

Title

Male and female vowels identified by visual inspection of raw complex waveforms

Presented June 4, 2001 at the 141st Acoustical Society of America Conference

Michael A. Stokes

MAS Enterprises

Abstract

A perceptual experiment involving visual identification of the raw complex waveforms of American English vowels from two male speakers was presented at the 131st meeting of Acoustical Society of America (Stokes, 1996). In that study, a subject correctly identified 5 out of 9 vowels for both male speakers using only a visual presentation of the raw complex waveforms. The present study replicates those earlier results with the same subject correctly identifying 4 out of 9 vowels and 6 out of 9 vowels for two new male speakers. Furthermore, 4 out of 9 vowels produced by a female speaker were correctly identified. These results demonstrate that the visual cues being used for vowel identification can be applied across genders, in addition to extending the work from 1996 to now show successful visual identification of vowels across 4 male talkers using only raw complex waveform displays. These results will be discussed, including an analysis of the errors made by the subject. A description of the visual cues and the new model of vowel perception and production that resulted from these cues can be found at <http://www.indy.net/~masmodel>.

Objective

The goal of this experiment was to add to the experimental evidence supporting the MAS Model of Vowel Perception and Production (Stokes, 1998). The foundation of the MAS Model is that categorical boundaries are established by the number of F1 cycles per pitch period, with each categorical pair being differentiated by a high and low F2 frequency. Stokes (1996) demonstrated the successful identification of raw complex waveform displays produced by 2 male talkers. Identification of 2 new male talkers would bring the total to 4 males, demonstrating the robustness of the visual cues found in raw complex waveforms. Furthermore, identification of the waveform displays of a female would demonstrate that the visual cues in waveforms can be used to identify vowels across genders.

Method

The productions of the two males and one female used in the visual experiment were taken from a database created and maintained by Dr. John Mullennix at the University of Pittsburgh-Johnstown. The database consists of 50 male and 50 female Midwestern speakers producing three tokens of h-vowel-d (hVd) words. The recordings were made using CSRE software and displayed using TFR. A Compaq Presario MV500 15 inch monitor was used to view the waveform displays.

In order to extend the results of Stokes (1996), the same methods were used. Subject MS was presented with 9 American English vowels from each speaker in random order with an experimenter present to insure no auditory cues were available to the subject. The vowel /a/ was excluded because of the inability of some Midwestern talkers to accurately produce this vowel. Subject MS selected approximately 20 ms from the entire vowel for inspection, and took notes on each vowel before writing a final answer. Each talker was presented to MS one at a time, with another experimenter scoring the responses.

Results

The organization of the vowel space described in the MAS Model is shown in Table 4 as a reference to understand the results. The results of the visual identification for the male talkers are displayed in Tables 1 and 2, and the female results are shown in Table 3. In Tables 1, 2, and 3 the vowel spoken by the speaker and the perceived vowel are displayed. By determining the number of categories that a response differed from the intended vowel, the misperception of F1 cycles can be determined. Similarly, the degree to which F2 frequency information has been interpreted correctly can be assessed (for example, perceiving a high F2 frequency when a low F2 was produced). Analyzing these errors will eventually lead to improved training methods for visual inspection of waveforms which will target areas of the visual displays that are particularly difficult for a subject. The errors from these three visual tests indicate difficulty with identifying vowels in the higher categories (these categories have more F1 cycles than the lower numbered categories). In the future, training sessions will concentrate on these particular vowels and the visual cues that are currently being missed.

Table 1
Male Talker 1

<u>Vowel Produced By Talker</u>	<u>Perceived Vowel by Subject MS</u>	<u>Comments</u>
1) head	hawed	same category, wrong F2 interpretation
2) heard	heard	correct
3) hid	hid	correct
4) hawed	hud	wrong category by 1, correct F2 interpretation
5) who'd	who'd	correct
6) hud	head	wrong category by 1, correct F2 interpretation
7) had	had	correct
8) heed	heed	correct
9) hood	hood	correct

Table 2
Male Talker 2

<u>Vowel Produced By Talker</u>	<u>Perceived Vowel by Subject MS</u>	<u>Comments</u>
1) heed	heed	correct
2) hid	hid	correct
3) hood	heard	confused for its nearest neighbor
4) heard	hood	confused for its nearest neighbor
5) had	had	correct
6) hud	head	wrong category by 1, wrong F2 interpretation
7) who'd	who'd	correct
8) head	hawed	same category, wrong F2 interpretation
9) hawed	hud	wrong category by 1, correct F2 interpretation

Table 3

Female Talker

<u>Vowel Produced By Talker</u>	<u>Perceived Vowel by Subject MS</u>	<u>Comments</u>
1) head	hood	wrong category by 2, wrong F2 interpretation
2) heard	heard	wrong category by 1, wrong F2 interpretation
3) hawed	hawed	correct
4) heed	heed	correct
5) who'd	who'd	correct
6) hid	head	wrong category by 2, correct F2 interpretation
7) hood	hud	wrong category by 3, correct F2 interpretation
8) had	had	correct
9) hud	heard	wrong category by 2, wrong F2 interpretation

Discussion

The results indicate that the vowels in Category 1 of the MAS Model have a 100% identification rate. The errors begin to appear and increase as the category number increases. This is a result of the increase in F1 cycles within a pitch period. This creates a more complex signal making it more difficult to interpret. Table 5 shows the auditory confusion data organized by the MAS Model (Stokes, 1998). From this table, it can be seen that auditory confusions also increase as the category number increases. Presumably, the increasingly complex waveform that is created by the increase in F1 frequency results in an increase in auditory errors as it does with visual identification errors. With few exceptions, the visual error pattern corresponds to the auditory error pattern.

Reducing the F2 interpretation errors and identifying the proper category would improve with training. For example, the limited experience with female hVd productions possessed by subject MS (he had only seen waveform displays of one female prior to testing) was apparent by several errors that would not be considered a "near miss." However, the visual cues organized by the MAS Model led to the correct identification of nearly half the female talker's vowel space. Future work will study the effect of training on identification accuracy, which will be analyzed and used to facilitate training of subjects other than MS. Future work will also include identifying the relationship of the auditory system with the organization of the vowel space described in the MAS Model.

References

Michael A. Stokes (1998), "MAS Model of Vowel Perception and Production," posted on the internet at: <http://www.indy.net/~masmodel>

Michael A. Stokes (1996), "Identification of vowels based on visual cues within raw complex waveforms," presented at the 131st meeting: Acoustical Society of America.

Table 4					
	<u>F0</u>	<u>F1</u>	<u>F2</u>	<u>F3</u>	<u>F1-F0/100</u>
/i/	136	270	2290	3010	1.35
/u/	141	300	870	2240	1.59
/ɪ/	135	390	1990	2550	2.55
/ʊ/	137	440	1020	2240	3.03
/er/	133	490	1350	1690	3.57
/ɛ /	130	530	1840	2480	4.00
/ɔ/	129	570	840	2410	4.41
/æ/	130	660	1720	2410	5.30
/ɹ/	127	640	1190	2390	5.13
/a/	124	730	1090	2440	6.06

Formants values from Peterson, G.E., and Barney, H.L. (1952). "Control methods used in the study of vowels", *Journal of the Acoustical Society of America*, 24,175-184.

Table 5							
Vowels Intended by Speaker	Vowels as Classified by Listeners						
	/i/ - /u/	/I/ - /U/	/ɛ/ - /ɔ/	/æ/ - /ʌ/			
/i/	10,267 ---	<u>4</u> ---	<u>6</u> 3	--- ---			
/u/	--- 10,196	--- <u>78</u>	1 ---	--- ---			
/I/	<u>6</u> ---	9,549 ---	<u>694</u> 1	2 ---			
/U/	--- <u>96</u>	--- 9,924	1 <u>51</u>	1 <u>171</u>			
/ɛ/	--- ---	<u>257</u> ---	9,014 3	<u>949</u> 2			
/ɔ/	--- <u>5</u>	--- <u>71</u>	1 9,534	2 <u>62</u>			
/æ/	--- ---	<u>1</u> ---	<u>300</u> 2	9,919 15			
/ʌ/	--- ---	1 <u>103</u>	1 <u>127</u>	8 9,476			

Error data from Peterson, G.E., and Barney, H.L. (1952). "Control methods used in the study of vowels", *Journal of the Acoustical Society of America*, 24,175-184.