CUE INTERACTION IN THE PERCEPTION OF INTERVOCALIC AND SYLLABLE-INITIAL VOICELESS FRICATIVE/AFFRICATE CONTRASTS

A. Faulkner, S Rosen[1], AM Darling, and M Huckvale
Department of Phonetics and Linguistics, University College London. and
Department of Audiology, North Western University, Evanston, Illinois[1]

ABSTRACT

The interaction of frication duration, frication rise-time and pre-frication silence interval has been studied for the English voiceless fricative/affricate contrast. Cue interaction is a pervasive feature of the results, and major differences were found between syllable-initial and intervocalic stimuli. These interactions prevent the data from being fitted by models based on either acoustic, auditory, and information integration principles. The interactions are ascribed to cognitive processes.

BACKGROUND

One of the major theoretical problems in speech perception arises from the existence of substantial trading relations or interactions between multiple acoustic cues to the identity of speech sounds. Traditional accounts of cue trading appeal to articulatory representations of speech [e.g., 1]. Others [e.g., 2] have approached cue trading at the level of information integration. We have been concerned with the contribution that auditory transformations of the acoustic speech signal may make in speech perception.

Amongst known auditory transformations are a number that may in themselves account for cue trading. For example, Delgutte [3] has proposed that the rapid adaptation that occurs in hair-cell mechanical to neural transduction results in an interaction in the neural coding of frication noise between frication rise time and the duration of the silence that precedes frication. This could account for perceptual interactions in the perception of the voiceless fricative/affricate contrast [4]. Other properties of the peripheral auditory system, such as the frequency dependence of the temporal resolution and group delay of cochlear filtering, may account for further interactions. For example, the effect of temporal voice onset time cues in the plosive voicing contrast depends on place of articulation. Since the frequency region in which temporal voice-onset time information is present depends on place of articulation, the auditory frequency channels by which this temporal information is processed will also depend on place of articulation. In consequence, the properties of different auditory frequency channels could influence the processing of such temporal information.

This study concerns the voiceless fricative/affricate contrast. The perception of the contrast between the voiceless fricative /ʃ/ and the voiceless affricate /tʃ/ (the initial consonants in the words "ship" and "chip") is generally supposed to be based on at least two perceptual cues. Both cues are associated with acoustic differences that are typical of natural speech tokens. One is the duration of the noise-excited frication part of the consonant. In /ʃ/ this is relatively long, whilst in /tʃ/ it is shorter. The second cue arises from the amplitude envelope of the frication, which has a gradual onset for /ʃ/, but a more rapid onset for /tʃ/, where there is often also a distinct brief initial burst of frication. When these consonants are preceded by a vowel, the duration of silence between the vowel and the onset of frication has been identified as a third cue; the silent interval is typically absent before /ʃ/ but present before /tʃ/.

While there has been debate about the relative importance of duration and time-amplitude envelope cues [5], there is clear evidence that these cues show trading relations [4,6].

EXPERIMENTAL STUDY

Existing published data are limited, and come from studies which differ in many respects. In order to model the interactions between these cues, it was necessary to collect a substantial body of empirical data from human listeners. The stimuli for the experiment described here were all based on a natural /ɑʃɑ/ token produced by a female British English talker. Both syllable-initial (/ʃɑ/ /tʃɑ/) and intervocalic (/ɑʃɑ/ /ɑtʃɑ/) contexts were investigated. The stimuli represented factorial combinations of total frication duration (120 to 220 ms), frication rise time (0 to 100 ms), and in the intervocalic case, the duration of the silent interval between the initial vowel and the onset of frication (0 to 80 ms). Stimulus manipulations were performed by digitally modifying the duration and amplitude of the frication noise from the natural /ɑʃɑ/ model and adding silence where necessary.

Nine listeners responded to a total of 6 repetitions of each of 193 stimuli. They were asked to label each stimulus as containing either of the consonants "sh" or "ch".

Major differences were found between the syllable-initial and intervocalic stimuli, which have not been compared in previous studies. A full report of the data appears in [7]. Selected data are shown in fig. 1. A logistic regression was used to establish significant main effects and cue interactions. The most striking interaction was that between rise time and the presence or absence of a preceding vowel. Shorter rise times increased the frequency of affricate labelling for a syllable-initial consonant (fig 1, panel d), while this effect was absent in the intervocalic case and indeed reversed where the silent interval between the initial vowel and consonant was short (fig 1, panels a and b). The data show other, more expected, properties of fricative/affricate perception in that shorter silence duration and longer frication duration both led to increased frequency of fricative labelling. The most pervasive feature of the data is that of interaction between cues. Interaction was found between frication duration and rise

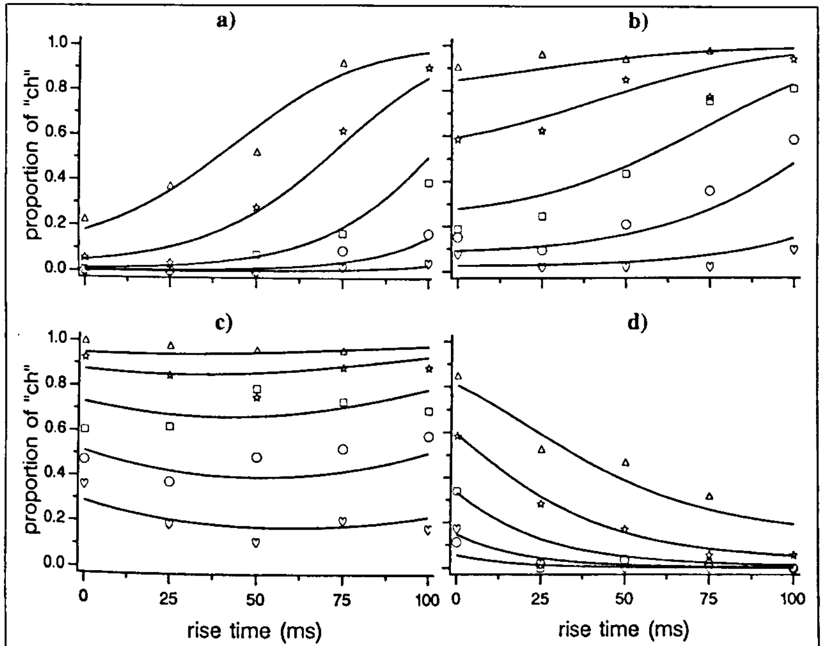

Figure 1. Proportion of affricate ("ch") responses as a function of frication rise time. Data are shown for the intervocalic stimuli with pre-frication silence durations of 0, 20, and 60 ms (panels a, b and c), and for the syllable-initial stimuli (panel d). The curves shown in the figure result from a logistic regression. Each line represents results from a particular frication duration, with values of 120, 145, 170, 195 and 220 ms, from top to bottom.

time, and between frication duration and the presence of a preceding vowel. Where the initial vowel was present, frication duration and rise time also interacted with silence duration.

MODELLING

An Auditory Model

To examine Delgutte's claim that rapid adaptation can account for interactions between silence duration and frication rise-time, an auditory model was developed. This made use of a gammatone filterbank [8] and a hair-cell model [9] to simulate peripheral auditory processing. A simplified model was used here which included only one auditory filter with a centre frequency of 3000 Hz. The output of this model was the probability of hair-cell firing as a function of time. Probability of firing is equivalent to the relative firing rate over the ensemble of auditory fibres driven by an auditory filter. The decision statistic was the peak hair-cell firing probability close to the onset of frication, relative to the subsequent probability of firing to the quasi-steady state frication. The greater the relative peak height, the stronger is the presumed evidence for an affricate consonant. A model of this sort has not previously been examined with truly speech-like inputs, and we were able to verify that the rapid adaptation exhibited by the hair-cell simulation does predict an interaction between silence duration and frication rise time. Longer silences following a vowel allow the hair-cell to recover from its response to the preceding vowel, and hence to exhibit a greater probability of firing at the onset of frication. However, the predicted interactions do not correspond to those observed. In particular, in the intervocalic context ,we observed that longer rise time leads to increased frequency of affricate responses at shorter silence durations (fig 1, panels a and b), where the model predicts the reverse effect of rise time.

An acoustic model based on rate of rise of frication

An acoustic model due to Weigelt et al. [10] was also examined. Here, the decision statistic was the peak rate of rise of the physical intensity of frication, where a high peak rate is assumed to be associated with affricate responses. Because the measurement of rate of rise necessarily entails the integration of information over time, the extracted rate of rise is affected not only by the rise time, but also by the silence interval. The rate of rise parameter is also affected by frication duration, but only for very short rise times (less than 25 ms). Hence, this model, like the auditory model above, can predict interactions between cues. However, the predictions do not account for the empirical data. As with the auditory model above, the rate of rise model cannot account for the finding that longer rise times in the intervocalic case lead to increased frequency of affricate responses.

It would be possible to consider an auditory model in which the decision statistic proposed by Weigelt et al. was applied to auditory firing probability rather than to the acoustic signal. It is difficult to imagine, however, that such a model would be able to account for our data.

FLMP model

While Massaro's fuzzy logical FLMP model [2] makes no assumptions about the relationship between acoustic or auditory parameters and the perceptual classification of speech sounds, it does claim to account for the integration of information from several sources and as such, can be fitted to data such as these. For our data, the FLMP model fails because it incorporates the assumption that features do not interact in the statistical sense, an assumption clearly contradicted by our data.

CONCLUSIONS

None of the models examined here can provide a satisfactory account of affricate/fricative perception. An articulatory account also appears untenable because the observed perceptual interactions in our intervocalic condition are not consistent with the interdependence of the corresponding acoustic properties found in speech production. Again, the problematic finding is that a shorter frication rise time increases the frequency of fricative labelling at the shorter silence durations. A shortening of frication noise rise times associated with fricative as opposed to affricate productions has never been observed in speech production [6,11]. Our perceptual results are not likely to arise from an artefact of the particular stimuli, as we have since replicated the same finding using different stimuli based on manipulations of a natural /ɑtʃɑ/ token.

A full account of the cue interactions in this consonant contrast cannot yet be provided. Auditory transformations in the time-domain do not (and probably cannot) account for all aspects of the data, although they may well have an important role. We have reached the same conclusion for the place of articulation dependence of the plosive voicing contrast. A related study [12], led to the conclusion that auditory processing cannot account for the effects of place of articulation on the interpretation of voice onset time. Our results appear to refute the claim made by Damper et al. [13] that properties of auditory frequency analysis play a key role in this phenomenon.

It may be necessary to consider cognitive rather than electro-physiological notions of auditory attributes relating for example to subjective duration and suddenness of onset. Further, the surrounding speech context of the consonant exhibits profound interactions with other cues. It may, for this reason, be necessary to take account of the phonetic or other linguistic properties of the context in order to fully understand these processes.

ACKNOWLEDGEMENT

Supported by JCI Cognitive Science/HCI grant SPG8920412.

REFERENCES

[1] Repp, B. H., Liberman, A. L., Eccardt, T., and Pesetsky, D (1978) Perceptual integration of acoustic cues for stop, fricative and affricate manner. J. Exp. Psychol., Human Percept. and Perf., 4, 621-637.

[2] Massaro, D.W. (1987) *Speech Perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, New Jersey: Lawrence Earlbaum Associates

[3] Delgutte, B. (1982) Some correlates of phonetic distinctions at the level of the auditory nerve. In: *The representation of speech in the peripheral auditory system,* eds. R. Carlson and B. Granstrom. Amsterdam: Elsevier, pp 131-149

[4] Dorman, M.F., Raphael, L.J. and Isenberg, D. (1980) Acoustic cues for a fricative-affricate contrast in word-final position. J. Phonetics, 8, 397-405

[5] Kluender, K.R. and Walsh, M.A. (1992) Amplitude rise time and the perception of the voiceless affricate/fricative distinction. Percept. Psychophys., 51, 328-333

[6] Howell, P., and Rosen, S. (1982) Production and perception of rise time in the voiceless affricate/fricative distinction. J. Acoust. Soc. Am, 73, 976-1984.

[7] Rosen, S., Darling, A.M., Faulkner, A. and Huckvale, M. (1993) Cue interaction in an intervocalic voiceless affricate/fricative contrast. Speech Hearing and Language, Work in Progress, Dept. Phonetics and Linguistics, UCL. Vol 7, 183-197.

[8] Darling, A.M. (1991) Implementing a Gammatone filter. Speech Hearing and Language, Work in Progress, Dept. Phonetics and Linguistics, UCL, Vol 5, 41-62.

[9] Meddis R (1988) Simulation of auditory-neural transduction: Further studies. J. Acoust. Soc. Am., 83, 1056-1063

[10] Weigelt, L.F., Sadoff, S.J. and Miller, J.D. (1990) Plosive/fricative distinction: the voiceless case. J. Acoust. Soc. Am., 87, 2729-2737.

[11] Howell, P., and Rosen, S. (1983) Closure and frication measurements and perceptual integration of temporal cues for the voiceless affricate/fricative contrast. Speech Hearing and Language, Work in Progress, Dept. Phonetics and Linguistics, UCL. Vol. 1, 109-117.

[12] Darling, A.M., Huckvale, M.A., Rosen, S., and Faulkner, A. (1992) " Phonetic classification of the plosive voicing contrast using computational modelling" In: *Speech and Hearing; Proc. Inst. Acoust. 1992 Conference,* Vol 14., part 6.

[13] Damper, R., Pont, M., and Elenius, K (1991) Representation of initial stop consonants in a computational model of the dorsal cochlear nucleus. STL-QPSR, 4, 7-41.