BRITISH ACOUSTICAL SOCIETY

"SPRING MEETING" at Chelsea College, London, S.W.3. on Wednesday 25th April / Friday 27th April, 1973.

SPEECH AND HEARING: Session 'A': Speech Production and Perception.

Paper No:

73SHA5

Durational Cues in the Perception of Certain Consonants

W.A. Ainsworth

Department of Communication, University of Keele, Staffordshire, England.

1. Introduction. There are many instances in the perception of speech where the duration of events allows discriminations to be made (1, 2, 3, 4). It has recently been demonstrated that the rhythm, or tempo, of a sequence of sounds can have an effect on the perception of vowel sounds which is equivalent to a change in vowel duration (3). The question arises as to whether changes in the tempo of precursive sequences can have a similar effect on the perception of consonants whose discrimination depends upon duration cues.

In the present study the effects of tempo and duration on the perception of syllables beginning with /b, w/ or /s, t, d/ have been examined. The duration of initial formant transitions was varied in order to synthesize (b, w/, and the duration of a burst of noise was varied to produce /s, t, d/.

2. Experimental Procedure. The method adopted was to present a series of synthesized sounds to a group of listeners for identification. The sounds were generated by a parallel formant speech synthesizer of the type described by Holmes et al (5). An 'on-line' computer system generated the control signals for the synthesizer and recorded the listeners' responses (6).

Seven volunteers, who spoke with British English accents, acted as listeners in the experiments. They were asked to classify each sound they heard into one of the categories specified by the labels on the boxes in front of them, and to press the appropriate switch.

3. Semivowels and Stops. Liberman et al (4) had previously found that the duration of the initial formant transitions was a sufficient cue for discriminating /b/ from /w/ and /g/ from /j/. The object of the present experiment was to determine whether the perceptual boundary between such pairs of phonemes is affected by the tempo of a sequence of sounds which precede the test syllable.

Each stimulus consisted of a precursor of three bursts of noise followed by the test syllable as shown in Figure 1. The syllable was a two formant voiced sound, the frequencies of whose formants began at 220 and 760 Hz, then increased linearly for a time T1 so that they reached values appropriate to one of the vowels /i, ɛ, a, ɜ/. They then continued at constant frequencies until the end of the syllable at a time T2 later. The sounds in the precursor were of a /ʃ/ quality, of duration T2, and separated by periods of silence T2 long. The duration of the formant transitions, T1, and the duration of the precursor sounds and vowel, T2, were varied randomly in the course of a listening session. T1 covered the range 10-60 msec. in 10 msec. steps, and T2 80-400 msec. in 80 msec. steps.

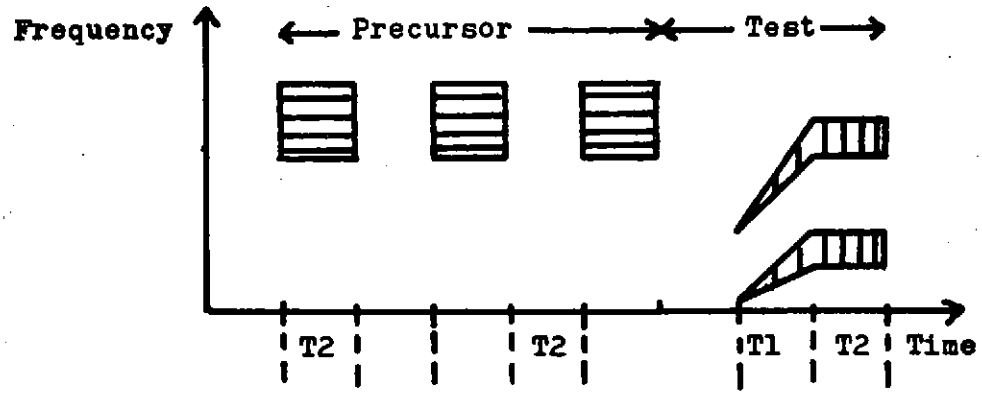

Figure 1. Spectrogram of stimulus sounds.

The listeners were asked to clasify the initial consonant in the syllable following the noise burst as 'B', and 'W', 'other consonant' or 'no consonant'. Each listener heard all of the stimuli three times, with the order of presentation being varied on each occasion.

The results showed that as the duration of the formant transitions were increased the proportion of listeners hearing /b/ diminished and the proportion hearing /w/ increased. About 50% of each were heard when T1 was 30-40 msec. long, agreeing with the results of Liberman et al. (4). There was a small shift in the location of the /b-w/ boundary as the duration T2 was increased. Figure 2 shows the number of sounds heard in each category for the shortest value of T2 (80 msec.) and the longest (400 msec.).

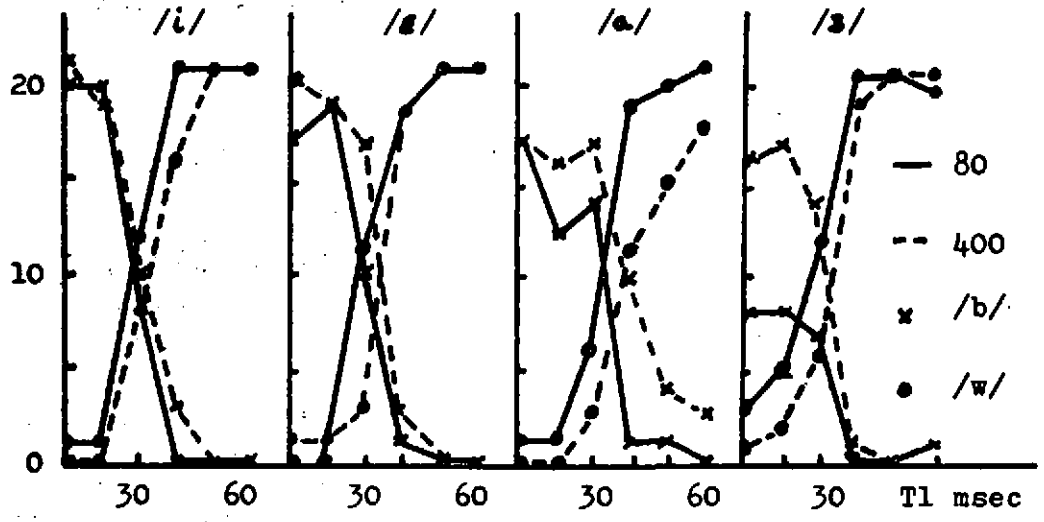

Figure 2. Number of /b,w/ responses as a function of formant transition duration T1.

The shift is in the expected direction, but appears to be only about 5 msec. in extent. This is probably too small to be significant for either synthesis or perception, and so fits well with the observation of Kelly et al. (7) that in normal speech the slopes of the second formant transitions in /b/ and /w/ are not affected by the overall tempo of the utterances in which they occur.

4. Fricatives and stop consonants. Spectrograms of /di/, /ti/ and /si/ are very similar except that the fricative part of the consonant increases in duration from /d/ through /t/ to /s/. This suggests that this duration may serve as a perceptual cue for distinguishing between these syllables. In order to test this possibility a set of sounds was generated consisting of a burst of noise in the 3-4 kHz. region, followed by a steady voiced sound with formant frequencies

of 220, 2440, and 3220 Hz. By adjusting the duration of the noise burst, clear examples of each of the above syllables could be produced.

In order to find out whether the perceptual boundaries between these sounds can be influenced by the tempo of a precursive sequence, the program for generating the random series of sounds was modified so that the test syllables consisted of a burst of noise of duration T1, followed immediately by an /i/ vowel of duration T2. The precursor was a sequence of three /ʃ/ sounds as before. The duration of the noise burst in the test syllables, T1, covered the range 10-120 msec. in 10 msec. steps, and the duration of the precursor sounds and the vowel, T2, 80-400 sec. in 80 msec. steps.

The same listeners were asked to classify the consonant part of the /i/ syllables as 'D', 'T', 'S', 'other consonant' or 'no consonant'. They heard the range of stimuli formed by every combination of T1 and T2 twice. The order of presentation was different on each occasion.

The results are shown in Figure 3 for the precursive sequences containing the 80 and 400 msec. sounds. It can be seen that there was a shift in the boundary between /d/ and /t/ of about 15 msec., and in the boundary between /t/ and /s/ of about 30 msec.

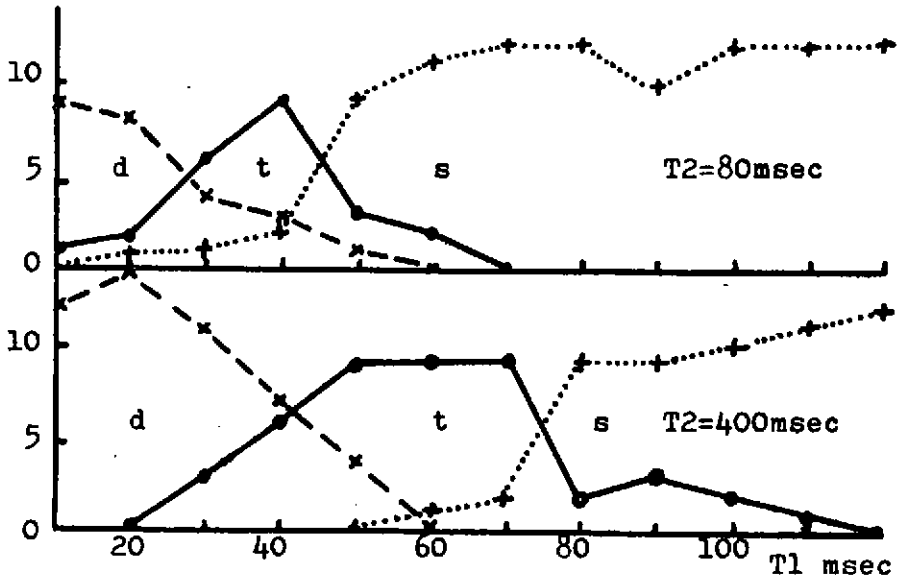

Figure 3. Number of /d,t,s/ responses as a function of noise duration T1.

5. Nature of the Precursive Sounds. In order to ascertain whether precursors of different types of sounds affect the boundaries in different ways, the above experiment was repeated with the /ʃ/ sounds of the precursor replaced by /i/ sounds, similar to the vowel in the test syllable, and by neutral /ə/ vowel sounds.

With the /ə/ sounds the results obtained were almost identical to those obtained with the /ʃ/ sounds. With the /i/ sounds the listeners reported that the consonants were more difficult to identify. Despite this the /d-t/ and /t-s/ boundaries occurred at almost the same points as with the other precursor sounds.

6. Preceding and Following Sounds. In the experiments described so far, the duration of the vowel in the text syllable was made identical to the duration of the sounds and silent periods in the precusor. It is possible that the duration of the sounds preceding or following the consonant may affect its categorisation. The experiments were therefore repeated, once with the sounds in the precursor kept constant at 240 msec. and the duration of the vowel in the text syllable varied, and once with the vowel duration constant and the precursor sound duration varied.

The results showed that both of these factors affect the

location of the perceptual boundaries, but the duration of the following vowel had a greater effect. Similar results were found with both the /ʃ/ and /ə/ precursors.

Haggard and Summerfield (8), studying the effect of speech rate on the perception of voicing, have obtained similar results with a more speech-like precursor.

7. Number of Precursor Sounds. Finally, in order to determine the number of syllables in the precursor which are needed to establish the tempo, the experiment was repeated with the number of /ə/ sounds varying from one to seven. It was found that the size of the shift in the perceptual boundaries was almost identical on each occasion. It appears that a single sound in the precursor is sufficient to establish the 'tempo'. It would thus be more sensible to speak of the duration of the sounds, rather than their tempo, influencing the location of the perceptual boundaries.

8. Conclusions. The experiments indicate that the tempo of an utterance may influence the perception of some consonants. The shifts in the /b-w/ boundaries were rather small, but the shifts in the /d-t/ and /t-s/ boundaries were relatively much greater. The results of this experiment, and those of a previous one with vowels (3), suggest that when the rate of an utterance is increased the perceptual mechanism acts in such a way that it expects those portions of the speech which have a relatively steady spectrum will be shortened, whereas those which have a changing spectrum will occupy the same period of time.

Broadbent and Ladefoged (9) and Fourcin (10) have shown that judgments in the frequency domain are made relative to the sounds in a precursor. The present results suggest a similar adaptation takes place in the time domain.

References.

1. D.C. Bennet: Language and Speech, 11, 65-85, 1968.
2. L.J. Gerstman: J. Acoust. Soc. Am., 28, 160(A), 1956.
3. W.A. Ainsworth: J. Acoust. Soc. Am., 51, 648-651, 1972.
4. A.M. Liberman, P.C. Delattre, L.J. Gerstman, and F.S. Cooper: J. Exptl. Psychol., 52, 127-137, 1956.
5. J.N. Holmes, I.G. Mattingley, and J.N. Shearme: Language and Speech, 7, 127-143, 1964.
6. W.A. Ainsworth and J.B. Millar, Behav. Res. Meth. Instr., 3, 21-24, 1971.
7. J. Kelly, J.K. Anthony, and E. Uldall; J. Acoust. Soc. Am., 25, 1113(A), 1965.
8. M.P. Haggard and A.Q. Summerfield: Speech Synthesis and Perception, 6, 1-12, 1972.
9. D.E. Broadbent and P. Ladefoged: Proc. Royal Soc. B., 151, 384-399, 1960.
10. A.J. Fourcin: IEEE Trans. Audio Electroacoustics, AU-16, 65-67, 1968.