Modeling the Acoustic-to-Auditory Transformation for Stop Consonant-Vowel Syllables

R. I. Damper and S. R. Gunn
Image, Speech and Intelligent Systems Research Group
Department of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ
UK

Abstract submitted to the 5th International Conference on Cognitive and Neural Systems,
Boston, MA, May 30–June 2, 2001.

Corresponding author:	R. I. Damper
Presenting author:	R. I. Damper
Emails:	{rid|srg}@ecs.soton.ac.uk
Telephone:	+44 23 80 594577
FAX:	+44 23 80 594498
First choice:	speech and language (B)
Second choice:	audition (B)

Modeling the Acoustic-to-Auditory Transformation for Stop Consonant-Vowel Syllables

R. I. Damper and S. R. Gunn
Image, Speech and Intelligent Systems Research Group
Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, UK
Emails: {rid|srg}@ecs.soton.ac.uk

A key question for the study of spoken language is: How is the *continuous* sound pressure wave of speech decoded by a listener into the *discrete* percept of a particulate linguistic message? It is well known that this decoding exhibits a high degree of invariance to the acoustic effects of speaker, phonetic context etc. Significant efforts have been, and still are, aimed at determining the mechanisms of sound-to-symbol decoding, with particular emphasis on stop consonant perception where the degree of invariance is especially marked. Unsurprisingly, the traditional research method employed in this field has been human perception studies. Less usually, signal analysis (e.g., Kewley-Port 1982; Soli 1983) and animal psychophysics (e.g., Kuhl and Miller 1978; Kluender, Diehl, and Killeen 1987; Sinnott and Williamson 1999) have been used. However, computational modeling offers a valuable, complementary, yet under-exploited approach to the study of speech perception.

In previous work over many years (Damper, Pont, and Elenius 1990; Damper 1998; Damper, Gunn, and Gore 2000; Damper and Harnad 2000), we have studied extensively the emergence of phonetic categories in a variety of computational models in which a biologically-inspired 'front end' produces an auditory time-frequency representation which acts as input to a connectionist 'back end' learning system. Past work has concentrated on modeling the perception of the voicing distinction in the particular synthetic stop consonant-vowel (CV) stimuli of Lisker and Abramson (1970). This has shown that realistic modeling of the movement of phoneme boundary with place of articulation (which is a characteristic of these stimuli) is dependent on realistic modeling of auditory preprocessing, confirmed the relative importance of formant transition information over onset information, suggested that the 'low-frequency property' (Stevens and Blumstein 1981) is primary and (together with the animal psychophysics work) comprehensively denied the necessity to invoke 'special mechanisms' to explain human speech perception.

Major limitations of the work so far are the restriction to the voicing distinction and the sole use of the Lisker and Abramson stimuli. In this paper, we will report on modeling the place distinction in CV stimuli. Not all synthetic CV series show the same boundary shift with place of articulation (e.g., Kluender 1991) and synthetic stimuli do not capture the context variability of real speech. Hence, we will also report on modeling work with the real speech data of Nossair and Zahorian (1991).

References

Damper, R. I. (1998). Auditory representations of speech sounds in a neural model: The role of peripheral processing. In *Proceedings of International Joint Conference on Neural Networks (IJCNN 98)*, Anchorage, AL, pp. 2196–2201.

Damper, R. I., S. R. Gunn, and M. O. Gore (2000). Extracting phonetic knowledge from learning systems: Perceptrons, support vector machines and linear discriminants. *Applied Intelligence 12*, 43–62.

Damper, R. I. and S. R. Harnad (2000). Neural network models of categorical perception. *Perception and Psychophysics 62*(4), 843–867.

Damper, R. I., M. J. Pont, and K. Elenius (1990). Representation of initial stop consonants in a computational model of the dorsal cochlear nucleus. Technical Report STL-QPSR 4/90, Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology (KTH), Stockholm. Also published in W. A. Ainsworth (Ed.), *Advances in Speech, Hearing and Language Processing, Vol. 3 (Part B)*, pp. 497–546. Greenwich, CT: JAI Press, 1996.

Kewley-Port, D. (1982). Measurements of formant transitions in naturally-produced stop consonant-vowel syllables. *Journal of the Acoustical Society of America 72*, 379–389.

Kluender, K. R. (1991). Effects of first formant onset properties on voicing judgements resulting from processes not specific to humans. *Journal of the Acoustical Society of America 90*, 83–96.

Kluender, K. R., R. L. Diehl, and P. R. Killeen (1987). Japanese quail can learn phonetic categories. *Science 237*, 1195–1197.

Kuhl, P. K. and J. D. Miller (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America 63*, 905–917.

Lisker, L. and A. Abramson (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of 6th International Congress of Phonetic Sciences, Prague, 1967*, pp. 563–567. Academia, Prague.

Nossair, Z. B. and S. A. Zahorian (1991). Dynamic spectral shape features as acoustic correlates for initial stop consonants. *Journal of the Acoustical Society of America 89*, 2978–2991.

Sinnott, J. M. and T. L. Williamson (1999). Can macaques perceive place of articulation from formant transition information? *Journal of the Acoustical Society of America 106*(2), 929–937.

Soli, S. D. (1983). The role of spectral cues in the discrimination of voice-onset time differences. *Journal of the Acoustical Society of America 73*(6), 2150 – 2165.

Stevens, K. N. and S. E. Blumstein (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas and J. L. Miller (Eds.), *Perspectives on the Study of Speech*, pp. 1–38. Hillsdale, NJ: Lawrence Erlbaum Associates.