THE RELATIVE CONTRIBUTION OF RISE TIME, STEADY TIME, AND OVERALL DURATION OF NOISE BURSTS TO THE AFFRICATE-FRICATIVE DISTINCTION IN ENGLISH: A RE-ANALYSIS OF OLD DATA

Vincent J. van Heuven, Phonetics Laboratory, Dept. of Linguistics, University of California at Los Angeles, Los Angeles, CA 90024/ Fonetisch Laboratorium, Vakgroep Algemene Taalwetenschap, Rijksuniversiteit Leiden, Leiden, The Netherlands.

The results of an unpublished study by Gerstman (1957) indicate that variations of the rise time of a noise burst provide a more powerful cue to the English affricate-fricative distinction than similar variations in steady state duration of the noise burst. Since this finding is counter to the general sensitivity of the human ear to such temporal phenomena (indicating that it is more sensitive to steady state duration differences), I propose a re-analysis of Gerstman's data. As both rise time and steady time increments add to the overall duration, I have factored out this shared property. It then appears that overall duration explains about 75% of the response variance, whereas rise time independently only accounts for an additional 3%.

It is my working hypothesis that the increment of a particular acoustic parameter necessary to create a cross-over between phonemic categories, is directly related to the magnitude of the Just Noticeable Difference (JND) for that parameter as established in non-speech psycho-acoustic experiments.

There are at least three acoustic parameters that are important for the affricate-fricative distinction, one of which, the duration of a pre-burst silent interval, is relevant only if the sibilant is not in utterance initial position. For initial bursts rise time (the duration of the portion of the noise during which the amplitude envelope generally increases) and steady time (the duration of the portion during which the amplitude remains relatively constant) have been claimed relevant cues.

Psycho-physical experiments have revealed that differences in steady time are perceived more accurately than differences in rise time. Van Heuven & Van den Broecke (1978a, b) found JND's for noise bursts with standard overall durations of 250 or 500 msec and ranging in rise time from 0 to 100 msec, to be on the order of 25% of the reference rise time. Comparing this with JND's for steady duration, obtained within the same experimental paradigm (method of adjustment), consistently smaller values are found, typically less than 10% of the reference duration (Zwicker, 1970).

Though a few studies have been published in which the role of rise time as a determinant of the release feature was investigated (Cutting & Rosner, 1974; Dorman *et al.*, 1978), these do not allow us to determine the exact nature of the psychometric curve. However, as early as 1957 Gerstman completed a dissertation in which the perceptual effects of rise time and steady time of noise bursts were investigated in a series of experiments, culminating in his experiment IV, in which a number of deficiencies that occurred in the earlier experiments were largely eliminated through the use of more sophisticated methodology and instrumentation. Gerstman (1957:86) concludes from this study that "rise time principally distinguishes fricatives from affricates (...). However, for a middle range of rise times, steady time is also a cue."

This apparent superiority of rise time over steady time does not reflect the general sensitivity of the human hearing mechanism to these acoustic phenomena. In an attempt to resolve this incompatibility of speech perception and general auditory behavior, I propose to re-analyse Gerstman's data. As his study has unfortunately never been published, I shall summarise his crucial experiment here, and then proceed with the re-analysis.

Method

Using the Voback (Borst & Cooper, 1957), a pattern playback also capable of generating non-periodic signals, a series of 97 different syllables were synthesised, each of which consisted of an [ɑ]-like vowel with a duration of 300 msec, preceded by a burst of filtered noise (2200 - 5000 Hz, 18 dB/oct roll off below 2200 Hz), separated from the vowel by a 10 msec silent interval. The intensity envelope of the stimulus could be controlled by the Amplituder (Borst & Cooper, 1957) by varying the width of a painted track running parallel to the stylised spectrogram that served as

input to the Voback. Rise times (the time needed for the intensity of the noise to reach its maximum level of 27 dB *re* background as a linear function of time) were varied between 10 and 100 msec, and combined with a range of steady state durations (where the intensity was maintained at a constant -3 dB *re* vowel intensity). The exact sampling of the stimulus space will be apparent from figure 1. Each stimulus was recorded twice in rapid succession with a 1500 msec separation between the members of a pair. The pairs were randomly ordered in blocks of 10, with 6 sec intervals between pairs, and 10 sec intervals between blocks.

48 native (American) English paid volunteers with previous experience in speech perception experiments at the University of Connecticut listened to the stimuli played through a loudspeaker in an acoustically treated room. They were instructed to assign each stimulus(pair) to one of five response categories in a forced choice paradigm: [ʒ, ʃ, dʒ, tʃ or pʃ].

Results

Figure 1 contains Gerstman's summary graph (his figure VI-4), specifying the stimulus space sampling, as well as boundary lines that divide the space into five areas corresponding to the response categories, according to a 50% identification criterion. The percentages in the graph refer to the maximum percentage of responses favoring that particular category.

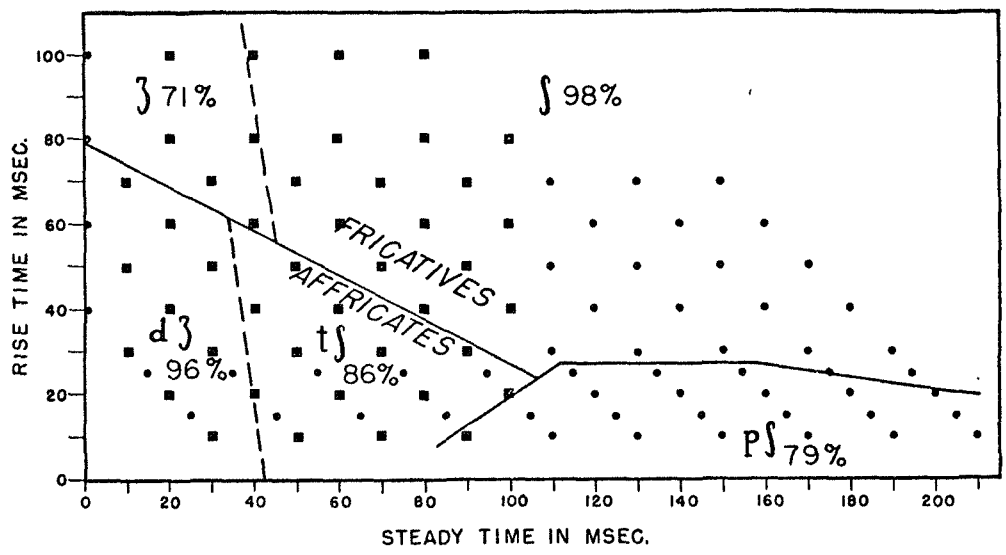

Figure 1: Summary of results of Gerstman (1957:experiment IV): classification of 97 stimuli on a 50% identification criterion as a function of rise and steady time. The squares represent the 43 sample points selected for further analysis in this paper.

The slope of the affricate-fricative boundary in figure 1 indicates that rise time is indeed a more effective cue than steady time. However, the deviation of the boundary line from a 45 degree angle is still moderate, which primarily points to additivity of the component durations, suggesting that overall duration (i.e. the sum of rise and steady time) could be the overriding cue. This is, in fact, a suggestion put forward by Gerstman (1957:78) himself in a discussion of the results of one of his preliminary experiments, but which was not followed up in this instance.

Controlling for overall duration

A straightforward way to ascertain whether rise and steady time have any role in signalling the affricate-fricative distinction other than merely adding to the overall duration of the noise burst, is to vary rise and steady time in negative proportion (i.e. one at the expense of the other), while keeping their sum constant. Let us assume that an increase in rise time leads to a preponderance of fricative judgements, and that the co-occurring decrement in steady time would result in a larger number of affricate judgements. If the two component durations are purely additive, then a shift in the rise/steady time ratio within a set of identical overall noise durations cannot result in a cross-over from fricative to affricate.

In order to have the ranges and sampling densities of rise and steady times as comparable as possible, steady times exceeding 100 msec and rise times at values other than whole multiples of 10 msec were excluded from further analysis, leaving a subset of 43 data points marked by squares in figure 1.

In figure 2 the percentage of fricative responses is plotted as a function of the rise/steady time division separated out for each of the 8 different overall noise burst durations.

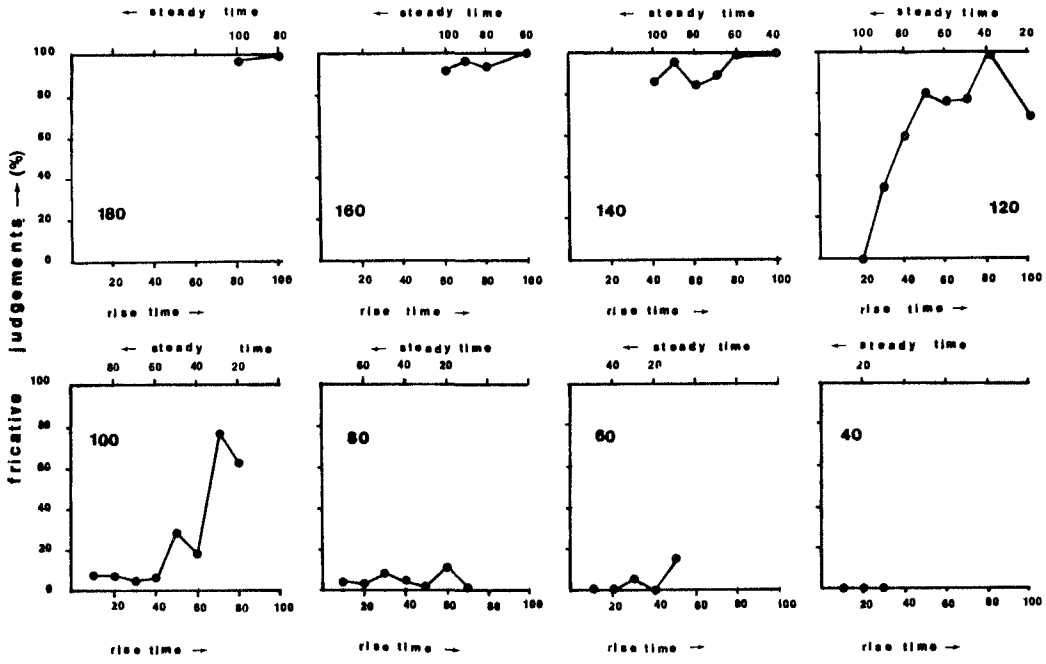

Figure 2: % perceived fricatives as a function of overall duration of the noise burst and the rise/steady time division. All durations are in msec.

A complete cross-over from affricate to fricative is effected when going from an overall noise duration of 80 to 140 msec. There are no differential effects due to a shift in the rise/steady time proportion for overall durations above 130 and below 90 msec. However, within the set of overall durations of 100 msec a not very convincing cross-over is reached by changing the rise/steady time division from 60/40 to 70/30 msec. A full cross-over is obtained for overall durations of 120 msec when changing the ratio from 20/100 to 80/40 msec.

Apparently rise time provides a useful cue only in the restricted range of overall durations between 90 and 130 msec, i.e. when overall duration by itself is an ambiguous cue. Overall noise durations of spoken sibilants are by no means limited to this range, as is evidenced by acoustic measurements of 192 spoken CV-syllables (Gerstman, 1957:figure VII-1), where C stands for [ʒ, ʃ, dʒ, tʃ] and V for [e, a, o]. In these data overall noise durations range from 30 to 220 msec, and the middle part of the range (80 - 140 msec) is not employed significantly more frequently than the adjacent 60 msec intervals: 34% *versus* 33% for the 20 - 80 and 32% for the 140 - 200 msec interval.

Quantifying the contribution of rise time, steady time, and overall noise duration

In order to establish the contribution of rise time, steady time, and overall duration more precisely, the 43 data points (figure 1) have been subjected to a stepwise multiple linear regression analysis. This model was adopted after a number of alternative solutions, including some with transformations to fit the data to sigmoid functions, had proven less adequate.

When only rise and steady time are entered as predictor variables, rise time is the better cue, explaining 44% of the response variance. However, steady time independently accounts for another 36%.

When the alternative view is tested, viz. that the primary cue is provided by overall duration, with a possible secondary cue for the rise/steady time division, the results of the analysis differ

radically: though the total percentage of explained variance is, of course, the same as before, the independent contribution of rise time, when compared with that of overall duration, turns out to be very small indeed: 3 *versus* 77%.

Conclusion

What we have done here is, in fact, to partial out the effect of rise time as such, and the inevitable circumstance that greater rise time leads, *ceteris paribus*, to longer duration. Clearly, once this duration incrementing artifact has been partialled out, Gerstman's conclusion (c.f. introduction) has to be rejected. A more adequate conclusion would be that affricates are distinguished from fricatives principally by the overall duration of the noise burst, and that rise time is a secondary cue only for a middle range of overall durations.

This, then, would be in line with the *a priori* difference in effectivity of cues provided by rise time, as opposed to duration in general, as predictable from known non-speech psycho-physical characteristics of these temporal phenomena.

Acknowledgements

This research was supported by a grant from the Dutch Organisation for the Advancement of Pure Research (Z.W.O.). I am grateful to Marcel van den Broecke, Peter Ladefoged, Sieb Nooteboom, and George Papçun for their comments on an earlier version of this paper.

References

Dorman, M.F., Rafael, L.J. & Liberman, A.M. (1978). "Some experiments on the sound of silence in phonetic perception," Unpublished manuscript at Haskins Laboratories.

Cutting, J.E. & Rosner, B.S. (1974). "Categories and boundaries in speech and music," Perception and Psychophysics *16*, 564-570.

Borst, J.M. & Cooper, F.S. (1957). "Speech research devices based on a channel vocoder," Journal of the Acoustical Society of America *29*, 777.

Gerstman, L.J. (1957). *Perceptual Dimensions for the Friction Portions of Certain Speech Sounds* (Unpublished Ph. D.-dissertation, New York University).

Heuven, V.J.J.P. van & Broecke, M.P.R. van den (1978a). "Auditory discrimination of rise and decay times in tone and noise bursts," Progress Report of the Institute of Phonetics Utrecht *3(2)*, 49-69.

Heuven, V.J.J.P. van & Broecke, M.P.R. van den (1978b). "Auditory discrimination of rise and decay times in tone and noise bursts, two experiments," submitted to Journal of the Acoustical Society of America.

Repp, B.H., Liberman, A.M., Eccardt, T. & Pesetsky, D. (1978). "Perceptual integration of acoustic cues for stop, affricate, and fricative manner," Journal of Experimental Psychology, Human Perception and Performance *4*, 621-637.

Zwicker, E. (1970). "Subjektive und objektive Dauer von Schallimpulsen und Schallpausen," Acustica *22*, 214-218.