DETECTION TIMES FOR VOWELS VERSUS CONSONANTS

Brit van Ooyen, Anne Cutler & Dennis Norris

MRC Applied Psychology Unit, 15 Chaucer Rd., Cambridge, CB2 2EF, U.K.

ABSTRACT

This paper reports two experiments with vowels and consonants as phoneme detection targets in real words. In the first experiment, two relatively distinct vowels were compared with two confusible stop consonants. Response times to the vowels were longer than to the consonants. Response times correlated negatively with target phoneme length. In the second, two relatively distinct vowels were compared with their corresponding semivowels. This time, the vowels were detected faster than the semivowels. We conclude that response time differences between vowels and stop consonants in this task may reflect differences between phoneme categories in the variability of tokens, both in the acoustic realisation of targets and in the representation of targets by subjects.

1. INTRODUCTION

The majority of data on perception of vowels and consonants comes from identification and discrimination experiments [1]. In a typical identification task, listeners classify a sound as a member of a particular category. In discrimination, they may be asked whether they can discriminate between two sounds belonging to the same or different phonemic categories. These tasks have been used extensively in the study of categorical perception of speech. There seems to be no qualitative difference in how vowels versus consonants are processed in these tasks [2] [3].

In contrast, there is evidence from recent studies that vowels and consonants *do* differ in how they are processed in the performance of another task, phoneme detection. This task [4] requires listeners to press a response key when they hear an occurence of a pre-specified target phoneme. The experimental variable is response time (RT), which is taken to reflect difficulty of processing. The task has been chiefly used to measure psycholinguistic variables which influence processing difficulty; little attention has been paid to the task in its own right. The choice of which phonemes to use as targets has often been motivated by which sounds are comparatively easy to locate in a speech signal. Most experiments have used stop consonant targets.

Studies of spontaneous slips of the ear [5] suggest that consonants are misperceived more often than vowels; in particular, vowels in stressed syllables seem to be accurately perceived. From this it might have been expected that vowels would prove to be easier as detection targets than consonants. Yet the very few phoneme detection results previously available for vowels suggested that RTs were longer for vowel than for consonant detection [6] [7].

Indeed, just this pattern was discovered in two previous experiments by the present authors. In two vowel detection experiments, one with real words and one with nonsense words, Cutler, Norris & van Ooyen [8] found that RTs were very long in comparison with RTs reported in previous work on consonants [1]. Moreover, error rates were high. It was concluded that vowels are difficult as targets in a detection task. A tentative explanation invoked the notion of confusibility between phonemes; it was suggested that vowels were more confusible than consonants, either within the vowel repertoire of the English language as a whole or within the experimental situations we had constructed.

Experiment 1 was designed to test the effects of confusibility of targets within the experimental situation. It also provided a direct test of the relative detectability of vowel and consonant targets within one experiment. We compared detection times for two vowels and two consonants. The two vowel targets were highly distinct, while the two consonantal targets were relatively confusible.

2. EXPERIMENT 1

Method

Materials. The target vowels were high front */i/* and low back /a/ plus the two stop consonants /p/ and /t/. These occurred in 144 mono and disyllabic words, 36 for each target phoneme. Of these 36 words, 12 had the target phoneme in word initial, 12 in word medial and 12 in word final position. For /a/ & /i/, 20 of the 36 words were monosyllabic, 16 were disyllabic. For /p/ & /t/, 16 of the 36 words were monosyllabic, 20 were disyllabic. Half of all disyllabic words had initial stress, half final stress. Target phonemes always occurred in stressed syllables. The words were matched for frequency for initial, medial and final means within each target phoneme. Forty words, 10 per target phoneme, were dummy target items. About 1000 words were fillers.

Experimental design. The material formed 4 blocks, one for each target phoneme. Each block consisted of 55 lists of

2 to 6 words in length. Of these, 36 lists had an experimental item in 3rd, 4th or 5th position, 10 lists had a dummy target item in 1st or 2nd position and 10 lists contained no occurrence of the target phoneme. Before each block the target was specified with examples. All materials were recorded by a male native speaker of British English. The blocks were presented in four different orders.

Subjects. Twenty four subjects between 18 and 32 years of age participated for payment in the experiment. All were native speakers of British English with normal hearing. Six subjects heard each order of presentation of the blocks.

Procedure. Subjects listened to taped instructions that requested them to press a single response key as soon as they detected a target phoneme, as specified in the examples, anywhere in a word. A timing mark, aligned with the onset of each experimental word, initiated response timing. The data were stored on a microcomputer. The 144 experimental words were digitized to measure word length, target phoneme duration, and the time between target phoneme onset and timing mark. RTs were adjusted for these measurements to give RTs from target phoneme onset.

Results

Two analyses of variance, with subjects and with words as random factors, were carried out. We report only effects significant in both. The mean RT for the two consonants was 517 ms, significantly shorter than the mean RT for the two vowels at 600 ms (F1 [1,20] = 33.53, p < .001; F2 [1,132] = 58.17, p < .001). Fig. 1 shows mean RTs in ms for the vowels vs. the stops. There was also a main effect of position of the target in the word (Fl [2,40] = 37.26, p < .001; F2 [2,132] = 40.18, p < .001); but this effect interacted with the vowel-consonant comparison (Fl [2,40] = 29.16, p < .001; F2 [2,132] = 9.95, p < .001). 7-tests showed that there was no significant difference between the means for vowels (589 ms) and consonants (568 ms): *t1* and *t2* > .25. The differences were, however, significant in both medial position (vowels 644 ms; consonants 561 ms; *t1* [23] = 7.4, p < .001; *t2* [23] = 8.67, p < .001) and final position (vowels 561 ms; consonants 427 ms; fl [23] = 4.58, p < .001; *tl* [23] = 3.44, p < .005).

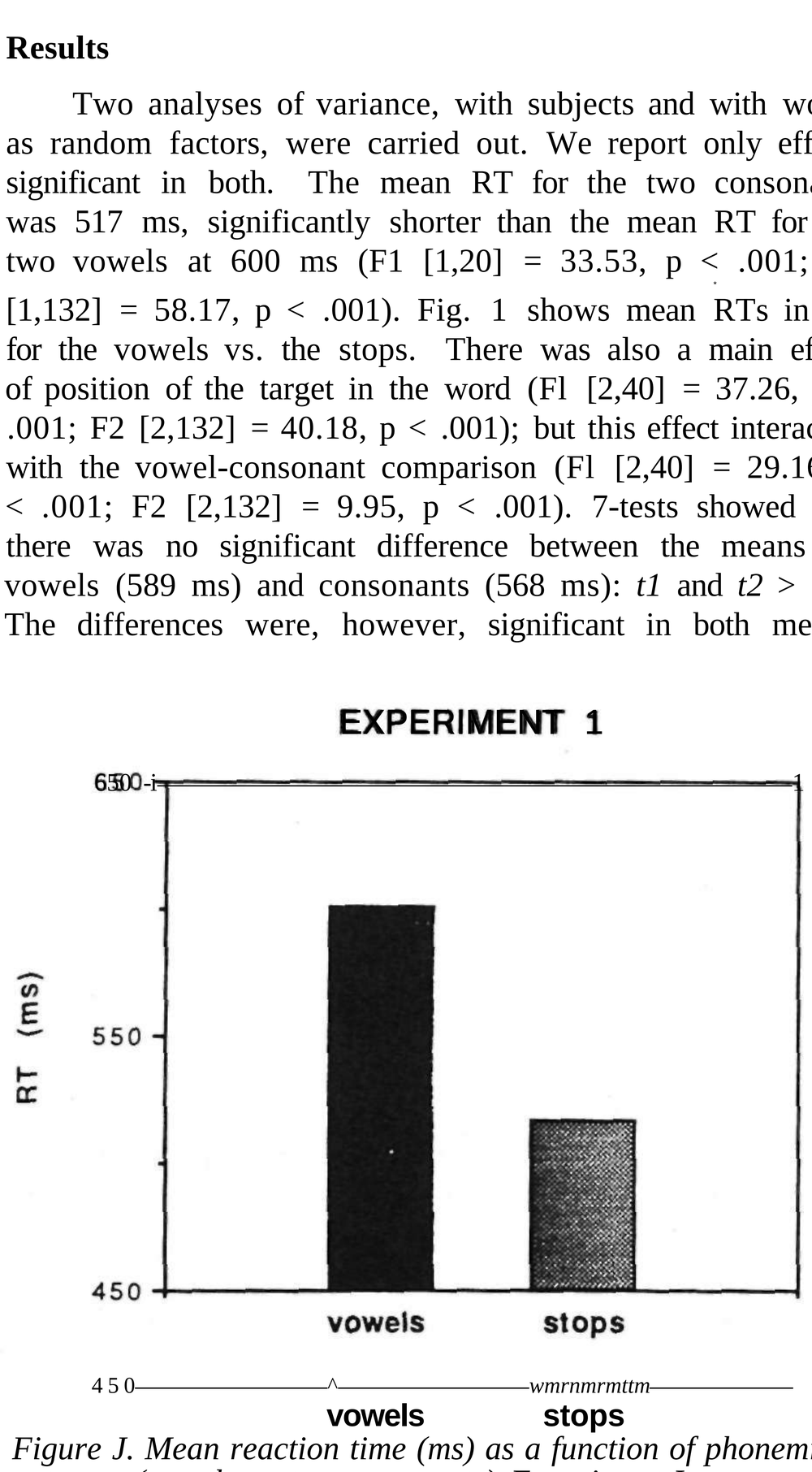

Figure J. Mean reaction time (ms) as a function of phonemic category (vowels or stop consonants). Experiment J.

The overall error rate was 6%, with consonants at 9% being missed significantly more often than vowels at 4% (Fl [1,20] = 8.93, p < .01; F2 [1,132] = 26.89, p < .01). Together with the mean RTs, this suggests a speed-accuracy tradeoff: the phonemes that are responded to most rapidly are also missed most frequently. A correlation analysis showed that vowel RT correlated negatively with target length: longer vowels produced faster responses. No such correlation was found for consonant RTs. (The same correlation was observed in the predecessor experiments [8]).

The results, therefore, confirm the previous finding that vowels are at a disadvantage in a detection task relative to stop consonants. In the present experiment, however, RTs were not as long as in the Cutler et al. [8] studies; even in medial position, vowel RTs were around 100 ms faster than in the previous experiments. Thus comparing only two vowel targets, which were highly distinct, has removed some of the difficulty of vowel detection; but it has not removed the difference between detection of vowels and detection of consonants. Only in initial position in a word do detection times for vowels approach those for stop consonants.

One difference between vowels and consonants, of course, is their role in the syllabic structure of a word. On this account, the RT differences which we have observed reflect differences in phonemic function rather than intrinsic or comparative acoustic/articulatory characteristics of phoneme categories.

This idea can be tested by comparing vowels with the semivowels /j/ and /w/. These phonemes have been characterized [9] as neither consonantal nor vocalic, in that they function in syllable structure like consonants, but are acoustically more similar to vowels. Indeed, Ladefoged [10] characterizes /j/ and /w/ as "non-syllabic versions of the English high vowels /i/ and /u/ respectively" (p. 209). This makes our choice of target vowels an obvious one.

If the obtained differences in RT between vowels and consonants are related to the functioning of a sound in syllable structure, we expect the semivowels, which function as consonants, to be faster than the vowels in the same way that stop consonants were. On the other hand, if the differences are due to acoustic structure of vowels versus consonants, then we would expect that semivowels, which in acoustic structure resemble vowels more closely than they resemble stop consonants, should produce a response pattern more similar to that of vowels.

3. EXPERIMENT 2

Method

Materials and design. Target phonemes were high front /i/ and high back /u/ plus their corresponding semivowels /j/ and /w/, respectively. Due to restrictions of the language, comparisons of /j/ & /i/ were limited to initial

and medial position and comparisons of /w/ & /u/ were limited to medial position only. The experiment was otherwise the same as Experiment 1.

Subjects. Twenty four native speakers of British English aged 19 to 46, participated for payment in the experiment. All reported normal hearing. Six heard each order of presentation of the blocks.

Results

This time, the two vowels were responded to significantly faster than the two consonants (Fl [1,20] = 106.06, p < .001; F2 [1,140] = 39, p < .001). Fig. 2 shows mean RTs in ms for the vowels (532 ms) vs. the semivowels (635 ms). The difference was significant for all sub-comparisons: medial /u/ (524 ms) versus medial /w/ (609 ms; t1 = 4.34, p < .001, *tl* = 4.9, p < .001); initial /i/ (507 ms) versus initial /j/ (649 ms; rl = 7.38, p < .001, *tl* - 2.07, p < .06); medial /i/ (563 ms) versus medial /j/ (668 ms; rl = 4.72, p < .001, r2 = 3.71, p < .001).

An error analysis showed that 17% of the semivowels were missed; this was a significantly higher percentage than the one for the vowels at 6% (Fl [1,20] = 16.67, p < .01; F2 [1,140] = 43.61, p < .01). There was. no significant difference within the vowels, but within the semivowels word-medial /j/ was missed more often than word-medial /w/ and initial /j/, (Fl [2,40] = 11.35, p < .01, F2 [2,66] = 31.55, p < .01). Again, a negative correlation between RT and vowel duration (the longer the vowel, the faster the RT) appeared, but only for /u/, not for the other phonemes.

This pattern of results conclusively rules out an explanation of the previous findings in terms of syllabic function. However, the fact that a significant RT difference was found - this time in *favour* of vowel RTs - suggests that an explanation in terms of acoustic/articulatory structure of phonemes may also not be a simple matter.

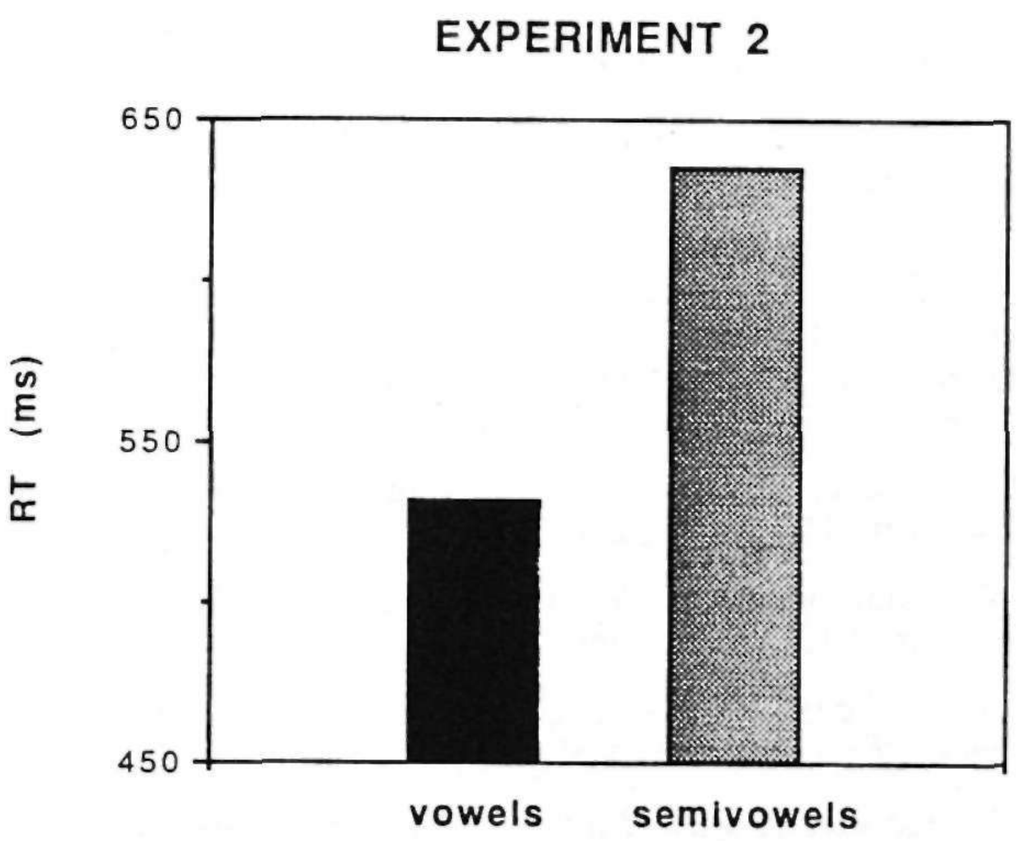

Figure 2. Mean reaction time (ms) as a function of phonemic category (vowels or semivowels). Experiment 2.

4. CONCLUSION

The two experiments have shown that English vowels are hard to detect relative to stop consonants, but easy relative to semivowels. Experiment 1 showed a strong disadvantage for vowels, replicating previous experiments. Experiment 2 showed that this disadvantage can not be explained in terms of the functioning of a sound as either a vowel or a consonant in the syllabic structure of a word, because in this case the target phonemes which function as consonants were responded to more slowly.

One factor which clearly plays some role in these findings is orthographic interference. Cutler et al. [8] found evidence for an orthographic effect in the detection of the vowel schwa: responses to this vowel were faster when the orthographic representation was "e", suggesting that "e" may act as the canonical orthographic representation for schwa. Similarly, orthography played a role in the large number of missed responses for word-medial /j/ in Experiment 2. In experimental words like "dune", "cubic" and "fuse", there is no corresponding grapheme for the phoneme /j/. (For instance, there is no difference in spelling to indicate the presence of the phoneme /j/ in British English "duty" as opposed to the absence of this sound in some varieties of American English.) If subjects indeed have a canonical orthographic representation of sounds, and this facilitates responses to target phonemes which are orthographically represented in the canonical form, then it is likely that responses will be even slower when there is no corresponding orthography whatsoever. The results of the error rate analysis support this suggestion: /j/ was significantly more often missed in word-medial position, where no orthographic symbol was available, than in word-initial position, where the orthographic representation was always "y". It would seem that subjects were not accustomed to making purely phonetic judgements.

However, orthography can not provide the entire explanation. In Experiment 1, for instance, /a/ was constantly represented by orthographic "a", yet it was detected more slowly than /i/, which had considerable orthographic variation ("e", "ee", "ea", "ie"). Moreover, if it can be argued, that orthographic "a" is an ambiguous symbol because it can also represent other vowel sounds such as in "back", the same argument should apply to the consonant targets used: orthographic "p" occurs in "photo" as well as in "pole", and "t" occurs in "thin" as well as in "tin". Yet there was no doubt that the stop consonant RTs were significantly shorter than the RTs to each vowel. The strongest evidence that there is more in these findings than can be explained by orthography comes, however, from our previous work: Cutler et al. [8] found long RTs for vowels in nonsense words. Subjects can have no prior orthographic representation for nonsense words; if they construct an orthographic representation in order to perform the phoneme detection task, then surely they must be free to construct it solely in terms of putative canonical representations.

We suggest that the explanation for the pattern of RT differences which we have observed must be sought in acoustic/articulatory characteristics of different phoneme categories. One interesting possibility concerns variation within individual phoneme categories. Ades [11] has pointed out that the effective range of any vowel category, as measured in number of just noticeable differences (JNDs), is larger than the effective range of any consonant category. (He used the notion of JND's to explain apparent differences in categoricality between vowels and consonants [1].) This means that the perceptual representation of a vowel varies over more JND's within one and the same category.

There are several ways in which such variability could affect performance in the phoneme detection task. Firstly, there could be variability in the realisations of individual tokens of the target in our stimulus materials. Consistent with this suggestion is the finding in Experiment 1 that RTs to vowels and to consonants were *not* significantly different in word-initial position, in which no prior context distorts phonemic realisation.

Secondly, there could be variability in the representation which subjects form of the specified target. Indeed, there exist independent arguments that memory representations for vowels change over time. Cowan [12], in an attempt to explain memory decay in vowel discrimination, has proposed that mental representations of vowels may gradually become more diffuse over time. Given the larger effective perceptual range of vowels, as noted by Ades, Cowan's suggestion would imply that vowel representations can drift further from their canonical representation, while still remaining in the same phonemic category, than consonants. If vowel representations indeed drift in this way, then again there will be greater variability in the degree to which any individual token in the stimulus materials matches the subject's representation. If this type of variability is exercising a large effect, we would predict a greater disadvantage for vowels when subjects are required to retain a target representation over long intervals. In the form of the detection task which we used, the interval between target specification and stimulus was up to 20 s.

It should be noted, however, that although the representation a subject uses for performing this particular task may change over time, it is not the case that the memory representation for a phonemic category disappears altogether. In a similar phoneme detection experiment we asked subjects to recall all of the target phonemes after the experiment; most of them did so quite easily.

Our explanation rests in part on the assumption that, like vowels, semivowels have a comparatively large perceptual range. This seems intuitively plausible; semivowels are acoustically closer to vowels than to consonants. However, it does not explain their added disadvantage in relation to the vowels in Experiment 2. Significantly, in Experiment 2 we even found a difference between the vowels and semivowels in word-initial position. As we have already noted, part of the difficulty of semivowel targets may be due to their lack of consistent orthographic representation; again, however, RTs to semivowels were significantly slower than to vowels even when the orthography was consistent. Semivowels were notably shorter than the vowels (the measured lengths for the semivowels ranged from 27 ms to 111 ms, those for the vowels from 108 ms to 329 ms, so there was almost no overlap in length); but again, we found *no* correlation between RT and duration for the semivowels. Further research will be necessary to determine whether phonemic realisations and representations tend to be yet more variable for semivowels than for vowels; the extension of our comparisons to further consonant categories is also desirable.

5. ACKNOWLEDGEMENTS

This research was supported by the ESPRIT BRA program [project P3207).

6. REFERENCES

[1] **A.** Cuder, "Detection of vowels and consonants", Technical Report ESPRIT BRA P3207 "ACTS", 1990.

[2] C.J. Darwin & A.D. Baddeley, "Acoustic memory and the perception of speech", *Cog. Psych.*, Vol. 6, pp.41-60, 1974.

[3] D.H. Whalen, "Vowel and consonant judgements are not independent when cued by the same information", *Perc. & Psychophys.*, Vol. 46, pp. 284-292, 1989.

[4] D.J. Foss, "Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times", *J, Verb. Learn. Verb. Behav.*, Vol. 8, pp. 457-462, 1969.

[5] Z.S. Bond & S. Games, "Misperceptions of fluent speech", in R. Cole (Ed.) *Perception And Production of Fluent Speech*, Hillsdale, NJ: Erlbaum, 1980.

[6] J. Mehler, J.-Y. Dommergues, U.H. Frauenfelder & J, Segui, "The syllable's role in speech segmentation", *J Verb. Learn. Verb. Behav.*, Vol. 20, pp. 298-305, 1981.

[7] D.T. Hakes, "Does verb structure affect sentence comprehension?", *Perc. & Psychophys.*, Vol. 10, pp. 229-232, 1971.

[8] A. Cutler, D. Norris & B. van Ooyen, "Vowels as phoneme detection targets", *Proceedings of the International Conference on Spoken Language Processing*, Kobe, Japan, Vol. 1, pp. 581-584, 1990.

[9] N. Chomsky & M. Halle, *The Sound Pattern of English*, New York: Harper & Row, 1968.

[10] P. Ladefoged, *A Course in Phonetics*, New York: Harcourt Brace Jovanovich, 1982.

[11] A.E. Ades, "Vowels, consonants, speech and non-speech", *Psychol. Rev.*, Vol. 84, pp.524-530, 1977.

[12] N. Cowan & P.A. Morse, "The use of auditory and phonetic memory in vowel discrimination", *J. Acoust. Soc. Amer.*, Vol. 79, pp. 500-507, 1986.