

Acoustic cues for a fricative-affricate contrast in word-final position

Michael F. Dorman

Arizona State University and Haskins Laboratories New Haven, Conn 06510, USA.

Lawrence J. Raphael

Herbert H. Lehman College of the City University of New York and Haskins Laboratories New Haven, Conn. 06510 U.S.A.

and

David Isenberg

Haskins Laboratories New Haven, Conn. 06510, U.S.A.

Received 1st March 1979

Abstract:

The experiments reported here were designed to identify some of the acoustic cues to the fricative-affricate contrast in word-final position (as in "dish" vs. "ditch"). Listening tests prepared from computer-edited natural speech tokens of "dish" and "ditch" reveal that each of the following variables can influence the identification of fricatives and affricates: the temporal and/or spectral characteristics of the vocalic interval; duration of a silent interval; presence or absence of a release burst; rise-time of the fricative noise and the duration of the fricative noise. These results indicate that aspects of qualitatively different acoustic information are integrated over a relatively long period of time when listeners identify fricatives and affricates in word-final position. This outcome suggests that neither a single acoustic property detector nor a single natural category can satisfactorily account for the perception of fricatives and affricates.

Introduction

In the several experiments reported here we investigate the acoustic variables which cue the contrast between fricative and affricate in word-final position. Our interest in these variables stems from several sources. One is that our knowledge of the relevant acoustic variables comes from very early unpublished experiments conducted at the Haskins Laboratories using the Pattern Playback as a speech synthesizer. Since the Pattern Playback did not have a proper noise source for fricatives¹ these early studies bear replication and extension. Another source is our interest in the temporal interval over which the cues for a contrast may be distributed (Dorman, Raphael & Liberman, 1979; Repp, Liberman, Eccardt & Pesetsky, 1978). Investigation of the fricative-affricate contrast in syllable-final position

¹ The pattern playback did not have a separate noise source. Fricatives were simulated by random selection of harmonic tone bursts within the bandwidth of the desired fricative noise (F. S. Cooper, pers. Comm.).

allows us to determine the perceptual salience of acoustic variables which occur before vocal tract closure for the affricate, the perceptual salience of the closure interval itself and the perceptual salience of aspects of the fricative noise which follows the closure interval.

Inspection of the natural speech utterances "dish" and "ditch" in Fig. 1 suggests three acoustic properties of the fricative noise which vary between the utterances and thus may serve as acoustic cues (Bailey & Summerfield, 1978). These are the release burst, the rise-time of the fricative noise and the duration of the fricative noise. Indeed, Gerstman (1957) has shown that, in absolute initial position, rapid rise-times and brief durations of fricative noise lead to the perception of affricates while slow rise-times and longer fricative noise durations lead to the perception of fricatives. The role of the release burst that can be seen in the spectrogram of "ditch" preceding the fricative noise has not been previously investigated. In Fig. 1 it is also evident that "ditch" has a silent closure interval before the onset of the fricative noise that is a consequence of stop articulation. There is no corresponding closure interval in "dish". Both Kuipers (1955) and Truby (1955) have shown that long closure intervals lead to the perception of affricates while short intervals lead to the perception of fricatives. However, since both of these studies used the Pattern Playback as a synthesizer they deserve replication. Finally in Fig. 1 we see that the vocalic portions of the two utterances differ in duration (Isenberg, 1978) as well as in the degree of formant movement preceding the closure. The perceptual salience of these variables has not been previously investigated.

The experiments which follow were designed to assess the effects of each of the acoustic variables described above on the perception of the fricative-affricate contrast. In summary, these variables are: (a) the vocalic portion of the utterance, (b) the duration of the closure interval, (c) the presence of a release burst, (d) the rise-time of the fricative noise amplitude envelope and (e) the duration of the fricative noise.

There is more, however, to our experiments than an investigation of the acoustic cues for a particular phonetic contrast. Our data bear on theories of speech perception that suggest that natural categories (Cutting & Rosner, 1974) or detectors of simple acoustic properties mediate the recognition of phonetic identity. If the outcome of our experiments is that each of the acoustic variables described above can cue the fricative-affricate contrast, then it will be clear that a single natural category or acoustic property detector cannot account for the perception of that contrast.

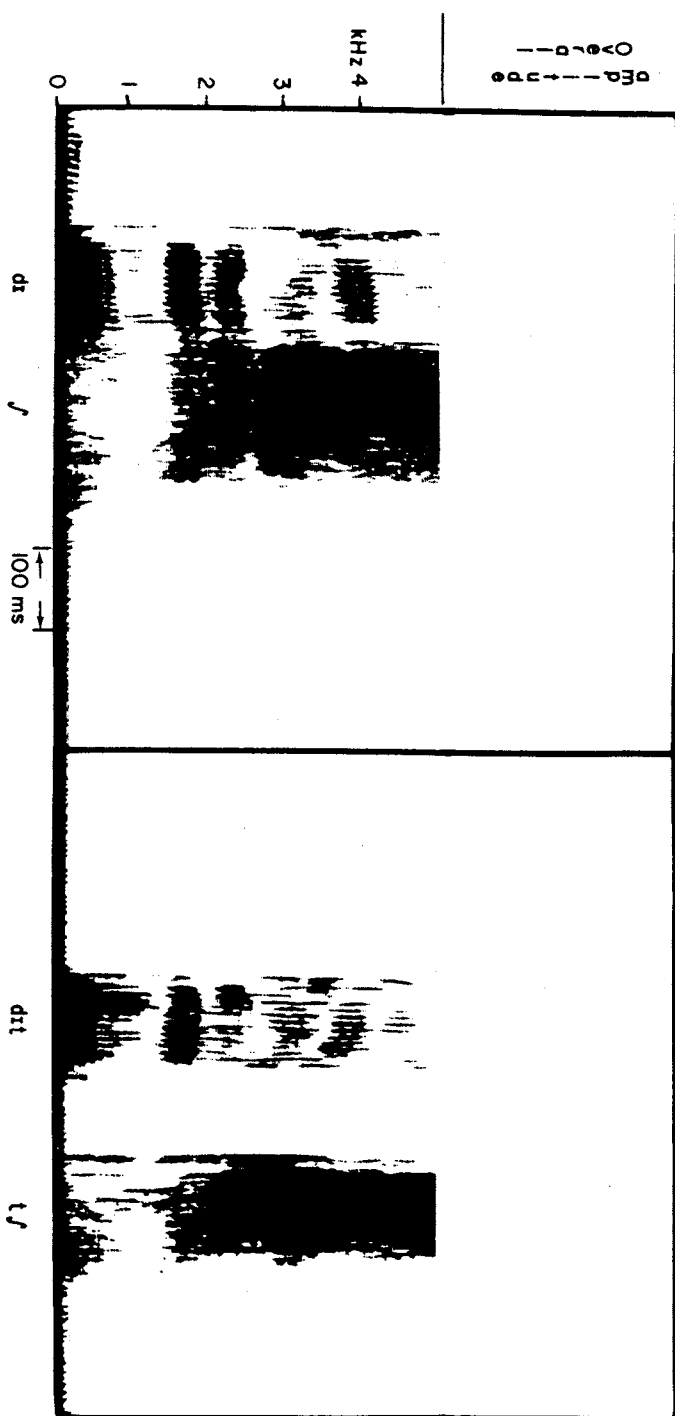
Experiment 1

The purpose of our first experiment was to determine whether two aspects of fricative noise onset—the release burst and the rise-time of the fricative noise— have perceptual salience in the contrast between fricative and affricate in word-final position.

Method

To assess the effect of the release burst, we first digitized and stored in computer memory a male speaker's recording of the utterance, "Put it in the ditch." To create the stimuli for one stimulus condition, we used the "ditch" stimulus as recorded (with release burst) and inserted intervals of silence between the vocalic portion and the burst. The silent intervals ranged in duration from 0–100 ms in 10 ms steps. To create the stimuli for another stimulus condition, we removed the burst from the "ditch" stimulus and then inserted the same intervals of silence between vocalic portion and fricative noise onset as were used in the first condition. Four tokens of each stimulus in each condition were generated and the resulting stimuli randomized (with a 3 s interstimulus interval) into a single test sequence.

The subjects were 16 undergraduates at Arizona State University. They were run in groups in a large sound attenuated room. The stimulus sentences were presented via tape recorder and headphones at a comfortable listening level. Subjects were instructed to identify the last word in each stimulus sentence as "dish" or "ditch".



Spectrograms of a male speaker's productions of "dish" (left) and "ditch" (right) showing contrastive acoustic features which may be relevant in perception of the fricative-affricate distinction (see text).

To assess the effects of fricative noise rise-time, a male speaker's recording of "Put it in the dish," was first digitized and stored in computer memory. For one stimulus condition we used the "dish" stimulus as recorded (with a 35 ms rise-time) and then inserted intervals of silence between the vocalic portion and the fricative noise onset (rise-time was defined as the interval between the onset of energy and the point of maximum signal amplitude). The silent intervals ranged from 20 – 150 ms in 10-ms steps. For a second stimulus condition we removed the first 30 ms of the /ʃ/ fricative noise leaving an effective rise-time of 0 ms². To compensate for the temporal reduction in frication, we embedded an additional 30 ms of full amplitude fricative noise in the center of the /ʃ/ fricative noise. Four tokens of each stimulus in each of the two stimulus conditions were then generated, placed within the original sentence frame and randomized with a 3 – s inter-sentence interval into a single test sequence.

The subjects were 16 undergraduates at Arizona State University. For all conditions the stimuli were reproduced at a comfortable listening level in a large sound attenuated room via tape recorder and loudspeaker. Subjects were instructed to identify the last word in each stimulus sentence as "dish" or "ditch".

Results

The effect of the release burst on the identification of "ditch" is shown in Fig. 2. For the stimuli with a release burst, the phoneme boundary between /ʃ/ and /tʃ/ (the point of 50% /ʃ/ and /tʃ/ responses) falls at approximately 18 ms of silence. For the stimuli without a burst, the boundary falls at approximately 28 ms. The difference in boundary is significant by a Wilcoxon sign-ranks test ($T = 3, P < 0.05$). We conclude that the burst has perceptual salience in natural speech for the fricative-affricate contrast in word-final position.

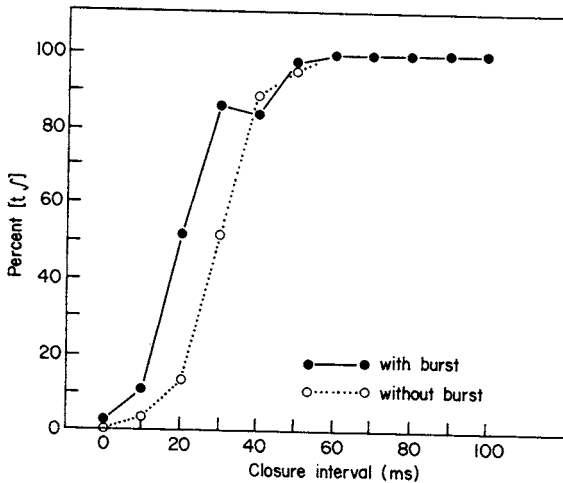


Figure 2

Effect of closure duration and presence or absence of a release burst on listeners' identifications of stimuli as "dish" or "ditch".

² The rise-time of the original fricative noise was 35 ms. When the initial 30 ms of the noise was removed, the rise-time of the fricative noise that remained was essentially instantaneous, although the overall amplitude continued to increase for another 5 ms. Therefore we refer to our stimuli as having rise-times of 35 ms and 0 ms, respectively, even though we removed only 30 ms from the onset of the original fricative noise.

The effect of fricative noise rise-time on the identification of "ditch" is shown in Fig. 3. For the stimuli with a 5 ms rise-time, the phoneme boundary falls at approximately 37 ms of silence, while for the stimuli with a 35 ms rise-time, the boundary falls at approximately 57ms. The difference in boundary is significant ($T = 2, P < 0.05$). We conclude from this

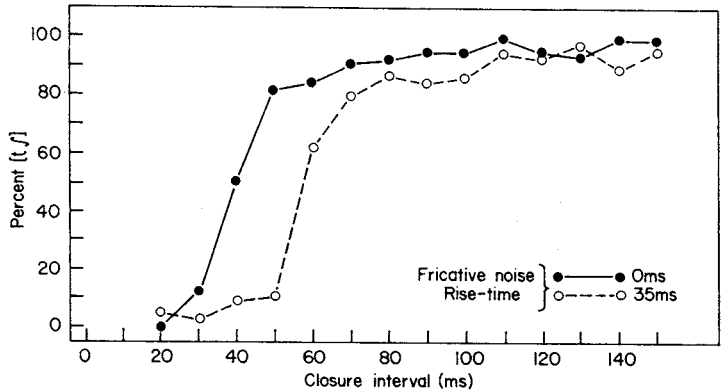


Figure 3 Effect of closure duration and fricative noise rise-time on listeners' identification of stimuli as "dish" or "ditch".

outcome that the rise-time of the fricative noise is also a cue to the fricative-affricate contrast in natural speech. Thus, it is the case for natural speech, as it is for synthetic speech, that the onset characteristics of the fricative noise play a significant role in the perception of the fricative-affricate contrast.

The duration of the closure interval also exerted a powerful influence on the identification of "ditch" and "dish". Indeed, regardless of whether the source utterance was "ditch" or "dish", at short closure intervals the stimuli were generally heard as "dish" while at long intervals the stimuli were heard as "ditch". Thus, at least two cues, the duration of the closure interval and the fricative noise onset timing, operate to specify the contrast between fricative and affricate in word-final position.

Experiment 2

Our first experiment demonstrated that the onset characteristics of the fricative noise and the duration of the preceding silent interval were cues to the fricative-affricate contrast. In Experiment 2 we assessed whether the duration of fricative noise is also a perceptual cue to the contrast in word-final position.

Method

The utterance, "Put it in the dish," was produced by a male speaker and then stored in digital form in computer memory. To assess the effect of fricative noise we created "dish" stimuli with fricative noise durations of 320, 240 and 160 ms. The stimulus with the longest fricative noise duration was the stimulus originally recorded. To create the stimulus with 240 ms of fricative noise, 80 ms of fricative noise was removed from the center of the 320 ms fricative noise. The onset and offset of the fricative noise were then recombined. A similar procedure with the 240 ms stimulus created the 160 ms stimulus. (We should note that since the fricative noise was aperiodic, the removal and recombination of the fricative noise did not produce audible transients.) For each stimulus condition, intervals of silence ranging from 0 – 100 ms in 10 – ms steps were inserted between the vocalic portion and fricative noise portion of the signal. Four tokens of each stimulus in each stimulus condition were generated and then randomized with a 3 s interstimulus intervals into a single test sequence.

are the proportions of responses from the two continua (54.1% for /dIt/ + /f/: 65% for /dI/ + /tj/. In fact, it is clear that the vocalic section has almost as great an effect in specifying this fricative-affricate contrast as does the fricative noise itself.

An analysis of variance was performed on the crossovers computed for each subject in which vocalic section and fricative noise were within-subjects factors. A significant effect was obtained for both the fricative noise [$F(1,9) = 21.96, P < 0.01$] and for the vocalic portion [$F(1,9) = 38.22, P < 0.01$], indicating that both sections were effective in specifying this fricative-affricate contrast. There was no interaction between these factors ($F < 1.0$). A similar pattern was obtained in an additional analysis of variance performed upon the proportion of "ditch" responses.

These data suggest that properties of the vocalic region are effective cues for the perception of an affricate. One might have predicted that the acoustic properties of the fricative noise would have contained the major cues. However, a comparison of the identification functions for /dIt/ + /f/ and /dI/ + /tj/ reveals that listeners make about as many affricate responses in the two conditions (59% and 62% respectively). Thus, the vocalic portion of these utterances seems to be about as effective as the fricative noise portion as a cue to the fricative-affricate contrast.

We should point out that a fricative was generally reported even when both vocalic and fricative noise portions of the signals were appropriate for an affricate, if the silent interval was very brief (less than 20 ms). The converse was also the case—when the closure was very long, an affricate was generally reported. (The failure of the /dI+/f/ function to reach 100% was due to two subjects who were unable to hear the affricate with these stimuli even at very long closure intervals.) This outcome suggests that silence is a very powerful cue for the perception of an affricate in syllable-final position.

We can suggest three possible accounts for this outcome. One possibility is that when the physical gap in the signal is less than 20 ms, the auditory system does not resolve the gap. On this view the signal arriving at central processing mechanisms would be without an important cue for stop manner. Therefore, no affricate would be reported. This seems unlikely since there is a systematic increase in the percentage of /tj/ responses as silent interval is increased from 10 to 20 ms. Another possibility is that the vocalic portion of the utterance (forward) masks the onset of the fricative noise. This could result in the perception of a more gradual (or fricative-like) fricative noise onset. This explanation, however, seems weak on several accounts. The most important is that we would expect little or no forward masking between signals with such different frequency spectra and source characteristics. Finally, a third possibility is that a very brief silent interval is inappropriate for the production of a stop consonant and is, therefore, not an appropriate cue for the perception of stop manner. This, we suggest, is the most likely account.

General discussion

In this series of experiments with natural speech we have found the following acoustic variables to have perceptual salience for the fricative-affricate contrast in syllable-final position: the vocalic portion of the utterance, the presence and duration of the silent closure interval, the release burst, the rise-time of the fricative noise and the duration of the fricative noise. We may account for the perceptual salience of this constellation of acoustic variables by reference to articulation: each variable is one of the distributed acoustic consequences of the gesture that differentiates fricative from affricate. In this context we should note Bailey & Summerfield's (1978) suggestion that "...it appears possible to demonstrate that some "cue-value" attends every acoustic detail that distinguishes two different phonetic events" (see also Lisker, 1977). We see in the present results a confirmation of this suggestion.

The importance of this outcome for theories of speech perception is that those theories must not only specify how the various elements of the acoustic signal are detected, but also

what events bear on the same phonetic percept. This latter requirement entails looking across some time-varying segment of the speech signal. We should wonder then how the recognition routines "know" what the relevant acoustic hallmarks for a given phoneme are, and how they are spread over time in the speech signal. It seems reasonable to suggest that a system that appreciates the several acoustic consequences of articulatory gestures would be most likely to know what to look for and over what period of the signal to look. This is simply to say that we expect in speech perception a link with speech production [see Bailey & Summerfield (1978) for an excellent discussion of this issue].

Finally, we should note that the kinds of experiments described here—that is, those that demonstrate trading relationships among the several cues to the fricative-affricate contrast—may well prove probative in investigations of the phylogeny of speech perception. For example, we would not expect non-human primates to appreciate the diverse acoustic consequences of the production of fricative and affricate and, thus, we would not expect them to show the trading relations in perception found in the experiments reported here (see also Liberman & Pisoni, 1977).

Research supported by Grant HD-01994 from the National Institute of Child Health and Human Development to Haskins Laboratories and by NIH Post-Doctoral Fellowship NS-05493 to D. Isenberg. For the sake of continuity we have included some data in this study which has been reported elsewhere (Dorman, Raphael & Liberman, 1979). Requests for reprints should be sent to Michael F. Dorman, Haskins Laboratories, 270 Crown Street, New Haven, CT 06510

References

- Bailey, P. & Summerfield, Q. (1978). Some observations on the perception of [s] + stop clusters. *Haskins Laboratories Status Report on Speech Research SR-53*, 25–60.
- Dorman, M. F., Raphael, L. J. & Liberman, A. M. (1979). Some experiments on the sounds of silence in phonetic perception. *Journal of the Acoustical Society of America*, **65**, 1518–1532.
- Cutting, J. & Rosner, B. (1974). Categories and boundaries in speech and music. *Perception & Psychophysics* **16**, 564–570.
- Gerstman, L. (1957). Perceptual dimensions for the fricative noise portions of certain speech sounds. Unpublished doctoral dissertation, New York University.
- Isenberg, D. (1978). Effect of speaking rate on the relative duration of stop closure and fricative noise. *Haskins Laboratories Status Report on Speech Research SR-55/56*, 63–79.
- Kuipers, A. (1955). Affricates in intervocalic position. *Haskins Laboratories Quarterly Progress Report* **15**, Appendix 6.
- Liberman, A. & Pisoni, D. (1977). Evidence for a special speech-processing subsystem in the human. In *Recognition of Complex Acoustic Signals*, (Bullock, T. H., Ed.). Berlin: Dahlem Konferenzen.
- Lisker, L. (1977). Closure hiatus: cue to voicing, manner and place of consonant occlusion. *Journal of the Acoustical Society of America* **61**, S-48(A).
- Repp, B. H., Liberman, A. M., Eccard, T. & Pesetsky, D. (1978). Perceptual integration of temporal cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance* **4**, 621–637.
- Truby, H. (1955). Affricates. *Haskins Laboratories Quarterly Progress Report* **11**, 7–8.

