20828	1	24.20828	0.240818	0.625
	manner	58.51203	1	58.51203	1.144094	0.289
	nasality	3420.603	1	3420.603	7.48419	**0.008**
	envelope	68.67803	1	68.67803	0.143837	0.706
	fricative	59.79188	1	59.79188	0.68798	0.410
vowel	total	848.5616	1	848.5616	11.62708	**0.001**
	voicing	116.7601	1	116.7601	0.49343	0.485
	place	2838.647	1	2838.647	28.23819	**0.000**
	manner	71.93357	1	71.93357	1.406527	0.240
	nasality	4039.35	1	4039.35	8.837993	**0.004**
	envelope	309.1969	1	309.1969	0.647572	0.424
	fricative	75.07212	1	75.07212	0.863798	0.356
noise	total	5142.162	1	5142.162	70.45841	**0.000**
	voicing	7004.546	1	7004.546	29.60132	**0.000**
	place	4502.69	1	4502.69	44.79169	**0.000**
	manner	6642.592	1	6642.592	129.8835	**0.000**
	nasality	41366.69	1	41366.69	90.50924	**0.000**
	envelope	859.7476	1	859.7476	1.800627	0.184
	fricative	4617.308	1	4617.308	53.12787	**0.000**
carrier * shift	total	6.496623	1	6.496623	0.089017	0.766
	voicing	31.26351	1	31.26351	0.13212	0.717
	place	0.339231	1	0.339231	0.003375	0.954
	manner	0.000277	1	0.000277	5.41E-06	0.998
	nasality	2097.43	1	2097.43	4.589123	0.036
	envelope	84.99212	1	84.99212	0.178005	0.675
	fricative	50.25489	1	50.25489	0.578245	0.450
carrier * vowel	total	66.35361	1	66.35361	0.909184	0.344
	voicing	16.66757	1	16.66757	0.070437	0.792
	place	94.77	1	94.77	0.94275	0.335
	manner	218.53	1	218.53	4.272946	0.043
	nasality	8782.073	1	8782.073	19.21495	0.000
	envelope	0.428431	1	0.428431	0.000897	0.976
	fricative	29.19003	1	29.19003	0.335868	0.564
shift * vowel	total	45.53453	1	45.53453	0.623919	0.433
	voicing	85.60689	1	85.60689	0.361776	0.550
	place	45.72188	1	45.72188	0.45483	0.503
	manner	20.84089	1	20.84089	0.407505	0.526
	nasality	3213.357	1	3213.357	7.03074	**0.010**
	envelope	95.52751	1	95.52751	0.20007	0.656
	fricative	4.121723	1	4.121723	0.047426	0.828
carrier * shift * vowel	total	29.76222	1	29.76222	0.407805	0.525
	voicing	0.3328	1	0.3328	0.001406	0.970
	place	392.8101	1	392.8101	3.907582	0.052
	manner	77.29923	1	77.29923	1.511442	0.223
	nasality	594.7991	1	594.7991	1.301405	0.258
	envelope	7.387692	1	7.387692	0.015473	0.901
	fricative	62.56849	1	62.56849	0.719928	0.399
carrier *	total	472.6917	1	472.6917	6.476869	**0.013**

noise						
	voicing	2.343877	1	2.343877	0.009905	0.921
	place	644.5824	1	644.5824	6.412153	**0.014**
	manner	753.6185	1	753.6185	14.7356	**0.000**
	nasality	406.4042	1	406.4042	0.889202	0.349
	envelope	55.74511	1	55.74511	0.116751	0.734
	fricative	146.4961	1	146.4961	1.68562	0.199
shift * noise	total	73.0197	1	73.0197	1.000523	0.321
	voicing	316.6543	1	316.6543	1.338186	0.252
	place	239.6825	1	239.6825	2.384305	0.128
	manner	5.822308	1	5.822308	0.113844	0.737
	nasality	209.1696	1	209.1696	0.457658	0.501
	envelope	184.0897	1	184.0897	0.385551	0.537
	fricative	293.7877	1	293.7877	3.380393	0.071
carrier * shift * noise	total	55.29047	1	55.29047	0.757596	0.387
	voicing	42.84308	1	42.84308	0.181056	0.672
	place	96.94231	1	96.94231	0.964359	0.330
	manner	4.800769	1	4.800769	0.09387	0.760
	nasality	222.2025	1	222.2025	0.486173	0.488
	envelope	110.8432	1	110.8432	0.232146	0.632
	fricative	94.23077	1	94.23077	1.084242	0.302
vowel * noise	total	239.0817	1	239.0817	3.275922	0.075
	voicing	290.3749	1	290.3749	1.227129	0.272
	place	636.72	1	636.72	6.33394	0.014
	manner	0.496277	1	0.496277	0.009704	0.922
	nasality	1048.433	1	1048.433	2.293944	0.135
	envelope	319.4241	1	319.4241	0.668991	0.416
	fricative	247.1248	1	247.1248	2.843478	0.097
carrier * vowel * noise	total	1.176008	1	1.176008	0.016114	0.899
	voicing	23.18228	1	23.18228	0.097969	0.755
	place	1.728277	1	1.728277	0.017192	0.896
	manner	166.6132	1	166.6132	3.25781	0.076
	nasality	2169.681	1	2169.681	4.747206	0.033
	envelope	0.295508	1	0.295508	0.000619	0.980
	fricative	2.806277	1	2.806277	0.03229	0.858
shift * vowel * noise	total	253.8848	1	253.8848	3.478755	0.067
	voicing	4.537108	1	4.537108	0.019174	0.890
	place	218.8581	1	218.8581	2.177149	0.145
	manner	52.88372	1	52.88372	1.034042	0.313
	nasality	41.7106	1	41.7106	0.091262	0.764
	envelope	93.3712	1	93.3712	0.195554	0.660
	fricative	0.847877	1	0.847877	0.009756	0.922
carrier * shift * vowel * noise	total	2.008623	1	2.008623	0.027522	0.869
	voicing	9.240123	1	9.240123	0.039049	0.844
	place	10.60212	1	10.60212	0.105467	0.746
	manner	4.189569	1	4.189569	0.081919	0.776
	nasality	853.9995	1	853.9995	1.868529	0.177
	envelope	4.396431	1	4.396431	0.009208	0.924
	fricative	4.212308	1	4.212308	0.048468	0.826

Error	total	4597.836	63	72.98152		
	voicing	14907.66	63	236.6295		
	place	6333.082	63	100.5251		
	manner	3221.99	63	51.1427		
	nasality	28793.76	63	457.0439		
	envelope	30080.68	63	477.4712		
	fricative	5475.288	63	86.90933		

Table B4. Summary of MANOVA from Experiment 3, as described in 5.3. Data from all 9 CI users in the experiment are included in the analysis. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were three factors: "noise condition", "stimulation rate" and " channel number", here summarised as "noise", "stimrate" and "channo", respectively. Significant effects at the *a priori* significance level (*p*≤0.05) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
noise	total	284.5869	1	284.5869	0.934738	0.339
	voicing	349.2347	1	349.2347	0.711424	0.403
	nasality	5719.427	1	5719.427	5.040231	**0.030**
	place	130.1356	1	130.1356	0.41796	0.521
	manner	736.8845	1	736.8845	1.948424	0.169
	fricative	14.02661	1	14.02661	0.018073	0.894
	envelope	497.2547	1	497.2547	1.629146	0.208
stimrate	total	28.38438	1	28.38438	0.09323	0.761
	voicing	0.171671	1	0.171671	0.00035	0.985
	nasality	848.0003	1	848.0003	0.747298	0.392
	place	17.21736	1	17.21736	0.055298	0.815
	manner	28.92533	1	28.92533	0.076483	0.783
	fricative	340.1374	1	340.1374	0.438269	0.511
	envelope	157.1275	1	157.1275	0.514794	0.477
channo	total	9.100278	1	9.100278	0.02989	0.863
	voicing	259.21	1	259.21	0.528035	0.471
	nasality	11.9025	1	11.9025	0.010489	0.919
	place	2.777778	1	2.777778	0.008921	0.925
	manner	107.1225	1	107.1225	0.283247	0.597
	fricative	20.85444	1	20.85444	0.026871	0.871
	envelope	196	1	196	0.642151	0.427
noise * stimrate	total	0.428824	1	0.428824	0.001408	0.970
	voicing	570.8172	1	570.8172	1.162809	0.287
	nasality	870.0361	1	870.0361	0.766717	0.386
	place	5.813926	1	5.813926	0.018673	0.892
	manner	95.81147	1	95.81147	0.253339	0.617
	fricative	332.5757	1	332.5757	0.428526	0.516
	envelope	52.23765	1	52.23765	0.171145	0.681
noise * channo	total	28.26694	1	28.26694	0.092844	0.762
	voicing	305.0844	1	305.0844	0.621486	0.435
	nasality	78.3225	1	78.3225	0.069022	0.794
	place	6.25	1	6.25	0.020073	0.888
	manner	2.4025	1	2.4025	0.006353	0.937
Error	total	14005	46	304.4564		
	voicing	22581.18	46	490.8951		
	nasality	52198.73	46	1134.755		
	place	14322.49	46	311.3586		

	manner	17396.98	46	378.1952		
	fricative	35700.26	46	776.0926		
	envelope	14040.31	46	305.2242		

Table B5. Summary of MANOVA from Experiment 3, as described 5.3. Only data from CI users with baseline consonant scores of 50% or more were included in the analyses (N=). Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were three factors: "noise condition", "stimulation rate" and " channel number", here summarised as "noise", "stimrate" and "channo", respectively. Significant effects at the *a priori* significance level ($p \le 0.05$) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
noise	total	20.1601	1	20.1601	0.116559	0.736
	voicing	162.9452	1	162.9452	0.498829	0.487
	nasality	2640.418	1	2640.418	3.779919	0.065
	place	10.98922	1	10.98922	0.037662	0.848
	manner	28.6225	1	28.6225	0.236235	0.632
	fricative	674.7006	1	674.7006	3.13045	0.091
	envelope	0.8836	1	0.8836	0.006506	0.936
stimrate	total	151.0618	1	151.0618	0.873391	0.360
	voicing	474.9507	1	474.9507	1.453982	0.241
	nasality	552.7922	1	552.7922	0.791356	0.383
	place	125.9067	1	125.9067	0.431505	0.518
	manner	60.025	1	60.025	0.495414	0.489
	fricative	53.43803	1	53.43803	0.24794	0.623
	envelope	13.689	1	13.689	0.100796	0.754
channo	total	50.2445	1	50.2445	0.290498	0.595
	voicing	12.9605	1	12.9605	0.039676	0.844
	nasality	10.082	1	10.082	0.014433	0.905
	place	24.642	1	24.642	0.084453	0.774
	manner	0.002	1	0.002	1.65E-05	0.997
	fricative	24.8645	1	24.8645	0.115365	0.737
	envelope	0.018	1	0.018	0.000133	0.991
noise * stimrate	total	0.784	1	0.784	0.004533	0.947
	voicing	787.0647	1	787.0647	2.409467	0.135
	nasality	265.3967	1	265.3967	0.379931	0.544
	place	10.37003	1	10.37003	0.03554	0.852
	manner	49.43211	1	49.43211	0.407986	0.530
	fricative	129.0007	1	129.0007	0.598532	0.447
	envelope	6.453444	1	6.453444	0.047518	0.829
noise * channo	total	11.4005	1	11.4005	0.065914	0.800
	voicing	7.8125	1	7.8125	0.023917	0.879
	nasality	8.712	1	8.712	0.012472	0.912
	place	11.552	1	11.552	0.039591	0.844
	manner	0.338	1	0.338	0.00279	0.958
Error	total	3805.12	22	172.96		
	voicing	7186.414	22	326.6552		
	nasality	15367.84	22	698.5383		
	place	6419.274	22	291.7852		
	manner	2665.546	22	121.1612		
	fricative	4741.624	22	215.5283		
	envelope	2987.802	22	135.8092		

Table B6. Summary of MANOVA from Experiment 3, as described in 5.3. Only data from CI users with baseline consonant scores of less than 50% were included in the analyses (N=). Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were three factors: "noise condition", "stimulation rate" and " channel number", here summarised as "noise", "stimrate" and "channo", respectively. Significant effects at the *a priori* significance level (*p*≤0.05) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
noise	total	394.805	1	394.805	5.623892	**0.029**
	voicing	1525.361	1	1525.361	5.989678	**0.025**
	nasality	3083.742	1	3083.742	2.976544	0.102
	place	196.02	1	196.02	3.713262	0.070
	manner	1164.031	1	1164.031	10.77848	0.004
	fricative	1094.34	1	1094.34	3.517951	0.077
	envelope	1125.751	1	1125.751	6.501929	**0.020**
stimrate	total	223.5025	1	223.5025	3.183733	0.091
	voicing	315.0625	1	315.0625	1.237165	0.281
	nasality	685.1306	1	685.1306	0.661314	0.427
	place	160.0225	1	160.0225	3.031351	0.099
	manner	82.81	1	82.81	0.766789	0.393
	fricative	608.8556	1	608.8556	1.957274	0.179
	envelope	291.5556	1	291.5556	1.683919	0.211
channo	total	155.0025	1	155.0025	2.207969	0.155
	voicing	405.0156	1	405.0156	1.590387	0.223
	nasality	2.640625	1	2.640625	0.002549	0.960
	place	64.8025	1	64.8025	1.227572	0.282
	manner	242.5806	1	242.5806	2.246204	0.151
	fricative	154.3806	1	154.3806	0.496284	0.490
	envelope	434.7225	1	434.7225	2.510799	0.130
noise * stimrate	total	0.7225	1	0.7225	0.010292	0.920
	voicing	76.5625	1	76.5625	0.30064	0.590
	nasality	716.9006	1	716.9006	0.69198	0.416
	place	51.84	1	51.84	0.98202	0.335
	manner	28.09	1	28.09	0.260103	0.616
	fricative	126.0006	1	126.0006	0.405051	0.533
	envelope	37.51563	1	37.51563	0.216677	0.647
noise * channo	total	17.64	1	17.64	0.251277	0.622
	voicing	532.4556	1	532.4556	2.090809	0.165
	nasality	274.7306	1	274.7306	0.26518	0.613
	place	0.0025	1	0.0025	0.000047	0.995
	manner	2.805625	1	2.805625	0.025979	0.874
	fricative			38.13062	0.122578	
	envelope			3.4225	0.019767	
Error	total	1263.625	18	70.20139		
	voicing	4583.968	18	254.6649		
	nasality	18648.26	18	1036.014		
	place	950.205	18	52.78917		
	manner	1943.925	18	107.9958		
	fricative	5599.318	18	311.0732		
	envelope	3116.54	18	173.1411		

Error	total	1263.625	18	70.20139		
	voicing	4583.968	18	254.6649		

Table B7. Summary of MANOVA from Experiment 4, as described in 5.4. Data across all acoustic model conditions were included. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: "channel interaction condition","noise condition", "stimulation rate" and " channel number", here summarised as "chanint","noise", "stimrate" and "channo", respectively. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
chanint	total	2407.776	2	1203.888	8.597039	**0.001**
	voicing	3461.736	2	1730.868	3.817841	0.058
	nasality	4861.775	2	2430.888	2.537777	0.118
	place	1045.277	2	522.6384	5.294059	**0.000**
	manner	1077.53	2	538.7649	6.878833	**0.002**
	fricative	1750.553	2	875.2766	4.669267	**0.003**
	envelope	1508.495	2	754.2477	5.662181	0.334
noise	total	310.99	1	310.99	2.220799	**0.000**
	voicing	6351.713	1	6351.713	14.01021	**0.000**
	nasality	8746.91	1	8746.91	9.131523	**0.000**
	place	2.521364	1	2.521364	0.02554	**0.000**
	manner	593.0004	1	593.0004	7.571301	**0.000**
	fricative	193.698	1	193.698	1.033305	**0.000**
	envelope	273.4158	1	273.4158	2.052548	**0.000**
stimrate	total	2.163712	1	2.163712	0.015451	**0.000**
	voicing	246.5467	1	246.5467	0.543817	**0.024**
	nasality	1978.964	1	1978.964	2.065981	0.082
	place	1.054848	1	1.054848	0.010685	**0.006**
	manner	6.6825	1	6.6825	0.085321	**0.001**
	fricative	136.2334	1	136.2334	0.726753	**0.011**
	envelope	58.4003	1	58.4003	0.438414	**0.004**
channo	total	1232.815	1	1232.815	8.803608	0.138
	voicing	450.2912	1	450.2912	0.993224	**0.000**
	nasality	2.677576	1	2.677576	0.002795	**0.003**
	place	2859.753	1	2859.753	28.96783	0.873
	manner	695.5227	1	695.5227	8.880283	**0.007**
	fricative	1737.464	1	1737.464	9.268707	0.311
	envelope	62.45939	1	62.45939	0.468886	0.154
chanint * noise	total	92.76771	2	46.38386	0.33123	0.901
	voicing	485.5128	2	242.7564	0.535457	0.462
	nasality	1561.375	2	780.6877	0.815016	0.152
	place	403.213	2	201.6065	2.042171	0.918
	manner	73.34892	2	36.67446	0.468252	0.771
	fricative	640.0766	2	320.0383	1.707282	0.395
	envelope	32.76458	2	16.38229	0.122983	0.509
chanint * stimrate	total	83.46424	2	41.73212	0.298012	**0.003**
	voicing	1652.502	2	826.251	1.822493	0.320
	nasality	261.8756	2	130.9378	0.136695	0.958
	place	14.37288	2	7.186439	0.072795	**0.000**
	manner	78.18682	2	39.09341	0.499136	**0.003**
	fricative	335.4668	2	167.7334	0.894794	**0.003**
	envelope	134.8783	2	67.43917	0.50627	0.494
noise * stimrate	total	86.5728	1	86.5728	0.618222	0.718
	voicing	308.5094	1	308.5094	0.680491	0.586
	nasality	254.537	1	254.537	0.265729	0.444
	place	3.030303	1	3.030303	0.030695	0.133

	manner	91.83341	1	91.83341	1.172509	0.627
	fricative	148.697	1	148.697	0.793242	0.184
	envelope	3.030303	1	3.030303	0.022749	0.884
chanint * noise * stimrate	total	9.640606	2	4.820303	0.034422	0.743
	voicing	13.59106	2	6.79553	0.014989	0.165
	nasality	2223.091	2	1111.546	1.160422	0.872
	place	110.5656	2	55.2828	0.559986	0.930
	manner	49.14591	2	24.57295	0.313742	0.608
	fricative	627.5814	2	313.7907	1.673954	0.411
	envelope	29.77652	2	14.88826	0.111767	0.604
chanint * channo	total	621.4911	2	310.7455	2.219053	0.433
	voicing	734.3838	2	367.1919	0.809929	0.411
	nasality	13.28015	2	6.640076	0.006932	0.607
	place	692.6309	2	346.3155	3.507998	0.861
	manner	47.91455	2	23.95727	0.305881	0.280
	fricative	1262.264	2	631.1321	3.366849	0.374
	envelope	15.69288	2	7.846439	0.058904	0.880
noise * channo	total	253.7045	1	253.7045	1.81172	0.966
	voicing	52.69364	1	52.69364	0.116228	0.985
	nasality	571.2512	1	571.2512	0.59637	0.316
	place	441.4694	1	441.4694	4.471859	0.572
	manner	13.1103	1	13.1103	0.16739	0.731
	fricative	36.5928	1	36.5928	0.195209	0.190
	envelope	186.7348	1	186.7348	1.401829	0.894
chanint * noise * channo	total	100.8005	2	50.40023	0.359911	0.112
	voicing	327.5568	2	163.7784	0.361252	0.447
	nasality	1822.511	2	911.2555	0.951325	0.993
	place	85.80061	2	42.9003	0.434558	**0.032**
	manner	17.06424	2	8.532121	0.108936	0.737
	fricative	276.5261	2	138.263	0.737581	**0.037**
	envelope	26.72379	2	13.36189	0.100308	0.943
Error	total	25206.33	180	140.0352		
	voicing	81605.35	180	453.3631		
	nasality	172418.5	180	957.8807		
	place	17769.9	180	98.72168		
	manner	14097.98	180	78.32213		
	fricative	33741.87	180	187.4548		
	envelope	23977.44	180	133.208		

Table B7. Summary of MANOVA from Experiment 4, as described in 5.4. Only data from the “no channel interaction” acoustic model are included. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were three factors: “noise condition”, “stimulation rate” and “ channel number”, here summarised as “noise”, “stimrate” and “channo”, respectively. Significant effects at the *a priori* significance level (*p*≤0.05) are highlighted. Because the factors “stimulation rate” and “channel number” overlapped, interactions involving both these factors could not be computed.

Source	**Dependent Variable**	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	***F***	***p***
noise	total	325.5819	1	325.5819	2.914374	0.093
	voicing	3979.855	1	3979.855	8.640582	**0.005**

	nasality	476.3152	1	476.3152	0.420152	0.519
	place	229.9978	1	229.9978	2.831881	0.098
	manner	358.9675	1	358.9675	6.875167	**0.011**
	fricative	341.1516	1	341.1516	1.998682	0.163
	envelope	202.0202	1	202.0202	1.419028	0.238
stimrate	total	7.445682	1	7.445682	0.066648	0.797
	voicing	576.7384	1	576.7384	1.252145	0.268
	nasality	1508.131	1	1508.131	1.330305	0.253
	place	3.494545	1	3.494545	0.043027	0.836
	manner	8.553636	1	8.553636	0.163825	0.687
	fricative	64.80818	1	64.80818	0.379687	0.540
	envelope	15.48205	1	15.48205	0.108749	0.743
channo	total	737.182	1	737.182	6.598722	**0.013**
	voicing	97.50568	1	97.50568	0.211693	0.647
	nasality	0.638409	1	0.638409	0.000563	0.981
	place	1838.258	1	1838.258	22.63382	**0.000**
	manner	415.4327	1	415.4327	7.956625	**0.006**
	fricative	357.9602	1	357.9602	2.097158	0.153
	envelope	2.800227	1	2.800227	0.019669	0.889
noise * stimrate	total	60.7475	1	60.7475	0.543768	0.464
	voicing	135.802	1	135.802	0.294837	0.589
	nasality	72.29455	1	72.29455	0.06377	0.801
	place	34.56818	1	34.56818	0.425626	0.517
	manner	0.073636	1	0.073636	0.00141	0.970
	fricative	754.4736	1	754.4736	4.420184	**0.040**
	envelope	10.50568	1	10.50568	0.073794	0.787
noise * channo	total	100.5057	1	100.5057	0.899654	0.347
	voicing	165.3657	1	165.3657	0.359022	0.551
	nasality	823.0475	1	823.0475	0.726001	0.398
	place	46.43273	1	46.43273	0.57171	0.453
	manner	1.312727	1	1.312727	0.025142	0.875
	fricative	186.142	1	186.142	1.090538	0.301
	envelope	35.46023	1	35.46023	0.249079	0.620
noise	total	325.5819	1	325.5819	2.914374	0.093
	voicing	3979.855	1	3979.855	8.640582	**0.005**
	nasality	476.3152	1	476.3152	0.420152	0.519
	place	229.9978	1	229.9978	2.831881	0.098
	manner	358.9675	1	358.9675	6.875167	0.011
	fricative	341.1516	1	341.1516	1.998682	0.163
	envelope	202.0202	1	202.0202	1.419028	0.238
stimrate	total	7.445682	1	7.445682	0.066648	0.797
	voicing	576.7384	1	576.7384	1.252145	0.268
	nasality	1508.131	1	1508.131	1.330305	0.253
	place	3.494545	1	3.494545	0.043027	0.836
	manner	8.553636	1	8.553636	0.163825	0.687
	fricative	64.80818	1	64.80818	0.379687	0.540
	envelope	15.48205	1	15.48205	0.108749	0.743
Error	total	6702.953	60	111.7159		
	voicing	27636.02	60	460.6004		
	nasality	68020.39	60	1133.673		
	place	4873.038	60	81.2173		
	manner	3132.731	60	52.21218		
	fricative	10241.3	60	170.6883		
	envelope	8541.909	60	142.3652		

Table B8. Summary of MANOVA from Experiment 4, as described in 5.4. Only data from the "1mm channel interaction" acoustic model are included. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were three factors: "noise condition", "stimulation rate" and " channel number", here summarised as "noise", "stimrate" and "channo", respectively. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
noise	total	40.90909	1	40.90909	0.350742	0.556
	voicing	1060.34	1	1060.34	2.61677	0.111
	nasality	5210.336	1	5210.336	6.191506	0.016
	place	38.40323	1	38.40323	0.408663	0.525
	manner	53.06187	1	53.06187	0.908331	0.344
	fricative	159.3018	1	159.3018	0.857645	0.358
	envelope	55.25838	1	55.25838	0.527088	0.471
stimrate	total	69.00023	1	69.00023	0.591587	0.445
	voicing	541.102	1	541.102	1.335364	0.252
	nasality	324.0082	1	324.0082	0.385023	0.537
	place	5.745682	1	5.745682	0.061142	0.806
	manner	28.64205	1	28.64205	0.490304	0.486
	fricative	257.7784	1	257.7784	1.387821	0.243
	envelope	21.98205	1	21.98205	0.209678	0.649
channo	total	1117.066	1	1117.066	9.577376	**0.003**
	voicing	536.2036	1	536.2036	1.323276	0.255
	nasality	15.24568	1	15.24568	0.018117	0.893
	place	1625.063	1	1625.063	17.29288	**0.000**
	manner	214.7236	1	214.7236	3.675712	0.060
	fricative	2638.102	1	2638.102	14.20295	**0.000**
	envelope	52.80091	1	52.80091	0.503647	0.481
noise * stimrate	total	23.1275	1	23.1275	0.198288	0.658
	voicing	50.8475	1	50.8475	0.125485	0.724
	nasality	2272.328	1	2272.328	2.700236	0.106
	place	57.04568	1	57.04568	0.607044	0.439
	manner	102.3275	1	102.3275	1.751677	0.191
	fricative	16.69114	1	16.69114	0.089861	0.765
	envelope	18.46023	1	18.46023	0.176085	0.676
noise * channo	total	251.0457	1	251.0457	2.152388	0.148
	voicing	104.4736	1	104.4736	0.257826	0.613
	nasality	438.482	1	438.482	0.521054	0.473
	place	378.2045	1	378.2045	4.024611	0.049
	manner	28.80364	1	28.80364	0.49307	0.485
	fricative	38.76568	1	38.76568	0.208706	0.649
	envelope	31.45091	1	31.45091	0.299998	0.586
noise	total	40.90909	1	40.90909	0.350742	0.556
	voicing	1060.34	1	1060.34	2.61677	0.111
	nasality	5210.336	1	5210.336	6.191506	0.016
	place	38.40323	1	38.40323	0.408663	0.525
	manner	53.06187	1	53.06187	0.908331	0.344
	fricative	159.3018	1	159.3018	0.857645	0.358
	envelope	55.25838	1	55.25838	0.527088	0.471
stimrate	total	69.00023	1	69.00023	0.591587	0.445
	voicing	541.102	1	541.102	1.335364	0.252
	nasality	324.0082	1	324.0082	0.385023	0.537

	place	5.745682	1	5.745682	0.061142	0.806
	manner	28.64205	1	28.64205	0.490304	0.486
	fricative	257.7784	1	257.7784	1.387821	0.243
	envelope	21.98205	1	21.98205	0.209678	0.649
Error	total	6998.153	60	116.6359		
	voicing	24312.56	60	405.2093		
	nasality	50491.78	60	841.5296		
	place	5638.376	60	93.97294		
	manner	3505.013	60	58.41688		
	fricative	11144.6	60	185.7433		
	envelope	6290.229	60	104.8372		

Table B9. Summary of MANOVA from Experiment 4, as described in 5.4. Only data from the "3.3mm channel interaction" acoustic model are included. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were three factors: "noise condition", "stimulation rate" and " channel number", here summarised as "noise", "stimrate" and "channo", respectively. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	**Dependent Variable**	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	***F***	***p***
noise	total	37.26672	1	37.26672	0.194347	0.661
	voicing	1797.032	1	1797.032	3.635659	0.061
	nasality	4621.634	1	4621.634	5.14407	**0.027**
	place	137.3334	1	137.3334	1.135223	0.291
	manner	254.32	1	254.32	2.045403	0.158
	fricative	333.3213	1	333.3213	1.618592	0.208
	envelope	48.90182	1	48.90182	0.320832	0.573
stimrate	total	9.182045	1	9.182045	0.047885	0.828
	voicing	781.2082	1	781.2082	1.580499	0.214
	nasality	408.7002	1	408.7002	0.4549	0.503
	place	6.1875	1	6.1875	0.051147	0.822
	manner	47.67364	1	47.67364	0.383422	0.538
	fricative	149.1136	1	149.1136	0.724089	0.398
	envelope	155.8145	1	155.8145	1.02226	0.316
channo	total	0.058182	1	0.058182	0.000303	0.986
	voicing	550.9657	1	550.9657	1.114684	0.295
	nasality	0.073636	1	0.073636	8.2E-05	0.993
	place	89.06273	1	89.06273	0.736209	0.394
	manner	113.2809	1	113.2809	0.911077	0.344
	fricative	3.665682	1	3.665682	0.0178	0.894
	envelope	22.55114	1	22.55114	0.147952	0.702
noise * stimrate	total	12.33841	1	12.33841	0.064345	0.801
	voicing	135.4509	1	135.4509	0.274037	0.603
	nasality	133.0057	1	133.0057	0.148041	0.702
	place	21.98205	1	21.98205	0.181708	0.671
	manner	38.57818	1	38.57818	0.31027	0.580
	fricative	5.113636	1	5.113636	0.024832	0.875
	envelope	3.840909	1	3.840909	0.025199	0.874
noise * channo	total	2.953636	1	2.953636	0.015403	0.902
	voicing	110.4111	1	110.4111	0.223378	0.638
	nasality	1132.233	1	1132.233	1.260222	0.266
	place	102.6327	1	102.6327	0.848381	0.361

	manner	0.058182	1	0.058182	0.000468	0.983
	fricative	88.21114	1	88.21114	0.428349	0.515
	envelope	146.5475	1	146.5475	0.961461	0.331
noise	total	37.26672	1	37.26672	0.194347	0.661
	voicing	1797.032	1	1797.032	3.635659	0.061
	nasality	4621.634	1	4621.634	5.14407	**0.027**
	place	137.3334	1	137.3334	1.135223	0.291
	manner	254.32	1	254.32	2.045403	0.158
	fricative	333.3213	1	333.3213	1.618592	0.208
	envelope	48.90182	1	48.90182	0.320832	0.573
stimrate	total	9.182045	1	9.182045	0.047885	0.828
	voicing	781.2082	1	781.2082	1.580499	0.214
	nasality	408.7002	1	408.7002	0.4549	0.503
	place	6.1875	1	6.1875	0.051147	0.822
	manner	47.67364	1	47.67364	0.383422	0.538
	fricative	149.1136	1	149.1136	0.724089	0.398
	envelope	155.8145	1	155.8145	1.02226	0.316
Error	total	11505.23	60	191.7538		
	voicing	29656.77	60	494.2795		
	nasality	53906.35	60	898.4392		
	place	7258.487	60	120.9748		
	manner	7460.24	60	124.3373		
	fricative	12355.97	60	205.9329		
	envelope	9145.302	60	152.4217		

Table B10. Summary of MANOVA combining all data from all CI users in experiment 3 and "no channel interaction" acoustic model conditions from experiment 4, as described in 5.5. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: "noise condition", "stimulation rate", " channel number" and "group", here summarised as "noise", "stimrate", "channo" and "group", respectively. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
noise	total	604.4302	1	604.4302	3.093962	0.081
	voicing	3104.016	1	3104.016	6.552052	**0.012**
	nasality	5067.214	1	5067.214	4.467881	0.037
	place	345.3527	1	345.3527	1.907079	0.170
	manner	1082.033	1	1082.033	5.586805	**0.020**
	fricative	225.486	1	225.486	0.520259	0.472
	envelope	682.7318	1	682.7318	3.204715	0.076
stimrate	total	4.860716	1	4.860716	0.024881	0.875
	voicing	241.1383	1	241.1383	0.509002	0.477
	nasality	13.88767	1	13.88767	0.012245	0.912
	place	18.93831	1	18.93831	0.10458	0.747
	manner	35.65891	1	35.65891	0.184116	0.669
	fricative	73.1437	1	73.1437	0.168763	0.682
	envelope	46.6013	1	46.6013	0.218745	0.641
channo	total	418.2323	1	418.2323	2.140851	0.146
	voicing	28.26056	1	28.26056	0.059653	0.808
	nasality	4.090909	1	4.090909	0.003607	0.952
	place	899.844	1	899.844	4.969045	0.028
	manner	455.7601	1	455.7601	2.353203	0.128
	fricative	258.5195	1	258.5195	0.596477	0.442

	envelope	132.3701	1	132.3701	0.62134	0.432
group	total	6288.825	1	6288.825	32.19128	**0.000**
	voicing	2681.944	1	2681.944	5.66113	**0.019**
	nasality	9407.414	1	9407.414	8.294736	**0.005**
	place	5173.869	1	5173.869	28.57072	**0.000**
	manner	6456.751	1	6456.751	33.33781	**0.000**
	fricative	9883.527	1	9883.527	22.80406	**0.000**
	envelope	4084.701	1	4084.701	19.17341	**0.000**
noise * stimrate	total	21.60952	1	21.60952	0.110615	0.740
	voicing	105.5045	1	105.5045	0.222702	0.638
	nasality	771.6535	1	771.6535	0.680385	0.411
	place	4.266835	1	4.266835	0.023562	0.878
	manner	51.52752	1	51.52752	0.266049	0.607
	fricative	1012.794	1	1012.794	2.336799	0.129
	envelope	10.85456	1	10.85456	0.050951	0.822
noise * channo	total	7.740626	1	7.740626	0.039623	0.843
	voicing	18.725	1	18.725	0.039525	0.843
	nasality	666.072	1	666.072	0.587291	0.445
	place	7.382227	1	7.382227	0.040766	0.840
	manner	0.145102	1	0.145102	0.000749	0.978
	fricative	448.0209	1	448.0209	1.033709	0.312
	envelope	47.4301	1	47.4301	0.222635	0.638
noise * group	total	0.545135	1	0.545135	0.00279	0.958
	voicing	765.1299	1	765.1299	1.61506	0.207
	nasality	1792.75	1	1792.75	1.58071	0.211
	place	2.129601	1	2.129601	0.01176	0.914
	manner	61.69521	1	61.69521	0.318548	0.574
	fricative	88.25043	1	88.25043	0.203618	0.653
	envelope	53.94478	1	53.94478	0.253215	0.616
stimrate * group	total	33.68948	1	33.68948	0.17245	0.679
	voicing	260.8703	1	260.8703	0.550653	0.460
	nasality	2256.486	1	2256.486	1.989597	0.161
	place	3.556315	1	3.556315	0.019638	0.889
	manner	4.466527	1	4.466527	0.023062	0.880
	fricative	367.5697	1	367.5697	0.848086	0.359
	envelope	144.4093	1	144.4093	0.677851	0.412
noise * stimrate * group	total	31.73085	1	31.73085	0.162424	0.688
	voicing	657.6273	1	657.6273	1.38814	0.241
	nasality	274.3112	1	274.3112	0.241867	0.624
	place	32.37983	1	32.37983	0.178805	0.673
	manner	56.79483	1	56.79483	0.293246	0.589
	fricative	19.44652	1	19.44652	0.044869	0.833
	envelope	57.31015	1	57.31015	0.269011	0.605
channo * group	total	255.2418	1	255.2418	1.306534	0.256
	voicing	344.6256	1	344.6256	0.727446	0.396
	nasality	9.576409	1	9.576409	0.008444	0.927
	place	757.644	1	757.644	4.183799	**0.043**
	manner	35.9641	1	35.9641	0.185692	0.667
	fricative	86.58455	1	86.58455	0.199775	0.656
	envelope	85.7501	1	85.7501	0.402507	0.527
noise *	total	113.8081	1	113.8081	0.582562	0.447

channo * group						
	voicing	465.697	1	465.697	0.983007	0.324
	nasality	160.8255	1	160.8255	0.141804	0.707
	place	41.28223	1	41.28223	0.227965	0.634
	manner	3.679102	1	3.679102	0.018996	0.891
	fricative	8.19092	1	8.19092	0.018899	0.891
	envelope	1.215102	1	1.215102	0.005704	0.940
Error	total	20707.95	106	195.358		
	voicing	50217.2	106	473.7472		
	nasality	120219.1	106	1134.143		
	place	19195.53	106	181.0899		
	manner	20529.71	106	193.6765		
	fricative	45941.56	106	433.4109		
	envelope	22582.22	106	213.0398		

Table B11. Summary of MANOVA combining all data from all CI users in experiment 3 and "1mm channel interaction" acoustic model conditions from experiment 4, as described in 5.5. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: "noise condition", "stimulation rate", " channel number" and "group", here summarised as "noise", "stimrate", "channo" and "group", respectively. Significant effects at the *a priori* significance level (*p*≤0.05) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	**Dependent Variable**	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	***F***	***p***
noise	total	8508.466	11	773.4969	3.903732	**0.000**
	voicing	5940.043	11	540.0039	1.220641	0.282
	nasality	33507.07	11	3046.097	3.144266	**0.001**
	place	8994.512	11	817.6829	4.342215	**0.000**
	manner	9303.186	11	845.7442	4.289011	**0.000**
	fricative	12511.27	11	1137.388	2.573669	**0.006**
	envelope	6122.306	11	556.5733	2.901879	**0.002**
stimrate	total	319436.2	1	319436.2	1612.151	**0.000**
	voicing	295755.2	1	295755.2	668.5339	**0.000**
	nasality	433633.4	1	433633.4	447.6084	**0.000**
	place	250375.7	1	250375.7	1329.592	**0.000**
	manner	615992.8	1	615992.8	3123.876	**0.000**
	fricative	419736.6	1	419736.6	949.7751	**0.000**
	envelope	558686.7	1	558686.7	2912.898	**0.000**
channo	total	285.2128	1	285.2128	1.43943	0.233
	voicing	1263.37	1	1263.37	2.85576	0.094
	nasality	10912.1	1	10912.1	11.26377	0.001
	place	19.95562	1	19.95562	0.105972	0.745
	manner	634.4333	1	634.4333	3.217393	0.076
	fricative	30.57287	1	30.57287	0.06918	0.793
	envelope	468.6815	1	468.6815	2.443626	0.121
group	total	89.93436	1	89.93436	0.453886	0.502
	voicing	245.0572	1	245.0572	0.553935	0.458
	nasality	100.3073	1	100.3073	0.10354	0.748
	place	22.0885	1	22.0885	0.117299	0.733
	manner	0.262663	1	0.262663	0.001332	0.971
	fricative	597.9064	1	597.9064	1.352936	0.247
	envelope	156.6057	1	156.6057	0.816515	0.368
noise * stimrate	total	608.004	1	608.004	3.068512	0.083
	voicing	754.8011	1	754.8011	1.706175	0.194

	nasality	26.81018	1	26.81018	0.027674	0.868
	place	799.656	1	799.656	4.246485	**0.042**
	manner	306.446	1	306.446	1.554076	0.215
	fricative	1431.995	1	1431.995	3.240302	0.075
	envelope	232.7804	1	232.7804	1.213678	0.273
noise * channo	total	5586.395	1	5586.395	28.19377	**0.000**
	voicing	1956.422	1	1956.422	4.422354	0.038
	nasality	13997.76	1	13997.76	14.44888	**0.000**
	place	4907.167	1	4907.167	26.05897	**0.000**
	manner	7124.482	1	7124.482	36.13029	**0.000**
	fricative	7347.47	1	7347.47	16.62577	**0.000**
	envelope	4924.709	1	4924.709	25.6766	**0.000**
noise * group	total	7.181239	1	7.181239	0.036243	0.849
	voicing	175.6843	1	175.6843	0.397122	0.530
	nasality	85.95084	1	85.95084	0.088721	0.766
	place	46.15925	1	46.15925	0.245124	0.622
	manner	0.469413	1	0.469413	0.002381	0.961
	fricative	121.2775	1	121.2775	0.274425	0.601
	envelope	68.33327	1	68.33327	0.356278	0.552
stimrate * group	total	44.70013	1	44.70013	0.225595	0.636
	voicing	37.17358	1	37.17358	0.084028	0.772
	nasality	56.00455	1	56.00455	0.057809	0.810
	place	125.2545	1	125.2545	0.66515	0.417
	manner	6.006011	1	6.006011	0.030458	0.862
	fricative	262.1456	1	262.1456	0.59318	0.443
	envelope	44.28041	1	44.28041	0.230871	0.632
noise * stimrate * group	total	0	0	.	.	.
	voicing	56.11813	1	56.11813	0.126851	0.722
	nasality	82.16097	1	82.16097	0.084809	0.771
	place	160.2042	1	160.2042	0.850747	0.358
	manner	242.1429	1	242.1429	1.227976	0.270
	fricative	124.3514	1	124.3514	0.281381	0.597
	envelope	139.8258	1	139.8258	0.729028	0.395
channo * group	total	2.173872	1	2.173872	0.010971	0.917
	voicing	225.9446	1	225.9446	0.510732	0.476
	nasality	1139.773	1	1139.773	1.176505	0.281
	place	2.364816	1	2.364816	0.012558	0.911
	manner	57.34151	1	57.34151	0.290795	0.591
	fricative	10.70854	1	10.70854	0.024231	0.877
	envelope	40.06052	1	40.06052	0.208869	0.649
noise * channo * group	total	13.42631	1	13.42631	0.067761	0.795
	voicing	513.5292	1	513.5292	1.160797	0.284
	nasality	2874.243	1	2874.243	2.966873	0.088
	place	10.04487	1	10.04487	0.053342	0.818
	manner	196.8231	1	196.8231	0.998146	0.320
	fricative	269.0257	1	269.0257	0.608748	0.437
	envelope	6.752606	1	6.752606	0.035207	0.852
Error	total	407.3655	1	407.3655	2.055917	0.155
	voicing	12.91314	1	12.91314	0.029189	0.865
	nasality	0.003682	1	0.003682	3.8E-06	0.998

	place	665.956	1	665.956	3.536486	0.063
	manner	4.640011	1	4.640011	0.023531	0.878
	fricative	965.2364	1	965.2364	2.184126	0.142
	envelope	30.34041	1	30.34041	0.15819	0.692

Table B12. Summary of MANOVA combining all data from all CI users in experiment 3 and "3.3mm channel interaction" acoustic model conditions from experiment 4, as described in 5.5. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: "noise condition", "stimulation rate", " channel number" and "group", here summarised as "noise", "stimrate", "channo" and "group", respectively. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	**Dependent Variable**	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	***F***	***p***
noise	total	278.7465	1	278.7465	1.158247	0.284
	voicing	1767.247	1	1767.247	3.586055	0.061
	nasality	10339.96	1	10339.96	10.32972	**0.002**
	place	0.669609	1	0.669609	0.003289	0.954
	manner	955.5835	1	955.5835	4.074947	**0.046**
	fricative	221.2748	1	221.2748	0.488077	0.486
	envelope	456.1596	1	456.1596	2.085471	0.152
stimrate	total	4.023364	1	4.023364	0.016718	0.897
	voicing	351.4404	1	351.4404	0.713135	0.400
	nasality	73.16496	1	73.16496	0.073092	0.787
	place	2.184881	1	2.184881	0.010732	0.918
	manner	73.90157	1	73.90157	0.315142	0.576
	fricative	480.334	1	480.334	1.059496	0.306
	envelope	311.7	1	311.7	1.42503	0.235
channo	total	5.755335	1	5.755335	0.023915	0.877
	voicing	766.5156	1	766.5156	1.555395	0.215
	nasality	5.648011	1	5.648011	0.005642	0.940
	place	57.25601	1	57.25601	0.281226	0.597
	manner	219.5003	1	219.5003	0.936027	0.336
	fricative	21.819	1	21.819	0.048127	0.827
	envelope	184.098	1	184.098	0.841659	0.361
group	total	1406.757	1	1406.757	5.845353	**0.017**
	voicing	0.712547	1	0.712547	0.001446	0.970
	nasality	2951.498	1	2951.498	2.948575	0.089
	place	2013.796	1	2013.796	9.891228	**0.002**
	manner	3141.25	1	3141.25	13.39541	**0.000**
	fricative	3702.905	1	3702.905	8.16768	**0.005**
	envelope	1301.949	1	1301.949	5.952253	**0.016**
noise * stimrate	total	3.329308	1	3.329308	0.013834	0.907
	voicing	105.7088	1	105.7088	0.214502	0.644
	nasality	886.6878	1	886.6878	0.88581	0.349
	place	1.638636	1	1.638636	0.008049	0.929
	manner	10.63095	1	10.63095	0.045334	0.832
	fricative	149.2251	1	149.2251	0.329153	0.567
	envelope	45.22759	1	45.22759	0.206771	0.650
noise * channo	total	7.784456	1	7.784456	0.032346	0.858
	voicing	400.0955	1	400.0955	0.811864	0.370
	nasality	848.8801	1	848.8801	0.848039	0.359
	place	24.42223	1	24.42223	0.119955	0.730
	manner	1.719557	1	1.719557	0.007333	0.932

	fricative	32.64801	1	32.64801	0.072013	0.789
	envelope	121.2874	1	121.2874	0.554502	0.458
noise * group	total	74.43879	1	74.43879	0.309308	0.579
	voicing	195.6059	1	195.6059	0.396919	0.530
	nasality	140.1777	1	140.1777	0.140039	0.709
	place	265.8875	1	265.8875	1.305968	0.256
	manner	96.75451	1	96.75451	0.412596	0.522
	fricative	85.62332	1	85.62332	0.188864	0.665
	envelope	146.7964	1	146.7964	0.671124	0.414
stimrate * group	total	36.03762	1	36.03762	0.149743	0.700
	voicing	328.4755	1	328.4755	0.666535	0.416
	nasality	1240.605	1	1240.605	1.239376	0.268
	place	22.65286	1	22.65286	0.111265	0.739
	manner	0.261814	1	0.261814	0.001116	0.973
	fricative	33.73278	1	33.73278	0.074406	0.786
	envelope	1.412601	1	1.412601	0.006458	0.936
noise * stimrate * group	total	7.890758	1	7.890758	0.032788	0.857
	voicing	657.1174	1	657.1174	1.333407	0.251
	nasality	212.1011	1	212.1011	0.211891	0.646
	place	24.05694	1	24.05694	0.118161	0.732
	manner	131.1938	1	131.1938	0.559457	0.456
	fricative	231.0046	1	231.0046	0.509538	0.477
	envelope	17.13816	1	17.13816	0.078352	0.780
channo * group	total	4.307335	1	4.307335	0.017898	0.894
	voicing	14.48456	1	14.48456	0.029392	0.864
	nasality	7.511011	1	7.511011	0.007504	0.931
	place	25.95601	1	25.95601	0.127489	0.722
	manner	0.287284	1	0.287284	0.001225	0.972
	fricative	4.420001	1	4.420001	0.009749	0.922
	envelope	51.79801	1	51.79801	0.23681	0.628
noise * channo * group	total	25.96746	1	25.96746	0.1079	0.743
	voicing	34.86746	1	34.86746	0.070752	0.791
	nasality	256.2841	1	256.2841	0.25603	0.614
	place	74.82223	1	74.82223	0.367507	0.546
	manner	0.975557	1	0.975557	0.00416	0.949
	fricative	335.426	1	335.426	0.739866	0.392
	envelope	27.33638	1	27.33638	0.124976	0.724
Error	total	25510.22	106	240.6625		
	voicing	52237.95	106	492.8108		
	nasality	106105.1	106	1000.991		
	place	21580.98	106	203.5942		
	manner	24857.22	106	234.5021		
	fricative	48056.23	106	453.3607		
	envelope	23185.61	106	218.7322		

Table B13. Summary of MANOVA combining all data from CI users with baseline consonant recognition scores of 50% or more from experiment 3 and “no channel interaction” acoustic model conditions from experiment 4, as described in 5.5. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: “noise condition”, “stimulation rate”, “ channel number” and “group”, here summarised as “noise”,

“stimrate”, “channo” and “group”, respectively. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted. Because the factors “stimulation rate” and “channel number” overlapped, interactions involving both these factors could not be computed.

Source	**Dependent Variable**	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	***F***	***p***
noise	total	182.38	1	182.38	1.423207	0.236
	voicing	540.0164	1	540.0164	1.271633	0.263
	nasality	3030.216	1	3030.216	2.97977	0.088
	place	120.2124	1	120.2124	0.872932	0.353
	manner	216.5491	1	216.5491	3.062466	0.084
	fricative	142.3217	1	142.3217	0.778912	0.380
	envelope	47.14907	1	47.14907	0.335327	0.564
stimrate	total	140.0997	1	140.0997	1.093272	0.299
	voicing	978.131	1	978.131	2.303306	0.133
	nasality	0.980641	1	0.980641	0.000964	0.975
	place	71.68766	1	71.68766	0.520566	0.473
	manner	24.69655	1	24.69655	0.349262	0.556
	fricative	109.9948	1	109.9948	0.60199	0.440
	envelope	27.38642	1	27.38642	0.194774	0.660
channo	total	86.50092	1	86.50092	0.675012	0.414
	voicing	6.426182	1	6.426182	0.015132	0.902
	nasality	4.779003	1	4.779003	0.004699	0.946
	place	394.0946	1	394.0946	2.861748	0.095
	manner	128.9791	1	128.9791	1.82404	0.181
	fricative	41.4991	1	41.4991	0.22712	0.635
	envelope	1.095571	1	1.095571	0.007792	0.930
group	total	115.6138	1	115.6138	0.902195	0.345
	voicing	125.3719	1	125.3719	0.295226	0.588
	nasality	40.43878	1	40.43878	0.039766	0.842
	place	131.2005	1	131.2005	0.952723	0.332
	manner	8.077138	1	8.077138	0.114228	0.736
	fricative	34.36444	1	34.36444	0.188073	0.666
	envelope	10.77407	1	10.77407	0.076626	0.783
noise * stimrate	total	11.79105	1	11.79105	0.092012	0.762
	voicing	303.6309	1	303.6309	0.714991	0.400
	nasality	335.2467	1	335.2467	0.329666	0.567
	place	34.47669	1	34.47669	0.250355	0.618
	manner	33.50075	1	33.50075	0.473772	0.493
	fricative	591.4684	1	591.4684	3.237047	0.076
	envelope	0.16416	1	0.16416	0.001168	0.973
noise * channo	total	7.866182	1	7.866182	0.061384	0.805
	voicing	23.72756	1	23.72756	0.055874	0.814
	nasality	184.6931	1	184.6931	0.181618	0.671
	place	0.982227	1	0.982227	0.007133	0.933
	manner	0.025102	1	0.025102	0.000355	0.985
	fricative	443.892	1	443.892	2.429376	0.123
	envelope	39.63132	1	39.63132	0.28186	0.597
noise * group	total	35.28184	1	35.28184	0.275323	0.601
	voicing	2002.144	1	2002.144	4.714657	**0.033**
	nasality	994.0442	1	994.0442	0.977495	0.326
	place	28.93229	1	28.93229	0.210094	0.648
	manner	32.50911	1	32.50911	0.459748	0.500
	fricative	1013.406	1	1013.406	5.546268	**0.021**
	envelope	71.40713	1	71.40713	0.507852	0.478
stimrate *	total	79.36707	1	79.36707	0.619343	0.434

group						
	voicing	30.35327	1	30.35327	0.071476	0.790
	nasality	1654.439	1	1654.439	1.626896	0.206
	place	109.6727	1	109.6727	0.796397	0.375
	manner	65.72964	1	65.72964	0.929557	0.338
	fricative	3.425266	1	3.425266	0.018746	0.891
	envelope	1.023545	1	1.023545	0.00728	0.932
noise * stimrate * group	total	24.28832	1	24.28832	0.189534	0.664
	voicing	895.6711	1	895.6711	2.10913	0.150
	nasality	84.40874	1	84.40874	0.083004	0.774
	place	0.19036	1	0.19036	0.001382	0.970
	manner	36.95572	1	36.95572	0.522633	0.472
	fricative	26.51737	1	26.51737	0.145127	0.704
	envelope	15.07495	1	15.07495	0.107214	0.744
channo * group	total	443.324	1	443.324	3.45949	0.066
	voicing	72.33556	1	72.33556	0.170336	0.681
	nasality	9.482753	1	9.482753	0.009325	0.923
	place	788.6996	1	788.6996	5.727203	**0.019**
	manner	130.6691	1	130.6691	1.84794	0.178
	fricative	216.4147	1	216.4147	1.184416	0.280
	envelope	0.679321	1	0.679321	0.004831	0.945
noise * channo * group	total	70.62556	1	70.62556	0.551128	0.460
	voicing	90.36818	1	90.36818	0.212799	0.646
	nasality	341.6906	1	341.6906	0.336002	0.564
	place	43.92223	1	43.92223	0.318945	0.574
	manner	1.260102	1	1.260102	0.017821	0.894
	fricative	33.81392	1	33.81392	0.18506	0.668
	envelope	0.131321	1	0.131321	0.000934	0.976
Error	total	10508.07	82	128.1472		
	voicing	34822.44	82	424.6638		
	nasality	83388.24	82	1016.93		
	place	11292.31	82	137.7111		
	manner	5798.277	82	70.71069		
	fricative	14982.92	82	182.7185		
	envelope	11529.71	82	140.6062		

Table B14. Summary of MANOVA combining all data from CI users with baseline consonant recognition scores of 50% or more from experiment 3 and "1mm channel interaction" acoustic model conditions from experiment 4, as described in 5.5. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: "noise condition", "stimulation rate", " channel number" and "group", here summarised as "noise", "stimrate", "channo" and "group", respectively. Significant effects at the *a priori* significance level (*p*≤0.05) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
noise	total	52.25497	1	52.25497	0.39663	0.531
	voicing	46.1292	1	46.1292	0.120086	0.730
	nasality	6753.733	1	6753.733	8.408887	**0.005**
	place	0.298582	1	0.298582	0.002031	0.964
	manner	71.09683	1	71.09683	0.944799	0.334
	fricative	822.6919	1	822.6919	4.246494	**0.043**

	envelope	10.32634	1	10.32634	0.091265	0.763
stimrate	total	35.00579	1	35.00579	0.265704	0.608
	voicing	34.97231	1	34.97231	0.091042	0.764
	nasality	103.7584	1	103.7584	0.129187	0.720
	place	66.97468	1	66.97468	0.455472	0.502
	manner	88.5371	1	88.5371	1.176562	0.281
	fricative	5.970871	1	5.970871	0.03082	0.861
	envelope	0.368869	1	0.368869	0.00326	0.955
channo	total	164.0046	1	164.0046	1.244843	0.268
	voicing	253.754	1	253.754	0.660587	0.419
	nasality	23.18878	1	23.18878	0.028872	0.865
	place	339.2647	1	339.2647	2.307225	0.133
	manner	66.49501	1	66.49501	0.883646	0.350
	fricative	604.0759	1	604.0759	3.118062	0.081
	envelope	17.41641	1	17.41641	0.153928	0.696
group	total	49.48048	1	49.48048	0.375571	0.542
	voicing	301.3049	1	301.3049	0.784375	0.378
	nasality	563.7636	1	563.7636	0.701927	0.405
	place	98.458	1	98.458	0.66958	0.416
	manner	37.78104	1	37.78104	0.502069	0.481
	fricative	290.0853	1	290.0853	1.497335	0.225
	envelope	70.40114	1	70.40114	0.622211	0.433
noise * stimrate	total	3.358245	1	3.358245	0.02549	0.874
	voicing	394.0685	1	394.0685	1.025862	0.314
	nasality	139.784	1	139.784	0.174041	0.678
	place	1.779497	1	1.779497	0.012102	0.913
	manner	0.257115	1	0.257115	0.003417	0.954
	fricative	54.66665	1	54.66665	0.282173	0.597
	envelope	19.79137	1	19.79137	0.174918	0.677
noise * channo	total	36.69556	1	36.69556	0.27853	0.599
	voicing	11.53473	1	11.53473	0.030028	0.863
	nasality	200.3114	1	200.3114	0.249402	0.619
	place	64.85592	1	64.85592	0.441063	0.508
	manner	6.341011	1	6.341011	0.084265	0.772
	fricative	286.3682	1	286.3682	1.478148	0.228
	envelope	37.22841	1	37.22841	0.329028	0.568
noise * group	total	0.113033	1	0.113033	0.000858	0.977
	voicing	800.8289	1	800.8289	2.084766	0.153
	nasality	19.31401	1	19.31401	0.024047	0.877
	place	37.59768	1	37.59768	0.255689	0.614
	manner	0.338768	1	0.338768	0.004502	0.947
	fricative	227.4455	1	227.4455	1.174007	0.282
	envelope	23.01331	1	23.01331	0.203394	0.653
stimrate * group	total	219.8881	1	219.8881	1.669015	0.200
	voicing	953.0018	1	953.0018	2.480911	0.119
	nasality	870.1518	1	870.1518	1.083402	0.301
	place	115.6813	1	115.6813	0.786709	0.378
	manner	13.45077	1	13.45077	0.178746	0.674
	fricative	218.5112	1	218.5112	1.12789	0.291
	envelope	31.78211	1	31.78211	0.280893	0.598
noise * stimrate * group	total	11.06932	1	11.06932	0.084019	0.773
	voicing	756.3388	1	756.3388	1.968946	0.164

	nasality	1546.078	1	1546.078	1.924979	0.169
	place	45.82424	1	45.82424	0.311635	0.578
	manner	129.0505	1	129.0505	1.71494	0.194
	fricative	138.6961	1	138.6961	0.715909	0.400
	envelope	0.0259	1	0.0259	0.000229	0.988
channo * group	total	603.2477	1	603.2477	4.578826	0.035
	voicing	99.19398	1	99.19398	0.258228	0.613
	nasality	0.202526	1	0.202526	0.000252	0.987
	place	710.2822	1	710.2822	4.830389	0.031
	manner	67.71001	1	67.71001	0.899792	0.346
	fricative	1078.927	1	1078.927	5.569101	**0.021**
	envelope	15.60891	1	15.60891	0.137953	0.711
noise * channo * group	total	135.8837	1	135.8837	1.031397	0.313
	voicing	64.50348	1	64.50348	0.167919	0.683
	nasality	85.71889	1	85.71889	0.106726	0.745
	place	187.4059	1	187.4059	1.274484	0.262
	manner	12.12601	1	12.12601	0.161141	0.689
	fricative	99.22756	1	99.22756	0.512183	0.476
	envelope	0.028409	1	0.028409	0.000251	0.987
Error	total	10803.27	82	131.7472		
	voicing	31498.97	82	384.1338		
	nasality	65859.62	82	803.1661		
	place	12057.65	82	147.0445		
	manner	6170.559	82	75.25072		
	fricative	15886.22	82	193.7344		
	envelope	9278.031	82	113.1467		

Table B15. Summary of MANOVA combining all data from CI users with baseline consonant recognition scores of 50% or more from experiment 3 and "3.3mm channel interaction" acoustic model conditions from experiment 4, as described in 5.5. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: "noise condition", "stimulation rate", " channel number" and "group", here summarised as "noise", "stimrate", "channo" and "group", respectively. Significant effects at the *a priori* significance level (*p*≤0.05) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	**Dependent Variable**	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	***F***	***p***
noise	total	50.00983	1	50.00983	0.267845	0.606
	voicing	146.1108	1	146.1108	0.325191	0.570
	nasality	6386.893	1	6386.893	7.560178	**0.007**
	place	12.40246	1	12.40246	0.074354	0.786
	manner	171.6018	1	171.6018	1.389655	0.242
	fricative	145.0758	1	145.0758	0.695783	0.407
	envelope	8.856889	1	8.856889	0.059858	0.807
stimrate	total	143.9548	1	143.9548	0.771001	0.382
	voicing	11.55069	1	11.55069	0.025708	0.873
	nasality	80.95275	1	80.95275	0.095824	0.758
	place	116.7274	1	116.7274	0.699797	0.405
	manner	8.03468	1	8.03468	0.065066	0.799
	fricative	0.145397	1	0.145397	0.000697	0.979
	envelope	12.77146	1	12.77146	0.086314	0.770
channo	total	32.97628	1	32.97628	0.176616	0.675
	voicing	259.4237	1	259.4237	0.577386	0.450
	nasality	6.155636	1	6.155636	0.007286	0.932

	place	1.344727	1	1.344727	0.008062	0.929
	manner	34.96041	1	34.96041	0.283114	0.596
	fricative	9.389557	1	9.389557	0.045032	0.832
	envelope	7.65023	1	7.65023	0.051703	0.821
group	total	544.171	1	544.171	2.914501	0.092
	voicing	2781.096	1	2781.096	6.189744	**0.015**
	nasality	808.2145	1	808.2145	0.956685	0.331
	place	112.4815	1	112.4815	0.674342	0.414
	manner	288.249	1	288.249	2.33428	0.130
	fricative	1391.849	1	1391.849	6.675302	**0.012**
	envelope	376.7263	1	376.7263	2.546056	0.114
noise * stimrate	total	1.292895	1	1.292895	0.006925	0.934
	voicing	303.9128	1	303.9128	0.676403	0.413
	nasality	397.4147	1	397.4147	0.470421	0.495
	place	27.38219	1	27.38219	0.16416	0.686
	manner	6.768534	1	6.768534	0.054813	0.815
	fricative	70.09435	1	70.09435	0.336172	0.564
	envelope	10.20955	1	10.20955	0.069	0.793
noise * channo	total	3.38148	1	3.38148	0.018111	0.893
	voicing	67.10114	1	67.10114	0.149344	0.700
	nasality	267.7422	1	267.7422	0.316927	0.575
	place	8.094727	1	8.094727	0.048529	0.826
	manner	0.380557	1	0.380557	0.003082	0.956
	fricative	67.10114	1	67.10114	0.321817	0.572
	envelope	94.74609	1	94.74609	0.640329	0.426
noise * group	total	0.243247	1	0.243247	0.001303	0.971
	voicing	1128.604	1	1128.604	2.511876	0.117
	nasality	44.32665	1	44.32665	0.05247	0.819
	place	82.93711	1	82.93711	0.497219	0.483
	manner	16.69339	1	16.69339	0.135185	0.714
	fricative	1006.105	1	1006.105	4.825276	0.031
	envelope	20.79186	1	20.79186	0.140519	0.709
stimrate * group	total	76.51135	1	76.51135	0.409784	0.524
	voicing	1114.614	1	1114.614	2.48074	0.119
	nasality	941.7011	1	941.7011	1.114693	0.294
	place	66.18283	1	66.18283	0.396775	0.531
	manner	104.9066	1	104.9066	0.849548	0.359
	fricative	161.7958	1	161.7958	0.775972	0.381
	envelope	96.40542	1	96.40542	0.651543	0.422
noise * stimrate * group	total	6.925125	1	6.925125	0.03709	0.848
	voicing	895.1871	1	895.1871	1.992372	0.162
	nasality	57.18237	1	57.18237	0.067687	0.795
	place	0.04104	1	0.04104	0.000246	0.988
	manner	85.84882	1	85.84882	0.695216	0.407
	fricative	116.6051	1	116.6051	0.559238	0.457
	envelope	1.193722	1	1.193722	0.008068	0.929
channo * group	total	36.14628	1	36.14628	0.193594	0.661
	voicing	102.7506	1	102.7506	0.228687	0.634
	nasality	7.753136	1	7.753136	0.009177	0.924
	place	88.20223	1	88.20223	0.528784	0.469
	manner	35.84291	1	35.84291	0.290261	0.592

	fricative	27.09018	1	27.09018	0.129924	0.719
	envelope	6.46898	1	6.46898	0.04372	0.835
noise * channo * group	total	14.14023	1	14.14023	0.075733	0.784
	voicing	12.64801	1	12.64801	0.02815	0.867
	nasality	451.8822	1	451.8822	0.534894	0.467
	place	71.93473	1	71.93473	0.431258	0.513
	manner	0.120557	1	0.120557	0.000976	0.975
	fricative	349.398	1	349.398	1.675711	0.199
	envelope	14.44609	1	14.44609	0.097632	0.755
Error	total	15310.35	82	186.7116		
	voicing	36843.18	82	449.3071		
	nasality	69274.2	82	844.8073		
	place	13677.76	82	166.802		
	manner	10125.79	82	123.4852		
	fricative	17097.6	82	208.5073		
	envelope	12133.1	82	147.9647		

Table B16. Summary of MANOVA combining all data from CI users with baseline consonant recognition scores of less than 50% from experiment 3 and "no channel interaction" acoustic model conditions from experiment 4, as described in 5.5. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: "noise condition", "stimulation rate", " channel number" and "group", here summarised as "noise", "stimrate", "channo" and "group", respectively. Significant effects at the *a priori* significance level (*p*≤0.05) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	**Dependent Variable**	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	***F***	***p***
noise	total	693.4384	1	693.4384	6.789389	**0.011**
	voicing	4359.032	1	4359.032	10.55259	**0.002**
	nasality	3460.321	1	3460.321	3.114218	0.082
	place	392.8727	1	392.8727	5.262372	**0.024**
	manner	1521.056	1	1521.056	23.37019	**0.000**
	fricative	1433.889	1	1433.889	7.060545	**0.010**
	envelope	1301.201	1	1301.201	8.705588	**0.004**
stimrate	total	129.808	1	129.808	1.270938	0.263
	voicing	7.832742	1	7.832742	0.018962	0.891
	nasality	5.573367	1	5.573367	0.005016	0.944
	place	139.1964	1	139.1964	1.86448	0.176
	manner	86.54697	1	86.54697	1.329746	0.252
	fricative	288.0903	1	288.0903	1.418572	0.237
	envelope	158.515	1	158.515	1.060533	0.306
channo	total	609.2164	1	609.2164	5.964779	**0.017**
	voicing	147.2546	1	147.2546	0.356482	0.552
	nasality	0.958367	1	0.958367	0.000863	0.977
	place	842.98	1	842.98	11.29138	**0.001**
	manner	569.4399	1	569.4399	8.749127	**0.004**
	fricative	416.5802	1	416.5802	2.051262	0.156
	envelope	350.4012	1	350.4012	2.344334	0.130
group	total	14450.14	1	14450.14	141.4799	**0.000**
	voicing	9490.776	1	9490.776	22.97582	**0.000**
	nasality	24221.39	1	24221.39	21.79876	**0.000**
	place	11458.21	1	11458.21	153.4781	**0.000**
	manner	17186.73	1	17186.73	264.0646	**0.000**
	fricative	29490.3	1	29490.3	145.2118	**0.000**
	envelope	10617.27	1	10617.27	71.03404	**0.000**

noise * stimrate	total	10.86983	1	10.86983	0.106425	0.745
	voicing	2.176379	1	2.176379	0.005269	0.942
	nasality	746.3537	1	746.3537	0.671703	0.415
	place	9.794182	1	9.794182	0.131189	0.718
	manner	19.34697	1	19.34697	0.297255	0.587
	fricative	566.2861	1	566.2861	2.788422	0.099
	envelope	12.75464	1	12.75464	0.085334	0.771
noise * channo	total	2.497515	1	2.497515	0.024453	0.876
	voicing	172.1253	1	172.1253	0.416691	0.520
	nasality	841.5115	1	841.5115	0.757343	0.387
	place	12.68523	1	12.68523	0.169914	0.681
	manner	0.710186	1	0.710186	0.010912	0.917
	fricative	152.112	1	152.112	0.749007	0.389
	envelope	21.70923	1	21.70923	0.145244	0.704
noise * group	total	59.25261	1	59.25261	0.580137	0.449
	voicing	0.753296	1	0.753296	0.001824	0.966
	nasality	1316.536	1	1316.536	1.184855	0.280
	place	17.28874	1	17.28874	0.231576	0.632
	manner	377.6387	1	377.6387	5.802209	0.018
	fricative	353.0904	1	353.0904	1.738635	0.191
	envelope	457.6452	1	457.6452	3.061842	0.084
stimrate * group	total	201.9667	1	201.9667	1.977436	0.164
	voicing	761.8527	1	761.8527	1.844337	0.178
	nasality	1803.621	1	1803.621	1.623222	0.206
	place	97.36705	1	97.36705	1.304192	0.257
	manner	39.46964	1	39.46964	0.606429	0.438
	fricative	639.4623	1	639.4623	3.148745	0.080
	envelope	277.357	1	277.357	1.855637	0.177
noise * stimrate * group	total	22.5885	1	22.5885	0.221162	0.639
	voicing	182.543	1	182.543	0.441911	0.508
	nasality	343.6577	1	343.6577	0.309285	0.580
	place	84.67418	1	84.67418	1.134177	0.290
	manner	21.89097	1	21.89097	0.336343	0.564
	fricative	20.90076	1	20.90076	0.102916	0.749
	envelope	47.87131	1	47.87131	0.320279	0.573
channo * group	total	11.28438	1	11.28438	0.110484	0.740
	voicing	498.7713	1	498.7713	1.207454	0.275
	nasality	3.255034	1	3.255034	0.002929	0.957
	place	232.468	1	232.468	3.113816	0.082
	manner	7.909186	1	7.909186	0.12152	0.728
	fricative	0.756852	1	0.756852	0.003727	0.951
	envelope	288.6852	1	288.6852	1.931427	0.169
noise * channo * group	total	76.97752	1	76.97752	0.753679	0.388
	voicing	697.0046	1	697.0046	1.687349	0.198
	nasality	0.385458	1	0.385458	0.000347	0.985
	place	12.08256	1	12.08256	0.161841	0.689
	manner	4.104852	1	4.104852	0.063069	0.802
	fricative	3.08867	1	3.08867	0.015209	0.902
	envelope	2.222561	1	2.222561	0.01487	0.903

Error	total	7966.578	78	102.1356		
	voicing	32219.99	78	413.0768		
	nasality	86668.65	78	1111.137		
	place	5823.243	78	74.65696		
	manner	5076.656	78	65.08533		
	fricative	15840.61	78	203.0848		
	envelope	11658.45	78	149.4673		

Table B17. Summary of MANOVA combining all data from CI users with baseline consonant recognition scores of less than 50% from experiment 3 and "1mm channel interaction" acoustic model conditions from experiment 4, as described in 5.5. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: "noise condition", "stimulation rate", " channel number" and "group", here summarised as "noise", "stimrate", "channo" and "group", respectively. Significant effects at the *a priori* significance level (*p*≤0.05) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	**Dependent Variable**	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	***F***	***p***
noise	total	412.8328	1	412.8328	3.897582	0.052
	voicing	2526.151	1	2526.151	6.818805	**0.011**
	nasality	7196.003	1	7196.003	8.118137	**0.006**
	place	77.25286	1	77.25286	0.914571	0.342
	manner	1087.578	1	1087.578	15.56838	**0.000**
	fricative	475.7201	1	475.7201	2.216099	0.141
	envelope	1060.876	1	1060.876	8.796681	**0.004**
stimrate	total	292.1346	1	292.1346	2.758062	0.101
	voicing	740.5164	1	740.5164	1.998866	0.161
	nasality	172.1253	1	172.1253	0.194182	0.661
	place	145.7	1	145.7	1.724894	0.193
	manner	25.29188	1	25.29188	0.362046	0.549
	fricative	865.62	1	865.62	4.032412	**0.048**
	envelope	290.4737	1	290.4737	2.408579	0.125
channo	total	779.5747	1	779.5747	7.360017	**0.008**
	voicing	852.1591	1	852.1591	2.300221	0.133
	nasality	11.61364	1	11.61364	0.013102	0.909
	place	767.8812	1	767.8812	9.090688	**0.003**
	manner	437.0041	1	437.0041	6.25559	**0.014**
	fricative	1381.133	1	1381.133	6.43388	**0.013**
	envelope	466.8727	1	466.8727	3.871263	**0.053**
group	total	13606.12	1	13606.12	128.4563	**0.000**
	voicing	8375.138	1	8375.138	22.60689	**0.000**
	nasality	29685.05	1	29685.05	33.48905	**0.000**
	place	11146.1	1	11146.1	131.9549	**0.000**
	manner	18027.17	1	18027.17	258.0538	**0.000**
	fricative	25928.2	1	25928.2	120.7842	**0.000**
	envelope	11650.12	1	11650.12	96.60167	**0.000**
noise * stimrate	total	3.081833	1	3.081833	0.029096	0.865
	voicing	14.52183	1	14.52183	0.039199	0.844
	nasality	2.847307	1	2.847307	0.003212	0.955
	place	101.3242	1	101.3242	1.199543	0.277
	manner	0.469333	1	0.469333	0.006718	0.935
	fricative	56.29176	1	56.29176	0.26223	0.610
	envelope	55.70919	1	55.70919	0.461935	0.499
noise * channo	total	21.02552	1	21.02552	0.198503	0.657
	voicing	209.7291	1	209.7291	0.566119	0.454

	nasality	11.42867	1	11.42867	0.012893	0.910
	place	101.7164	1	101.7164	1.204186	0.276
	manner	1.787761	1	1.787761	0.025591	0.873
	fricative	72.30364	1	72.30364	0.33682	0.563
	envelope	20.07274	1	20.07274	0.166441	0.684
noise * group	total	188.0328	1	188.0328	1.77523	0.187
	voicing	276.5588	1	276.5588	0.746511	0.390
	nasality	105.6647	1	105.6647	0.119205	0.731
	place	230.7249	1	230.7249	2.731474	0.102
	manner	647.9672	1	647.9672	9.275467	**0.003**
	fricative	1214.273	1	1214.273	5.65658	**0.020**
	envelope	619.6967	1	619.6967	5.138464	**0.026**
stimrate * group	total	72.46923	1	72.46923	0.684187	0.411
	voicing	10.16305	1	10.16305	0.027433	0.869
	nasality	1005.537	1	1005.537	1.134392	0.290
	place	92.06402	1	92.06402	1.089915	0.300
	manner	111.4385	1	111.4385	1.595211	0.210
	fricative	164.85	1	164.85	0.767939	0.384
	envelope	148.865	1	148.865	1.234374	0.270
noise * stimrate * group	total	10.3125	1	10.3125	0.097361	0.756
	voicing	124.8885	1	124.8885	0.33711	0.563
	nasality	2260.515	1	2260.515	2.55019	0.114
	place	5.132182	1	5.132182	0.060758	0.806
	manner	95.304	1	95.304	1.36425	0.246
	fricative	137.4111	1	137.4111	0.640117	0.426
	envelope	9.159186	1	9.159186	0.075947	0.784
channo * group	total	43.53068	1	43.53068	0.410976	0.523
	voicing	27.83909	1	27.83909	0.075146	0.785
	nasality	0.390307	1	0.390307	0.00044	0.983
	place	193.8626	1	193.8626	2.295074	0.134
	manner	33.30009	1	33.30009	0.476681	0.492
	fricative	252.28	1	252.28	1.175223	0.282
	envelope	198.8807	1	198.8807	1.649099	0.203
noise * channo * group	total	138.7375	1	138.7375	1.30983	**0.256**
	voicing	626.9251	1	626.9251	1.69225	0.197
	nasality	625.3667	1	625.3667	0.705504	0.404
	place	99.99638	1	99.99638	1.183824	0.280
	manner	17.68909	1	17.68909	0.253214	0.616
	fricative	4.296307	1	4.296307	0.020014	0.888
	envelope	1.720742	1	1.720742	0.014268	0.905
Error	total	8261.778	78	105.9202		
	voicing	28896.53	78	370.4683		
	nasality	69140.03	78	886.4107		
	place	6588.581	78	84.46899		
	manner	5448.938	78	69.85818		
	fricative	16743.92	78	214.6656		
	envelope	9406.769	78	120.5996		

Table B18. Summary of MANOVA combining all data from CI users with baseline consonant recognition scores of less than 50% from experiment 3 and "3.3mm channel interaction" acoustic

model conditions from experiment 4, as described in 5.5. Seven dependent variables (six feature transmission values and total percentage correct) were entered. There were four factors: "noise condition", "stimulation rate", " channel number" and "group", here summarised as "noise", "stimrate", "channo" and "group", respectively. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted. Because the factors "stimulation rate" and "channel number" overlapped, interactions involving both these factors could not be computed.

Source	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
noise	total	406.741	1	406.741	2.484624	0.119
	voicing	3062.103	1	3062.103	6.975435	**0.010**
	nasality	6832.735	1	6832.735	7.345548	**0.008**
	place	35.25824	1	35.25824	0.335028	0.564
	manner	1402.655	1	1402.655	11.6339	**0.001**
	fricative	1425.563	1	1425.563	6.192824	**0.015**
	envelope	1046.106	1	1046.106	6.654488	**0.012**
stimrate	total	126.2844	1	126.2844	0.771423	0.382
	voicing	878.148	1	878.148	2.000411	0.161
	nasality	143.4069	1	143.4069	0.15417	0.696
	place	91.16983	1	91.16983	0.866307	0.355
	manner	129.011	1	129.011	1.070042	0.304
	fricative	752.7478	1	752.7478	3.270029	0.074
	envelope	443.866	1	443.866	2.823519	0.097
channo	total	116.34	1	116.34	0.710676	0.402
	voicing	861.7306	1	861.7306	1.963012	0.165
	nasality	1.566095	1	1.566095	0.001684	0.967
	place	138.4626	1	138.4626	1.315688	0.255
	manner	354.7134	1	354.7134	2.942063	0.090
	fricative	135.2296	1	135.2296	0.587454	0.446
	envelope	412.3801	1	412.3801	2.623232	0.109
group	total	7664.443	1	7664.443	46.81913	**0.000**
	voicing	3318.885	1	3318.885	7.560381	**0.007**
	nasality	14955.18	1	14955.18	16.0776	**0.000**
	place	7378.958	1	7378.958	70.11577	**0.000**
	manner	12567.06	1	12567.06	104.2336	**0.000**
	fricative	20047.83	1	20047.83	87.09025	**0.000**
	envelope	6608.456	1	6608.456	42.0377	**0.000**
noise * stimrate	total	1.179409	1	1.179409	0.007205	0.933
	voicing	2.199409	1	2.199409	0.00501	0.944
	nasality	834.3003	1	834.3003	0.896916	0.347
	place	14.02188	1	14.02188	0.133238	0.716
	manner	1.772182	1	1.772182	0.014699	0.904
	fricative	71.31409	1	71.31409	0.309797	0.579
	envelope	39.15237	1	39.15237	0.249056	0.619
noise * channo	total	7.339636	1	7.339636	0.044835	0.833
	voicing	634.3541	1	634.3541	1.445051	0.233
	nasality	996.6699	1	996.6699	1.071472	0.304
	place	27.81856	1	27.81856	0.264335	0.609
	manner	2.430307	1	2.430307	0.020157	0.887
	fricative	0.191761	1	0.191761	0.000833	0.977
	envelope	61.3965	1	61.3965	0.390555	0.534
noise * group	total	192.1819	1	192.1819	1.173965	0.282
	voicing	133.5091	1	133.5091	0.304132	0.583
	nasality	154.9584	1	154.9584	0.166588	0.684
	place	325.4822	1	325.4822	3.092772	0.083
	manner	440.2283	1	440.2283	3.65134	0.060

	fricative	357.2401	1	357.2401	1.551895	0.217
	envelope	631.0767	1	631.0767	4.014404	**0.049**
stimrate * group	total	206.4164	1	206.4164	1.260918	0.265
	voicing	0.588015	1	0.588015	0.001339	0.971
	nasality	1079.425	1	1079.425	1.160438	0.285
	place	146.8298	1	146.8298	1.395195	0.241
	manner	17.86964	1	17.86964	0.148214	0.701
	fricative	219.7678	1	219.7678	0.954698	0.332
	envelope	66.85	1	66.85	0.425246	0.516
noise * stimrate * group	total	6.460742	1	6.460742	0.039466	0.843
	voicing	182.3327	1	182.3327	0.415352	0.521
	nasality	288.0903	1	288.0903	0.309712	0.579
	place	73.73388	1	73.73388	0.700628	0.405
	manner	60.00152	1	60.00152	0.497664	0.483
	fricative	116.2141	1	116.2141	0.504848	0.479
	envelope	17.91903	1	17.91903	0.113987	0.737
channo * group	total	111.028	1	111.028	0.678227	0.413
	voicing	26.14064	1	26.14064	0.059548	0.808
	nasality	2.346095	1	2.346095	0.002522	0.960
	place	4.081227	1	4.081227	0.03878	0.844
	manner	61.48803	1	61.48803	0.509994	0.477
	fricative	93.15031	1	93.15031	0.404656	0.527
	envelope	237.2401	1	237.2401	1.509131	0.223
noise * channo * group	total	20.10764	1	20.10764	0.12283	0.727
	voicing	205.4668	1	205.4668	0.468051	0.496
	nasality	10.12585	1	10.12585	0.010886	0.917
	place	26.92256	1	26.92256	0.255821	0.614
	manner	1.71564	1	1.71564	0.01423	0.905
	fricative	102.7791	1	102.7791	0.446485	0.506
	envelope	21.78183	1	21.78183	0.138559	0.711
Error	total	12768.85	78	163.7032		
	voicing	34240.74	78	438.9838		
	nasality	72554.61	78	930.1873		
	place	8208.692	78	105.2396		
	manner	9404.165	78	120.5662		
	fricative	17955.29	78	230.196		
	envelope	12261.84	78	157.2031		