The recognition of isolated words and words in sentences: Individual variability in the use of sentence context

Ken W. Grant[a] and Philip F. Seitz[b]
Walter Reed Army Medical Center, Army Audiology and Speech Center, Washington, DC 20307-5001

(Received 8 March 1999; revised 27 October 1999; accepted 1 November 1999)

Estimates of the ability to make use of sentence context in 34 postlingually hearing-impaired (HI) individuals were obtained using formulas developed by Boothroyd and Nittrouer [Boothroyd and Nittrouer, J. Acoust. Soc. Am. **84**, 101–114 (1988)] which relate scores for isolated words to words in meaningful sentences. Sentence materials were constructed by concatenating digitized productions of isolated words to ensure physical equivalence among the test items in the two conditions. Isolated words and words in sentences were tested at three levels of intelligibility (targeting 29%, 50%, and 79% correct). Thus, for each subject, three estimates of context ability, or k factors, were obtained. In addition, auditory, visual, and auditory–visual sentence recognition was evaluated using natural productions of sentence materials. Two main questions were addressed: (1) Is context ability constant for speech materials produced with different degrees of clarity? and (2) What are the relations between individual estimates of k and sentence recognition as a function of presentation modality? Results showed that estimates of k were not constant across different levels of intelligibility: k was greater for the more degraded condition relative to conditions of higher word intelligibility. Estimates of k also were influenced strongly by the test order of isolated words and words in sentences. That is, prior exposure to words in sentences improved later recognition of the same words when presented in isolation (and vice versa), even though the 1500 key words comprising the test materials were presented under degraded (filtered) conditions without feedback. The impact of this order effect was to reduce individual estimates of k for subjects exposed to sentence materials first and to increase estimates of k for subjects exposed to isolated words first. Finally, significant relationships were found between individual k scores and sentence recognition scores in all three presentation modalities, suggesting that k is a useful measure of individual differences in the ability to use sentence context.
[S0001-4966(00)03802-9]

PACS numbers: 43.71.An, 43.71.Gv, 43.71.Ky [JMH]

INTRODUCTION

Most speech communication occurs in circumstances that permit both auditory and visual processing of the speech signal. With few exceptions, listeners are able to integrate the visual cues derived from speechreading (i.e., lipreading) with audition to improve their speech recognition (Sumby and Pollack, 1954; ANSI, 1969).

The recognition performance when optical and acoustic phonetic information are available is determined by at least three factors: (1) the ability to extract cues from the auditory and visual signals, (2) the ability to integrate these cues, and (3) the ability to use one's language knowledge to constrain the number of possible response alternatives (Grant and Walden, 1996; Grant *et al.*, 1998; Grant and Seitz, 1998). This last factor includes knowledge of what constitutes a well-formed word within the language, knowledge of word frequency and word familiarity, and the use of morphosyntactic, semantic, and situational cues (collectively denoted as ''context''). The ability to make use of sentence context also involves working memory capacity, information processing speed, vocabulary size, and inference-making skills (Rönnberg *et al.*, 1998). In the study described here, we focused on an individual's ability to use context in sentence identification, and how individual differences in the ability to make use of sentence context may relate to auditory, visual, and auditory–visual recognition of sentence materials. Although it is important ultimately to delineate how each of these different context factors differs across individuals, the present study is much more limited in scope. We felt that it was important first to determine whether estimated differences in subjects' ability to make use of context could account for a significant proportion of variance in typical speech-recognition measures obtained under auditory, visual, and auditory–visual conditions. If an individual's ability to make use of context failed to explain significant amounts of the variability in speech recognition typically observed across hearing-impaired listeners, then the desire to further explore the various context factors might seem less well motivated.

The extent to which contextual cues may assist in speech recognition depends upon the nature of the speech sample, as well as potential individual differences in the ability to use contextual information. For example, identifying *nonsense* consonant–vowel–consonant (CVC) syllables requires that sufficient stimulus cues for each consonant and vowel segment be received accurately. However, with *meaningful*

---

[a] Electronic mail: grant@tidalwave.net

[b] Current affiliation: United States Department of State, Foreign Service.

CVC words, lexical constraints make it possible to identify words correctly without having to resolve all of the individual segments (e.g., /buk/ is not a real word, whereas /bʊk/ is, thus restricting the choice of back-rounded vowel). Similarly, words presented in isolation under auditory or auditory–visual conditions are usually harder to identify than words presented in sentences (Miller *et al.*, 1951; Boothroyd and Nittrouer, 1988). In order to understand the relationship between segment, word, and sentence-recognition performance, these contextual variables must be quantified.

One of the most familiar tests of the effects of linguistic context on speech intelligibility is the Speech Intelligibility in Noise (SPIN) test, designed to distinguish between the reception of acoustic cues and the ability to make use of linguistic information stored in long-term memory (Kalikow *et al.*, 1977). The SPIN test sought to control the amount of linguistic information by testing speech intelligibility under conditions of controlled word predictability. Sentence contexts with high (PH) and low (PL) word predictability were generated, and the difference in intelligibility between the word score in PH sentences and PL sentences (i.e., PH-PL) was used as a measure of the individual's use of context. Earlier studies using the SPIN test by Hutchinson (1989) and Schum and Matthews (1992) described results where a number of elderly listeners had PH-PL scores that were lower than expected based on normative data. The finding that some subjects demonstrate less facilitation from contextual information than others must be considered when interpreting any observed speech communication difficulties.

SPIN word scores are based on the intelligibility of the final word from each PH and PL sentence. Therefore, one has to perceive enough of the initial portion of the sentence in order to make use of semantic and syntactic information. In the SPIN test, the words comprising the low- or high-context portion of each sentence typically are subjected to the same speech-to-babble distortion as are the target words at the end of each sentence. If two listeners obtain different PH-PL scores at the same signal-to-noise (S/N) level, it does not necessarily mean that the listener with the greater difference score is able to make better use of context than the other. It is possible that the two listeners are affected differently by the speech babble interference, and the listener with the smaller PH-PL difference was not able to resolve the initial context portion of the sentence as well as the other listener. Stated differently, SPIN performance may be affected by factors related to audibility and other auditory processing differences, and not solely to the use of contextual information (Humes *et al.*, 1994).

A second problem in measuring a listener's ability to use contextual information with SPIN sentences is accounting for the interaction between PL scores and PH-PL difference scores. Because of the interdependence between PL and PH-PL scores, very low or very high PL scores necessarily limit PH-PL difference scores. Bilger *et al.* (1984) noted that raw scores reflect two different skills: the ability to listen in noise and the ability to use context, and suggested that it might be useful to transform raw scores into their normal deviates prior to any subtraction.

To circumvent some of the problems with the SPIN test, a more general approach to quantifying the effects of lexical, semantic, and morpho-syntactic context was suggested by Boothroyd and Nittrouer (1988). Using probability theory, phonemic and semantic redundancies inherent in a speech corpus can be represented quantitatively by simple power-law equations. In Eq. (1), the probability of recognizing a word is assumed to be equal to the joint probability of recognizing its component parts, or segments. If each of these parts is statistically independent and equally recognizable, then

$$P_w = P_p^n, \tag{1}$$

where P_w is the probability of recognizing the whole word, P_p is the probability of recognizing each independent segment, and n is the number of parts, or segments, in the word. However, in the case of real words, the segments are not independent due to coarticulation and structural properties of the lexicon, and the recognition of the whole word does not require that all segments be received. Therefore, for real words

$$P_w = P_p^j, \tag{2}$$

where $1 \leqslant j \leqslant n$. Application of this equation to several sets of data showed that for monosyllabic CVC words, j was approximately 2.5 (Boothroyd and Nittrouer, 1988; Rabinowitz *et al.*, 1992). Hence, for these simple three-segment stimuli, subjects responded as if the words consisted of about 2.5 independent parts (each part containing the equivalent of 1.2 phonemes) rather than three independent parts.

Equation (3), developed by Boothroyd and Nittrouer (1988), relates word recognition in isolation to word recognition in sentences,

$$P_s = 1 - (1 - P_w)^k, \tag{3}$$

where P_s is the recognition probability for words in sentences, P_w is the recognition probability for words in isolation, and k is a free parameter reflecting the degree of predictability or context of the sentence materials. The exponent k represents the effective proportional increase in the number of channels of independent information due to contextual constraints. To obtain an estimate of k, one acquires speech-recognition scores for words presented in isolation (or in nonsense sentences) and for words presented in sentences. Thus, a criticism applied to SPIN measures, namely that the recognition of the final word is affected by the audibility of previous words, and not necessarily due to contextual information, is circumvented because all of the words in the sentence are treated equally in contributing to the overall amount of contextual information. Further, because k is a ratio of two logarithms, rather than a simple difference score (as in the SPIN test), the dependence of contextual information on isolated word intelligibility is theoretically avoided.[1]

Figure 1 shows the predicted relationship between word recognition in isolation and word recognition in meaningful sentences for different k factors. The figure predicts that word recognition in meaningful sentences is predicted to be better than isolated word recognition for all values of $k > 1$

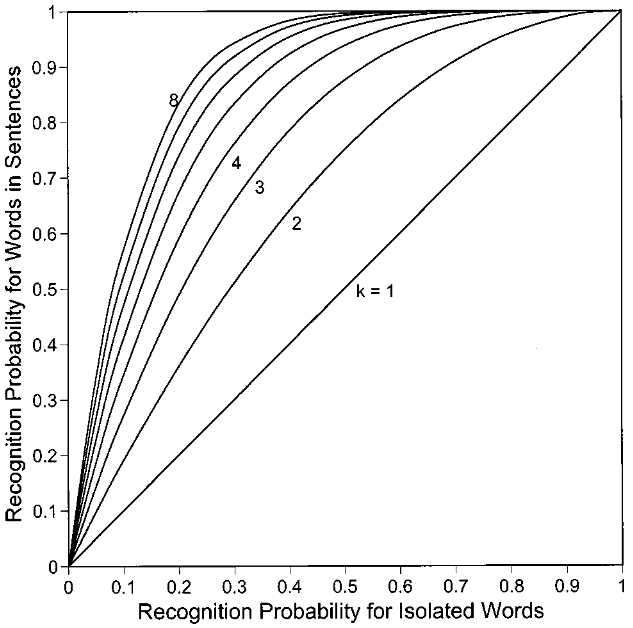

FIG. 1. Predicted isolated word recognition and word recognition in meaningful sentences for different *k* factors (after Boothroyd and Nittouer, 1988).

(with the exception of isolated word-recognition scores of zero and 100% correct) and that the difference in percent correct between words in sentences and isolated words grows more rapidly as *k* increases. Thus, for highly contextual materials, small improvements in isolated word recognition translate to large improvements in the recognition of words in sentences.

In the literature, *k* has been interpreted primarily as a property of the stimulus materials. Word and sentence-recognition scores collected from many subjects for a given corpus of speech materials are plotted and iteratively fit so as to determine a single *k* value that minimizes the variability across subjects. A lingering question, however, is whether *k* can be used to differentiate between *individual subjects* regarding their ability to use context in everyday communication. One practical concern in addressing this question is revealed by the convergence of the family of curves at either very low or very high isolated word-recognition accuracy. The theoretical curves displayed in Fig. 1 suggest that in order to demonstrate differences in *k* among individual subjects or different speech materials, it is important to control for the overall intelligibility of words presented in isolation. Isolated word-recognition scores near 30% yield the greatest range of intelligibility scores for words in meaningful sentences. On the other hand, word-recognition scores below 20% or above 80% correct show a greatly compressed range of word-in-sentence scores. It is important to point out that because *k* is a ratio of logarithms, only isolated word-recognition scores of zero or 100% correct are theoretically problematic. For example, even with isolated word-recognition scores as high as 95% correct, a large range of *k* values is possible, although very difficult to show statistically (e.g., for this case, the word-in-sentence recognition score would have to be 99.99% correct or higher to obtain *k* values greater that 3.0). Therefore, it is prudent to control for the range of isolated word-recognition scores making it feasible to demonstrate a range of *k* for individual subjects.

A second concern with treating *k* as a property of an individual rather than a property of the stimulus is that the traditional method for obtaining *k* values assumes that *k* is constant across various levels of isolated word intelligibility for a fixed speech corpus. In our application of *k* as an attribute of individuals and not materials, it follows that the amount of contextual knowledge applied to the recognition of words in sentences under unfavorable listening situations would be identical to that obtained under highly favorable listening situations. That is, an individual with a specified *k* would be predicted to perform according to the appropriate iso-*k* contour line as shown in Fig. 1. However, there is no *a priori* reason to adopt this assumption, and, in fact, it is reasonable to think that listeners may be able to modulate the degree to which contextual information is applied depending on the quality of the incoming speech signal. Thus, the recognition of speech stimuli with easily accessible information (such as clear speech presented to normally hearing listeners) might be less dependent on internal linguistic constraints than speech stimuli with less accessible information (such as conversational speech presented to hearing-impaired listeners). We will return to this point later.

Two recent studies have made use of the model of Boothroyd and Nittrouer and are germane to the present investigation. Rabinowitz *et al.* (1992) studied consonant recognition, vowel recognition, NU6 words, low-context sentences (IEEE, 1969) and high-context sentences (Boothroyd *et al.*, 1985) in 20 postlingually deafened cochlear implant users. Significant correlations across subjects were found between segment (consonants and vowels) and word intelligibility, and between word and sentence intelligibility ($r > 0.85$). For low-context sentence materials such as IEEE/Harvard sentences, the average *k* value was found to be approximately 1.14 (recall that a *k* of 1.0 implies no contextual information). For sentence sets with a higher degree of predictability [City University of New York (CUNY) sentences], *k* was approximately 4.5 (Rabinowitz *et al.*, 1992). Thus, *k* is dependent on the different morpho-syntactic and semantic characteristics of the test materials, and any variability in *k* values due to individual differences would likely be located around the stimulus-dependent *k* value.

Recently, Olsen *et al.* (1997) investigated the relationship between phoneme, isolated word, and sentence recognition in noise for listeners with normal and impaired hearing. Sentence lists were constructed from lists of CVC words (2–8 words per sentence) so that the key words comprising the sentence measures were the same words tested in isolation (at least orthographically).[2] The sentences were syntactically correct but were of low context, making them similar to SPIN-PL sentences. During testing, subjects were first presented with a list of words, immediately followed by the corresponding sentences made from those same words. Thus, although subjects were not informed that the same words would be presented in sentences, it is highly likely that prior exposure to the isolated word materials inflated the scores for the corresponding sentence materials. Because the calculation of *k* involves forming the ratio of the logarithms of the error terms for sentences and isolated words, procedural effects that increase sentence recognition relative to word recognition necessarily will increase the estimate of *k*. Presumably, a reverse ordering of the test conditions would cause a reduction in the estimate of *k*. The design of the present study will allow us to examine this issue in more detail.

The Olsen *et al.* study also demonstrated large individual differences across both normal-hearing and hearing-impaired subjects. For example, normal-hearing subjects who recognized isolated words with about 40% accuracy showed markedly different abilities when recognizing words in sentences (scores between 20%–85%). Olsen *et al.* (1997) offered no specific explanation regarding these apparent individual differences, but the data show that large individual differences exist even for subjects who acquired language under normal conditions.

In summary, context factors in language processing (e.g., the use of syntactic and semantic information to facilitate speech understanding) are often cited as important determinants of both unimodal and bisensory speech recognition (e.g., Massaro, 1987; Boothroyd and Nittrouer, 1988; Erber, 1996; Montgomery and Demorest, 1988; DeFilippo, 1990; Nittrouer and Boothroyd, 1990). In past studies on auditory–visual speech recognition in hearing-impaired subjects, large individual differences have been readily observed. Whether some of these differences in performance across individuals are due to differences in the ability to make use of context is not known. In this study, the ability of hearing-impaired individuals to use sentence context was investigated. However, no attempt is made to control for the various factors that comprise contextual information or the resources required on the part of individual to use context (such as memory, vocabulary, and speed of processing). Instead, we simply quantify possible individual differences in context use among hearing-impaired subjects by obtaining estimates of individual *k* values (essentially the difference in speech-recognition performance for isolated words and for words in sentences). This was achieved by measuring isolated word and words-in-sentence recognition at several different levels of isolated word intelligibility. Finally, individual estimates of *k* were used to predict auditory, visual, and auditory–visual recognition for sentences presented in a background of continuous speech–shaped noise. These latter conditions were included for two main purposes. First, they provide baseline measures on sentence-recognition performance to test the hypothesis that the ability to make use of context is a primary factor in determining individual differences in speech recognition. Second, if *k* is predictive of speech-recognition performance across modality, it would lend support to the idea that *k* is a measure of a subject's ability to deploy higher-level knowledge sources, independent of the modality in which the speech is presented. Overall, the results obtained from this study should be useful in evaluating whether an individual's ability to use context is an important factor in determining speech-recognition performance.

I. METHODS

This study included three parts: (1) establishing three different conditions of intelligibility (targeting 29%, 50%, and 79% correct) separately for each hearing-impaired listener using isolated words, (2) obtaining estimates of *k* for individual HI subjects at each of the three intelligibility levels, and (3) measuring auditory, visual, and auditory–visual sentence recognition. For auditory and auditory–visual sentence recognition, a continuous background of speech-shaped noise was presented.

A. Subjects

The subjects were 34 hearing-impaired (HI) adults (32 male, 2 female) between the ages of 33 and 85 years ($\bar{\chi} = 67.7$; s.d. = 11.3). Subjects had a wide variety of hearing losses and configurations ranging from normal hearing to moderate hearing loss in the low frequencies (average pure tone thresholds at 0.5, 1, and 2 kHz between 3 and 52 dB HL in the better ear) and mild-to-severe hearing loss in the high frequencies (average pure-tone thresholds at 3, 4, 6, and 8 kHz between 18 and 88 dB HL in the better ear re: ANSI, 1989). One subject had normal hearing through 6 kHz in the left ear and a moderate-to-severe loss in the right. Although the exact causes of the hearing losses were not known, most of the subjects had a history of noise exposure due to military service. All hearing losses were of sensorineural origin (as confirmed by air- and bone-conduction testing and middle-ear admittance testing) and occurred postlingually. There was no history of chronic middle-ear disease or retrocochlear signs and all subjects were fitted with hearing aids at least 1 year prior to participation in the study. All subjects were native speakers of American English with normal or corrected-to-normal vision (static visual acuity equal to or better than 20/30 as measured with a Snellen chart from a distance of 20 ft). All testing was conducted binaurally using headphones. Unfiltered speech levels (see below) were set at approximately 40 dB SL relative to the subject's better ear, or at the subject's most comfortable listening (MCL) level, whichever was lower. Subjects provided informed consent and were paid for their participation.

B. Stimuli

Speech materials consisted of isolated words and words in sentences taken from the IEEE/Harvard (1969) sentence set. This set consists of 720 phonetically balanced low-context sentences each containing five key words. The sentences are organized into 72 lists with 10 sentences per list. Estimates of *k* for individual subjects required a comparison between word recognition in isolation and word recognition in meaningful sentences. Approximately 2000 isolated key words from the IEEE set (40 lists), as well as all of the necessary non key words to construct sentences were recorded separately by a male talker, low-pass filtered (8.5 kHz), and digitized at 20 kHz with 16-bit amplitude quantization. Sentences were constructed by concatenating key words and non key words together so that in the main comparison between isolated words and words in context, physically identical key words were used. This process ensured that any differences observed in key-word recognition scores between the two sets of materials could only be due to morpho-syntactic and semantic context and not due to coarticulation, prosodic information, or any other physical signal alteration that occurs between words spoken in isolation or in sentence context.

In creating concatenated sentence materials, care was taken to make the resultant sentences sound as natural as possible.[3] Thus, when recording the non key words, phrases rather than isolated words were recorded whenever possible. For example, for the sentence ''The birch canoe slid on the smooth planks,'' the non key word ''The'' and the non key phrase ''on the'' were recorded along with the remaining five key words (''birch,'' ''canoe,'' ''slid,'' ''smooth,'' and ''planks''). The concatenated sentence thus consisted of the pieces the, birch, canoe, slid, on the, smooth, and planks. The intensity of all non key words and nonkey word phrases was scaled to be 0.7 times the original level. This amplitude scaling was chosen to make concatenated sentences appear as natural as possible. Concatenated sentences were approximately 1.2 times longer than their fluently produced counterparts. Whereas the test sentences sounded somewhat unnatural because of the absence of appropriate intonation and stress, they were nevertheless quite intelligible and easily understood by our HI subjects.

To test sentence recognition in noise, naturally produced IEEE sentences spoken by a female talker were used. All 720 IEEE sentences were originally videotaped at the Massachusetts Institute of Technology, and the auditory and visual images were transferred to an optical disk recorder (Panasonic TQ-3031F). The audio portion of each sentence was digitized (16-bit A/D, 20-kHz sampling rate), normalized in level so that all sentences had the same overall rms, and stored on computer for later playback. The computer independently controlled the visual and audio playback for each sentence, allowing for precise alignment of the audio and visual signals (± 2 ms). The noise used to mask the sentence materials was a 20-s sample of white noise that was shaped according to the long-term average magnitude spectrum of the IEEE sentences. The duration of the noise sample used on each trial was equal to the sentence duration plus 100 ms. This was accomplished by positioning a pointer randomly in the noise file and extracting the appropriate duration of shaped noise.

C. Procedures

1. Establishing equivalent intelligibility conditions across listeners for isolated words

The intelligibility of auditorily presented isolated words was controlled by bandpass filtering. Three bandpass filters with different bandwidths, each centered at 1.0 kHz, were determined independently for each subject using an adaptive tracking procedure that controlled the bandwidth of the filter (TDT PF1). Three different tracking algorithms, each run separately, were used: 1-down 2-up (targeting 29% correct—filter 1), 1-down 1-up (targeting 50% correct—filter 2), and 3-down 1-up (targeting 79% correct—filter 3). Subjects were presented randomized lists of isolated key words (up to 200 unique IEEE key words for each filter bandwidth estimate) and were required to repeat back verbally what they heard. Scoring was performed online by the experimenter. The filter bandwidth initially was increased or decreased symmetrically in semitone steps according to the specific decision rule for the three different adaptive tracks (getting narrower after correct responses and broader after incorrect responses). After four track reversals, the step size was decreased to a quarter tone and continued at the smaller step size until an additional eight reversals were recorded. The final bandwidth was computed from the geometric mean of the last eight reversals. Each subject was tested for a minimum of nine tracks (three tracks per filter). A fourth track was run if any of the three filter bandwidth estimates differed by more than five quarter tones and the final average was then based on all four estimates. The order of tracks was randomized separately across subjects.

2. Estimating individual *k* factors

To estimate k, the recognition of words in isolation and words in sentences was tested under each of the three filter conditions described above (hereafter labeled filter 1, filter 2, and filter 3). Five hundred key words were tested per filter condition. Because some of the IEEE key words are repeated within and across lists of sentences, the total number of unique key words presented was 1048 (instead of 1500). Half of the subjects ($n=17$) received isolated words prior to being tested with the concatenated sentence materials, whereas the other half received sentences first. Within each group (words first or sentences first), the test materials were randomized separately for each subject, although the same words were used for all subjects in a given filter condition. Thus, each subject obtained three isolated word scores (based on the average of 500 words each) and three sentences scores (based on the same 500 words used for isolated words). The three estimates of k for each subject (one per filter condition) were computed using Eq. (3).

3. Measuring auditory, visual, and auditory–visual speech recognition in noise

Speech recognition in continuous speech-shaped noise was evaluated following the estimation of individual k factors. Fifteen lists of IEEE sentences (50 key words per list) not previously presented were used to evaluate each of the three receiving modalities (five lists per modality). The signal-to-noise ratio for auditory and auditory–visual conditions was fixed at 0 dB S/N for all subjects. Whereas isolated words and concatenated sentences used to estimate k were spoken by a male speaker, the IEEE sentences used for speech recognition in noise were spoken by a female speaker. For visual and auditory–visual conditions, the subjects were seated approximately 5 ft away from a 19-in. television monitor (SONY PVM 2030). Listening was conducted binaurally under headphones (Beyer model DT 770) at approximately 85 dB SPL. The subjects gave their responses verbally and the number of correct key words was scored online by the experimenter. The final score for each condition (auditory, visual, and auditory–visual) was the average of five lists (50 key words per list).

TABLE I. Average bandwidth for three filter conditions targeting 29% (filter 1), 50% (filter 2), and 79% (filter 3) correct isolated word recognition, respectively, for 34 hearing-impaired subjects. Filters were logarithmically centered at 1 kHz.

Subject	Bandwidth (Hz) Filter 1	 Filter 2	 Filter 3
1	76	209	730
2	76	176	730
3	142	451	1370
4	586	1666	3915
5	443	1292	4461
6	1439	2051	2686
7	461	1413	2051
8	118	1402	2652
9	161	466	3624
10	76	339	900
11	85	281	1056
12	122	330	2701
13	108	549	1687
14	146	414	1225
15	231	1277	3782
16	218	980	2925
17	133	504	1800
18	62	269	1077
19	192	1038	4758
20	593	1056	2814
21	76	561	1042
22	168	723	1095
23	58	198	693
24	80	296	1529
25	395	832	1905
26	118	612	2220
27	146	514	1518
28	163	560	1957
29	93	273	872
30	73	367	1081
31	261	634	1289
32	313	820	1608
33	85	687	1873
34	870	1353	3084
Mean	246.1	723.3	2020.9
Standard error	48.3	81.2	191.8

II. RESULTS

A. Filter bandwidths

Filter conditions corresponding to roughly 29%, 50%, and 79% intelligibility, established for each subject during the tracking task with isolated words, resulted in average bandwidths of 246, 723, and 2021 Hz, respectively (see Table I). However, because there was a wide range of hearing loss across subjects, there was also a great amount of variability in the final filter bandwidths. To determine whether various characteristics of the subject's hearing loss were good predictors of filter bandwidth, a canonical correlation was used with bandwidth as the dependent measure and hearing loss at single frequencies (1 and 2 kHz) and average frequencies (0.5, 1, and 2 kHz; 2 and 4 kHz; 2, 3, and 4 kHz; 2, 3, 4, 6, and 8 kHz; and 3, 4, 6, and 8 kHz) as the independent measures. As might be expected, the bandwidth for the narrowest filter condition was most highly correlated with hearing loss measured at 1 kHz ($r=0.69$, $p<0.0001$) since the filters were centered at 1 kHz. For the two broader filter conditions, the hearing loss at higher and lower frequencies became important in determining the final bandwidth. For example, for filter condition 2 (targeting 50% word intelligibility), the average pure-tone three-frequency loss at 0.5, 1, and 2 kHz ($r=0.64$, $p<0.0001$) was the best predictor of bandwidth, whereas for filter condition 3 (targeting 79% word intelligibility), the hearing loss at 2 kHz ($r=0.79$, $p<0.0001$), and the average three-frequency loss at 2, 3, and 4 kHz ($r=0.72$, $p<0.0001$) were the most important frequencies for determining filter bandwidth. Thus, for broader filters with higher intelligibility, the high-frequency hearing loss, and especially the loss at 2 kHz, became increasingly more important in determining the final filter bandwidth.

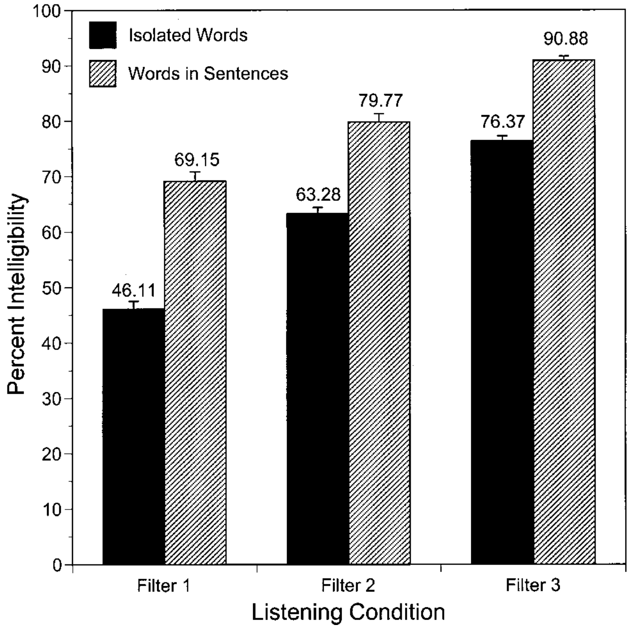

FIG. 2. Isolated word and words in sentence scores obtained for the three filter conditions. Error bars show +1 standard error.

B. Estimates of *k*

Isolated word and word-in-sentence scores obtained for the three filter conditions are shown in Fig. 2. The mean isolated word scores, based on 500 words each (10 blocks of 50 words), were 46%, 63%, and 76% correct, respectively. The corresponding words-in-sentence scores were 69%, 80%, and 91% correct, respectively. Thus, even though the sentence materials were constructed by concatenating words together (and therefore lacked normal prosody and coarticulation among syllables), the meaningful sequence of the words and well-formed grammatical structure of the sentences significantly enhanced word intelligibility.

Prior exposure to either isolated words or words in sentences facilitated the recognition of speech materials presented later. Figure 3 shows the effect of presentation order on the intelligibility of isolated words and words in sentences. Recall that half the subjects (group 1; $n=17$) were exposed to words in isolation prior to receiving concatenated sentence materials, whereas the remaining subjects (group 2; $n=17$) were exposed to concatenated sentences prior to iso-

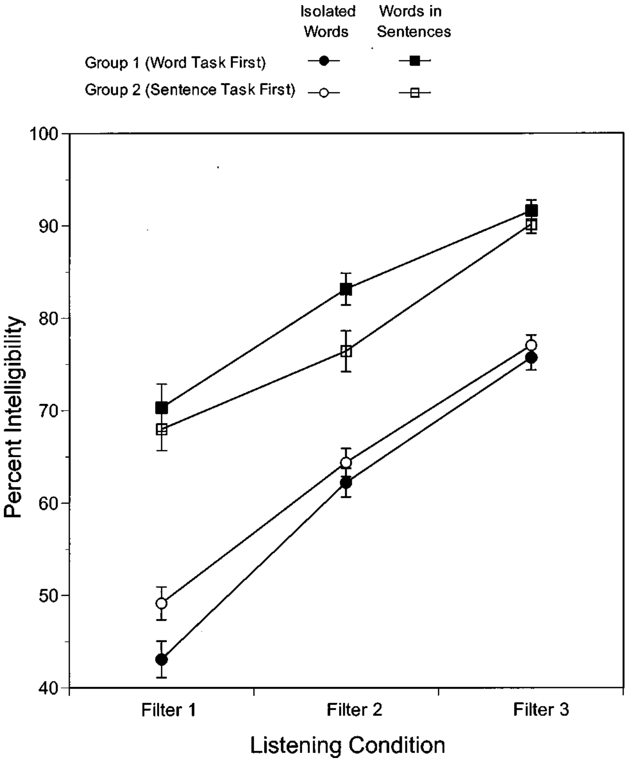

FIG. 3. Same as Fig. 2 but broken out for the two groups of subjects. Group 1 (filled symbols) received isolated words prior to sentences, group 2 (open symbols) received sentences prior to being exposed to isolated words. Square symbols show percent correct scores for words-in-sentences, circle symbols show percent correct scores for isolated words. Error bars show ±1 standard error.

lated words. The facilitating effects of prior exposure can be measured by comparing the difference in intelligibility scores for words received first (group 1) versus words received second (group 2) and for sentences received first (group 2) versus sentences received second (group 1). Averaged across filter conditions, the facilitation effect on isolated word recognition due to prior exposure to concatenated sentences was 3.7%, and the facilitation effect on word-in-sentence recognition due to prior exposure to isolated words was 3.2%.

Although these differences in intelligibility due to test order may seem small at first, they have a significant effect on the estimates of individual k factors. Because k is related directly to the difference in intelligibility between isolated word and word-in-sentence scores, experimental variables, like test order, that affect intelligibility will also affect estimates of k. By testing concatenated sentences after exposure to isolated words (group 1), the word-in-sentence scores were enhanced relative to what they would have been had there been no prior exposure, and the estimated k values were relatively large. This is illustrated in Fig. 3 by the greater separation between isolated word and word-in-sentence scores for group 1 subjects. Conversely, by testing isolated words after exposure to concatenated sentences (group 2), the word scores were enhanced relative to what they would have been had there been no prior exposure (thereby making the difference between isolated words and concatenated sentences smaller), and the estimated k values were relatively small. Figure 4 shows the estimated group mean k values for each filter condition. A repeated measures analysis of variance was carried out on the k values with one between-subjects factor (test order) and one within-subjects factor (filter condition) at three levels. The results showed significant main effects of test order $[F(1,32)=15.74,\ p<0.001]$ and filter condition $[F(2,64)=20.60,\ p<0.001]$. The analysis also showed a significant interaction between test order and filter $[F(2,64)=5.04,\ p<0.01]$. This interaction may be seen in Fig. 4 as the more similar performance for the two groups with filter 3 than with the other two filter conditions.

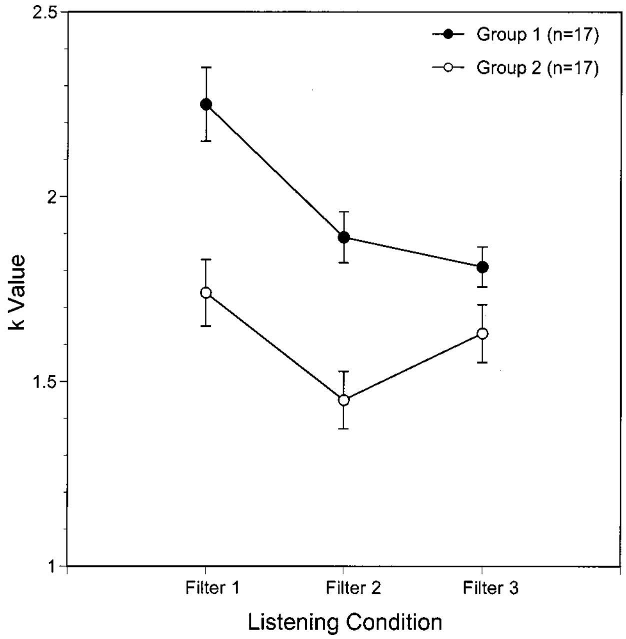

FIG. 4. Group mean k values for the three filter conditions. Filled circles =group 1 (words first); filled triangle=group 2 (sentences first). Error bars show ±1 standard error.

Post hoc analyses (paired t-tests with Bonferonni adjustment) of the data revealed that the estimated k values obtained for filter 2 were significantly smaller than for filter 1 for both groups of subjects. However, whereas the k values for filter 3 were essentially the same as those measured for filter 2 for group 1, the k values for filter 3 were significantly larger than those measured for filter 2 for group 2. At present, we have no explanation for this result, especially considering that the words tested across groups for any given filter condition were identical.

With the exception of filter condition 3, group 2, the average k values obtained for the two groups of subjects decreased with increasing signal clarity, at least over the range of intelligibility tested. That is, as more stimulus information becomes available due to a broadening of the listening band, there appears to be less contextual information employed by the subjects. Note that this result is counter to the theoretical predictions of Boothroyd and Nittrouer (1988). Under their formulation, k is expected to be constant regardless of the level of isolated word intelligibility, as long as the speech materials remain fixed. However, given that contextual constraints are used to reduce ambiguity of missing or distorted information due to signal degradation, it may be that as the intelligibility of the signal increases, the relative importance of contextual processing diminishes. This is not to imply that there is no benefit of context for signals with relatively high intelligibility. According to Eq. (3), *k* values above 1.0 indicate some information due to context. As Fig. 4 shows, all mean *k* values were well above 1.0, even for group 2, whose *k* values are probably smaller than those typically reported given the prior exposure to sentence materials. Averaged across filter conditions, the mean *k* value for group 1 was 1.98, whereas the mean value for group 2 was 1.61. According to Boothroyd and Nittrouer (1988), this means that the morpho-syntactic information contained in IEEE sentences represents an effective increase of 60%–100% in the number of channels of statistically independent information available from the stimulus alone.

An important question regarding the potential use of *k* as an indication of individual skill in using contextual information is whether estimates of *k* are stable across different listening conditions. To address this question, Pearson correlation coefficients for all pairs of *k* across filter conditions were computed. A significant correlation across filter conditions indicates that a subject's ranking with regard to contextual ability is consistent relative to other subjects. The correlation between adjacent filter conditions (i.e., filter 1 versus filter 2, and filter 2 versus filter 3) was highly significant ($r_{12} = 0.78$, $p_{12} < 0.0001$ and $r_{23} = 0.69$, $p_{23} < 0.0001$, respectively), accounting for roughly 50%–60% of the variance across subjects. The correlation for filter 1 versus filter 3 was more modest ($r_{13} = 0.46$, $p < 0.01$), but nevertheless significant. The significance of these correlations is critical for the hypothesis that *k*, a presumed measure of subject ability to make use of contextual information, tends to travel with the subject from one task to the next. That is, a subject with a high *k* value on one task is likely to have a high *k* value on some other task.

C. Relation between *k* and auditory, visual, and auditory–visual sentence recognition

Sentence recognition in noise measured at 0 dB S/N for auditory and auditory–visual conditions, as well as sentence recognition for speechreading alone, is shown in Fig. 5. Auditory sentence recognition, shown along the abscissa, ranged from 9% to 87% ($\overline{X} = 49\%$). Visual sentence recognition (filled circles) ranged from 0% to 24% ($\overline{X} = 4\%$), whereas auditory–visual sentence recognition (filled triangles) ranged from 56% to 99% ($\overline{X} = 84\%$). The solid line represents the unity function auditory=visual and auditory=auditory–visual. The fact that all data for the auditory–visual condition lie well above the unity line indicates that all subjects derived benefit from combining auditory and visual cues.

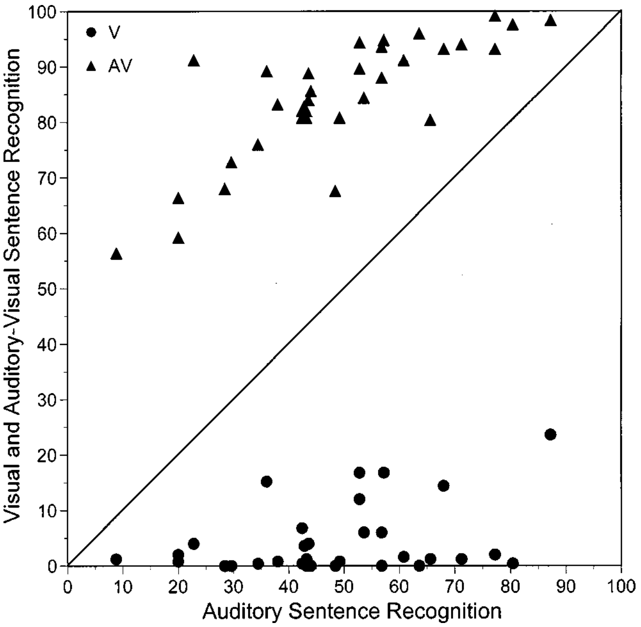

FIG. 5. Sentence recognition in noise measured at 0 dB S/N for auditory, visual, and auditory–visual conditions. Filled circles=*V* condition; filled triangles=auditory–visual condition. The location of any symbol with respect to the abscissa shows the score for the auditory condition.

Our original hypothesis concerning the role of context ability in sentence recognition was that subjects with greater *k* values would, in general, have higher recognition scores for words in sentences than subjects with lower *k* values. However, sentence scores are influenced by a number of factors other than context, such as a subject's ability to extract auditory and visual segmental and prosodic cues and the ability to integrate these cues (for auditory–visual recognition). Therefore, to determine whether individual *k* factors influence a subject's sentence recognition score, these other potentially confounding factors have to be addressed. In the present study, no direct data were obtained regarding segmental or prosodic cue extraction ability in either auditory, visual, or auditory–visual modalities. However, at least for auditory speech recognition, this ability is known to be determined primarily by audibility of specific speech cues, which in turn may be estimated by the subjects hearing loss, especially for high frequencies. This estimate of auditory cue extraction ability can be further augmented by data obtained in the filter-tracking phase of the current study, in that one can assume that listeners who can identify isolated words through narrow filters have better suprathreshold abilities to extract auditory cues than listeners who require broader filter bandwidths to achieve the same level of word-recognition performance (Noordhoek *et al.*, 1998, 1999).

To determine whether an individual's *k* value is useful in explaining some of the individual differences typically observed in sentence-recognition tasks, three separate stepwise multiple linear regression analyses were carried out with sentence recognition for each of the three modalities (auditory, visual, and auditory–visual) as dependent variables and with *k*, age, high-frequency hearing loss (3–8 kHz), and filter bandwidth determined in the tracking tasks as independent variables. Not surprisingly, auditory factors presumably related to acoustic cue extraction (i.e., high-frequency pure-tone thresholds and filter bandwidth) were significantly correlated with both auditory and auditory–visual sentence recognition. For the visual-only condition, age was the most important factor whereas hearing loss and filter bandwidth did not contribute significantly to the amount of variance explained. Importantly, however, the mean *k* values for each

TABLE II. Predicting sentence recognition (P_s) in noise. Equations indicate final subset model from a multilinear forward stepwise linear regression ($\alpha = 0.15$). Predictor variables were age, average high-frequency hearing loss from 3–8 kHz (HL), mean k value, filter 1 bandwidth ($F1$), filter 2 bandwidth ($F2$), and filter 3 bandwidth ($F3$). The amount of variance explained by each factor is also shown (in percent).

	Predictors	Variance	F value	p value
I. Auditory sentence recognition in noise				
$P_s = C + 21.106k - 0.558\text{HL} - 0.012F2$				
($r^2 = 0.498$, $p < 0.001$)	HL	33.3	11.5562	0.0019
	k	13.3	7.9260	0.0085
$C = 51.248$	$F2$	3.2	4.2737	0.0474
II. Visual sentence recognition				
$P_S = C + 5.252k - 0.262\text{age}$				
($r^2 = 0.316$, $p < 0.003$)	Age	23.6	10.4735	0.0029
	k	8.0	3.6111	0.0667
$C = 12.724$				
III. Auditory–visual sentence recognition in noise				
$P_S = C + 14.704k - 0.370\text{HL} - 0.009F1$				
($r^2 = 0.564$, $p < 0.0001$)	HL	33.8	17.8620	0.0002
	k	19.4	13.3227	0.0010
$C = 81.158$	$F1$	3.2	3.1186	0.876

subject (averaged over the three filter conditions) remained a significant factor in the final subset model for all three modalities.[4] Approximately 50% of the variance in the auditory-alone condition and approximately 56% of the variance in the auditory–visual condition could be accounted for with each individual subject's average k, better ear high-frequency hearing loss (average pure-tone thresholds from 3–8 kHz), and filter 2 bandwidth and filter 1 bandwidth, respectively (see Table II). In both of these conditions, hearing loss was the far more important factor accounting for between 33%–34% of the variance. The proportion of total variance accounted for by k was 19.4% for the auditory–visual condition and 13.3% for the auditory condition. In the visual-only condition, age and average k accounted for roughly 32% of the variance. Here too, k accounted for relatively little variance (8%) with age accounting for 23.6% of the total variance explained.

III. DISCUSSION AND CONCLUSIONS

There can be little doubt that sentence context facilitates word recognition and examples are readily found in the literature. For instance, in the ANSI standard for calculating the articulation index (ANSI, 1969), Fig. 15 (p. 23) shows that the recognition of words when presented in isolation versus meaningful sentences at a fixed signal-to-noise ratio can vary by over 50 percentage points, depending on the speech materials. Even when the words are physically identical, as in the present experiment, recognition rates for words in sentences were between 14%–30% better than for words presented in isolation. As Boothroyd and Nittrouer (1988) noted, this facilitation occurs for at least two basic reasons: (1) within a word, lexical redundancy serves to reduce the ambiguity of a given phoneme based on a knowledge of its position within the word and the phonemes surrounding it, and (2) within a sentence, the identity of a target word can be guessed with better than chance accuracy because of semantic and morpho-syntactic constraints imposed by the words preceding and following it. Lexical redundancy is primarily responsible for facilitating word recognition relative to nonsense words or isolated syllables (Boothroyd and Nittrouer's j factor), whereas semantic and morpho-syntactic redundancy is primarily responsible for facilitating the recognition of words in sentence relative to words in isolation (Boothroyd and Nittrouer's k factor).

Past work on the facilitating effects of intra- and interword redundancy has focused primarily on differences across materials. Far fewer studies have focused on the possibility that individuals differ with respect to the amount of facilitation derived from these different sources of linguistic redundancy (*viz.*, Lyxell and Rönnberg, 1987a, 1987b, 1989; Rönnberg *et al.*, 1998 for exceptions). In the present study, isolated-word and word-in-sentence recognition scores based on 500 items each were used to derive k estimates. Further, three such estimates were made for each subject at different levels of isolated word intelligibility. The large number of trials per subject, although clinically unfeasible, was nevertheless deemed necessary to stabilize individual performance.

Given our overall objective of obtaining estimates of subject ability to make use of sentence context and applying this information to predict sentence-recognition scores, the first point to be established was that individuals do indeed vary in the amount of facilitation derived from sentence context, and that the positioning among individual subjects (good to poor users of context) was relatively stable for different conditions of distortion. Results showed a range of individual k values averaged across intelligibility (i.e., filter condition) between 1.2 and 2.5. This represents a large range in the amount of information derived from sentence context when viewed in terms of additional channels of information (Boothroyd and Nittrouer, 1988). Thus, individuals clearly differ with regard to the amount of contextual facilitation they can achieve.

An important question regarding the estimates of subjects' ability to make use of sentence context pertains to the stability of the measurements. If k reflects a specific factor or primary ability of subjects to use context to facilitate word recognition, then the rank ordering of individuals with respect to *k* should not vary with changes in task difficulty. The significant correlation of *k* values across filter conditions implies that the ordering of subjects with regard to their ability to make use of context cues was fairly stable. The correlation between filter conditions 1 and 2 ($r = 0.78$) has the greatest reliability because the isolated word scores fell within the region most sensitive for demonstrating differences in *k* values (see Fig. 1). Correlations involving filter condition 3 are more suspect because several of the *k* estimates for this filter condition were compromised by ceiling effects (due to perfect performance for the recognition of words in sentences). Nevertheless, the remaining two correlations (filter 1 versus filter 3 and filter 2 versus filter 3) were also significant, though somewhat more modest. Thus, using *k* as a measure of an individual subject's ability to use context to facilitate word recognition appears to be stable across several conditions of distortion.

Like most measures of human perceptual performance, *k* is subject to the effects of different test methodologies. Thus, whereas the relative position of an individual with respect to other subjects was fairly stable (good context users were good regardless of test condition), the absolute amount of contextual facilitation depended on how the measurements were made. Even with no feedback, a large stimulus set size, and an open response format, subjects showed a strong learning effect depending on whether they received isolated words first or sentences first. In the recent study by Olsen *et al.* (1997), where sentence tests were administered immediately following isolated word tests, we would have to conclude that their *k* estimates may be somewhat high. One implication of the learning effects associated with test order is that the estimates of isolated words correct and words in sentences should be carried out with lists that do not contain the same words but that have been pretested to be of equal expected mean performance levels.

In addition to being subject to the effects of test order, *k* estimates also depend on the isolated word-recognition score, unlike the theoretical functions described by Boothroyd and Nittrouer (1988). Specifically, *k* decreased as isolated word recognition improved. A decrease in *k* was observed for 32 of 34 subjects when comparing the results for filter 1 versus filter 2. Thus, it appears that listeners compensate for reduced stimulus information by increasing their reliance on contextual information. Similar conclusions have been reached for read and spoken word recognition, especially under conditions of stimulus distortion (Stanovich and West, 1979; Cohen and Faulkner, 1983; Wingfield *et al.*, 1985; Madden, 1988; Ben-Dror *et al.*, 1991). For example, Stanovich and West (1979) measured good and poor third-grade readers' ability to rapidly spot real words, pseudowords, and nonwords. Overall, search times were significantly longer for poor readers than for good readers. Moreover, the effect of orthographic structure, measured by the difference in search time between spotting real words versus pseudo-or nonwords, tended to be greater for the poorer readers. In another study, Stanovich and Pachella (1977) demonstrated that reaction times were more affected by stimulus probability (e.g., context) when the clarity of the stimulus was decreased. These results, as with the present results for aurally presented words, suggest that the various factors contributing to word recognition (visual or auditory peripheral analyses, use of lexical redundancy, semantic constraints, morpho-syntactic constraints, etc.) may be employed flexibly so that a reduction in one process can be compensated for by greater reliance on other processes. In other words, as the stimulus becomes degraded (reduced bottom-up information) there is a tendency for subjects to rely more heavily on top-down information. The fact that the vast majority (32 of 34) of subjects had higher *k* values for filter 1 than for filter 2 suggests that the tendency to rely less on contextual information as signal clarity improves is common to individual hearing-impaired listeners in spite of large differences in the amount of hearing loss and age.

In a review article, Elliott (1995) equated the ability to use contextual information in speech to that of verbal auditory closure. Elliot further suggested that if this ability represented a primary factor of human perceptual behavior, it should provide useful information regarding an individual's speech comprehension skill. In terms of the present experiment, this means that *k*, to the extent that it is modality independent, should help explain some of the differences in auditory, visual, and auditory–visual speech recognition. This was demonstrated by multiple stepwise regression in which an attempt was made to predict auditory, visual, or auditory–visual sentence-recognition scores using a number of different subject variables such as age, hearing loss, filter bandwidth, and *k*. The final subset model for each modality included *k*, indicating that the ability to use semantic and morpho-syntactic constraints in sentences is a significant subject factor in predicting speech recognition regardless of the test modality. By including the average *k* value with the other factors listed in Table I, the amount of variance accounted for was increased by approximately 13%, 8%, and 19% for auditory, visual, and auditory–visual conditions, respectively. Whereas the amount of variance that can further be explained by adding context to other subject variables may be relatively small, it is nevertheless a significant factor in explaining some of the intersubject variability in sentence recognition.

This last point deserves additional comment. In speech recognition and in reading, the ability to use one's knowledge of the language to compensate for missing or incomplete signal information has been regarded as an important factor in explaining individual differences in performance. Whereas individual *k* values were significant factors in determining speech-recognition performance, the amount of variance explained was moderate to small. This may have been due to the conditions under which the speech-recognition scores were obtained. Recall that the speech materials were presented at approximately 85 dB SPL and 0 dB S/N for all subjects with no adjustments for the varying amounts of hearing loss across the study group. Thus, it is likely that individual differences in signal audibility existed across subjects for the auditory and auditory–visual conditions. Because differences in audibility have been shown repeatedly to be a dominant factor in predicting speech-recognition performance, the importance of intersubject differences in the ability to use contextual information might have been reduced. Had the sentence-recognition tests been conducted at different levels depending on the subjects hearing status, so that the speech was equally audible for all subjects, the amount of variance explained by *k* might have been larger.

Another important factor to consider when interpreting the results of the regression analyses is that *k* values were obtained under rather constrained conditions (auditory only, concatenated ''unnatural'' sentences). It is possible that in addition to being dependent on the specific speech corpus and perceiver attributes, *k* may also be dependent on modality. Therefore, predicting visual and auditory–visual performance from auditory derived *k* factors may have led to an underestimate of the association. However, at present we know of no data to suggest that the knowledge base used when employing contextual information in speech recognition is in any way different across modalities. In fact, the premise of the present experiment that led us to construct concatenated sentence materials was that once certain stimulus variables, like coarticulation and prosody, were controlled so that the stimulus information for words in isolation and words in sentences were identical, any differences observed in the recognition rates would have to be ascribed to use of higher-order processes that are essentially signal independent (and modality independent).

The amount of additional variance accounted for in the speechreading condition by individual *k* factors was only 8%. Given that all subjects had normal or corrected-normal vision, why might this be so? One possibility is that the speechreading scores were extremely low overall and rather skewed. Thus, the limited range of speechreading scores may have compromised the regression analysis and produced an artificially low correlation between *k* values and speechreading scores. Another possibility is that speechreaders have particular difficulty extracting prosodic cues, such as word boundaries, creating greater ambiguity when confronted with sentences rather than isolated word productions (Erber and McMahon, 1978). Thus, the problem for the speechreader is primarily related to the extraction of relevant signal cues and not one of deployment of higher-level knowledge sources regarding semantic and syntactic information. Under this interpretation, it is not surprising that measures pertaining to a subject's ability to make use of context would have little to do with overall speechreading scores.

It is also well established that individual differences in speechreading ability are known to be quite substantial across a variety of test materials (Demorest and Bernstein, 1992; Watson *et al.*, 1996) despite ''normal'' visual acuity. In these studies, normal or corrected-normal visual acuity was assumed based on subject reports. In the present study, subjects passed a visual screening (Snellen chart) with 20/30 or better visual acuity. However, this measure was taken at a distance of 20 ft, whereas all visual speech testing was done at approximately 5 ft. It is possible that the subjects may have exhibited different degrees of visual acuity at this closer distance.

In addition to possible visual acuity deficits, a predominant factor in visual speech recognition seems to be age. In the present study, age accounted for approximately 24% of the variance in *V* scores. Aging is known to have detrimental effects on visual function, especially with regard to motion detection and contrast sensitivity (Wood and Bullimore, 1995). It is possible that there are substantial individual differences related to these visual measures within the study group. Thus, heterogeneity in basic visual function impacts speechreading performance and limits the extent to which we can observe the effects of differences in individual *k* factors. In future studies, it may be necessary to screen more carefully for visual anomalies in elderly subjects before presuming normal visual acuity.

IV. SUMMARY

Word recognition in sentences is determined by a number of factors including the ability to recognize phonemes (auditorily and /or visually), the ability to integrate cues across modality (under auditory–visual conditions), the ability to use intraword redundancy (lexical redundancy), and the ability to use semantic and morpho-syntactic constraints in sentence contexts (Boothroyd and Nittrouer's *k* factor). The present study focused on the role of sentence context in facilitating word recognition and whether understanding an individual subject's ability to use context would help explain some of the variability commonly observed in speech communication. The results indicated that the use of sentence context to facilitate word recognition varies substantially across hearing-impaired subjects. Furthermore, subjects with high *k* factors performed slightly better than subjects with lower *k* factors on sentence recognition in all three receiving modalities (auditory, visual, and auditory–visual). The *k* factor for an individual depends to a large degree on the speech materials used in the test, as shown in earlier studies. But it also depends on the order in which isolated words or words in sentences are presented, as well as on the clarity of the signal information available to each subject, at least for isolated word recognition scores between 29%–75% correct.

ACKNOWLEDGMENTS

Portions of this research were presented at the 134th Meeting of the Acoustical Society of America, San Diego, CA, 3 December 1997 [Grant, K. W., and Seitz, P. F. (1997). ''The recognition of isolated words and words in sentences: Individual variability in the use of sentence context,'' J. Acoust. Soc. Am. **102**, 3132.] This research was supported by Grant Numbers DC00792 and DC01643 from the National Institute of Deafness and Other Communication Disorders to Walter Reed Army Medical Center, and by the Clinical Investigation Service, Walter Reed Army Medical Center, under Work Unit # 2548. All subjects participating in this research provided written informed consent prior to beginning the study. We would like to thank James Hillenbrand, Arthur Boothroyd, and two anonymous reviewers for their helpful comments and suggestions. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the

views of the Department of the Army or the Department of Defense.

---

[1] Solving for k in Eq. (3) gives $k=\log(1-P_s)/\log(1-P_w)$, where the quantity $(1-P_s)$ is the error rate for words presented in sentences and $(1-P_w)$ is the error rate for words presented in isolation.

[2] Words spoken in isolation and in sentences differ with respect to many variables. In sentences, words tend to be shorter in duration, are influenced by coarticulation and prosody, and undergo numerous phonological transformations (e.g., vowel neutralization and phonetic substitutions, as in the transformation from ‘‘Did you eat?’’ [/dɪd yu it/] to ‘‘Jueet?’’ [/dʒit/]). This study minimized these differences across materials by concatenating isolated words to form sentences, thus eliminating these complex transformations and ensuring that differences in the intelligibility of isolated words and sentences were due mainly to semantic and morpho-syntactic context.

[3] Another option for keeping words in isolation and words in sentences physically identical was to excise words from naturally spoken sentences (Wingfield *et al.*, 1994). However, subjective impressions of the two methods (concatenated words versus excised words) led us to believe that the former method would be better tolerated and appear more natural to hearing-impaired subjects.

[4] Of special interest is the fact that hearing loss (pure-tone averages at 0.5, 1, and 2 kHz, 2–8 kHz, 3–8 kHz, or individual tonal thresholds at 1 and 2 kHz) and average k were uncorrelated $(-0.017<r<0.059)$. This lends further support that k reflects an ability of individual subjects to make use of higher-level language constraints independent of stimulus quality.

American National Standards Institute (ANSI) (**1969**). ‘‘American National Standard Methods for the Calculation of the Articulation Index,’’ ANSI S3.5-1969, American National Standards Institute, New York.

American National Standards Institute (ANSI) (**1989**). ‘‘Specifications for Audiometers,’’ ANSI S3.6-1989, American National Standards Institute, New York.

Ben-Dror, I., Pollatsek, A., and Scarpati, S. (**1991**). ‘‘Word identification in isolation and in context by college dyslexic students,’’ Brain Lang. **40**, 471–490.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeckowski, C. (**1984**). ‘‘Standardization of a test of speech perception in noise,’’ J. Speech Hear. Res. **27**, 32–48.

Boothroyd, A., Hnath-Chislm, T., and Hanin, L. (**1985**). ‘‘A sentence test of speech perception: Reliability, set-equivalence, and short-term learning,’’ City University of New York, Report Number RCl10.

Boothroyd, A., and Nittrouer, S. (**1988**). ‘‘Mathematical treatment of context effects in phoneme and word recognition,’’ J. Acoust. Soc. Am. **84**, 101–114.

Cohen, G., and Faulkner, D. (**1983**). ‘‘Word recognition: age differences in contextual facilitation effects,’’ Br. J. Psychol. **74**, 239–251.

DeFilippo, C. L. (**1990**). ‘‘Speechreading training: Believe it or not!,’’ ASHA **32**, 46–48.

Demorest, M. E., and Bernstein, L. E. (**1992**). ‘‘Sources of variability in speechreading sentences: A generalizability analysis,’’ J. Speech Hear. Res. **35**, 876–891.

Elliott, L. L. (**1995**). ‘‘Verbal auditory closure and the speech perception in noise (SPIN) Test,’’ J. Speech Hear. Res. **38**, 1363–1376.

Erber, N. P. (**1996**). *Communication Therapy for Adults with Sensory Loss*, 2nd ed. (Clavis, Clifton Hill, Victoria, Australia).

Erber, N. P., and McMahon, D. A. (**1978**). ‘‘Effects of sentence context on the recognition of words through lipreading by deaf children,’’ J. Speech Hear. Res. **19**, 112–119.

Grant, K. W., and Seitz, P. F. (**1998**). ‘‘Measures of auditory–visual integration in nonsense syllables and sentences,’’ J. Acoust. Soc. Am. **104**, 2438–2450.

Grant, K. W., and Walden, B. E. (**1996**). ‘‘Evaluating the articulation index for auditory–visual consonant recognition,’’ J. Acoust. Soc. Am. **100**, 2415–2424.

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). ‘‘Auditory–visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory–visual integration,’’ J. Acoust. Soc. Am. **103**, 2677–2690.

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., and Lee, L. (**1994**). ‘‘Factors associated with individual differences in clinical measures of speech recognition among the elderly,’’ J. Speech Hear. Res. **37**, 465–474.

Hutchinson, K. M. (**1989**). ‘‘Influence of sentence context on speech perception in young and older adults,’’ J. Gerontol. **44**, 36–44.

Institute of Electrical and Electronic Engineers (**1969**). *IEEE Recommended Practice for Speech Quality Measures* (IEEE, New York).

Kalikow, D. H., Stevens, K. N., and Elliott, L. L. (**1977**). ‘‘Development of a test of speech intelligibility on noise using sentence materials with controlled word predictability,’’ J. Acoust. Soc. Am. **61**, 1337–1351.

Lyxell, B., and Rönnberg, J. (**1987a**). ‘‘Guessing and speechreading,’’ Br. J. Audiol. **21**, 13–20.

Lyxell, B., and Rönnberg, J. (**1987b**). ‘‘Necessary cognitive determinants for speechreading skills,’’ in *Adjustment to Acquired Hearing Loss: Analysis, Change, and Learning*, edited by J. G. Kyle (Anthony Row, Chippenham, Witshire), pp. 48–54.

Lyxell, B., and Rönnberg, J. (**1989**). ‘‘Information processing skills and lipreading,’’ Br. J. Audiol. **23**, 339–347.

Madden, D. J. (**1988**). ‘‘Adult age differences in the effect of sentence context and stimulus degradation during visual word recognition,’’ Psychol. Aging **3**, 167–172.

Massaro, D. M. (**1987**). *Speech Perception by Ear and Eye: a Paradigm for Psychological Inquiry* (Erlbaum, Hillsdale, NJ).

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). ‘‘The intelligibility of speech as a function of the context of the test material,’’ J. Exp. Psychol. **41**, 329–335.

Montgomery, A. A., and Demorest, M. (**1988**). ‘‘Issues and developments in the evaluation of speechreading,’’ in *New Reflections on Speechreading*, edited by C. DeFilippo and D. Sims (A. G. Bell Association for the Deaf, Washington, DC), pp. 193–214.

Nittrouer, S., and Boothroyd, A. (**1990**). ‘‘Context effects in phoneme and word recognition by young children and older adults,’’ J. Acoust. Soc. Am. **87**, 2705–2715.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1999**). ‘‘Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners,’’ J. Acoust. Soc. Am. **105**, 2895–2902.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1998**). ‘‘Minimum bandwidth required for speech reception by normal-hearing and hearing-impaired listeners,’’ J. Acoust. Soc. Am. **103**, 3051.

Olsen, W. O., Van Tasell, D. J., and Speaks, C. E. (**1997**). The Carhart Memorial Lecture, American Auditory Society, Salt Lake City, Utah 1996. ‘‘Phoneme and word recognition for words in isolation and in sentences,’’ Ear Hear. **18**, 175–88.

Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (**1992**). ‘‘Relations among different measure of speech reception in subjects using a cochlear implant,’’ J. Acoust. Soc. Am. **92**, 1869–1881.

Rönnberg, J., Samuelsson, S., and Lyxell, B. (**1998**). ‘‘Conceptual constraints in sentence-based lipreading in the hearing-impaired,’’ in *Hearing by Eye II: The Psychology of Speechreading and Audiovisual Speech*, edited by R. Campbell, B. Dodd, and D. Burnham (Erlbaum, Mahwah, NJ), pp. 143–153.

Schum, D. J., and Matthews, L. J. (**1992**). ‘‘SPIN test performance of elderly hearing-impaired,’’ J. Am. Acad. Audiol. **3**, 303–307.

Stanovich, K. E., and Pachella, R. G. (**1977**). ‘‘Encoding, stimulus-response compatibility, and stages of processing,’’ J. Exp. Psychol. Hum. Percept. Perform. **3**, 411–421.

Stanovich, K. E., and West, R. F. (**1979**). ‘‘The effect of orthographic structure on the word search performance of good and poor readers,’’ J. Exp. Child Psychol. **28**, 258–267.

Sumby, W. H., and Pollack, I. (**1954**). ‘‘Visual contribution to speech intelligibility in noise,’’ J. Acoust. Soc. Am. **26**, 212–215.

Watson, C. S., Weiguang, Q., Chamberlain, M. M., and Xiaofeng, L. (**1996**). ‘‘Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition,’’ J. Acoust. Soc. Am. **100**, 1153–1162.

Wingfield, A., Poon, L. W., Lombardi, L., and Lowe, D. (**1985**). ‘‘Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time,’’ J. Gerontol. **40**, 579–585.

Wingfield, A., Alexander, A. H., and Cavigelli, S. (**1994**). ‘‘Does memory constrain utilization of top-down information in spoken word recognition? Evidence from normal aging,’’ Lang. Speech **37**, 221–235.

Wood, J. M., and Bullimore, M. A. (**1995**). ‘‘Changes in the lower displacement limit for motion with age,’’ Ophthalmic Physiol. Opt. **15**, 31–36.