Robustness against distortion of fundamental frequency cues in simulated electro-acoustic hearing

Arthur Vermeulen[1]

Hearing and Balance Centre

Institute of Sound and Vibration Research

University of Southampton

Highfield, Southampton SO17 1BJ

U.K.

af.vermeulen.02@mindef.nl

Carl Verschuur

University of Southampton Auditory Implant Service

Highfield, Southampton SO17 1BJ

U.K.

cav@isvr.soton.ac.uk

Running title: Distortion of fundamental frequency cues

[1] Present address: Netherlands Defence Academy, 1781AC Den Helder, The Netherlands

Abstract

Speech recognition by cochlear implant users can be improved by adding an audible low frequency acoustic signal to electrical hearing; the resulting improvement is deemed "electro-acoustic (EAS) benefit". We assessed the role of fundamental frequency (F0) information as a predictor of EAS benefit. Normal hearing listeners were presented with vocoded speech tokens with differing manipulations of the F0 signal, specifically: a pure tone with the correct mean F0 but with smaller variations around this mean (i.e. a smaller modulation depth), or a narrowband of white noise centered around F0, at varying bandwidths; a pure tone down-shifted in frequency but keeping overall frequency modulations. Speech recognition thresholds significantly improved (p<0.05) when tones with reduced frequency modulation, or noise bands maintaining F0 information via their centre frequency were presented alongside the vocoded speech. Addition of a pure tone downshifted by 50 Hz or only a tone to indicate voicing, showed no significant EAS benefit. These results confirm that the presence of the target's F0 is beneficial for electro-acoustic hearing in a noisy environment and they indicate that the benefit is robust to a certain decrease in frequency selectivity, so long as mean F0 and frequency modulations of F0 are preserved.

PACS numbers: 43.71.Ky, 43.66.Hg, 43.66.Sr, 43.66.Ts

I. INTRODUCTION

Users of a cochlear implant (CI) have significant problems understanding speech in a noisy environment, despite their good performance in quiet (Wilson and Dorman, 2008a,b). Speech intelligibility in noise can be improved by adding an (amplified) low frequency acoustic signal to the electric stimulation of the CI to the same or to the not-implanted ear (Cullington and Zeng, 2010). The combination in the same ear is called electro-acoustic stimulation (EAS) and its use has proved beneficial even if only frequencies below a range as low as 125 Hz – the voice fundamental frequency (F0) for some male speakers – are present via acoustic stimulation (Zhang et al., 2010). In some cases, EAS benefit is synergistic, e.g. performance with combined electrical and acoustic hearing in the same ear is greater than the sum of electric-only and acoustic-only hearing (Wilson and Dorman, 2008a,b). One reason for EAS benefit is likely to be improved ability to undertake sound source segregation (Carlyon, 2004). For monaural listening, differences in both F0 and onset (and, to a lesser extent, offset) between target and masker improve performance (Carlyon, 2004), and F0 cues in particular are likely to be better represented by acoustic hearing than electrical hearing (Turner et al., 2008).

A general approach among studies evaluating the role of F0 in EAS is to compare the speech performance of electric hearing (alone) with electric hearing and an additional pure tone (‘a carrier’) which has been manipulated so that its frequency varies and equals the (instantaneous) fundamental frequency of the speech. The presentation of the carrier tone can also be varied according to either the presence or absence of voiced speech (this gives a fixed amplitude for voiced speech fragments and a zero-amplitude for voiceless fragments), or an amplitude which is modulated with the (instantaneous) energy in the low frequency region -

the carrier follows the ‘envelope’ of the low frequency. Kong and Carlyon (2006, 2007) used normal hearing listeners in EAS simulations and also found the benefit of the low-frequency acoustic hearing component. Simulated EAS benefit was obtained at a low SNR (5 dB) when the low pass acoustic signal was replaced with a tone with frequency equal to F0 and an amplitude that tracked the envelope of the low frequency signal. Benefit was also obtained with a tone that tracked amplitude changes but was fixed at mean F0. No benefit was observed for both tones if SNR was increased to 10 or 15 dB. They concluded that speech understanding in simulated EAS is not improved via the F0 cue of frequency variation, but by providing low frequency phonetic information on the variation in the low frequency spectrum over time. However, Carroll et al. (2011) found that there was a significant improvement to speech recognition at a fixed SNR of + 10 dB by adding a low-pass (500 Hz) acoustic signal (+8.3%) or an acoustic F0 modulated tone (+5.3%). The same authors also measured the SNR for a 50% correct score in the case of one competing talker (of the opposite sex) with simulated EAS. They found an improvement over electrical hearing for EAS of 6.3 dB and an improvement of about 4 dB if an F0 modulated tone was presented. This benefit remained if only frequency variation was considered (i.e. amplitude fixed) and disappeared if only the amplitude variation related to the energy was considered (i.e. frequency fixed). Brown and Bacon (2009a) also showed that a similar benefit of simulated EAS can be obtained with low-pass filtered acoustic stimulation (500 Hz) as an acoustic stimulation consisting of a tone that is modulated in frequency (F0) and in amplitude with the amplitude envelope of the low pass target speech. Brown and Bacon (2009b) repeated the experiment with real CI users and confirmed the result of a benefit of the additional FM/AM modulated tone over electric hearing alone.

While the majority of studies have suggested that F0 frequency is a key determinant of EAS benefit, few studies have evaluated how robust F0 is to frequency manipulations. Brown et al. (2010) measured speech recognition in noise with simulated EAS with a shifted tone which followed the variation of the fundamental frequency. They showed that an FM tone and an AM/FM tone both improve speech intelligibility, and this improvement remains by shifting the tone down to a frequency as much as 75 Hz below the mean F0 frequency of the talker. They concluded that segregation does not depend on the exact value of the fundamental frequency but rather the variation of F0 (in combination with voicing information).

Presenting the variations of the fundamental frequency in EAS seems to be sufficient for obtaining a significant benefit in speech recognition in a noisy environment. A good representation of the fundamental frequency (mean or variation) in the auditory system through acoustic stimulation bypasses the poor frequency selectivity of electric hearing (Wilson and Dorman, 2008a,b) and thus seems to be beneficial. However, a good representation is only possible if the frequency selectivity of the auditory system is sufficient for acoustic hearing and this might not be the case for people with a severe low frequency hearing loss. The present research project looks into the relation between frequency selectivity and the potential benefit of simulated EAS if only a tone with the instantaneous fundamental frequency is presented as acoustic stimulation. Assessing this relation might be important in explaining the large variability in speech intelligibility in noise for EAS by CI users (Wilson and Dorman, 2008a) and it might be helpful to predict the benefit of EAS for the CI user.

The aim of the present study was to assess the perceptual attributes of F0 that are most important to providing EAS benefit. To address this, we determined the extent to which speech recognition improves for simulated cochlear implant listeners in a multi-talker babble background when a manipulated additional acoustical tone is presented with a frequency corresponding to the instantaneous fundamental frequency of the speech signal (F0). In particular, the influence of a decrease in frequency selectivity for the additional tone was investigated. This was done by changing (1) the frequency range of F0 (i.e. the modulation depth of the F0-contour), and (2) the frequency specificity (i.e. replacing the pure tone by a narrowband noise signal centred at F0) and (3) the absolute value of F0 (by down-shifting).

II. EXPERIMENT

1. Subjects

Thirteen normal hearing listeners participated in the experiment (5 males, 8 females, mean age 34, age range 22 to 40). All participants were fluent in English and had audiometric air conduction thresholds better than 20 dB HL across frequencies 250 to 8000 Hz bilaterally. Participants with tinnitus were excluded and the necessary approvals (N^{o} 8804) were given by the Faculty of Engineering and the Environment Health and Safety officer and the Ethics Committee concerning ethics and risk assessment.

2. Stimuli

The target signal was the Bench-Kowal-Bamford (BKB) speech sentences spoken by a male speaker (Bench et al., 1979). These sentences are grouped in 21 lists of 16 sentences representing everyday life speech. The noise signal consisted of (8-talker) babble noise (male and female) which represents everyday situations with little opportunity to ‘listen in the gaps’. This type of noise has also been used by Brown and Bacon (2009a) and Verschuur et al.

(2013), for example. The level of noise was determined by the signal-to-noise ratio (SNR) as needed in the procedure, see below.

Vocoded speech and the low frequency signals were calculated separately and recombined off-line before the experiment. The target and noise signal were directly added in the time domain for vocoded speech – the electrical hearing component of the EAS simulation. This signal was subsequently high-pass filtered with a cut-off frequency of 250 Hz (-6 dB) which was just above the highest value of F0 of the target signal. The noise signal started 0.2 s before the target signal and ended after the target signal in order to prevent onset (and offset) clues. The SNR of the mixed signal was calculated with the (average) RMS values of the target speech (entire list of BKB sentences because of variation between lists) and the noise signal (total duration 15 s). The part of the noise signal used during the target sentence was chosen at random (uniform distribution) within the (possible) duration of the total available noise signal. Consequently each sentence was corrupted with a different noise profile, and every presentation of the sentence used a different profile. Each sentence lasted 2.5 seconds. The high-pass filtered speech was further processed by a vocoder to mimic CI speech processing strategies so that normal hearing listeners could hear how CI subjects perceive the processed speech. The vocoder used is described in Verschuur et al. (2013). It uses the processing strategies of the Nucleus 24 CI device and different options can be chosen for pre- and post-processing of the speech signals. In the present research a pre-emphasis filter of 9 dB (for the high frequencies) was chosen to simulate a CI microphone and the ACE coding strategy was selected. Within this strategy, only the 10 channels (out of 22) with the greatest amplitude were used (n-of-m strategy) as modulators of the noise bands generated with a series of fourth-order Butterworth filters. Centre frequencies of the noise bands corresponded to the centre frequencies of the 22 channels and were related to the Greenwood map

(Greenwood, 1990) to allow for the dependency of the critical bandwidth on the (centre) frequency (Moore, 2012). The inverse middle-ear filter of the program was used as a post-processing stage to cancel the middle-ear characteristics of the normal hearing listeners who were listing to the stimuli through a headphone, and the vocoder signal was again filtered by a high-pass filter (-6 dB at 250 Hz) to guarantee the absence of low frequency components. A high number of channels was chosen in order to obtain a good score in quiet for the vocoder only condition. This was necessary for the adaptive procedure used, see below, because the participant must be able to achieve a 100% score for speech in quiet. The vocoder signals are identical for all conditions studied.

The low frequency signal corresponded to the acoustic hearing component of the EAS stimulation. The (speech) target signal without noise was low-pass filtered (-6 dB at 250 Hz) and the energy in this signal (expressed by the RMS value) was calculated. Non-manipulated speech or a frequency modulated (FM) tone related to F0 was used. The FM tone – also called carrier – was different for each condition. It was low-pass filtered again to ensure that no higher frequencies were present. Next, its amplitude was set to about three times the value of the initial low-frequency signal. In the pilot study, this increase turned out to be necessary in order to hear the tone. Finally, the amplitude of the tone was set to zero in the case of voiceless speech.

Four versions of the FM tone were considered: (1) an FM tone (pure sine wave) with a frequency given by the instantaneous value of F0, (2) a distorted version of the FM tone to simulate a decrease in frequency selectivity by 'smearing' the pure tone in the frequency domain. The bandwidth of this smearing was another parameter of the study and two values were considered. (3) A distorted version of the FM tone to simulate a change in frequency selectivity by changing the possible depth of modulation; i.e. the frequency range over which the frequency variation of F0 was varied. (4) An FM tone (pure sine wave) with a frequency given by a shifted value of F0; i.e. the frequency was given by F0+Fc where Fc was a constant value. The latter signal was used to duplicate the experiment of Brown et al. (2010).

The calculation of the fundamental frequency was performed with the YAAPT algorithm described in Zahorian and Hu (2008) because of its accuracy (see Zahorian and Hu, 2008). The output of the YAAPT algorithm consists of the instantaneous value of F0 and a Boolean variable which indicates if the speech is voiced or voiceless. The sampling rate of the output was below the sample rate of the speech but F0 was estimated at each time instant of the original speech signal by interpolation. The BKB sentences without noise were used for the calculation of F0 because of the low accuracy of the methods available for an SNR smaller than 10 dB (Zahorian and Hu, 2008). This commonly used choice (e.g. Brown and Bacon, 2008a,b) does not reflect a real situation because that would require the use of the combined speech and noise signals. However, it facilitated the interpretation of the results because the (gross) estimation error in F0 would be relatively small (about 5%) and would not depend on the SNR.

Changing the frequency selectivity might be performed in a simulation by pre-processing the speech signal. In order to simulate supra-threshold changes due to hearing loss (i.e. widening of the tuning curves, see Moore, 2012), we used a novel approach in which it is assumed that the CI user does not hear the pure tone as a pure tone but as a narrow band of frequencies around F0 because more auditory filters (than for a normal hearing listener) pick up the fundamental frequency; the frequency specificity was decreased and the fundamental frequency was smeared in the frequency domain. The smearing of the fundamental frequency in the frequency domain was simulated with a band-pass filter centred at the instantaneous value of F0. The input of the filter was white noise and its output was narrowband noise with a changing centre frequency. The bandwidth of the filter was assumed to simulate the amount of smearing in the frequency domain. In the present work 39 Finite Infinite Response (FIR) filters of order 200 with centre frequencies between 55 Hz and 240 Hz (step size 5 Hz) were used and at each time instant the filter coefficients of the nearest centre frequency were used to calculate the narrowband noise carrier.

An alternative approach to simulated frequency selectivity is presenting the correct value of the mean F0 but the wrong value of the (frequency) variation around this value to the participant, so the frequency range of F0 is changed (Binns and Culling, 2007). Detection of the variation of the fundamental frequency can then be made easier or more difficult depending on the magnitude of the variation in the frequency domain, which is also called the modulation depth *m* of the frequency modulation or F0-contour. For example, a modulation depth of 1 gives the original FM signal, a modulation depth of 0.4 reduces the variation in frequency with 60%, and a modulation depth $m = 0$ gives no modulation in the frequency at all so that the signal is reduced to a pure tone with a frequency equal to the mean F0 of the target speech signal (144 Hz in the present study). The original F0-contour (i.e. the variation around its mean value) is thus multiplied by the value of *m*, as illustrated in Fig. 1. The value $m = -1$ corresponds to an inversion of the contour around its mean and has been used by Binns and Culling (2007) for example. A value $0 < m < 1$ reflects a decrease in frequency sensitivity and a value $m > 1$ is considered an increase of sensitivity.

3. Conditions

Eleven different conditions were included in the experiment, see Table I. Conditions I and II were the extreme conditions with no low frequency cue, and availability of all low frequency cues respectively. The latter contained the low pass filtered target speech (≤ 250 Hz) without noise. The bandwidth of the narrowband noise used for the ‘mildly smeared’ and ‘seriously smeared’ were 50 and 75 Hz respectively. The modulation depth *m* for conditions VI-VIII is indicated in the abbreviation and equal to 0.7, 0.4 and 1.5 respectively, and it equals -1 for the inverted modulation. A schematic illustration of the spectrogram of conditions III-IX is given in Fig. 1. Condition X corresponded with a pure tone which has a frequency equal to the mean value of F0 and it indicated voicing in order to assess the contribution of voicing in all other conditions. Finally, the frequency shift (Fc) applied in the last condition (XI) equalled -50 Hz.

4. Procedure

All participants were tested in the 11 listening conditions. For each condition they were asked to listen to the 16 sentences (one full list from the BKB sentence test) without a break. The stimuli were presented via circumaural headphones (Sennheiser HDA 300) to one ear at a level of 68 dB (A) for target speech (without the noise) in a quiet surrounding. Before starting the experiment, the equipment was calibrated once with a Bruel & Kjaer Soundlevel meter (G-4 type 2250) connected to an artificial ear (Bruel & Kjaer type 4153) and a sound level calibrator (Bruel & Kjaer type 4153). The participant was free to choose their right or left ear but they could not change during the experiment. After each sentence was played, participants were asked to repeat the sentence back to the tester as best as they could (no feedback was given). The tester scored the number of keywords correctly identified in each sentence on the computer (this was not visible to the participant) and the next sentence was played with an SNR related to the performance in the previous sentence. In a correct response all keywords were correctly identified in the sentence.

The performance measure used in this study was the signal-to-noise ratio (SNR) at which participants correctly identify 50% of the sentences; sometimes called the speech-reception threshold. Some authors use the abbreviation SRT but, as explained by Carroll et al (2011), this would not be correct in the present context because the SRT is initially defined as the absolute level in quiet, and not as a relative level in noise. The advantage of the measure chosen is that there is no need to present several lists with a different SNR for each listing condition, meaning that more conditions can be evaluated, and the SNR required for a 50% correct performance can be assessed without ceiling effects within the length of one list of the BKB sentences with an adaptive procedure. The adaptive procedure used has been inspired by the procedures of Turner et al. (2004) and Carroll et al. (2011); it starts at an ‘easy’ SNR with a step size of 5 dB and after one reversal the step size is decreased to 2 dB, and after four reversals the step size is decreased further to 1 dB. The first sentence of the first list was presented at a fixed SNR of 5 dB and the subsequent lists started 5 dB above the final SNR of the previous list. The second decrease in step size was included because a step size of 2 dB gives a final SNR which depends strongly on the number of reversals (which is relatively low in our case) and the SNR obtained at the end of the 5 dB-step size. The participants were able to take a short break between the different conditions if they wanted. Participants were given 20 sentences at the beginning of the session to familiarise themselves with processed speech. These sentences were randomly selected from the manipulated BKB sentence lists which were not used in the experiment. Conditions and the choice of the 11 (out of 21) BKB sentence lists were randomized to cancel out the influence of fatigue and training effects.

III. RESULTS

The final SNR scores for the 11 conditions are shown in Fig. 2 and repeated-measures ANOVA shows that scores are significantly different: $F(10,120) = 3.87$, $p < 0.001$, $\omega^2 = 0.12$. In order to indicate which conditions differed significantly from the others, post hoc tests Field, 2009) have been used. The results of such tests - pairwise comparisons (no adjustment for multiple comparisons have been used: LSD) - are summarized in Table II by showing the difference between the mean values of the conditions (row's condition minus column's condition) and the cell is grey if the difference is significant ($p < 0.05$). The values in Table II are symmetrical; the upper right triangle is the mirror of the lower left triangle (with the opposite sign). The *p*-value of the significant results was always so low that the results not only differed significantly (two-tailed test) but also confirm hypotheses about an increase or decrease of the score between both conditions (one-tailed test).

IV. DISCUSSION

Speech recognition in noise was measured with normal hearing listeners in simulated electro-acoustic hearing. Vocoded speech consisted of the high-pass filtered (250 Hz) target speech mixed with a multi-talker babble noise. The low frequencies of the target speech were presented as acoustical stimulation or replaced by a manipulated tone with the instantaneous fundamental frequency. The aim of the study was to determine the improvement in speech recognition caused by the manipulated acoustical stimulation, and to study the robustness for several manipulations deforming the pure tone. Manipulations involved consisted of changing the modulation depth, replacing the pure tone by a narrowband noise signal centred at F0, and down-shifting the tone in frequency.

A. Benefit of the fundamental frequency cue in simulated EAS

In the present study, the addition of F0 information to vocoded speech improved the SNR for a 50% correct score by about 3.5 dB (from the initial 4.7 dB for the CI simulation to about 1.2 dB in the EAS setting). Few other studies have used this performance measure so it is difficult to compare values; Chang et al. (2006) found a score of 0 dB for the EAS setting (down from 10 dB for the 4-channel CI simulation), and Carroll et al. (2011) report a score of about 5 dB for the EAS setting with a cut-off frequency of 500 Hz (down from 11.6 dB for the 6-channel CI simulation). Results of Kong and Carlyon (2007) reflect SNRs which are similar to the values measured by Carroll et al. (2011). Although the score for the EAS setting in our experiment is close to scores reported by the other authors, the benefit observed is smaller. This is probably due to the larger number of channels in our CI simulation (which gives a better baseline performance for the CI only condition) and the use of babble noise instead of a single competing talker (which facilitates glimpsing the target, see Li and Loizou, 2008) of the other sex (which results in a very different F0 and an easier setting).

Our results indicate that the low-frequency signal can be replaced with a pure tone with a frequency equal to the fundamental frequency, and with an amplitude which is fixed in level during voiced speech (and set to zero for voiceless speech). Although the improvement in performance as compared with the low-frequency signal was non-significant (1.3 dB), the improvement with only the CI simulation remains significant: 2.2 dB. Carroll et al. (2011) found a similar pattern (2.3 dB and 4 dB respectively), although both comparisons were significant. The amplitude of the pure tone indicates voicing but our results show that the presence of only the voicing cue has no significant benefit over the CI simulation – this is in line with the results of Carroll et al. (2011) and Brown and Bacon (2010). Consequently, our study confirms that the presence of the target's fundamental frequency, and, in particular, its variation in frequency is an important cue for improving speech recognition in noise.

B. Robustness to alterations of F0 manipulations

One of our aims was to determine the robustness of simulated EAS benefit to manipulations of the F0 contour in the frequency domain. Binns and Culling (2007) manipulated the F0 contour of normal speech by multiplying the contour by a modulation factor (*m*). They showed that normal hearing listeners have the same speech recognition performance (SNR is about -4 dB for a 50% correct score) for no modulation ($m = 0$), a reduced modulation ($0 < m < 1$) and a normal modulation ($m = 1$) in speech-shaped noise. The performance drops by another 1.3 dB when the contour is inverted ($m = -1$); which corresponds to giving incorrect information. We have performed a similar experiment for the acoustic tone in a simulated EAS setting. The average scores are a few dB higher because of the CI simulated speech. Our results show that the performance was robust against a change of the modulation index *m* because the performance did not change significantly between $m = 0$, 0.4, 0.7, 1 but, interestingly, this was also true for 1.5 (an increase) and -1 (an inverted contour). The latter observation does not agree with the observation of Binns and Culling (2007) for normal hearing listeners. However, a different picture emerges if the performance for the different modulation factors is compared with the performance of the CI only condition, because only the correct F0 contour and a reduction of the contour ($0 < m \leq 1$) give a significant benefit over the CI only condition. This is in line with the observation of several authors, and discussed in the previous section, that the frequency variation ($m \neq 0$) is a crucial cue of the fundamental frequency for speech recognition in noise. Despite the smaller range of possible modulation factors, the performance is still quite robust against changes of the modulation index. Consequently, performance can be considered robust against a reduction of the frequency selectivity so long as some frequency variation over time remains.

We also undertook a different approach to assessment of frequency selectivity by replacing the pure tone with narrowband noise whose centre frequency was determined by the instantaneous F0 value. The aim of using this simulation condition is that this mimics the widening of the tuning curves and a reduction of frequency specificity by smearing the tone in the frequency domain. Our results show no significant difference in performance between the pure tone and the two narrowband noise signals (which differ in bandwidth of the narrowband noise). This indicates that the auditory system is also robust against changes in frequency specificity when considering speech recognition in noise, and it suggests that centre frequency, rather than bandwidth, of the F0 signal is critical to EAS benefit, at least in simulation conditions. Interestingly, this mirrors findings of Verschuur et al. (2013) which show that simulated EAS benefit from the first formant signal does not depend on bandwidth, but can be obtained so long as centre frequency is available.

The bandwidths used in our experiment were around 50 Hz and 75 Hz. These values are only slightly larger than the Equivalent Rectangular Bandwidth (see Moore, 2012) of about 30-50 Hz for frequencies below 250 Hz, so one filter can be considered to have a maximal stimulation. Difficulties arise for higher values of the bandwidth in our approach because the fundamental frequency varies substantially. The lowest value for F0 which has been found in our data set is 50 Hz and this value results in a cut-off frequency (-3 dB point) of about 15 Hz for the widest narrowband noise condition; this value is quite close to the lowest possible frequency (0 Hz). Normal hearing listeners can successfully judge that there are two pure tones present if the frequency difference between the tones is 0.5-1 Hz (Moore, 2012) for frequencies below 500 Hz. It is likely that these signals stimulated a large range of auditory filters at each time instant, such as a pure tone would do if the auditory filters had been widened.

C. No evidence for benefit of tone downshifted in frequency

Brown et al. (2010) studied speech recognition in an EAS setting by using a pure tone which follows the fundamental frequency as an acoustic component. They showed that the tone could be downshifted in frequency by a value up to 100 Hz with little impact on the performance. This was a promising observation because it would make it possible to present useful information about the target's fundamental frequency to individuals with very limited steep sloping residual hearing (e.g. only acceptable aided hearing thresholds below 200 Hz). Our results do not support their observation. The performance of the frequency shifted tone was worse than the unshifted tone by 1.9 dB, although this comparison was not statistically significant. There was also no significant improvement over the vocoder only signal (the difference was only about 0.25 dB). Furthermore the performance of three distorted versions of the original F0-tone was significantly better (> 2.6 dB) than the shifted tone. Consequently, it seems that not only the perception of frequency variations of the fundamental frequency is important but also that the mean value of this frequency must be correct in order to obtain a significant benefit.

V. CONCLUSIONS

Speech recognition in a multi-talker babble background has been studied in a simulated EAS setting with normal hearing listeners. The present study was focused on the role of the fundamental frequency, in particular its variations in frequency, and the impact on the performance for a reduction in frequency selectivity. We found that speech recognition scores improved significantly when an additional acoustical tone with a fixed amplitude and frequency equal to the target's fundamental frequency was added to the CI simulation. Moreover, similar speech recognition scores were also obtained for some changes in the pure tone which mimic a reduction in frequency selectivity. This is true if these changes corresponded to either a decrease in frequency modulation (resulting in a reduction of the frequency range), or the use of narrowband noise instead of a pure tone (resulting in a reduction of frequency specificity). This finding suggests that the auditory system has a certain degree of robustness in its mechanism to use the information about instantaneous variations in the fundamental frequency of the target speech. However, we also found that the mean F0 value must be correct to have the benefit of presenting an additional pure tone in the EAS setting, as simulated EAS benefit was lost with down-shifting of the F0 value. This implies that, if results are replicated in EAS users, it would not be beneficial to present the variations of the fundamental frequency at a lower frequency, e.g. via frequency compression, for those with very limited residual hearing (i.e. only residual hearing below 200 Hz for example).

REFERENCES

Binns, C. and Culling, J.F. (**2007**) The role of fundamental frequency contours in the perception of speech against interfering speech. J. Acoust. Soc. Am. **122**(3), 1765-1776.

Bench, J., Kowal, A. and Bamford, J. (**1979**) The BKB (Bamford-Kowal-Bench) sentences lists for partially-hearing children. Br. J. Audiol **13**(3), 108-112.

Brown, C.A. and Bacon, S.P. (**2009a**) Low-frequency speech cues and simulated electric-acoustic hearing. J. Acoust. Soc. Am. **125**(3), 1658-1665.

Brown, C.A. and Bacon, S.P. (**2009b**) Achieving electric-acoustic benefit with a modulated tone. Ear Hear. **30**(5), 489-493.

Brown, C.A. and Bacon, S.P. (**2010**) Fundamental frequency and speech intelligibility in background noise. Hear. Res. **266**(1), 52-59.

Brown, C.A., Scherrer, N.M and Bacon, S.P. (**2010**) Shifting fundamental frequency in simulated electric-acoustic listening. J. Acoust. Soc. Am. **128**(3), 1272-1279.

Carlyon, R.P. (**2004**) How the brain separates sounds. Trends Cogn. Sci. **8**(10), 465-471.

Carroll, J., Tiaden, S. and Zeng, F.-G. (**2011**) Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing. J. Acoust. Soc. Am. **130**(4), 2054-2062.

Chang, J.E., Bai, J.Y. and Zeng, F.-G. (**2006**) Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise. IEEE Trans. Biomed. Eng. **53**(12), 2598-2601.

Cullington, H.E. and Zeng, F.-G. (**2010**) Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear. Ear Hear. **31**(1), 70-73.

Field, A. (**2009**) *Discovering Statistics using SPSS*. 3rd edition, SAGE Publications, Londen.

Greenwood, D.D. (**1990**) A cochlear frequency position function for several species – 29 years later. J. Acoust. Soc. Am. **87**(6), 2592-2605.

Kong, Y.-Y. and Carlyon, R.P. (**2006**) Acoustic cues for improved speech recognition in combined acoustic and electric hearing. J. Acoust. Soc. Am. **119**(5), 3238-3238.

Kong, Y.-Y. and Carlyon, R.P. (**2007**) Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. J. Acoust. Soc. Am. **121**(6) 3717-3727.

Li, N. and Loizou, P.C. (**2008**) A glimpsing account for the benefit of simulated combined acoustic and electric hearing. J. Acoust. Soc. Am. **123**(4), 2287-2294.

Moore, B.C.J. (**2012**) *An Introduction to the Psychology of Hearing*. 6th edition, Emerald, Bingley.

Turner, C.W., Gantz, B.J., Vidal, C., Behrens, A. and Henry, B.A. (**2004**) Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing. J. Acoust. Soc. Am. **115**(4), 1729-1735.

Turner, C.W., Reiss, L.A.J. and Gantz, B.J. (**2008**) Combined acoustic and electric hearing: preserving residual acoustic hearing. Hear. Res. **242**(1): 164-171.

Verschuur, C., Boland, C. and Frost, E. (**2013**) The role of first formant information in simulated electro-acoustic hearing. J. Acoust. Soc. Am. **133**(6), 4279-4289.

Wilson, B.S. and Dorman, M.F. (**2008a**) Cochlear implants: A remarkable past and a brilliant future. Hear. Res. **242**(1): 3-21.

Wilson, B.S. and Dorman, M.F. (**2008b**) Cochlear implants: Current designs and future possibilities. J. Rehabil. Res. Dev. **45**(5), 695-730.

Zahorian, S.A. and Hu, H. (**2008**) A spectral/temporal method for robust fundamental frequency tracking. J. Acoust. Soc. Am. **123**(6), 4559-4571.

Zhang, T., Dorman, M.F. and Spahr, A. (**2010**) Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. Ear Hear. **31**(1), 63-69.

TABLE I. Eleven listening conditions used in the experiment with abbreviation and description.

Condition	Description	Abbreviation
I	Vocoder signal alone	V
II	Vocoder signal and low pass filtered signal	V+LPF
III	Vocoder signal and the pure FM tone	V+F0
IV	Vocoder signal and the ‘mildly smeared’ FM tone	V+NB_small
V	Vocoder signal and the ‘seriously smeared’ FM tone	V+NB_wide
VI	Vocoder signal and FM tone with large modulation depth	V+mod_07
VII	Vocoder signal and FM tone with small modulation depth	V+mod_04
VIII	Vocoder signal and FM tone with extended modulation depth	V+mod_15
IX	Vocoder signal and FM tone with inverted modulation	V+mod_inv
X	Vocoder signal with tone indicating voicing	V+voice
XI	Vocoder signal and FM tone with shifted mean value	V+shifted

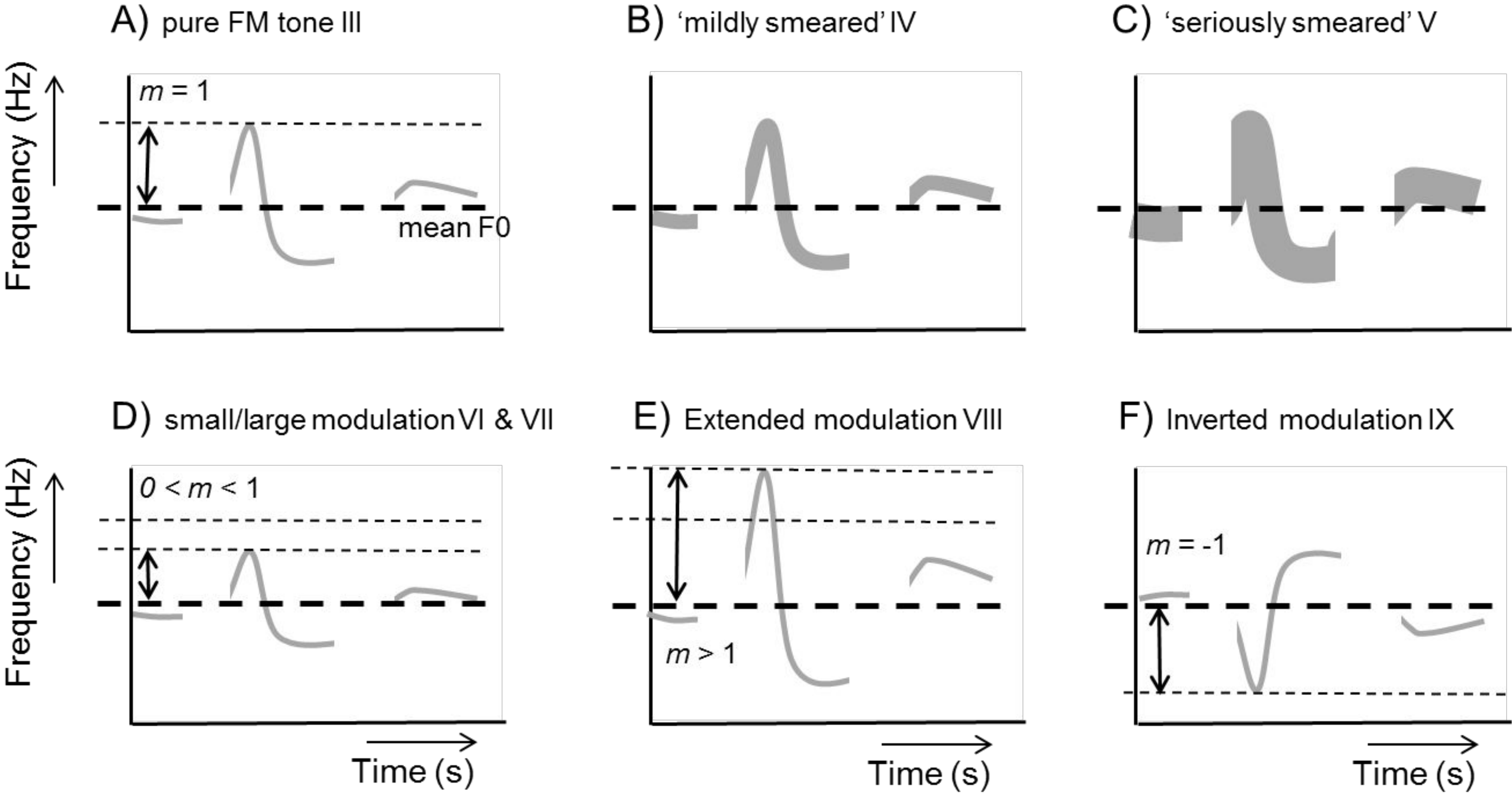
TABLE II. Overview of post hoc test: pairwise comparisons of the conditions. The value in each cell is the mean value for the horizontal condition minus the mean score of the vertical condition. The cell is grey if the difference is significant ($p < 0.05$). The 2nd and 3rd columns give the mean score (M) and standard deviation (SD) for each condition.

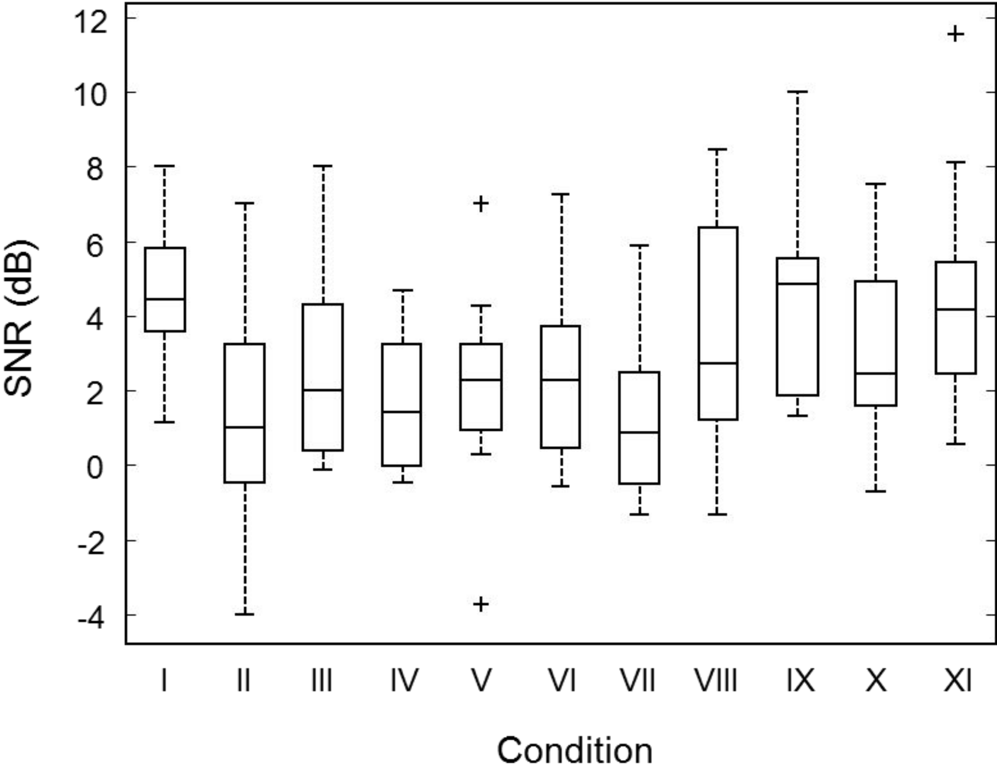
	M	SD	Condition										
Condition			I	II	III	IV	V	VI	VII	VIII	IX	X	XI
I: V	4.66	1.97		3.47	2.15	2.88	2.57	2.42	3.26	1.43	0.11	1.60	0.25
II: V+LPF	1.19	3.07	-3.47		-1.32	-0.59	-0.90	-1.05	-0.21	-2.04	-3.36	-1.87	-3.22
III: V+F0	2.51	2.48	-2.15	1.32		0.73	0.42	0.26	1.11	-0.72	-2.05	-0.55	-1.90
IV: V+NB_small	1.78	1.79	-2.88	0.59	-0.73		-0.31	-0.46	0.39	-1.45	-2.77	-1.28	-2.63
V: V+NB_wide	2.09	2.48	-2.57	0.90	-0.42	0.31		-0.15	0.69	-1.14	-2.46	-0.97	-2.32
VI: V+mod_07	2.24	2.29	-2.42	1.05	-0.26	0.46	0.15		0.85	-0.99	-2.31	-0.81	-2.16
VII: V+mod_04	1.40	2.11	-3.26	0.21	-1.11	-0.39	-0.69	-0.85		-1.84	-3.15	-1.66	-3.01
VIII: V+mod_15	3.23	3.16	-1.43	2.04	0.72	1.45	1.14	0.99	1.84		-1.32	0.17	-1.18
IX: V+mod_inv	4.55	2.67	-0.11	3.36	2.04	2.77	2.46	2.31	3.15	1.32		1.49	0.14
X: V+voice	3.06	2.70	-1.60	1.87	0.55	1.28	0.97	0.81	1.66	-0.17	-1.49		-1.35
XI: V+shifted	4.41	3.07	-0.25	3.22	1.90	2.63	2.32	2.16	3.01	1.18	-0.14	1.35	

Figure headings

FIG. 1. Schematic illustrative spectrogram of six different versions of the FM tone used in the experiment: (A) a pure tone following the fundamental frequency F0, (B,C) ‘smeared’ versions of the tone by using narrowband noise, (D-F) as (A) but with a variation in modulation depth *m*. The amplitude of all signals is set to zero for voiceless speech, as illustrated by the gaps in the spectrogram.

FIG. 2. Boxplot of the final SNR for the 11 conditions, see Table I for an explanation of the conditions.


A) pure FM tone III
B) 'mildly smeared' IV
C) 'seriously smeared' V
Frequency (Hz)
m = 1
mean F0
D) small/large modulation VI & VII
E) Extended modulation VIII
F) Inverted modulation IX
Frequency (Hz)
0 < m < 1
m > 1
m = -1
Time (s)
Time (s)
Time (s)


SNR (dB)
12
10
8
6
4
2
0
-2
-4
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
Condition