

Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners^{a)}

Diane Kewley-Port,^{b)} T. Zachary Burkle,^{c)} and Jae Hee Lee^{d)}

Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 26 March 2007; revised 27 July 2007; accepted 30 July 2007)

The purpose of this study was to examine the contribution of information provided by vowels versus consonants to sentence intelligibility in young normal-hearing (YNH) and typical elderly hearing-impaired (EHI) listeners. Sentences were presented in three conditions, unaltered or with either the vowels or the consonants replaced with speech shaped noise. Sentences from male and female talkers in the TIMIT database were selected. Baseline performance was established at a 70 dB SPL level using YNH listeners. Subsequently EHI and YNH participants listened at 95 dB SPL. Participants listened to each sentence twice and were asked to repeat the entire sentence after each presentation. Words were scored correct if identified exactly. Average performance for unaltered sentences was greater than 94%. Overall, EHI listeners performed more poorly than YNH listeners. However, vowel-only sentences were always significantly more intelligible than consonant-only sentences, usually by a ratio of 2:1 across groups. In contrast to written English or words spoken in isolation, these results demonstrated that for spoken sentences, vowels carry more information about sentence intelligibility than consonants for both young normal-hearing and elderly hearing-impaired listeners.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2773986]

PACS number(s): 43.71.Es, 43.71.Ky, 43.66.Sr [ARB]

Pages: 2365–2375

I. INTRODUCTION

It is widely believed that speech sounds can be separated into two broad categories, vowels and consonants. Vowels are generally higher in intensity, longer in duration, lower in frequency, and the result of slower movement in the articulators than are consonants. It is often noted that the more rapid change observed for consonant articulation can be considered overlaid on the slower vowel gestures (Perkell *et al.*, 2004). In speech perception, consonants are commonly thought to carry more information about sentence intelligibility than vowels. However, this position was challenged in a preliminary report by Cole *et al.* (1996). Results from their noise-replacement paradigm showed that more information about sentence intelligibility was carried by vowels when only vowel segments were present compared to when only consonant segments were present. In order to examine the role of consonant versus vowel information in spoken sentences by persons with hearing impairment, the present research adapted the noise replacement method described by Cole *et al.* (1996).

Evidence that consonants are processed relatively independently from vowels is found in many areas of research, but is also challenged in others. In seminal work in speech perception, Fletcher (1929) [as discussed in Allen (1996)]

argued for independent processing of consonants and vowels. His approach led to the establishment of the articulation index, AI (French and Steinberg, 1947), which today is still one of the better predictors of speech intelligibility performance by persons with hearing impairment. Arguments for the separate decoding of consonants and vowels have been strongly stated by the proponents of the motor theory of speech perception (Liberman *et al.*, 1967; Liberman and Mattingly, 1985). Consonants are stated to be categorically perceived by specialized speech perception processors, while vowels are processed by more general auditory mechanisms. More recently, Caramazza *et al.* (2000) reported evidence that vowels and consonants are processed by distinct neural mechanisms. Caramazza studied errors in speech produced by two Italian-speaking aphasics who had complimentary difficulties producing vowels and consonants and concluded that vowels and consonants have independent status in language production. This independence, however, is clearly challenged by approaches to language based on the syllable as the primary unit of speech (see Studdert-Kennedy, 1998 and Studdert-Kennedy and Goldstein, 2003).

Two recent reports have used experimental techniques to investigate the role of consonant and vowel information in identifying words in sentences by normal-hearing listeners. The first one provided the inspiration for the present investigation. It was a conference proceeding by Cole *et al.* (1996), and will be referred to as Cole96 throughout this report. Cole96 used the TIMIT database and sorted the phonetic segments into three groups of sounds: consonants, vowels, and weaksons. [Weaksons or weak sonorants consisted of liquids, glides, and nasals.] Cole96's first experiment re-

^{a)} Portions of the data were presented at the 147th Meeting of the Acoustical Society of America (2004, J. Acoust. Soc. Am. 115, 2601).

^{b)} Electronic mail: kewley@indiana.edu

^{c)} Current address: Cumberland Hearing Aids, 402 Bogle St. #3, Somerset, KY 42503. Electronic mail: zach@entusa.com

^{d)} Electronic mail: jaejalee@indiana.edu

placed the information associated with one of the three sound groups with either white noise or a periodic sound composed of sinusoids from 200 to 4000 Hz. Thirty-five high school students listened to altered and unaltered sentences through headphones at a comfortable level. Participants could listen to each sentence up to five times and typed the words they heard into a computer. Almost twice as many words were recognized when vowels (and weaksons) (87.4%) were present opposed to when consonants (and weaksons) (47.9%) were present. There was no effect of whether segments were replaced with noise or a periodic sound.

Two additional experiments were reported in Cole96. Experiment 2 demonstrated that sentence intelligibility was poor (3.1%, chance) when only weak sonorants were present. In experiment 3, Cole96 determined that the location of the segment boundaries in the TIMIT database did not strongly alter the large difference found between vowels and consonants in sentence intelligibility. Cole96 concluded overall that vowels are more important than consonants to the recognition of words in sentences such that there is a 2:1 benefit for vowel-only information compared to consonant-only information.

A second study by Owren and Cardillo (2006) also investigated the role of consonant and vowel information in speech. Methods and results contrasted in several important and interesting ways with Cole96. The hypothesis that Owren and Cardillo explored was different, namely to determine if vowels carry more indexical information about the speaker (gender, identity, age, etc.), while consonants carry linguistic cues that signal meaning. Owren and Cardillo (2006), here after "Owren2006," carefully selected word pairs that differed either by meaning (synonyms or not) or talker (same or different talker within gender). In contrast to Cole96, Owren2006 replaced the vowel and consonant segments, including observable formant transitions, with silence. Also, their task was a same-different discrimination paradigm. Although it is difficult to compare Owren2006 to Cole96, consider the data in Owren2006's experiment 2 where meaning was judged. Owren2006 reported a modest, but significant, improvement in discriminating differences in meaning when consonant-only words were presented compared to vowel-only words. Owren2006 carefully considered many reasons for the fact that their results are opposite to those found in Cole96 (see the following discussion for experiment 1). Owren2006's overall conclusion was that their results are consistent with the hypothesis that the dynamic acoustic cues available in consonants are the ones that signal linguistic meaning in isolated words.

While these recent reports have examined the contributions of vowel and consonantal information in spoken English for normal-hearing listeners, only a few have examined how cochlear hearing impairment may affect these contributions. A study by Owens *et al.* (1968) compared the performance of hearing-impaired adults listening to sets of words in which responses permitted only vowel confusions or only consonant confusions. They developed their own test with five responses, four likely vowels or "other vowel." The consonant task was the open set W-22 word list (Hirsh *et al.*, 1952) presented at 40 dB SL. Their results indicated vastly

more errors for the consonant task (46% correct) than for their own vowel task (94% correct). Thus they concluded that vowel errors were infrequent compared to consonant errors for hearing-impaired listeners. While this is not a direct test of the contribution of vowel versus consonant information to word identification, these results have been taken to support the typical statement that consonants are more important to speech recognition by hearing-impaired listeners than vowels. Apparently such word recognition results have been generalized to overall speech intelligibility, whereas the present experiment will examine intelligibility in sentences.

It is obvious that for a person with a high-frequency sloping audiogram, the audibility of higher-frequency consonant energy is sacrificed, often being completely inaudible. Miller (1951) stated that the weakness in intensity of the consonant sounds is unfortunate because the consonants are more critical for the correct interpretation of speech. This position motivated, in part, a number of studies that manipulated the relative intensity of the consonant to the vowel in short syllables, known as consonant-to-vowel ratio (CVR) studies. Increasing the CVR by increasing consonant intensity has produced some improvement in recognition. However the approach taken by Sammeth *et al.* (1999) was to maintain the level of three stop consonants and reduce the vowel intensity by 6 or 12 dB. In addition, in some conditions, just the isolated portion of the stop consonant was presented without the vowel, i.e., suggesting an extreme value for the CVR. Two notable results are that identification of the stops by hearing-impaired and normal-hearing listeners was considerably better when the vowel was present than absent, and second, that little difference in performance was obtained between the normal syllables and the ones in which the vowel intensity was reduced. The outcome of this study suggests that the presence of vowels is important for the correct identification of stops in syllables. These results challenge both the independence of vowel and consonant information, and the relative priority that consonants have in the processing of speech by hearing-impaired listeners. However, the outcome of these CVR studies may not be easily generalized to processing of words in sentences.

The purpose of this study was to compare the contribution of information provided by vowels versus consonants to sentence intelligibility when hearing is compromised in the elderly with typical hearing loss. The noise replacement paradigm developed by Cole96 was used in a modified form. In the present experiment, only two groups of sounds were investigated, either consonants or vowels, because weak sonorants were classified as consonant sounds. Sentences were presented in three conditions, with either the vowels or the consonants replaced with speech shaped noise, or unaltered. Two experiments were conducted. The first examined whether the preliminary results of Cole *et al.* (1996) would be replicated for a group of young normal-hearing listeners at a 70 dB SPL signal level (YNH70). These results provided the baseline data for the main experiment using elderly hearing-impaired listeners and a higher signal level, 95 dB SPL, to assure audibility of the sentences. Thus, in experiment 2 the differences between two groups of listeners, elderly hearing-impaired (EHI95) and young normal-hearing

(YNH95), in identifying words in altered and unaltered sentences were compared. It was hypothesized that: (a) both elderly hearing-impaired listeners and young normal-hearing listeners would perform better with unaltered sentences compared to sentences with either vowels or consonants replaced by noise; (b) the elderly hearing-impaired listeners would have lower performance than young normal-hearing listeners; and (c) young normal-hearing listeners would identify more words correctly for vowel-only compared to consonant-only sentences.

II. GENERAL METHODS

A. Overview of experimental design

The contribution of vowel versus consonant information was investigated using a noise replacement paradigm with sentences from the TIMIT database (Zue *et al.*, 1990; Garfalo *et al.*, 1990). Three listener groups were used. Two groups consisted of young normal-hearing participants and the other group of elderly hearing-impaired participants. In experiment 1 the signal level was set close to a conversation level at 70 dB SPL in order to replicate the Cole *et al.* (1996) study. The signal level for experiment 2 was set to a high level (95 dB SPL) so that the sentences would be reasonably audible for hearing-impaired listeners. Criteria for thresholds for the hearing-impaired listeners were selected to make audibility between vowels and consonants somewhat even over the 4000 Hz bandwidth. Three phonetic conditions were tested: (a) unaltered TIMIT sentences (Full); (b) sentences in which all of the vowels were replaced by noise (i.e., only consonants or Cin); and (c) sentences in which all of the consonants were replaced by noise (i.e., only vowels or Vin). Participants identified the words in the sentences during one test session 1–1/2 to 2 h in length.

B. Participants

All participants were American English native speakers from the North Midland dialect region (DR3) (Garfalo *et al.*, 1990), including Indianapolis and north. A speaker's dialect region was defined as the geographical area of the United States where he or she lived during their childhood years (age 2–10).

Two young normal-hearing groups ($N=8$, $N=16$) served as listeners in this study. They were all students at Indiana University, ranging in age from 20 to 35 years of age, and were paid to participate. Participants had pure-tone thresholds of 20 dB HL or better, at octave intervals from 250 through 8000 Hz, and normal tympanometric tracings. The right ear was used as the test ear.

Elderly hearing-impaired participants ($N=16$) were recruited through the Indiana University Hearing Clinic. Their ages ranged from 65 to 80 years ($M=72.9$ years) and they were paid to participate. Different audiometric criteria were needed to select the elderly hearing-impaired (EHI) listeners such that the vowels and consonants in the sentences would have somewhat balanced audibility. Participants were required to have a long-standing, bilaterally symmetrical, moderate sensorineural hearing loss, believed to be of cochlear origin based on acoustic reflex testing, case history, normal

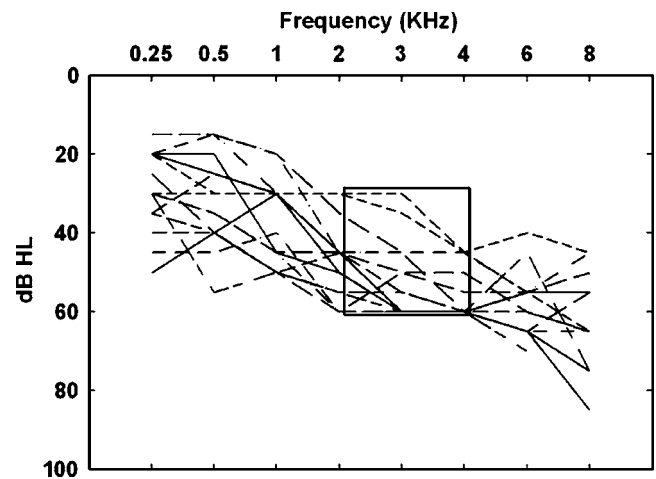


FIG. 1. Hearing thresholds for the 16 subjects in the EHI95 group. The box indicates the hearing criteria used for this study.

tympanograms, and symmetry of air and bone conduction thresholds. Hearing status criteria for subject selection were chosen such that there would be some impairment for vowels at 2000 Hz and good audibility for the consonants between 2000 and 4000 Hz. We selected 35 dB HL as the minimum loss that might affect vowel perception. Criteria for consonants were based on acoustic measurements of the standard long-term speech shaped spectrum (LTSS) (ANSI, 1969). MATLAB was used to generate samples of LTSS noise. Using a calibration vowel and the calibration procedures described in Section E, the 1/3 octave levels of the noise were adjusted to match the low-frequency spectrum of the calibration vowel set to 95 dB SPL. The 1/3 octave band level for the noise at 2000 Hz was 85.6 dB SPL and at 4000 Hz was 67.4 dB SPL. Thus the final criteria for the hearing status of the EHI listeners were thresholds between 35 and 60 dB HL at 2000 and 4000 Hz, yielding audibility for the LTSS of no more than 25 dB at 2000 Hz and no less than 7 dB at 4000 Hz.

The right ear of each EHI participant met all the criteria and was chosen as the test ear. Average thresholds for 0.25, 0.5, 1, 2, and 4 kHz were 29, 32, 37, 48, and 57 dB HL, respectively. Figure 1 shows individual pure-tone thresholds for these 16 EHI participants at octave intervals from 250 through 8000 Hz.

C. Stimuli

The stimuli used in this experiment were extracted from the Texas Instruments/Massachusetts Institute of Technology (TIMIT) corpus (Zue *et al.*, 1990; Garfalo *et al.*, 1990) of read speech. This database contains three types of sentences from 630 speakers representing eight major dialect divisions of American English. We selected the SI sentences from the North Midland dialect region, DR3, that was the best match for listeners from Indiana, Indianapolis, and further north. Forty-eight sentences containing 6–10 words each were chosen randomly with an average of 8.16 words per sentence. For the test sentences, one sentence was picked from each talker, (21 male and female sentences each). The remaining six sentences were used as training/familiarization sentences.

TABLE I. Distribution of sentences over gender and conditions.

	Number of sentences Male	Number of sentences Female	Number of words
Full	7	7	115
Vin	7	7	114
Cin	7	7	114
Total	21	21	343

The 42 test sentences were divided randomly into three test conditions ($N=14$, see Table I). Each condition had about 114 words that averaged four phones/word.

The TIMIT database provides a phonetic classification for American English consonants and vowels. We used this classification with the exception of our following three rules, which were applied in less than 5% of the altered sentences.

- (1) Stop closure symbols (i.e., “bc1”) were combined with stop (i.e., “b”) and treated as a single consonant. This yielded the 32 unique consonants in Table II.
- (2) Syllable final V+[r] as in “beer” were transcribed in TIMIT as a vowel plus the consonant [r]. These V+[r] transcriptions were considered as a single rhoticized vowel in this study.
- (3) The glottal stop [q] was a separate consonant in TIMIT, generally realized by silence. When the glottal stop was transcribed as occurring between two vowels, [VqV], it was treated as a vowel in this study.

Thus 32 consonants and 20 vowels were used in this experiment as shown in Table II. This is somewhat different than Cole96 in which 12 weak sonorants were placed into a separate category called “weaksons.”

D. Sentence processing procedures

In this experiment there were three conditions. “Full” had full phonetic information, i.e., unaltered sentences. “Vin” had the consonants replaced by speech shaped noise. “Cin” had the vowels replaced by speech shaped noise. The digital noise, based on the ANSI (1969) standard idealized long-term average speech spectrum (LTSS), was generated in MATLAB. The sample rate for the noise was the same as the TIMIT sentences, 16 000 Hz. The shape of the LTSS signal (flat from 0 to 500 Hz followed by a -9 dB/oct roll-off) was verified by measurement through the ER-3A insert earphones in a 2 cc coupler with the Larson Davis LD2800 spectrum analyzer.

The level of the LTSS replacement noise relative to the overall average level of phonemes in the sentences was set to

TABLE II. Classification of TIMIT phonetic symbols into consonants and vowels.

Group	Phone	Total
Consonant	b d g p t k dx q jh ch s sh z zh f th v dh hh hv l r y w el eng nx m n ng em en	32
Vowel	iy ih eh ey ae aa aw ay ah ao oy ow uh uw ux er ax ix axr ax-h	20

preserve level differences that normally occur for consonants versus vowels. The level differences were measured from eight sentences from the SI group (4 male and 4 female talkers). First, silence was edited out of all sentences. Then using the phonetic boundaries provided in the TIMIT database, the average level of either the consonants or the vowels was calculated. The difference in the average level of consonants relative to vowels was -10 dB. Because the replacement noise was a speech shaped noise, the noise level was referenced to the average level (avl) of all the vowels in the sentences in just the 0–500 Hz passband (approximately 95 dB SPL at the high signal level). Next, informal listening was used to determine that the consonants were not severely masked for various levels of the replacement noise for vowels. For example, equating the noise level to avl was clearly too loud. Eventually, a level of 6 dB less than avl was selected for the vowel replacement noise, and correspondingly the consonant replacement noise level was 16 dB less than avl.

The noise replacement was done using MATLAB scripts using the TIMIT .phn files. When there were adjacent consonants (e.g., CCC as in /str/) or vowels, the noise replaced the entire string. Thus, the duration of consonants and vowels (either singletons or strings) was preserved, and the average replacement noise level was at least 6 dB less than the average levels measured for the vowels or consonants in these sentences. All sentences were normalized to the average rms value measured across all 42 sentences.

E. Calibration

The signal level was set to ensure that no speech segments would be presented at a level greater than 100 dB SPL. To achieve this, first the most intense vowel (MIV_i) in each previously normalized sentence, i , was measured digitally as the rms value for a 50 ms window located around the peak amplitude. Over the 42 sentences, vowels in two sentences had the same highest value of MIV_i (both were highly stressed within their sentences). The vowel / Λ / was picked from the word “study” as the calibration vowel. A wave form piece 50 ms long around the peak was duplicated to make a 4 s long calibration signal (calvow4). The rms levels for the other MIV_i values ranged over 11 dB. However, the distribution of the MIV_i showed a narrow peak (mode) such that 71% of the MIV_i values were within ± 2 dB of the distribution peak.

All stimuli were played through TDT system II hardware and ER-3A insert earphones at a sample rate of 16 000 Hz. A low-pass filter was designed to reduce the effects of variability in high-frequency EHI audiograms. The filter was flat until 4000 Hz, had the -3 dB cutoff at 4400 Hz, and then a steep slope that fell 50 dB by 5000 Hz (>200 dB/octave). For calibration, the level was measured for the filtered calvow4 vowel played through TDT DA1 16 bit D/A to a HB6 headphone buffer. A 100 dB SPL level was set with ER-3A insert earphones in a 2 cc coupler using the Larson-Davis LD2800 spectrum analyzer and the linear weighting scale. The sentence with the calvow4 (in the word “study”) was not too loud in informal listening. With this

TABLE III. Mean and standard deviation (s.d.) of the words identified correctly by each group in each condition are shown in percent correct.

Group	Full		Vin		Cin	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
YNH70	98.58	(2.66)	74.01	(4.39)	33.99	(9.96)
YNH95	99.18	(0.98)	65.06	(5.43)	51.59	(9.77)
EHI95	93.49	(7.75)	40.13	(15.17)	20.02	(16.71)

calibration, the level for the modal peak of the MIV_i vowels was 95 dB SPL; thus, this level is referred to as “95 dB.” A second level of “70 dB” was set by attenuating the signal level by 25 dB.

Background noise was used in this study to mask any sentence editing transients. This uniform noise was generated by the TDT WG2 and was also filtered by the 4400 Hz low-pass filter and was continuously present. It was calibrated using the above-mentioned earphones and equipment to have a level less than -50 dB re the 100 dB SPL calvw4.

F. Procedures

Sentences were presented at the 70 or 95 dB level to the right ear of all participants. Listeners practiced with six sentences that were not in the test set, with feedback before beginning the experiment. Half of the participants in each group were assigned to one of two randomizations of the sentence list.

Each sentence was presented twice to the participants at their own pace. Participants were asked to verbally repeat as many of the words as they could identify. Before testing began, simple instructions were given noting that some of the sentences have some noise and sounded a little “choppy.” Participants were encouraged to guess. An experimenter sat behind the listeners and scored the responses by hand on a prepared form. A digital recorder was on during the session as backup to assist in scoring. Because the scoring task was quite easy using the form, recordings were rarely consulted.

The procedure to compute a score for each sentence was the same for all listener groups. Words circled on the paper form were correct only if they were spoken entirely correct by the listeners, e.g., no missing suffixes. For each of the two presentations, the experimenter circled the words correct, and any words heard correct in either the first or second presentation were counted. Thus the semantic meaning of the sentences was not considered in this experiment, only correct word recognition.

III. EXPERIMENT 1

The purpose of experiment 1 was to replicate some of the results in the report by Cole96. Cole96’s data were obtained at the participants’ most comfortable level (MCL). The level we selected for experiment 1 with YNH listeners was 70 dB SPL, a level that is slightly louder than normal conversational loudness but probably similar to one Cole96’s listeners selected in the laboratory.

The present experiment attempted to match several experimental variables in the Cole96 study. For example,

Cole96’s participants were high school students with no reported hearing problems. As a comparable match, eight young listeners who were students at Indiana University and passed our hearing screening participated in our experiment (called the YNH70 group). Otherwise, the stimuli and procedures for our noise replacement study were designed to be similar to Cole96 (see Sec. II).

A. Results

The young normal-hearing listeners identified words in the Full, Cin, and Vin conditions. Sentence intelligibility was scored as the total number of words correct and converted to percentage values for tables and figures. Intelligibility results for YNH70 listeners are shown in the top row of Table III. As expected, the intelligibility of unaltered sentences (Full) was nearly perfect (98.6%). The percent correct intelligibility for vowel-only and consonant-only sentences was 74.0% and 34.0%, respectively. Clearly the information available in vowels is overwhelmingly superior to that in consonants for the recognition of words in sentences by young normal-hearing listeners.

B. Discussion of experiment 1

Cole96 reported that sentence intelligibility depends more on vowel information than consonant information. Although this experiment differs from that of Cole96 in many details, such as choice of sentences, noise and noise level, and assignment of TIMIT phonetic symbols to consonant or vowel classes, the similarities of the results between the two experiments are striking. In several different noise replacement conditions reported by Cole96, higher scores were observed in the Vin conditions compared to the Cin conditions that typically yielded a benefit ratio about 2:1. In our experiment 1, the ratio of Vin to Cin was 2.2, very similar to Cole96’s 2:1 ratio.

A more careful examination of Cole96’s paper shows other results relevant to ours. Procedures for the YNH70 group most closely replicated Cole96’s experiment 1 when vowels or consonants were replaced by white noise. Interestingly, Cole96’s participants performed 13% better in both of their Vin and Cin test conditions than ours. This could be due to the addition of the 12 weak sonorants to both the Vin and Cin scores, although in their experiment 2 word identification from only weak sonorants was just 3%, indicating a small contribution to overall performance. Another reason for the higher Cole96 scores could be due to presentation level differences. Cole96’s participants had control over their own presentation level and perhaps the levels they chose

yielded higher performance than our 70 dB SPL (loud conversational speech) level. In spite of higher overall performance in Cole96, the difference between Vin and Cin was 40.0% in our experiment, which was very close to Cole's experiment 1, where the difference between Vin and Cin was 40.8%.

While there is good agreement between our results and those of Cole96, the results of Owren2006 are on the surface quite the opposite: Words with only consonants provide more linguistic information about meaning than words with only vowels. Owren2006 have already provided an excellent discussion of the many reasons why the results between their study with words and ours with sentences are discrepant. (Note that "ours" refers to both the present study and that of Cole96, henceforth to be called Cole/KP.) Four reasons will be discussed here. The Owren2006 stimuli were words with few prosodic cues, while the Cole/KP sentences were rich in prosodic and syntactic cues. Thus listeners could use considerable top-down processing in the sentence intelligibility task, but had to rely on primarily bottom-up acoustic cues to identify just single words. Given that Cole/KP replaced the segments with noise while Owren2006 replaced with silence, perhaps there was some type of masking that interfered with consonant processing in sentences. A slightly different conjecture about the noise replacement paradigm was that the noise might permit phoneme restoration differentially for vowels versus consonants. As noted by Owren2006, these possible reasons for different results between our experiments are testable.

The fourth point of discussion about these experiments deserves more scrutiny: There are many differences in how the boundaries between segments were defined across experiments. In experiments 1 and 2 of Cole96, as well as the present experiment 1, the TIMIT boundaries stored in files in the TIMIT database were used for the noise replacement studies. These boundaries were originally determined by three phoneticians and were intended to assign about half of the formant transitions to the vowels and half to the consonants. Perhaps better performance with vowels as opposed to consonants in word recognition was because the coarticulatory information available in vowels provides enough information about the surrounding consonants to enable the listeners to recognize the intended words. We are examining this in a new study wherein the consonant-vowel boundaries are being systematically shifted. In fact, Cole96 examined the effect of boundary placement in his experiment 3 that included four new experimental conditions such that the vowel or consonant boundaries were either expanded or reduced by 10 ms before replacing the segments with noise. Expanding consonants resulted in a significant improvement in word recognition, whereas expanding vowels improved recognition, but not significantly. Overall, however, the disproportionately high word recognition performance for vowel information over consonant information was maintained.

Experiments in Owren2006 were more explicitly concerned about the role of formant transitions between consonants and vowels. The importance of these "vowel margins" has been well established by Strange and her colleagues

(Strange *et al.*, 1976) in their series of silent-center vowel studies. To reduce the effects of transitions, Owren2006 edited all stimuli such that "visible formant transitions" were deleted between segments. Concerned there might still be dynamic information in the vowel-only segments, in their experiment 3 an additional 50% of the vowel signal was deleted (but consonant-only stimuli remained the same). Results showed a strong reduction in performance for the remaining vowel-only acoustic segments. The conclusions reached by Owren2006 and Cole96 were similar: Formant transitions at the edges of vowels contain more information about the neighboring consonants than the edges of consonants contain about neighboring vowels.

While the importance of the role of formant transitions versus more steady-state acoustic information needs further exploration (see Lee and Kewley-Port, 2006a, b), there are important implications for our experiment 2 that used hearing-impaired listeners in our noise replacement paradigm. The common wisdom in audiology is that the most important segments for word recognition are the consonants, not the vowels. However the present results suggest just the opposite. Namely, vowels may play a more important role for young normal-hearing listeners because they are more intense, longer than consonants, and furthermore contain considerable information about neighboring consonants in their formant transitions. Experiment 2 was designed to explicitly investigate if the role of consonants and vowels in sentence intelligibility would be the same or different for elderly hearing-impaired listeners.

IV. EXPERIMENT 2

A. Methods

Section II described the participants, stimuli, calibration, and procedures used in experiment 2. The criteria for hearing status and sound level used in this experiment were 95 dB SPL such that the vowels and consonants should have been reasonably audible to the elderly hearing-impaired listeners (EHI95). Although this level was somewhat loud for the young listeners (YNH95), this was the level of the loudest vowels in rather short sentences and informal listening indicated that the sound level was not uncomfortable.

In addition to the pure-tone audiometry and tympanometry given to the YNH70 listeners, further clinical evaluations were administered to the 16 EHI95 and 16 YNH95 participants in this experiment. Because the experiment involved repeating sentences, it was necessary to assess participants for possible memory deficits. The Mini-Mental Status Examination (MMSE) (Folstein *et al.*, 1975) and the forward and backward digit span tests were administered. The EHI95 and YNH95 participants all passed the MMSE, and the forward and backward digit span test.

In order to compare results of our experiment with common speech-recognition instruments used in clinical settings, the following speech recognition measures were given in the test ear: speech reception thresholds (SRT), word recognition scores (WRS), and the speech in noise test (SPIN, Bilger *et al.*, 1984). The SRT is a measure of sensitivity (threshold) for spondaic words (ANSI, 1996). Threshold in this study is

defined as the intensity required for 50% recognition of the spondaic words. Word-recognition abilities were assessed in quiet using CID W-22 full lists (50 words) at a presentation level of 40 dB SL [re: speech reception threshold (SRT)]. The SPIN sentences contained 25 high predictability and 25 low predictability sentences that were presented at a level of 95 dB SPL with a signal to babble ratio of +8 dB.

B. Results

1. General results

Average recognition performance for Full, Vin, and Cin sentence conditions for the listener groups are presented in Table III. Consistency of performance was assessed by considering the rank ordering of intelligibility scores across the three tasks. Spearman's rho coefficients were calculated from the YNH95 and EHI95 data that each had 16 listeners, but not the YIH70, which had too few listeners ($N=8$). As expected, rank order coefficients across tasks were small ($r < 0.3$) and not significant for YNH95 listeners because they exhibited little variability as a group. However, the rank order between tasks was moderate ($0.59 < r < 0.69$) and significant for EHI95 listeners indicating reasonable consistent performance within listeners.

Overall, as expected, both the YNH95 and EHI95 listeners performed significantly better in the Full condition compared to Vin and Cin. Clearly, listener groups had no difficulty identifying words in the unaltered sentences at 95 dB SPL. Only one of the participants in the EHI95 group scored poorly on the Full condition, 67.8% correct, whereas the range for the other 15 EHI95 listeners was 88.6%–99.1% correct such that there was considerable overlap with the YNH95 group (range 97.4%–100%). To test if sentence intelligibility in the Full condition was the same for both groups, first a two-way analysis of variance (ANOVA) with two listener groups and three sentence conditions as the repeated measures factor and number of words correct as the dependent variable was implemented. Given the expected significant group \times condition interaction ($F(2, 60)=23.84$, $p < 0.001$), the Tukey HSD post hoc analysis was applied and showed that for the Full condition, YNH95 and EHI95 listeners performed the same ($p < 0.91$). Similar results were obtained in a separate ANOVA when the number of words correct was transformed to rationalized arcsine units (RAU) to stabilize the error variance (Studebaker, 1985). The fact that these two groups scored equally well for the Full condition verified that the sentences were reasonably audible for the EHI95 group and the listening/repeating task was not difficult to perform. The Full condition was not considered in further analyses.

2. Comparison of YNH95 and EHI95 listeners

Performance by the YNH95 and EHI95 groups to identify the number of words in sentences (NW) in the vowel-only (Vin) and consonant-only (Cin) conditions is shown in Fig. 2 in percent correct. As expected, overall the YNH95 group ($M=58.4\%$) performed significantly better than the EHI95 group ($M=30.0\%$) according to a two-way analysis of variance with two listener groups and two conditions as

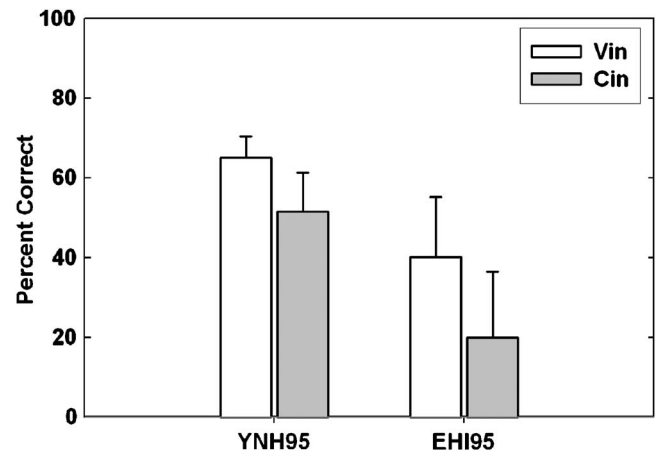


FIG. 2. Percentage of words scored correct in Vin and Cin conditions from YNH95 and EHI95 listeners in experiment 2. Error bars indicate standard deviation.

the repeated measures factor with NW as the dependent measure [$F(1, 1)=52.9$, $p < 0.001$]. Significantly better performance was obtained for the Vin condition ($M=52.6\%$) than the Cin condition ($M=35.6\%$) [$F(1, 30)=62.1$, $p < 0.001$]. There was a significant decrease in performance for both groups from Vin to Cin conditions with no significant interaction [$F(1, 30)=2.45$, $p=0.128$]. According to the Tukey HSD post hoc analysis ($p < 0.001$), performance in the Vin condition (65.1%) was significantly higher in the Cin condition (51.6%) for the YNH95 group. Also, in the EHI95 group, percent-correct performance in the Vin condition (40.1%) was significantly higher than in the Cin condition (19.9%).

C. Discussion of experiment 2

All three of our hypotheses were confirmed for experiment 2. First, both YNH95 and EHI95 listeners recognized more words in the unaltered sentences than the sentences with vowels or consonants replaced by noise. Second, YNH95 listeners performed significantly better than the EHI95 listeners, on average by 28%. As Fig. 2 clearly shows both YNH95 and EHI95 listeners found the Cin task more difficult than Vin. Even though the EHI95 group's average Cin performance was 20%, this is still well above chance given that word identification in sentences is an open response task. EHI95 participants had a wide distribution of scores for the Cin condition (0%–54.4%), and four scored less than 5% correct. Our third hypothesis was also confirmed, namely that the YNH95 group would perform better when vowel-only information was available compared to consonant-only information. Although no specific prediction about hearing-impaired listeners' performance was made, the 2:1 advantage of Vin over Cin conditions for the EHI95 listeners was unexpectedly similar to the typical 2:1 ratio found for normal-hearing listeners in Cole96 and in experiment 1.

The 2:1 ratio was not obtained, however, for the YNH95 listeners at 95 dB SPL where the ratio was 1.26:1. To examine this change in Vin advantage, a comparison of the YNH data for experiments 1 and 2 was made. In experiment 1 the signal level was set near conversational level (70 dB SPL)

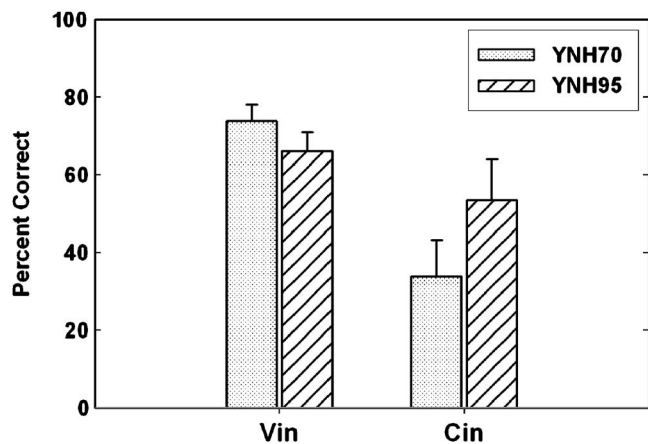


FIG. 3. Percentage of words scored correct in Vin and Cin conditions from YNH70 in experiment 1 and YNH95 in experiment 2. Error bars indicate standard deviation.

for the young normal-hearing listeners to roughly match Cole96's procedures wherein participants set their own most comfortable level (MCL). The higher signal level used in experiment 2 was set to 95 dB SPL to assure audibility for EHI listeners even though it may have introduced some distortions in auditory processing of complex speech sounds (Studebaker *et al.*, 1999). To examine possible negative effects of the high signal level on the YNH95 group ($N=16$), a comparison was made with data from the YNH70 group ($N=8$) in experiment 1. Specifically, data from the YNH70 group ($N=8$) and the first eight participants in the YNH95 group were compared (as shown in Fig. 3). According to a two-way analysis of variance with two listener groups (YNH70, YNH95) and two conditions (Cin, Vin) as the repeated measures factor, overall performance was similar between groups [$F(1, 14)=3.15$, $p=0.10$] ($M=54%$ at 70 dB, $M=58.4%$ at 95 dB) for different presentation levels. However, both condition [$F(1, 14)=111.1$, $p<0.001$] and the interaction [$F(1, 14)=3.15$, $p<0.097$] were significant. According to the Tukey HSD post hoc analyses, there was no significant difference in performance (see Fig. 3) between groups in the Vin condition ($p=0.38$). However, the YNH95 group performed significantly better (51.6%) in the Cin condition compared with the YNH70 group (34%) ($p<0.001$).

These results were unexpected. Several recent studies of the "rollover" effects of high presentation levels have shown a decrease in speech intelligibility at higher signal levels (Studebaker *et al.*, 1999; Molis and Summers, 2003; Hornsby *et al.*, 2005). While important differences exist between these studies, a reasonable generalization is that intelligibility for consonants (with higher frequency information) showed greater negative effects of increased level than for vowels where performance was less affected (Molis and Summers, 2003). The present results are consistent with results demonstrating little change for vowels as signal level increases, but are the opposite of the consonant results where the higher signal level increased performance in the Cin condition significantly by 15.6%. The above-cited experiments all used simultaneous noise, whereas the present experiments used replacement noise for the adjacent vowels. Perhaps our procedure allowed the auditory system to get more noise-free

looks at the consonant information that actually resulted in improved performance at the higher sensation level for consonants, although this is very speculative.

Summarizing, for young normal-hearing listeners, a high 95 dB SPL signal level versus a 70 dB SPL level did not change word recognition performance when vowel-only information was present in sentences, and actually improved performance for the consonant-only sentences. This improvement of consonant-only recognition resulted in a reduced vowel advantage of 1.26:1 at the 95 dB SPL level compared to the 70 dB SPL level. We note that broadened auditory filters of EHI listeners may have contributed to their overall reduced performance at 95 dB SPL compared to YNH, but based on results from Richie *et al.* (2005) it is unlikely to account for much of the large 28% difference we observed. The present experiment did not attempt to test EHI listeners at a 70 dB SPL level because the consonants would have had very poor audibility. Thus, it is reasonable to suggest that the vowel advantage at 70 dB SPL for EHI listeners would be extremely high (possibly, greater advantage than 50:1), given that consonant-only recognition was already poor at 20% at 95 dB SPL. Results for EHI95 listeners, however, demonstrated a 2:1 advantage for vowels, indicating a large relative improvement of consonant-only recognition over vowel-only recognition. Heeding the cautionary remarks of Studebaker *et al.* (1999) that audibility and signal level will interact in speech intelligibility, the result that YNH70 and EHI95 listeners have a similar 2:1 vowel advantage in the present noise-replacement task is likely serendipitous. This does not detract, however, from the overall result that listeners with typical hearing impairment will benefit from vowels more than consonants in identifying words in sentences.

D. EHI95 performance in relation to clinical tasks

Standard clinical instruments were used to assess the participant's ability to understand speech in both quiet and in noise in order to relate these clinical scores to scores obtained in the noise replacement experiment. SPIN and WRS were administered to the YNH95 and the EHI95 participants. The SPIN test has three scores, high probability, low probability, and the combination score of both. As expected, the YNH95 group performed better than the EHI95 group on all scores: 96.5% vs 90.8% on the high probability SPIN sentences; 56% vs 34.6% on the low probability sentences, and 76.3% vs 62.7% on the combined scores. For the WRS, the YNH95 group again performed better (98.1%) than the EHI95 group (88.9%).

For the elderly hearing-impaired listeners, the relationship between our novel noise replacement tasks and standard audiologic clinical measures, including audiometric thresholds, was examined using correlational analysis. Because there were many measures near 100% (SPIN High probability sentences and the word scores for Full sentences), all percentage scores were transformed to RAU. First product-moment correlation coefficients were calculated between the

TABLE IV. Correlations between standard clinical instruments and experimental conditions Cin, Vin, and Full (* $p < 0.05$).

	WRS	SRT	Spin High	Spin Low	Cin	Vin
SRT	-0.10					
Spin High	0.55*	-0.24				
Spin Low	0.42	-0.52*	0.66*			
Cin	0.64*	-0.58*	0.38	0.61*		
Vin	0.74*	-0.33	0.55*	0.66*	0.58*	
Full	0.62*	-0.30	0.51*	0.72*	0.60*	0.64*

Full, Cin, and Vin RAU scores and the clinical scores for the WRS, SRT, and SPIN High and Low probability sentences as shown in Table IV.

As expected, all speech perception measures were positively correlated with one another and negatively correlated with the SRT. The correlation coefficient, r , for $N=16$ must be greater than 0.49 to be significant at the $p < 0.05$ level. Scores for all three experimental conditions (Full, Vin, and Cin) were significantly correlated at a moderate level with the WRS and with the SPIN low probability score. The full and Vin scores were significantly correlated with the SPIN high probability score, while only the Cin scores were significantly correlated with the SRT.

Correlational analysis was also used to examine the relation among the hearing thresholds in dB HL, age, and performance separately for each of the three experimental conditions. The Cin scores were highly correlated with 2000 and 3000 Hz ($r < -0.85$), while for Vin only moderate correlations ($r < -0.60$) were obtained for 2000 and 3000 Hz. For Full, a moderate correlation of $r = -0.69$ was found for 2000 Hz. No significant correlations ($r < |0.36|$) were found between age and scores for the three experimental conditions. A better characterization of the effects of hearing status for the EHI group was obtained from various pure-tone averages as shown in Table V. The pure-tone averages of 2000, 3000, and 4000 Hz, as well as 3000 and 4000 Hz, were highly correlated, and significant, with the Cin condition. The pure-tone averages of 2000, 3000, and 4000 Hz, as well as 3000 and 4000 Hz, were correlated at a moderate level with the Vin scores.

Because high correlations were observed between the many variables, several linear regression analyses were undertaken to determine how the clinical variables were associated with the scores in our three experimental tasks. Forward, stepwise linear regression in STATISTICA (www.statsoft.com) was selected as the analysis tool. The first linear regression that focused just on audibility was con-

ducted on either the Cin or Vin scores. Only the individual hearing thresholds at 250, 500, 1000, 2000, 3000, and 4000 Hz were used as the variables in these regressions because the sentences had been low-pass filtered at 4000 Hz. Interestingly, the threshold at 2000 Hz accounted for 80% of the variance for the Cin condition, with an additional 13% contributed by thresholds at 3000, 250, and 500 Hz [total 93% of variance, $F(4, 11) = 32.2$, $p < 0.001$]. For the Vin condition, only the thresholds at 1000 and 2000 Hz contributed significantly and accounted for a total of 67% of the variance [$F(5, 10) = 4.17$, $p < 0.026$].

The second regression examined clinical variables more generally for each of the noise replacement conditions. Examining the correlation matrices in Tables IV and V, seven clinical variables were included in the regressions. First the two pure-tone averages that appeared to best represent the speech information, the "PTA Vowel" average of thresholds for 500, 1000, and 2000 Hz, and the "PTA Consonant" average of thresholds 3000 and 4000 Hz were selected. Thus, four clinical tests were included, the measures for the WRS, SRT, SPIN High, and SPIN Low sentences, and finally the elderly listeners' ages.

In the results of the linear regression between the Full RAU score and the seven clinical variables, only the SPIN Low sentence score contributed significantly [$F(1, 14) = 14.7$, $p < 0.002$] with variance accounted for as 51% ($R^2 = 0.51$). Although 12% additional variance was accounted for by WRS, it was not significant. These results are interpreted to mean that for elderly hearing-impaired persons, the identification of words in ordinary sentences, such as the Full sentences, is moderately related to the clinical tasks of scoring one word in a low predictability sentence (SPIN Low), and the identification of single words.

For the linear regression on the Vin RAU scores, only two variables significantly contributed to the regression, WRS with 55% of the variance, and the SPIN Low scores with additional variance 15% [$F(2, 13) = 15.3$, $p < 0.001$]. Notably the WRS and SPIN low scores are the same two variables that were associated with the Full sentence task, although the order of their contributions is reversed. This can be interpreted in a similar way, namely that processing of vowel-only information by elderly hearing-impaired listeners is more related to abilities to processing sound in other speech tasks, such as identification of individual words, whether in lists (WRS) or SPIN Low sentences, than to audiometric and other variables examined.

TABLE V. Correlations between pure-tone averages and experimental conditions Cin, Vin, and Full (* $p < 0.05$).

Pure-tone average	Cin	Vin	Full
500, 1000, 2000	-0.75*	-0.46	-0.42
3000, 4000	-0.81*	-0.57*	-0.43
2000, 3000, 4000	-0.9*	-0.65*	-0.53*

The linear regression of the Cin scores presented very different results. The variable accounting for 66% of the variance was the PTA for the 3000 and 4000 Hz frequencies. Thus, the poor audibility for the consonantal spectral information was the most important variable contributing to understanding consonant-only sentences. Significant additional variance of 15% was obtained with the PTA of 500, 1000, and 2000 Hz and the SRT in dB (6% variance). Together the performance by elderly hearing-impaired listeners could be predicted from these audibility factors with 87% of the variance, the highest in the regressions for the three experimental conditions.

The results of these regressions suggest that for elderly hearing-impaired listeners, audibility in the higher frequencies is a better predictor of sentence intelligibility from consonant-only information. Of interest is that variability in the pure-tone thresholds of 250, 500, 2000 and 3000 Hz together predicted the Cin scores with very high accuracy, accounting for 93% of the variance. Conversely, the WRS and the SPIN-low variables were better predictors of intelligibility from vowel-only conditions compared to measures of audibility. For the limited age range (65–80 years) spanned by these listeners, aging was not a factor contributing to reduced sentence processing. Naturally the important question of possible effects of cognitive decline with age in sentence processing in this challenging task should be systematically examined in future experiments by using aged-matched listener groups. Because the vowel-only scores were better predicted using speech evaluations, while the consonant-only scores were better predicted by audibility, it can be speculated that vowels and consonants are cognitively processed differently by the auditory system. This is similar to the finding reported by Caramazza *et al.* (2000) suggesting that vowels and consonants are likely processed by two independent neural mechanisms.

V. SUMMARY AND CONCLUSIONS

This study examined the contribution of information provided by vowels versus consonants to sentence intelligibility in young normal-hearing listeners and in elderly listeners with typical high-frequency hearing loss. A noise replacement paradigm replicated the Cole *et al.* (1996) finding that for young listeners at conversational levels, there is a 2:1 benefit for sentences that preserve only vowels compared to sentences that preserve only consonants. This same 2:1 benefit for vowel-only sentences was obtained for elderly hearing-impaired listeners in this study at a signal level where our calibration procedures assured audibility of some consonants at frequencies up to 4000 Hz. While three different experiments, Cole *et al.* (1996), and our experiments 1 and 2 with both young normal-hearing and elderly hearing-impaired listeners, have reported a large benefit for vowel-only relative to consonant-only information, numerous questions about the generality of these results can be raised. We have conducted two additional studies to answer some questions and some results are forecast here. Lee and Kewley-Port (2006a, b) examined issues related to how much information was preserved in the acoustic signal by presenting

portions of the sentences that spanned consonants and vowels (for example, onset transitions from consonants into vowels). It appears that listeners do best when they hear the vowel-only information, and that the acoustic properties of the phonemes (such as stressed versus unstressed vowels) had no effect on the results. It is not possible to speculate, however, if the large vowel benefit found for American English would be obtained for languages with quite a different phonological structure, or for whispered or distorted speech. Fogerty and Kewley-Port (unpublished) moved the consonant-vowel boundaries to extremes and found that formant transitions contribute differently to information available in consonants versus vowels for sentence-recognition tasks. Thus it is quite possible that this series of experiments may have far reaching consequences for theories of speech perception, particularly in relation to our understanding of essential roles that consonant and vowel segments play in speech understanding.

The majority of early research comparing information in vowels versus consonants previously reported focused on CV (consonant-vowel), VC (vowel-consonant), and CVC (consonant-vowel-consonant) monosyllabic words and nonsense syllables. Fletcher's early work (1929) in this area was motivated by the need for a transmission system following World War I. In his experiments, he concentrated on articulation and error patterns made by consonants and vowels. His conclusion that consonants provide more information than vowels for the recognition of isolated words was recently confirmed in Owren and Cardillo (2006). However, we note that linguistic processing of monosyllables relies on segmental, bottom-up information, while our sentence intelligibility task incorporates considerable predictive information from top-down processing. The phoneme error analysis used in Fletcher (1929) was based on discrete symbols to represent the monosyllables. However, incorrectly identified phonemes cannot reveal the actual effect that a specific phoneme has on syllable identification because acoustic information is distributed across phonemes (coarticulation). The noise replacement task removes specific acoustic information and directly measures the accuracy of word identification from the remaining distributed information. Thus our conclusion that vowels carry more information than consonants was assessed directly from acoustic information processed in the auditory system without any specific intervening symbol analysis.

Humans communicate using fluent speech (sentences), and the pitch contours, amplitude envelope, and durational cues in sentences aid in speech intelligibility. Individual words or syllables only contribute to a portion of the intelligibility in a spoken message. In the present research, vowel information was found to have a 2:1 benefit over consonant information in fluent sentences for both young normal-hearing and elderly hearing-impaired listeners. Thus, contrary to popular belief by scientists and clinicians alike, young and elderly listeners performed significantly better when the vowel information was preserved in sentences as opposed to the consonant information. A reasonable implica-

tion of this result is that design of hearing aids should verify that critical vowel information is preserved by the signal processing algorithms.

ACKNOWLEDGMENTS

We are grateful for the many contributions Larry Humes made to this study, which was a masters thesis in audiology done by T. Zachary Burkle. This research was supported by the National Institute of Health Grant No. DC-02229 to Indiana University.

- Allen, J. B. (1996). "Harvey Fletcher's role in the creation of communication acoustics," *J. Acoust. Soc. Am.* **99**, 1825–1839.
- American National Standards Institute. (1969). "American National Standards for Calculation of the Articulation Index," *ANSI S3.5, 1969* (American National Standards Institute, New York). NY.
- American National Standards Institute. (1996). "Specifications for Audiometers," *ANSI S3.6, 1996* (American National Standards Institute, New York, NY).
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (1984). "Standardization of a test of speech perception in noise," *J. Speech Hear. Res.* **27**, 32–48.
- Caramazza, A., Chialant, D., Capasso, R., and Miceli, G. (2000). "Separable processing of consonants and vowels," *Nature (London)* **403**, 428–430.
- Cole, R., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (1996). "The contribution of consonants versus vowels to word recognition in fluent speech," Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'96), Atlanta, GA, May, 1996, pp. 853–856.
- Fletcher, H. (1929). *Speech and Hearing* (Van Nostrand, New York).
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatr. Res.* **12**, 189–198.
- French, N., and Steinberg, J. (1947). "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* **19**, 90–119.
- Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (1990). "DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM," National Institute of Standards and Technology, NTIS Order No. PB91-505065.
- Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, G., Eldert, E., and Benson, R. W. (1952). "Development of materials for speech audiometry," *J. Speech Hear. Disord.* **17**, 321–337.
- Hornsby, B. W., Trine, T. D., and Ohde, R. N. (2005). "The effects of high presentation levels on consonant feature transmission," *J. Acoust. Soc. Am.* **118**, 1719–1729.
- Lee, J. H., and Kewley-Port, D. (2006a). "Sentence comprehension when formant transitions are present or absent by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **119**, 3340.
- Lee, J. H., and Kewley-Port, D. (2006b). "The use of subsegmental information in sentence comprehension with or without formant transitions by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **120**, 3347.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the speech code," *Psychol. Rev.* **74**, 431–461.
- Lieberman, A. M., and Mattingly, I. G. (1985). "The motor theory of speech perception revised," *Cognition* **21**, 1–36.
- Miller, G. A. (1951). *Language and Communication* (McGraw-Hill, New York).
- Molis, M. R., and Summers, V. (2003). "Effects of high presentation levels on recognition of low- and high-frequency speech," *ARLO* **4**, 124–128.
- Owens, E., Talbot, C., and Schubert, E. (1968). "Vowel discrimination of hearing-impaired listeners," *J. Speech Hear. Res.* **11**, 648–655.
- Owren, M. J., and Cardillo, G. C. (2006). "The relative roles of vowels and consonants in discriminating talker identity versus word meaning," *J. Acoust. Soc. Am.* **119**, 1727–1739.
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (2004). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," *J. Acoust. Soc. Am.* **116**, 2338–2344.
- Richie, C., Kewley-Port, D., and Coughlin, M. (2005). "Vowel perception in equal audibility conditions by young adults with normal and impaired hearing," *J. Acoust. Soc. Am.* **118**, 1101–1110.
- Sammeth, C. A., Dorman, M. F., and Stearns, C. J. (1999). "The role of consonant—vowel intensity ratio in the recognition of voiceless stop consonants by listeners with hearing impairment," *J. Speech Lang. Hear. Res.* **42**, 42–55.
- Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (1976). "Consonant environment specifies vowel identity," *J. Acoust. Soc. Am.* **60**, 213–224.
- Studdert-Kennedy, M. (1998). "The particulate origins of language generativity: from syllable to gesture," in *Approaches to the Evolution of Language*, edited by J. Hurford, M. Studdert-Kennedy, and C. Knight (Cambridge University Press, Cambridge, UK), pp. 202–221.
- Studdert-Kennedy, M., and Goldstein, L. (2003). "Launching language: The gestural origin of discrete infinity," in *Language Evolution: The States of the Art*, edited by M. H. Christiansen and S. Kirby (Oxford University Press, Oxford), pp. 235–254.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (1999). "Monosyllabic word recognition at higher-than-normal speech and noise levels," *J. Acoust. Soc. Am.* **105**, 2431–2444.
- Zue, V. W., Seneff, S., and Glass, J. (1990). "Speech database development at MIT: TIMIT and beyond," *Speech Commun.* **9**, 351–356.