MINIMAL DURATION FOR PERCEPTION OF FULL-SPECTRUM VOWELS

Rudolf Weiss

Western Washington
University
Bellingham, WA, USA

W. S. Brown, Jr.

University of Florida
Gainesville, FL, USA

Shaw N. Gynan

Western Washington
University
Bellingham, WA, USA

ABSTRACT

This study developed a perception test to determine the minimal duration threshold levels of vowels on the basis of short bursts of complete waveshapes (full spectral cues) of five vowels [i e a o u].

1. INTRODUCTION

Although considerable and meaningful work has been done in the area of vowel perception over the last several decades, recently developed and fairly accessible instrumentation is now available which allows for relatively easy access and manipulation of the speech signal [3]. Different approaches have been used such as bursts of vowels at set-time intervals with manipulation of F_1 and F_2 frequencies [2,3]. Other studies have involved masking techniques [5] and still others have dealt with vowel formant transitions for vowel vs. consonant vs. semi-vowel identification [4]. Most vowel perception experiments share in common the fact that they use synthesized vowels with manipulations of F_1, F_2 and/or F_3 relative to each other in frequency, band-width and/or synchrony. Shortcomings of some of these models have been shown by Bladon [1].

Since hitherto most experiments have dealt with synthesized vowels and manipulations of the spectra in efforts to isolate specific functions of distinct acoustic cues, it was decided to experiment with complete waveshapes (full spectral cues) of steady-state portions of vowels to determine on the basis of short bursts the minimal durational thresholds for consistent vowel classification. It was hoped that acoustic cues, it was decided to experiment with complete waveshapes (full spectral cues) of steady-state portions of vowels to determine on the basis of short bursts the minimal durational thresholds for consistent vowel classification. It was hoped that we could also thereby ascertain something about the degree of difficulty in vowel perception as the time duration of bursts decreased, i.e., to verify through other means that high vowels [i] and [u] are generally easier to classify as maintained in Liebermann [5] and as found in previous cross-language studies by Weiss [7,8], showing that durational variation affects the high vowel [i] less than other vowels.

2. PROCEDURE

Five vowels [i e a o u] were produced in steady-state fashion by a male speaker (F_0 = 100 Hz ± 2 Hz) and a female speaker (F_0 = 201 Hz ± 3 Hz). These vowels were digitized using the MacSpeech Lab II/MacAudio II hardware/software program. A sampling rate of 44 KHz was used in the recording of the utterances which yielded a frequency response ceiling of 20 KHz. Using built-in routines of the MacSpeech Lab program, the utterances were equalized in amplitude and segmented on the basis of full-wave displays. They were then seg-mented first into 300 ms segments (which served as the reference cue in the perception tests) and then into smaller whole-wave units. The formant distribution figures (LPC) for both the male and female utterances are given below:

Male:	F_0	F_1	F_2
[i]	100	285	2405
[e]	101-102	408	2242
[a]	98-99	652	1019
[o]	101-102	489	775
[u]	99-101	285	775

Female:	F_0	F_1	F_2
[i]	201-204	285	2691
[e]	198-201	449	2405
[a]	201-203	530	1223
[o]	199-200	245	571
[u]	201-203	408	775

Segments were cut from the mid-point of each vowel. From the male speaker sample segments of increments from one to four complete cycles yielded four samples in duration from 10 to 40 ms. A parallel procedure was followed for the female speaker. However, since the F_0 was twice that of the male, one to eight complete cycles yielded samples in duration from 5 to 40 ms. In addition, a one-half cycle segment of each vowel beginning with the first positive rise of the wave was isolated, yielding additional segments of 5 ms for the male and 2 ms for the female. Thus the male voice yielded five segments of each vowel for a total of 25 segments. Two tests were developed: one for each voice, in which each token occurred three times. This resulted in two perceptual test tapes: one of 75 tokens for the male voice and one of 135 tokens for the female voice. The tokens were randomized and rerecorded at five-second intervals to minimize the effect of short auditory memory. For reference purposes, two repetitions of 300 ms tokens of each vowel for the male and female voice were given at the onset of each test.

Both tests were administered individually to 38 phonetically unsophisticated subjects, 16 males and 22 females, at the University of Florida. The mean age of the subjects was 20. The order of presentation of the two tests was reversed for half of the subjects.

3. EQUIPMENT

Digitizing was performed with a Mac II with 4 mb. RAM and a 68020 microprocessor with a Mac Speech Lab II/ MacAudio II hardware-software package. Analog samples from the digitized utterances were made with a Teac V-570 cassette deck. The listening tests were administered individually using a Teac W370C cassette deck in conjunction with a Technics SU-V450 integrated amplifier and a Technics Model SB-C36 two-way speaker system for the reference samples.

4. RESULTS

The results indicated a high degree of accuracy in perception of vowels of most durations. Variations in responses to individual vowels were significant only for the shortest durations. Even a one full-spectral wave cue (female - 5 ms/male - 10 ms) was long enough for fairly consistent classification. The lengthy interval of 5 ms between cues no doubt enhanced categorical perception by minimizing short auditory memory as predicted by Repp [5]. There was still sufficient cue information even if only half the spectral information for one wave form was given to enable fairly consistent identification of vowels.

It is questionable how meaningful a ranking order of vowel difficulty might be due to the high degree of correct classification of responses. However, based on a possible 1026 correct classifications of each female vowel and 570 possible correct classifications of each male vowel, the ranking order from easiest to most difficult vowel for each voice is indicated below. Percentage indicates the total errors made by all subjects to each vowel.

Male		Female	
[o]	2.1%	[o]	7.6%
[u]	5.6%	[a]	8.0%
[i]	5.8%	[i]	9.7%
[a]	5.9%	[e]	24.0%
[e]	14.9%	[u]	35.5%

It is obvious from the above statistics that the most difficult male vowel to categorize was [e], with 14.9% errors, and the most difficult female vowel to categorize was [u] with 35.5% errors. Thus prior findings that [i] and [u] are among the easiest vowels to classify are not supported by this study.

It is also apparent that the female vowels posed much greater perceptual difficulties even if only vowels of the same duration are compared. The table below illustrates comparable male/female token values. For each time variation there were 114 tokens for 38 subjects. Errors are indicated as a percentage.

TABLE 1: PERCEPTION ERRORS OF COMPARABLE M/F VALUES

ms	[i]		[e]		[a]		[o]		[u]	
	M	F	M	F	M	F	M	F	M	F
5	13.1	21.9	18.4	37.7	8.7	9.6	0	10.5	7.0	51.7
10	3.5	1.7	11.4	14.9	8.7	7.8	4.3	7.0	7.8	59.6
20	4.3	0.8	9.6	12.2	3.5	11.4	0	3.5	4.3	35.0
30	6.1	7.0	23.6	11.4	5.2	7.0	4.3	4.3	7.0	28.0
40	1.7	1.7	11.4	11.4	3.5	1.7	1.7	5.2	1.7	18.4

The study shows that in general errors in perception increase as the vowel duration decreases. An exception is the male [o] which posed no difficulty for the listeners even at the shortest duration of 1/2 wave cycle (5 ms). Recognition levels for the shortest durations were as follows:

Male (tokens for 1/2, 1 and 2 cycles)

5 ms:	90.4% (81.6-100%)
10 ms:	92.9% (88.6-96.5%)
20 ms:	95.7% (90.4-100%)

Female (tokens for 1/2, 1-5 cycles)

2 ms:	74.6% (49.2-84.3%)
5 ms:	74.7% (48.3-90.4%)
10 ms:	79.7% (40.4-93.0%)
15 ms:	84.8% (57.9-98.2%)
20 ms:	87.4% (65.0-99.2%)
25 ms:	92.8% (84.3-99.2%)

For context independent recognition of vowels the male voice obviously yields the best response. With the exception of [e] all vowels could be truncated to one wave form (10 ms) and still have 90-100% recognition. For the female voice even 2 wave forms (10 ms) would yield only 40% recognition for [u] but 85-93% for all other vowels.

This study shows that overall best results for vowel recognition occurs for two wave shapes (20 ms) for the male voice with recognition level of 95.7% (minimum of 90.4% for any vowel); for the female voice the best results are with five wave shapes (25 ms) with a recognition level of 92.8% (minimum of 84.3% for any vowel). Thus it appears that duration, not number of complete cycles, is an overriding factor in determining minimal threshold levels in perception. The threshold for highly accurate classification seems to be located at between 20-25 ms.

Analysis of variance failed to establish significant correlations regarding vowel formant spread or the effect of order of presentation. Nor could statistically significant differences between male and female subjects in accuracy of vowel identification be established. A larger data base would be necessary to confirm this finding.

5. CONCLUSION

The degree of persistence of full-spectrum cues through the shortened time window was unexpected. A high degree of accuracy in vowel perception remained even to the shortest burst which allowed perceptual/auditory access only to half of a wave shape, i.e., a time duration of little more than 2 ms cue. Optimum results were obtained in the 20-25 ms token range. The implication of these preliminary findings is that if full-spectral cues are given, an exceedingly small time frame will suffice for fairly consistent and reliable perception and classification of vowels. More than twice as many errors were made in classifying the female tokens which correlated closely to the increase of the fundamental frequency of the female voice. We plan to expand our study to allow for a larger data base in forthcoming endeavors.

N.B. This research was made possible through the use of the research facilities at IASCP, University of Florida.

7. REFERENCES

[1] BLADON, A., (1983), "Two-formant models of vowel perception: shortcomings and enhancements", *Speech communication* 2, 305-313.

[2] CHISTOVICH, I., et al. (1987). "Interval of spectral information accumulation in perception of non-stationary vowels", *Proceedings XIth ICPhS*, 1, 262-265.

[3] CHISTOVICH, L.A. (1985), "Central auditory processing of peripheral vowel spectra", *JASA,* 77 (3), 789-805.

[4] DANILOFF, R.G. (1985), *"Speech science"*, San Diego: College Hill Press, 146 pp.

[5] LIEBERMANN, P. and S. BLUMENSTEIN (1988), *"Speech physiology, speech perception, and acoustic phonetics"*, New York: Cambridge University Press, 175 pp.

[6] REPP, B.H., et al. (1979), "Categories and context in the perception of isolated steady-state vowels", *Journal of experimental psychology, human perception and performance*, 5 (1), 129-145.

[7] WEISS, R. (1976), "The role of perception in teaching german vowels to american students", *Proceedings of the IVth international congress of applied linguistics*, 3, Stuttgart: Hochschulverlag, 513-523.

[8] WEISS, R. and H.H. WAENGLER (1975), "Experimental approach to the study of vowel perception in German," *Phonetica*, 32 (3), 180-199.