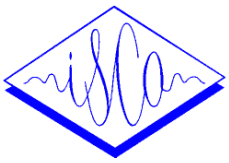

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

# THE RELATIONSHIP BETWEEN SPECTRAL DETAILS IN NATURALLY PRODUCED VOWELS AND IDENTIFICATION ERRORS IN NOISE AND REVERBERATION

Anna K. Nabelek

Department of Audiology and Speech Pathology
The University of Tennessee Knoxville, Tennessee 37996-0740

## ABSTRACT

Vowel identification was tested in quiet, noise and reverberation with 20 normal-hearing subjects and 20 hearing-impaired subjects. Stimuli were 15 English vowels spoken in a /b-t/ context by six male talkers. Each talker produced five tokens of each vowel. For hearing-impaired subjects, some vowels were confusable even in undegraded listening conditions and for normal-hearing subjects some vowels became confusable in degraded listening conditions. Apparently, both degradation of listening conditions and perceptual limitations introduced by hearing impairment caused reduction of available information in the vowels. Examination of error clusters allowed an indication of which segments of the vowels were masked by noise or by reverberation and which segments were not perceived by hearing-impaired subjects. Spectral analysis of the vowel stimuli revealed differences in detailed structure among vowels produced by various talkers and among the five tokens produced by each talker. As a result of these differences, error clusters were dependent upon talker and upon particular vowel token.

10.21437/ICSLP.1992-301

## I. INTRODUCTION

It is reasonable to assume that vowel identification errors in listening conditions degraded by either noise or reverberation and errors made by listeners with impaired hearing are a consequence of reduced information available to the listeners. The amount of information which remains in the degraded vowel or in the vowel perceived by A hearing-impaired listener is probably related to the spectrogram of the original vowel. Recent studies indicate that vowel identity is not only defined by values of formant frequencies in the vowel nucleus but also by formant trajectories or at least by values of formant frequencies of the initial, middle, and final segments of the vowel [1, 2,3]. The goal of the present study was to formulate relationships between vowel identification errors and information contained in the vowels recorded by various talkers and in various tokens of a vowel produced by one talker.

## II. METHOD

### 2.1. Test Materials

The test materials were produced by six male talkers and consisted of 15 English monophthongs and diphthongs, collectively called "vowels," /i, I, e, ɛ, æ, ɝ, ʌ, u, ʊ, o, ɔ, ɑ, ɔɪ, aʊ, aɪ/, spoken in a /b-t/ context without a carrier sentence. Each vowel was produced five times by each talker.

Formant frequencies and durations of the vowels were measured from spectrograms made with a Kay Sona-Graph Model DSP 5500. The formant frequencies were measured for the initial, middle, and final segments of the vowels when the vowels were divided into three equal segments. The initial, middle, and final values of F1 and F2 for all tokens by the six talkers were plotted in an F1-F2 space. Additional details about the study can be found elsewhere [4].

To obtain lists of vowels degraded by reverberation, the six lists were rerecorded in an auditorium. The auditorium had a volume of 4000 $m^3$ and a mean reverberation time of 1.5 s. To obtain lists of vowels degraded by noise, a noise with an envelope shaped according to a speech-like spectrum, "speech-shaped noise," was added at a speech-to-noise ratio, S/N, of 0 dB.

### 2.2. Subjects

There were two groups of subjects in this study: 1) 20 normal-hearing and 2) 20 hearing-impaired. Subjects in Group 2 had mild to moderate bilateral sensorineural hearing losses.

### 2.3. Procedures

Subjects were tested individually in a sound-treated room. The test tapes were presented monaurally at a comfortable level to the right ear for each normal-hearing subject and to the preferred ear of each hearing-impaired subject through a TDH-50 earphone.

Before data collection, subjects were informed about the procedures and were presented with practice items. During data collection, subjects were asked to repeat aloud each /b-t/ syllable and to point to their responses on a written list of the syllables. The test tapes consisted of the lists of 75 syllables (5 tokens of each of the 15 vowels). Each subject listened to each talker in quiet, noise, and reverberation, for a total of 18 test lists. The lists by the six talkers in three listening conditions were counterbalanced. Altogether, there were 100 presentations of each vowel to hearing-impaired and to normal-hearing subjects (20 subjects x 5 tokens) in each listening condition.

## III. ANALYSIS OF ERRORS

The errors were arranged in stimulus-response matrices. Each cell of the matrices was divided into six subcells corresponding to data for the six talkers. The matrices were for the three conditions and two groups of subjects.

### 3.1. Errors in Quiet by Normal-Hearing Subjects

All tokens of vowels in quiet had previously been accepted by judges as the intended targets. The few errors made by the normal-hearing subjects probably indicated that the subjects were less attentive listeners than the judges, or permitted less variability within a vowel category.

This was the case for /ʊ/ by talker DV. His tokens of /ʊ/ in /b-t/ context had different formant frequencies from his target /ʊ/ in the word "book" and from formant frequencies for this vowel reported by other authors [5]. Apparently the subjects estimated the variability within the /ʊ/ category to be unacceptably large and made identification errors.

There were frequent errors for the vowels /ɔ/ and /ɑ/. Most of the errors were confusions between /ɔ/ and /ɑ/. There were two possible sources of these errors: 1) spectral similarities of /ɔ/ and /ɑ/ and 2) poor discrimination of these two vowels by some subjects. Examination of locations of these vowels in F1-F2 space by the six talkers, revealed a range of spectral similarities

between the vowels from the overlapping of the middle segments to a considerable distance between F1-F2 locations. The numbers of /ɔ - ɑ/ and /ɑ - ɔ/ errors correspond to the degree of spectral similarities between the vowels for both groups of subjects. Individual subjects differed in their ability to discriminate /ɔ/ from /ɑ/, with the number of identification errors ranging from none to all possible (0 to 5). Reduced /ɔ/ and /ɑ/ distinctiveness in degraded listening conditions probably played only a marginal role since the numbers of either /ɔ - ɑ/ or /ɑ - ɔ/ errors were not generally greater in degraded than in undegraded listening conditions.

### 3.2 Errors When Information is Reduced by Listening Conditions or Limited by Hearing Impairment.

An examination of the matrices for the degraded listening conditions indicated that the errors were dependent upon the talker, listening condition, and subject group. Therefore, the errors had to be examined for each vowel separately.

*Talkers.* Dependence of the identification errors upon talker were anticipated because vowels by various talkers differ in location in F1-F2 space. Vowels also differ in distribution of energy along the duration of vowels, relative intensity of formant peaks, and temporal changes in formant frequency. It is logical to assume that the most important cues for identification of vowels degraded by either noise or reverberation were the same as the cues for vowel identification in quiet, namely, the frequencies of the first two formants of the initial, middle, and final vowel segments.

If the formant frequencies of the unmasked sections resembled those of another vowel, the errors were likely to be clustered. An arbitrary criterion of at least 14 was chosen to identify a cluster. There were two possible reasons why errors did not form clusters: 1) a general loss of distinctive information may have caused the degraded stimulus not to resemble any other vowel and 2) a token of a vowel may have been relatively close to the F1-F2 locations for several other vowels. The following discussion of differences in error patterns and possible sources of these differences was primarily based on clustered errors.

Vowels by some talkers were highly identifiable in degraded listening conditions, whereas the same vowels by other talkers were confusable. This was the case with /æ/. For /æ/ by talker JA, there were clusters of /ɛ/ responses in quiet, noise, and reverberation, respectively: 4, 27, and 19 by the normal-hearing subjects and 10, 32, and 44 by the hearing-impaired subjects. For /æ/ by other talkers such responses were infrequent. Examination of locations of the middle sections of /æ/ and /ɛ/ in F1-F2 space by the six talkers indicates that, although these locations for JA overlapped the locations for other talkers were relatively distant.

Errors were sometimes different for the same vowel produced by various talkers. This occurred in reverberation for the hearing-impaired subjects with the vowel /I/. For talkers MC, JS, and DV there were clusters of /e/ responses (21, 18, and 18, respectively), for talker SC there was a cluster of /ɛ/ responses (14), and for talker BS there was a cluster of /i/ responses (14). These errors were related to locations of the initial segments of /I/ stimuli and those of the responses. The subsection *Listening Conditions* explains why errors in reverberation tend to be based upon the F1-F2 proximity of initial segments of the stimulus and the response.

*Listening Conditions.* Both noise and reverberation reduce the information available in the vowels. In noise, the information is primarily reduced because formant frequency peaks may be masked in some segments of the vowel. In reverberation, formant frequency peaks are smeared in time. An illustration of spectral changes in noise and reverberation is given in Fig. 1. The three-dimensional spectrograms of /æ/ by talker JA were obtained using the Signal Technology, Inc. ILS computerized system. For the given example, the number of errors by the normal-hearing subjects were 4, 28, and 20, and the number of errors by the hearing-impaired subjects were 10, 34, and 48 in quiet, noise, and reverberation, respectively.

The difference between masking by noise and by reverberation apparently relates to the segment of the vowel which is affected. Noise can mask any segment (initial, middle, or final) which is less intense than the remaining segments. As talkers differ in distribution of energy along the duration of the vowel, either the initial, middle, or final segment of the vowel may avoid masking and remain available for identification. Reverberation overlaps the energy of a preceding phoneme over the following vowel and smears energy along the

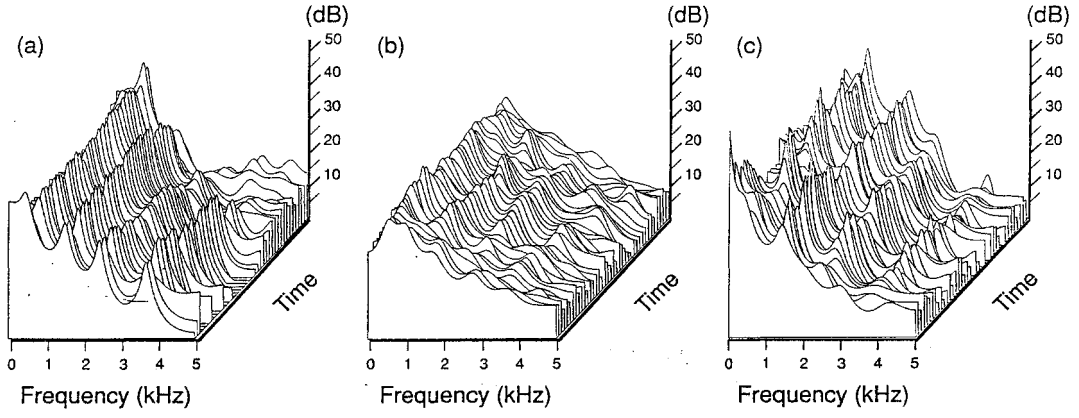


*Fig. 1. Three-dimensional spectrograms of /æ/ by talker JA in (a) Quiet, (b) Noise, and (c) Reverberation.*

duration of the vowel [6]. In this study, the effect of overlapping was negligible because the energy of the preceding consonant /b/ was low, relative to the energy of the following vowels. Reverberant degradation of the vowels was, therefore, mainly due to temporal smearing. The smearing effect tends to be greater for the final than initial segments of stimuli. As a consequence, the sections unmasked by noise (the most intense) and by reverberation (initial or middle) might carry different information about the stimulus, and the errors in the two degraded listening conditions might be different.

Responses of /i/ to /e/ produced by JA also reflected the unmasked vowel segments. The final segments of his /i/ and /e/ tokens overlapped in F1-F2 space. One of his /e/ tokens (Fig. 2a) had relatively lower intensity of the initial than the final segment. In noise, the intense final segment of this token was probably less masked than other segments and both the normal-hearing and the hearing-impaired subjects gave /i/ responses (17 and 23, respectively). This error did not occur in reverberation probably because the final segment, although more intense than the initial segment, was masked and did not play a role in identification. The error of /i/ for /e/ occurred rarely for other tokens (example shown in Fig. 2b) produced with even intensity along the duration of the vowel.

The opposite case occurred for /e/ by talker MC, an example of which is shown in Fig. 2c. His /e/ tokens were sufficiently distinct for correct identification in noise. In reverberation, however, there were 70 /ɛ/ responses by the normal-hearing subjects. Tokens of /e/ by MC were characterized by intense initial segments and declining intensities along durations of the vowels. The initial segments of /e/ and /ɛ/ were proximal in F1-F2 space, although the middle and final segments were not. When reverberation masked the final segments of /e/, identification apparently was based on the initial segments which resembled /ɛ/.

*Type of Subject*. The errors made by the hearing-impaired subjects can be viewed as a consequence of the perceptual limitations of the listeners. Several perceptual limitations such as decreased sensation level, impaired frequency resolution and impaired temporal resolution probably play a role in vowel identification by the hearing-impaired subjects. Exactly how all these and other aspects of hearing affect vowel perception is not yet known. The simplest perceptual limitation is the inability to hear less intense segments of the vowels and to extract peaks in the vowel spectra.

In degraded listening conditions, the perceptual limitations of the hearing-impaired subjects were combined with the reduced information carried by the signals. The spectrograms in Fig. 1 illustrate that formant peaks can become very small in noise and smeared in time by reverberation. As a consequence, the number of errors made by the hearing-impaired subjects were usually greater than the number of errors made by the normal-hearing subjects.

When the normal-hearing subjects made clustered errors, the hearing-impaired subjects usually made the same errors. However, the number of errors was greater for the hearing-impaired than for the normal-hearing subjects.

The diphthongs were generally more confusable for the hearing-impaired than for normal-hearing subjects, especially in reverberation. Most of the errors made by both groups were diphthongs identified as their initial monophthongs. The hearing-impaired subjects also made confusion errors between diphthongs. For example, (/aɪ/ was perceived as /au/. Diphthong confusions were very

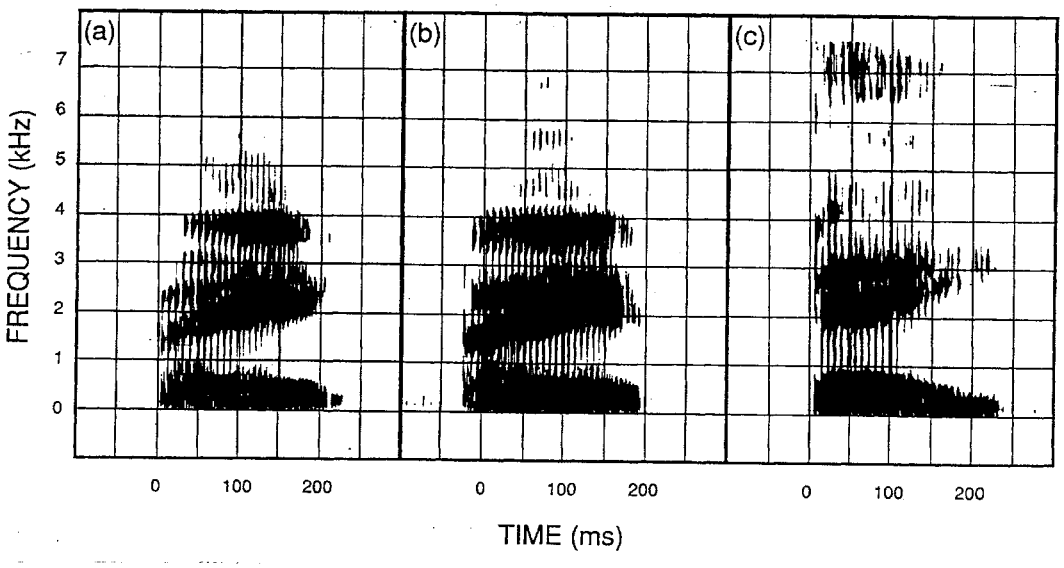


*Fig. 2. Spectrograms of /e/: (a) by talker JA with low intensity of the initial segment, (b) by talker JA with high intensity of the initial segment, and (c) by talker MC.*

infrequently made by the normal-hearing subjects.

Cluster errors by the normal-hearing subjects could be related to overlap or proximity of segments of the stimulus and response in F1-F2 space. Some cluster errors by the hearing-impaired subjects, however, appeared to be related to proximity of only F2 of the stimuli and responses, while F1s were distant.

## IV. DISCUSSION

The comparison of errors for individual vowels by the six talkers indicate that the errors depended upon the spectral details of the vowels. In quiet, most monophthongs and all diphthongs were correctly identified as the intended targets by the normal-hearing subjects. The frequent errors for /ʊ/ by talker DV were probably related to peculiarities in these tokens which did not match his own target /ʊ/ in the word, "book," or /ʊ/ reported by Peterson and Barney [5]. The errors for /ɔ/ and /ɑ/ by one talker were likely a consequence of spectral similarities between these two vowels. The errors for /ɑ/ by the remaining five talkers appeared to be related to poor discrimination between /ɑ/ and /ɔ/ by many subjects.

Errors in either noise or reverberation and errors made by subjects with impaired hearing seemed to be a consequence of reduced information available for the listeners which was insufficient for correct identification. When the available information resembled another vowel, the errors were clustered. When the available information did not resemble any other vowel or resembled several other vowels, the errors were unclustered. For the monophthongs, the most frequent errors were other monophthongs which had F1 and F2 similar to the formant frequencies of the unmasked segments of the stimuli. For the diphthongs, the most frequent errors were monophthongs resembling the initial and middle segments of the diphthongs. The errors which did not form clusters were more likely made by the hearing-impaired than by the normal-hearing subjects. Number and type of errors were talker dependent.

Generally, for the monophthongs, the errors in noise differed from errors in reverberation apparently because different segments of the vowel were masked and, in consequence, different segments were available for identification. Occasionally, errors were also made for some tokens of a vowel which were slightly different from the remaining tokens. Sometimes errors appeared only in noise or only in reverberation. Diphthong responses to monophthong stimuli were infrequent. The hearing-impaired subjects made the same errors as the normal-hearing subjects, although they made more of these errors. In addition, the hearing-impaired subjects made errors which were infrequently or never made by the normal-hearing subjects.

For the diphthong tokens, errors were made in both noise and reverberation or only in reverberation. There were never errors made only in noise. The clusters of errors were usually the same in noise and reverberation. As with the monophthongs, the hearing-impaired subjects made the same although more frequent errors as the normal-hearing subjects. Hearing-impaired subjects also made errors which were infrequently or never made by the normal-hearing subjects. Only in reverberation did the hearing-impaired subjects confuse some diphthongs. The confusions were dependent on the talker and mostly involved /ɑI/ perceived as /ɑu/.

The models of vowel perception based on spectral representation of vowels would predict most errors in degraded listening conditions for normal-hearing subjects, provided that reduction of information in the vowels can be determined. The amount of information that is reduced depends upon the distribution of intensity along vowel duration and upon relative intensities of formants. When the available information is sufficient for correct identification, no errors are made.

At the present time, it is not known what constitutes sufficient information for correct vowel identification. Vowel identification data in noise and reverberation provided examples of: 1) vowels which remained highly identifiable in one or both of the degraded conditions, 2) vowels which were frequently perceived as another vowel, causing error clusters, and 3) vowels which became confusable, but errors did not form clusters.

In the case of listeners with impaired hearing, the information is perceptually limited and the identification consequences are similar to those in degraded listening conditions for the listeners with normal hearing. Hearing impairment and degraded listening conditions are two factors which, in combination, drastically reduce available information in vowels.

The data of this study have demonstrated that the patterns of errors may not only be talker specific but also specific to one utterance. Vowel identification errors were related to details in spectra of the undegraded vowels. Although these details may be irrelevant for vowel identification in undegraded listening conditions by normal-hearing listeners, they may significantly affect vowel identification in degraded listening conditions and by hearing-impaired listeners.

**Research supported by the National Institute for Deafness and Communicative Disorders.**

**References**

[1] W. Strange. "Dynamic Specification of Coarticulated Vowels spoken in Sentence Context," *Journal of the Acoustical Society of America*, vol. 85, pp. 2135-2153, 1989.

[2] C. B. Huang. "An Acoustic and Perceptual Study of Vowel Formant Trajectories in American English." *Research Laboratory of Electronics*, MIT, Cambridge, MA, 1961.

[3] J. E. Andruski and T. M. Nearey. "On the Sufficiency of Compound Target Specification of Isolated Vowels and Vowels in /bVb/ Syllables," *Journal of the Acoustical Society of America*, vol. 91, pp. 390-410, 1992.

[4] A. K. Nabelek, Z. Czyzewski, and L. A. Krishnan. "Identification of Vowels Produced by Different Talkers," *Journal of the Acoustical Society of America*, in print, 1992.

[5] G. E. Peterson and H. L. Barney. "Control Methods Used in a Study of Vowels," *Journal of the Acoustical Society of America*, vol. 24, pp. 175-184, 1952.

[6] A. K. Nabelek, T. R. Letowski, and F. M. Tucker. "Reverberant Overlap- and Self-Masking in Consonant Identification," *Journal of the Acoustical Society of America*, vol. 86, pp. 1259-1265, 1989.