INITIAL SPEECH SOUND PROCESSING IN SPOKEN WORD RECOGNITION

PHILLIP DERMODY KERRIE MACKIE RICHARD KATSCH

Speech Communication Research Section
National Acoustic Laboratories
126 Greville St., Chatswood. N.S.W. 2067. AUSTRALIA.

ABSTRACT - The present study uses the gating paradigm to investigate the initial speech sound (ISS) in spoken word recognition. Results are presented for spoken words and consonant-vowel (CV) syllables, which both show recognition of the ISS in the first 30 msec. Acoustic analyses of the ISS show similarities between the words and syllables and are consistent with the templates proposed by Stevens & Blumstein (1). It is suggested that the time course of ISS perception indicates the need to change present models of spoken word recognition.

INTRODUCTION

In a previous investigation of initial speech sound processing we confirmed that listeners could identify the initial speech sound in CV syllables, consisting of a stop consonant plus the vowel /a/ in the first 30 msec after onset (2). This was paralleled by similar performance for initial consonant recognition during the first 30 msec in words. The present study investigates the same effects but examines in more detail the time course for speech sound recognition during the initial 30 msec for individual sounds.

STIMULI

A set of 12 English consonants (including the six stops) plus the vowel /a/ were recorded by a male speaker onto a computer based speech storage/ editing system. In addition, a set of words beginning with similar consonants were recorded. A subset of these words included words which began with a stop consonant plus the vowlel /a/ (eg. tartan; gardener; particle). The stimuli were digitised at 36KHz sampling rate with filters for input/output set at 12KHz. Each syllable and word was visually displayed and segments in increments of 10 msec, were marked and labelled. The endpoint of each gated stimulus was made to the nearest zero crossing point to avoid audible clicks. The gated stimuli were output to audio tape in sequential order (ie. 10, 20, 30 etc msec presented in order) with 4 seconds between each presentation. For example, the first 10 msec stimulus was produced (gate 1) followed 4 seconds later by the 20 msec stimulus (gate 2) etc. For the CV syllables, gates in increments of 10 msec from 10 to 100 msec were recorded. For the spoken words, 10 msec gate increments were used for the first 6 gates (ie. to 60 msec of the word after onset) and then incremented in 30 msec gates to the completion of the word. That is, for the words, gate 7 was 90 msec in duration from onset while gate 8 was 120 msec etc.

PROCEDURE

All subjects were given minimal practice to familiarize them with the task. The words and syllables were given in separate test sessions using different sets of subjects. The subjects were University undergraduates with normal hearing (N= 23 per test) who were paid for their participation. Subjects were instructed that they would hear short bits of a speech sound which would increase on succeeding trials and that they should write down their response after each presentation. Subjects were instructed to guess if unsure since recording a response per trial was mandatory. They were also told that the syllable or word could begin with any consonant sound in English and their task was to identify the initial consonant for the CV syllables and the initial consonant and the whole word for the word stimuli.

RESULTS

While the full set of CV syllables and words were presented to the subjects only the results for the stimuli beginning with the stop consonants are presented here. Figure 1 shows the results for the 6 stop consonants plus the vowel /a/. The cumulative percentage of subjects who obtained correct initial consonant recognition without subsequent errors are shown for each gate duration. For comparison, the results of our previous experiment (2) for tha same CV gates presented as part of a closed set of 6 alternative responses are also presented. The figure shows similar performance patterns for open and closed set responses. The open set performance reaches an asymptope slightly later (at about 40 msec)than the closed set performance which reaches an asymptope at the 30 msec gate.

Figure 2 shows the contribution of two of the CV syllables to the curve including the best result (from /ta/) and the worst result (from /ga/). The figure presents the percentage of subjects who obtain correct recognition as a function of gate duration. In the case of /ta/ all subjects have correctly recognised the initial sound correctly at the 20 msec gate, while for the /ga/ only 4% have recognised the sound at the 20 msec gate and at the 40 msec gate 80% of subjects have identified the initial /g/.

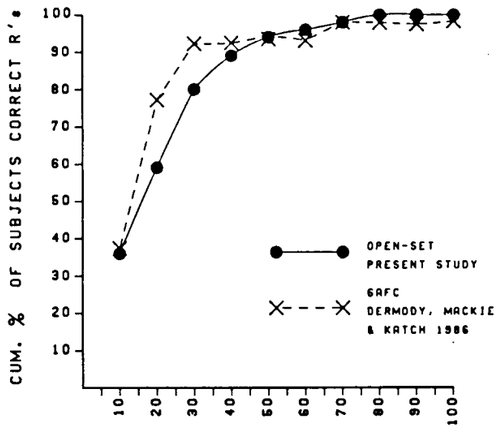

Figure 1
Initial consonant recognition point for 6 stops

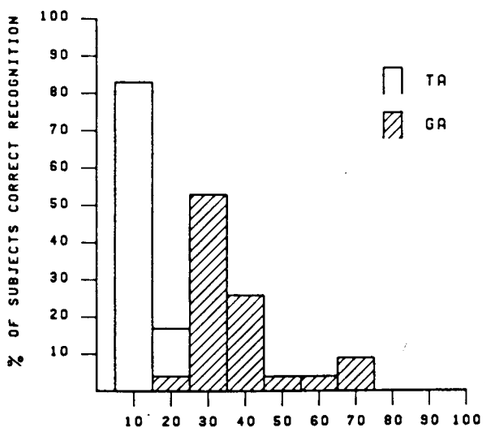

Figure 2
Initial consonant recognition point for /ta/,/ga/.

The results for the identification of the ISS in words were similar to those for the CV syllable. Figures 3 and 4 present the results for "tartan" and "gardener" for comparison with /ta/ and /ga/ in figure 2. In figure 3 -"tartan"-, the initial /t/ is correctly recognised by the majority (84%) of subjects at the first gate (10 msec), while the /g/ in "gardener" requires about 60 msec for equivalent performance levels. This shows a similar pattern as shown for gate recognition in syllables. In both /ta/ and "tartan", the initial /t/ is correctly recognised by most subjects at 10 msec while the /g/ requires about 60 msec for "gardener" and about 40 msec for /ga/.

Figures 3 and 4 also show the vowel and word recognition points for the words. The overall results for the word gates is similar in pattern to the word gates reported by Grossjean (3). Figure 3 shows the /a/ sound in "tartan" is recognised by only 4% of sujects by the 60 msec gate while in figure 4 , about 30% of subjects recognise the vowel at 60 msec. Both words however,require 90 to 120 msec for the majority of subjects to correctly identify the /a/. Despite similar word durations "tartan" is recognised by the majority of subjects by about 570 msec, while recognition for gardener is delayed until 690 to 720 msec.

DISCUSSION

In a previous study (2) we reported that recognition of the initial speech sound occurred for both syllables and words containing initial stops in about 30 msec. Acoustic analysis revealed the patterns suggested by Stevens & Blumstein (1) and Blumstein & Stevens (4) for these durations. It was also noted that the same acoustic patterns were also present for the 10 and 20 msec gate stimuli where subjects were able to recognise the initial sounds at better than chance level. Similar templates were also found for the stimuli in the present study which is consistent with the reports by Blumstein & Stevens (4) and by Kewley- Port, Pisoni, & Studdert-Kennedy (5) that the information is sufficient to recognise the initial speech sound. The present study extends these findings to the case of word recognition and indicates that the initial speech sound could play and important role in lexical access. These results suggest that the word recognition process begins much earlier than the 100 to 200 msec after word onset as suggested by Marslen-Wilson (6). That is, at least for stops in the word initial position, lexical access may begin around 30 msec after word onset. This finding probably needs to be incorporated into theories of the time course of lexical access and word recognition and would certainly provide a time allowance sufficient for elaborate search procedures in these processes.

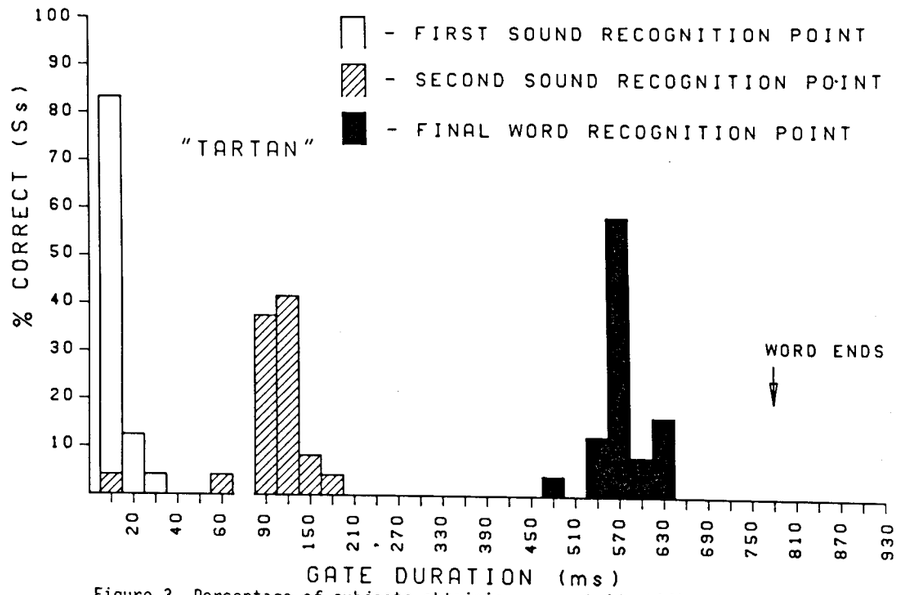

Figure 3 Percentage of subjects obtaining correct identification at each gate.

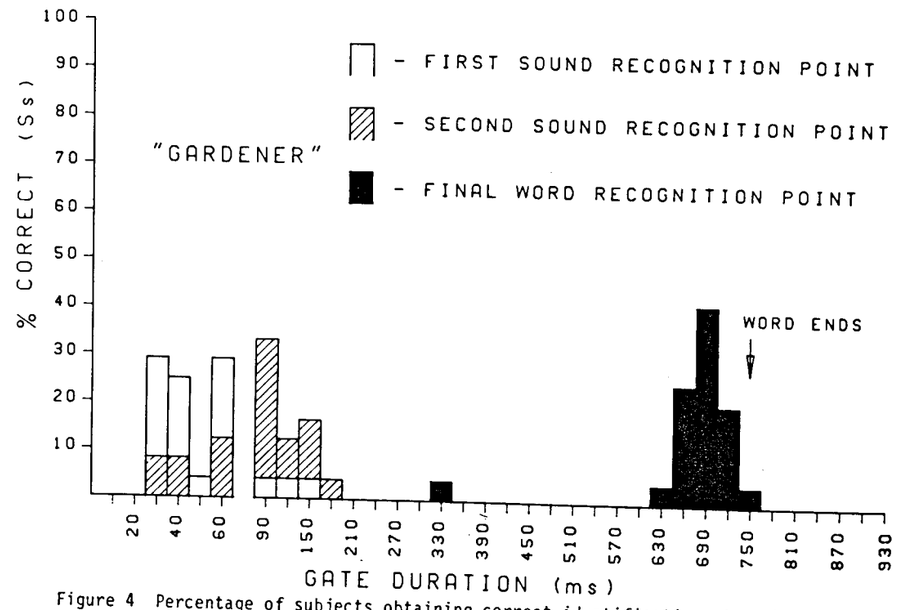

Figure 4 Percentage of subjects obtaining correct identification at each gate.

REFERENCES
(1)Stevens, K. & Blumstein, S. (1978) "Invariant cues for place of articulation in stop consonants". J. Acoustical Society of America, 64, 1358-1368.
(2)Dermody, P., Mackie, K. and Katsch, R. (1986) "Initial speech sound processing in spoken word recognition." Proceedings of the 1st Australian Conference on Speech Science & Technology, Canberra: Australian National Univ.
(3)Grossjean, F. (1980) "Spoken word recognition processes and the gating paradigm", Perception & Psychophysics, 28, 267-283.
(4)Blumstein, S. & Stevens, K. (1980) "Perceptual invariance and onset spectra for stop consonants in different vowel environments" J. Acoustical Society of America, 67, 648-662.
(5)Kewley-Port,D., Pisoni,D. & Studdert-Kennedy, M. (1983) "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants", J. Acoustical Society of America, 73, 1779-1793.
(6)Marslen-Wilson, W. & Tyler, L. (1980) "The temporal structure of spoken language understanding". Cognition, 8, 1-71.