Speechreading and the structure of the lexicon: Computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness

Edward T. Auer, Jr. and Lynne E. Bernstein
Spoken Language Processes Laboratory, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 10 October 1996; revised 19 August 1997; accepted 31 August 1997)

A lexical modeling methodology was employed to examine how the distribution of phonemic patterns in the lexicon constrains lexical equivalence under conditions of reduced phonetic distinctiveness experienced by speechreaders. The technique involved (1) selection of a phonemically transcribed machine-readable lexical database, (2) definition of transcription rules based on measures of phonetic similarity, (3) application of the transcription rules to a lexical database and formation of lexical equivalence classes, and (4) computation of three metrics to examine the transcribed lexicon. The metric percent words unique demonstrated that the distribution of words in the language substantially preserves lexical uniqueness across a wide range in the number of potentially available phonemic distinctions. Expected class size demonstrated that if at least 12 phonemic equivalence classes were available, any given word would be highly similar to only a few other words. Percent information extracted (PIE) [D. Carter, Comput. Speech Lang. **2**, 1–11 (1987)] provided evidence that high-frequency words tend not to reside in the same lexical equivalence classes as other high-frequency words. The steepness of the functions obtained for each metric shows that small increments in the number of visually perceptible phonemic distinctions can result in substantial changes in lexical uniqueness. *© 1997 Acoustical Society of America.* [S0001-4966(97)05112-6]

PACS numbers: 43.71.An, 43.71.Es, 43.71.Ma [WS]

INTRODUCTION

Spoken word recognition depends on the process of selecting a word candidate from a set of word patterns stored in memory. Current models of auditory spoken word recognition agree that the speed and ease of this selection process is partially a function of the lexical properties of the target word (Lahiri and Marslen-Wilson, 1991; Luce *et al.*, 1990; McClelland and Elman, 1986; Norris, 1994). Spoken word recognition is facilitated for frequently used words and for words perceptually similar to few other words (i.e., words with few neighbors). The effects of these lexical properties are particularly important when the phonetic information in the speech signal is degraded (Luce *et al.*, 1990). In speechreading (lipreading), where spoken word recognition occurs on the basis of viewing rather than listening to the talker, the optical speech signal typically affords less phonetic distinctiveness[1] than the acoustic speech signal. The target word's frequency and perceptual similarity to other words will likely be a substantial contributor to the speed and ease of word recognition for the speechreader. In the current study, we modeled visual perceptual similarity among words as a function of the number of perceptually available phonemic distinctions.

Even under optimal perceptual conditions (e.g., adequate lighting, moderate speaking rate, highly visible articulatory gestures), phonetic information is inadequate to specify all the phonemes in the speechreader's language. For example, speechreaders may not perceive any distinctions among productions of the consonants /b/, /p/, and /m/ (Scheinberg, 1988). The degree to which phonemic distinctions are unavailable is to a certain extent due to the articulatory characteristics of the talker (Kricos and Lesner, 1982; Lesner, 1988), the conditions (lighting, viewing angle, distance), the place of articulation, and the perceptual abilities of the speechreader (Jeffers and Barley, 1971). Current spoken word recognition models (Lahiri and Marslen-Wilson, 1991; Luce *et al.*, 1990; McClelland and Elman, 1986; Norris, 1994) imply that loss of phonemic distinctions by itself does not predict word intelligibility: because words are hypothesized to be recognized within the lexical context of perceptually similar words, the distribution of phoneme patterns for words in the language is also determinative of intelligibility.

For example, the English word ''bought'' remains lexically distinct from all other words after the loss of the distinctions between /b/, /p/, and /m/. ''Pought'' and ''mought'' are not words; therefore, a misperception of ''bought'' as either of these nonwords could nevertheless result in an accurate identification of the only existing word ''bought.'' Alternatively, under the same conditions, ''bat'' would not have a similar advantage versus ''pat'' and ''mat.''

Nitchie (1916) and Berger (1972) attempted to estimate the loss of perceptual uniqueness (i.e., homopheny) of English words experienced by speechreaders. According to Nitchie, approximately 50% of words in colloquial English speech are perceptually unique for the average speechreader. Berger estimated that, taking frequency of use into account, between 51% and 55% of the common words in English remain unique. Although the authors never fully specified their methods, it is clear that they used a fixed estimate of the

Downloaded 11 Jul 2013 to 128.95.155.147. Redistribution subject to ASA license or copyright; see http://asadl.org/terms

number of available phonemic distinctions, which did not take into account the occurrence of variation in visual phonetic distinctiveness due to a range of factors (some of which were mentioned earlier). Furthermore, their analyses focused solely on percentage of unique words in the lexicon. Because the number of words perceptually similar to the target word influences its speed and ease of auditory spoken word recognition (Luce *et al.*, 1990), the number of words a speechreader must discriminate among during recognition must be taken into account in order to obtain informative estimates. A computational modeling study of lexical uniqueness in English was undertaken in which the number of available phoneme distinctions was systematically varied to model a range of visual-phonetic information. Several different quantitative measures were also examined in order to understand better how the distribution of word phoneme patterns in English is affected by the number of available phonemic distinctions.

I. LEXICAL MODELING

A. Methods

The methods employed for the current study were originally developed in automatic speech recognition to assess the feasibility of using broad phonetic transcription to select a small subset of lexical candidates from a large lexical database (Huttenlocher and Zue, 1984; Carter, 1987) and have also been applied to the study of human spoken word recognition (see Altmann, 1990; Altmann and Carter, 1989; Pisoni *et al.*, 1985).

The methodology was applied as follows: First, a phonemically transcribed machine-readable lexical database was selected to serve as a representative sample of the words in the language. Along with a phonemic transcription, each word in the database had an estimate of its frequency of occurrence in the language. Second, transcription rules were defined on the basis of measures of phonetic similarity. The transcription rules were in the form of single-symbol substitutions for all phonemes in phonemic equivalence classes. A *phonemic equivalence class* comprised the set of phonemes rendered equivalent by the loss of phonetic distinctiveness. (For example, if /b/, /p/, and /m/ belong to a single phonemic equivalence class, then a rule was defined to transcribe each occurrence of /b/, /p/, and /m/ into a symbol representing the equivalence class.) Third, the lexical database was then transcribed according to the rules. *Lexical equivalence classes* were formed by collapsing across identically transcribed words. (For example, under the phoneme equivalence class definition given above, ''pat'' and ''bat'' would both belong to the same lexical equivalence class.) Lastly, metrics were computed to compare the distribution of patterns in the newly transcribed lexicon with the distribution of patterns in the original lexicon.

1. Lexical database

The method described above was applied to the PhLex database (Seitz *et al.*, 1995). PhLex's entries include the 19 052 most frequent words in the Brown corpus (Kucera and Francis, 1967), the 19 750 words listed in the Hoosier Mental Lexicon (Nusbaum *et al.*, 1984), and 1173 words extracted from stimulus and response sets from our laboratory. In total, after accounting for overlapping entries in the source lists, the PhLex database contains 32 377 unique orthographic entries. In addition to orthographic transcriptions, all of PhLex's entries have baseform phonemic transcriptions[2] that include stress and syllabification symbols, estimates of frequency of usage (Kucera and Francis, 1967), and subjective familiarity ratings (Nusbaum *et al.*, 1984). When an estimate of a word's frequency of occurrence was not available, it was assumed to be equal to 1. All frequencies were log-transformed to the base 10.

2. Transcription rules

Sets of transcription rules were developed using previously published perceptual data on speechreading (Eberhardt *et al.*, 1990; Montgomery and Jackson, 1983). Consonant transcription rules were based on visual consonant identification data collected in a Consonant-/ɑ/ environment (Eberhardt *et al.*, 1990), with overall percent correct of approximately 34.6%. Vowel transcription rules were based on visual vowel identification data collected in the environment /h/-Vowel-/g/ (Montgomery and Jackson, 1983), with overall percent correct of 54.2%. Estimates of visual phonetic similarity were obtained from multidimensional scaling solutions of the obtained consonant and vowel confusion matrices (Bernstein *et al.*, 1994). The analyses employed 40 phonemes, 17 vowels and 23 consonants. The similarity estimates were submitted to separate hierarchical cluster analyses using the average linkage between groups method (Aldenderfer and Blashfield, 1984; SPSS, 1990). Hierarchical cluster analysis was used to algorithmically generate nested sets of phonemic equivalence classes by incrementally joining phonemes based on their estimated similarity.

Sets of transcription rules (see Table I) were generated to model a range of visual phonetic distinctiveness due to a range of viewing conditions, talker characteristics, and speechreaders' abilities. Transcription rule sets were generated by hierarchically varying the total number of phonemic equivalence classes between 1 and 40. Thus, when there was only one equivalence class, all the consonants and vowels were transcribed as a single symbol. When there were 40 equivalence classes, each one contained a unique phoneme of English. Because perceptual data were not available for /ə j ŋ/, /ə/ was assumed to be most similar to /ʌ/, /j/ was assumed most similar to /ɑ/, and /ŋ/ was assumed most similar to /g/. Vowels and consonants were assumed to be maximally dissimilar, except for the consonant /j/ which was included in the vowel confusion matrix. Thus, except for the one equivalence class that contained all the phonemes and the ones that contained /j/, no phonemic equivalence class contained both consonants and vowels.

Table I lists the sets of phonemic equivalence classes employed in the current study. The table shows that the number of vowel and consonant equivalence classes was allowed to increase at the same rate. One heuristic for determining the appropriate number of clusters to represent a perceptual

Downloaded 11 Jul 2013 to 128.95.155.147. Redistribution subject to ASA license or copyright; see http://asadl.org/terms

TABLE I. Phonemic equivalence classes employed in the development of transcription rule sets. All phonemes within a set of brackets were considered perceptually equivalent.

Number of phonemic equivalence classes	Phonemic equivalence classes
28	{u} {ʊ} {ər} {o} {aʊ} {ɪ,i} {e,ɛ} {æ} {ɔɪ} {ɔ} {aɪ} {ə,ɑ,ʌ,j} {b,p,m} {f,v} {l} {n} {k} {ŋ,g} {h} {d} {t} {s,z} {w,r} {ð,θ} {ʃ} {t,ʃ} {ʒ} {dʒ}
19	{u,ʊ,ər} {o,aʊ} {ɪ,i} {e,ɛ} {æ} {ɔɪ} {ɔ} {aɪ,ə,ɑ,ʌj,j} {b,p,m} {f,v} {l} {n,k} {ŋ,g} {h} {d} {t,s,z} {w,r} {ð,θ} {ʃ,tʃ,ʒ,dʒ}
12	{u,ʊ,ər} {o,aʊ} {ɪ,i,e,ɛ,æ} {ɔɪ} {ɔ,aɪ,ə,ɑ,ʌ,j} {b,p,m} {f,v} {l,n,k,ŋ,g,h} {d,t,s,z} {w,r} {ð,θ} {ʃ,tʃ,ʒ,dʒ}
10	{u,ʊ,ər} {o,aʊ} {ɪ,i,e,ɛ,æ} {ɔɪ,ɔ,aɪ,ə,ɑ,ʌ,j} {b,p,m} {f,v} {l,n,k,ŋ,g,h,d,t,s,z} {w,r} {ð,θ} {ʃ,tʃ,ʒ,dʒ}
2	{u,ʊ,ər,o,aʊ,ɪ,i,e,ɛ,æ,ɔɪ,ɔ,aɪ,ə,ɑ,ʌ,j} {b,p,m,f,v,l,n,k,ŋ,g,h,d,t,s,z,w,r,ð,θ,ʃ,tʃ,ʒ,dʒ}
1	{u,ʊ,ər,o,aʊ,ɪ,i,e,ɛ,æ,ɔɪ,ɔ,aɪ,ə,ɑ,ʌ,j,b,p,m,f,v,l,n,k,ŋ,g,h,d,t,s, z,w,r,ð,θ,ʃ,tʃ,ʒ,dʒ}

data set is to look for large increases in the distance coefficients over which clusters are being formed at each stage of the analysis. The logic behind this strategy is that a large increase in the distance coefficient reflects the joining of two relatively dissimilar clusters, and therefore the appropriate cluster solution includes all the clusters formed up to, but not including, that stage (Aldenderfer and Blashfield, 1984). Examination of the distance coefficients obtained in the current cluster analyses suggested that 12 phonemic equivalence classes best approximated the phonemic distinctions perceived by the speechreaders whose data generated the confusion matrices. Furthermore, these 12 phonemic equivalence classes correspond to previous estimates of visually available phonemic distinctions (see Jackson, 1988). In addition, 28, 19, and 10 phonemic equivalence classes were selected to represent $\frac{3}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ of total number of vowels and consonants.

3. Application of transcription rules

Following transcription, lexical equivalence classes were formed by collapsing across identically transcribed words. Table II displays examples of the application of rules to the words ''tan'' and ''cat.'' The first column in Table II gives the number of equivalence classes in the transcription set from which each set of relevant rules for the example is taken. (The complete sets are in Table I.) The third and fourth columns list the transcriptions of ''tan'' and ''cat,'' after the application of the transcription rules. The two words enter into the same lexical equivalence class for all rule sets with ten or fewer phonemic equivalence classes.

The distribution of perceptually similar word patterns could be influenced by lexicon size or method of subsampling the lexicon. Specifically, common and rare words differ in phonemic composition, and common words tend to be

TABLE II. Transcription examples for ''tan'' and ''cat.'' All phonemes within a set of brackets were rewritten into a single symbol represented here by upper-case letters.

Number of phonemic equivalence classes in transcription rule set	Relevant transcription rules	Transcriptions of	
		''tan''	''cat''
19	{æ}⇒A {t,s,z}⇒T {n,k}⇒N	TAN	NAT
12	{ɪ,i,e,ɛ,æ}⇒A {d,t,s,z}⇒D {l,n,k,ŋ,g,h}⇒N	DAN	NAD
10	{ɪ,i,e,ɛ,æ}⇒A {l,n,k,ŋ,g,h,d,t,s,z}⇒D	DAD	DAD
2	{u,ʊ,ər,o,aʊ,ɪ,i,e,ɛ,æ,ɔɪ,ɔ,aɪ,ə,ɑ,ʌ,j}⇒V {b,p,m,f,v,l,n,k,ŋ,g,h,d,t,s,z,w,r,ð,θ,ʃ,tʃ,ʒ,dʒ}⇒C	CVC	CVC
1	{u,ʊ,ər,o,aʊ,ɪ,i,e,ɛ,æ,ɔɪ,ɔ,aɪ,ə,ɑ,ʌ,j,b,p,m,f,v,l,n, k,ŋ,g,h,d,t,s,z,w,r,ð,θ,ʃ,tʃ,ʒ,dʒ}⇒P	PPP	PPP

Downloaded 11 Jul 2013 to 128.95.155.147. Redistribution subject to ASA license or copyright; see http://asadl.org/terms

more like other common words (Landauer and Streeter, 1973). Thus, transcriptions and metrics were applied to three partitionings of the lexical database. One corresponded to the 1052 most frequent words in the Brown Corpus (Kucera and Francis, 1967), henceforth, ''1K;'' the second to the 4995 most frequent words, ''5K;'' and the third to the entire 31 081 word lexicon, ''30K.''[3] Several words at the 1000-word and 5000-word cutoff points had identical frequencies of occurrence.

The presence or absence of stress and syllabification in the transcriptions was manipulated to explore effects of suprasegmental information. The PhLex database contains two levels of marked stress and a third unmarked level. When stress and syllabification were taken into account, two words were considered equivalent only when their phonemic, stress, and syllabification patterns were identical. For example, the noun ''convert'' and the verb ''convert'' were not equivalent when stress was taken into account in the analysis.

4. Quantitative analyses

Following transcription, three commonly employed metrics were computed on the results (see Altmann, 1990). Each metric examined a different aspect of lexical equivalence. The first metric, *frequency-weighted percent words unique*, estimated the proportion of unique words in the transcribed partition of the lexicon. The second metric, *frequency-weighted expected class* size, estimated the average number of words in lexical equivalence classes in the transcribed partition of the lexicon. The third metric, *percent information extracted*, or *PIE* (Carter, 1987), estimated information remaining in the lexicon. The PIE values are a function of the size of resulting lexical equivalence classes and the distribution of frequencies for words within lexical equivalence classes.

Frequency-weighted percent words unique was computed as

$$\%\mathrm{WU}=\frac{\Sigma_{a=1}^{n_U} F_U}{\Sigma_{i=1}^{n_L} F_i}\times 100, \tag{1}$$

where n_U was the total number of unique entries after transcription, F_U was the frequency of occurrence for unique words in the transcribed lexical partition, n_L was the number of words in the lexicon, and F_i was the frequency of occurrence of words in the original lexical partition. The frequency-weighted metric was intended as an estimate of the extent to which unique words would be encountered in everyday language by a deaf speechreader.

Frequency-weighted expected class size was computed as

$$\mathrm{ECS}=\frac{1}{F_L}\sum_{a=1}^{n_E} I_a F_a, \tag{2}$$

where F_L was the sum of frequencies of occurrence of all words in the lexical partition, n_E was the total number of lexical equivalence classes, I_a was the number of words in equivalence class a, and F_a was the sum of frequencies of occurrence of the words in equivalence class a. This metric was intended as an estimate of the average size of the frequency-weighted equivalence classes that would typically be encountered in everyday language.

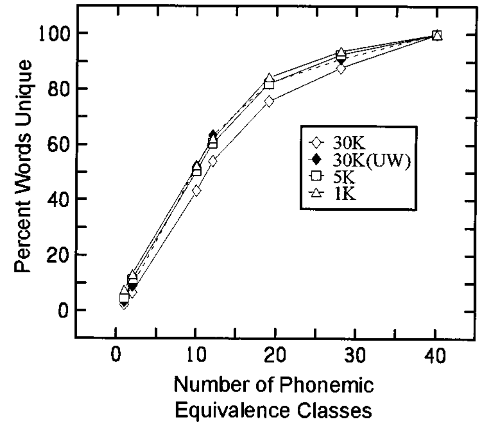

FIG. 1. The log-frequency-weighted percentage of words unique for 1–40 phonemic equivalence classes is plotted for 30 K-, 5 K-, and 1 K-word partitions of the lexicon. The unweighted percentage of words unique for 1–40 phonemic equivalence classes is plotted for the 30 K partition of the lexicon.

The PIE (Carter, 1987) was computed as

$$\mathrm{PIE}=\frac{\Sigma_{a=1}^{n_E} p_a \log p_a}{\Sigma_{i=1}^{n_L} p_i \log p_i}\times 100, \tag{3}$$

where n_E was the total number of equivalence classes after transcription, p_a was the sum of the probabilities of occurrence for words in equivalence class a, n_L was the number of words in the lexicon, p_i was the probability of occurrence of word i, and log was taken to the base 2. The probability of a word's occurrence was computed by dividing its raw frequency of occurrence by the total number of occurrences in the Brown corpus (Kucera and Francis, 1967).

The PIE is an information theoretic metric, developed by Carter (1987), of the amount of *information* extracted from the lexicon when the number of available phonemic distinctions is reduced. In essence, PIE is the number of binary units (bits) required to code the lexicon **after transcription** divided by the number of bits required to code the **original lexicon** multiplied by 100. For example, an original lexicon, containing four equally frequent words, could be coded by two bits. One bit dividing the entire lexicon in half and a second bit dividing each half of the lexicon in half again. A transcription of that lexicon resulting in two equally frequent lexical equivalence classes could be coded by a single bit. The resulting PIE would be equal to 50%. In contrast to expected class size, PIE is sensitive to the distribution of frequencies within lexical equivalence classes. Specifically, PIE is high when all the equivalence classes are roughly equal in frequency which is better approximated when frequent words are in different equivalence classes (see Carter, 1987, for a detailed discussion of PIE).

B. Results and discussion

The results of the analyses for each log-frequency-weighted metric are plotted in Figs. 1–3 as a function of the

Downloaded 11 Jul 2013 to 128.95.155.147. Redistribution subject to ASA license or copyright; see http://asadl.org/terms

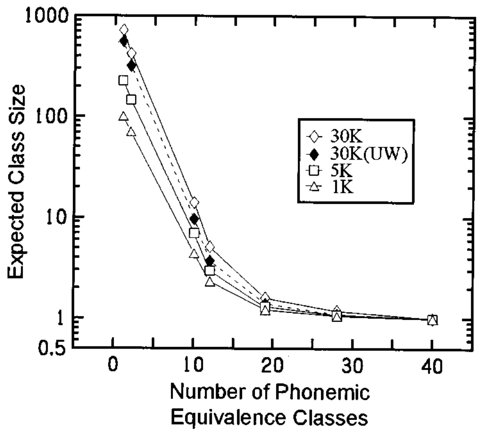

FIG. 2. The log-frequency-weighted expected class size for 1–40 phonemic equivalence classes is plotted for 30 K-, 5 K-, and 1 K-word partitions of the lexicon. The unweighted expected class size for 1–40 phonemic equivalence classes is plotted for the 30 K partition of the lexicon.

number of phonemic equivalence classes in a transcription rule set. Because results were unaffected by the presence versus absence of stress and syllabification, in Figs. 1–3 only the results for analyses with stress and syllabification were plotted. In addition, the unweighted statistics are also plotted in Figs. 1 and 2 for the 30K partition of the lexicon. Only small differences were observed between the unweighted and the log-frequency-weighted data for the 5K and 1K partitions, thus the unweighted 5K and 1K data were not plotted in Figs. 1 and 2. Examination of raw frequency-scaled analyses (i.e., frequencies were not log-transformed), in which all lexical partitions were influenced equivalently by the scaling, suggested that the small effect of log-frequency scaling on the 5K and 1K partitions was related to a trivial interaction of the log scaling and the distribution of frequencies in these lexical partitions.

In Figs. 1–3, 12 phonemic equivalence classes represented the best fit by the cluster analyses for both the consonant and vowel data. Beyond 12 equivalence classes, the distance coefficients in the cluster analyses were very small. This means that additional equivalence classes that were formed as a result of hierarchically separating larger ones into smaller ones represented perceptual distinctions not reliably made by the subjects who viewed the syllables. Thus, if the abscissas of the figures were translated into perceptual distance, they would be highly compressed as the number of equivalence classes increases beyond 12.

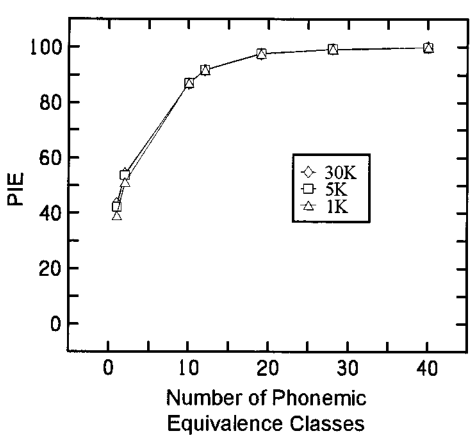

FIG. 3. The percent information extracted for 1–40 phonemic equivalence classes is plotted for 30 K-, 5 K-, and 1 K-word partitions of the lexicon.

Frequency-weighted percent words unique (Fig. 1) increased as the number of visible phonemic equivalence classes increased. Only a small effect of lexical partition size was obtained. These results demonstrate that the distribution of phoneme patterns constituting words in English substantially preserves visual lexical uniqueness, even when the transcription employs only 12 phonemic equivalence classes (i.e., far fewer than half the number of phonemes in the language). This number of equivalence classes corresponds to 54%–63% words unique across partitions. Berger's (1972) frequency-weighted lexical uniqueness estimate of 51% falls below our estimates.

Frequency-weighted expected class size (Fig. 2) decreased as the number of phonemic equivalence classes increased. Expected class size increased slightly as the size of the lexical partition increased. For 12 phonemic equivalence classes, frequency-weighted expected class size was 5.1 words for the 30K lexicon. On average, any given word is predicted to be equivalent with only a few other words, when at least 12 equivalence classes are available.

The PIE (Fig. 3) increased as the number of phonemic equivalence classes increased. However, the largest change in PIE was between 0 and 12 equivalence classes: above 12 equivalence classes, most of the information has been extracted from the lexicon. For 12 phonemic equivalence classes, 92% of the information was extracted. The PIE values were essentially identical, independent of lexical partition size.

The high PIE values obtained even with few equivalence classes provide evidence that high-frequency words tend to reside in equivalence classes with lower frequency words. An implication of this finding is that a speechreader could optimize word recognition accuracy by selecting the most frequent word in a lexical equivalence class. The current results are somewhat surprising given the results Landauer and Streeter (1973) (Pisoni *et al.*, 1985; but see Bard and Shillcock, 1993) have reported indicating that common words tend to be more like other common words. However, the present study differed from previous investigations not only in terms of the nature of similarity among words (visual versus auditory), but also in terms of the operational definition of similarity among words, the lexical entries employed in the analyses, and the metrics employed to assess the similarity of words. Therefore, further analyses are required to assess whether the current results are idiosyncratic to speechreading or have general implications for spoken word recognition.

II. GENERAL DISCUSSION

A. Subject characteristics

The present results demonstrate that lexical uniqueness varies substantially with the number of available phonemic distinctions. Thus, the level of performance in the nonsense

Downloaded 11 Jul 2013 to 128.95.155.147. Redistribution subject to ASA license or copyright; see http://asadl.org/terms

syllable confusion matrices used to estimate phonemic equivalence must be considered in interpreting results of the type presented here.

Berger's (1972) previous estimate of 51% words unique for the average speechreader is below our estimate of 54%–63% for 12 phonemic equivalence classes obtained on the basis of data from relatively inaccurate hearing speechreaders. The consonant confusion data employed here were obtained from matrices with 34.6% correct responses. These subjects also speechread CID Everyday Sentences (Davis and Silverman, 1970) and had mean scores of 24.9% keywords correct at pretest and 35.3% at posttest. Although these scores are typical in the speechreading literature for hearing adult subjects, this performance level is low in comparison with that observed for expert deaf speechreaders. Bernstein *et al.* (1996) reported that expert deaf speechreaders (in the upper quartile of subjects) scored between 61% and 79% words correct on the same sentence materials. These expert deaf speechreaders have achieved 41% correct (with a range from 37%–47% correct) with the same CV materials (Eberhardt *et al.*, 1990; Bernstein *et al.*, 1996). The consonant confusion data employed in the current study is comparable to that used by Berger (1972). However, the current study may represent a conservative estimate of phonemic/lexical uniqueness in the skilled speechreader. We are currently collecting consonant and vowel confusion data from skilled deaf speechreaders to investigate this hypothesis.

B. Lexical characteristics

The high obtained PIE values provide evidence that high-frequency words do not reside in the same lexical equivalence classes as other high-frequency words (Carter, 1987). As suggested earlier, a speechreader might optimize word recognition accuracy by selecting the most frequent word in a lexical equivalence class. Current models of spoken word recognition predict a bias toward selection of high-frequency words in the absence of other discriminating information. We predict that word recognition in speechreaders employs the same biasing processes as those for auditory spoken word recognition, a hypothesis that awaits empirical testing.

C. Modeling assumptions

Two simplifying assumptions employed in our analyses are relevant when considering the relationship of our modeling results to human performance. First, vowels and consonants were operationally defined to be equally distinct and to change their distinctiveness at equal rates. The decision to increase the number of vowel equivalence classes at the same rate as the consonant equivalence classes was a pragmatic one. In the literature, phoneme identification studies have not been conducted in such a way as to provide information on how visibility of vowels and consonants covaries for individual speechreaders. Such data are needed for accurately modeling changes in visibility across the entire phonemic inventory of a language.

Second, phonetic similarity was estimated based on identification of phonemes in monosyllabic nonsense syllables. Coarticulation effects arising from variation in surrounding phonetic contexts have been demonstrated to alter phoneme identification by speechreaders (Benguerel and Pichora-Fuller, 1982; Jackson, 1988). The use of phoneme identifications with monosyllabic nonsense syllables could lead to either an under-or overestimation of the phonetic information available to the speechreader. Although coarticulation is frequently thought to reduce phonetic information, Church (1987) has argued (for acoustic speech) that it is informative for identifying words. We are currently collecting consonant and vowel identification data within a wide variety of disyllabic nonsense word contexts to further refine our estimates of the visually available perceptual information. Preliminary analyses suggest that the distribution of words in the language substantially preserves uniqueness even when variability due to phonetic context is taken into account (Auer *et al.*, 1997).

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (DC02107). The authors thank Paul Iverson, Jennifer A. Johnson, Margaret MacEachern, Helena M. Saldana, Paula E. Tucker, and Robin S. Waldstein for their comments on earlier versions of this manuscript, and Paul A. Luce and Dominic W. Massaro for their reviewer's comments.

[1]We use *phonetic* or *visual phonetic distinctiveness* to refer to the discriminability of visual speech stimuli and *phonemic distinctions* to refer to the phonological differences that distinguish words from each other.

[2]Baseform transcriptions are basic phonemic representations that may involve the undoing of phonological rules. For example, the baseform transcription of the word ''writer'' would be /ˈraɪtər/, which would involve the undoing of an alveolar tapping rule from the pronunciation /ˈraɪɾər/.

[3]Due to homophony, the total number of entries in the analysis, 31 081, was reduced from the total number of orthographic entries in the lexicon, 32 377.

Aldenderfer, M. S., and Blashfield, R. K. (**1984**). *Cluster Analysis* (Sage, Beverly Hills, CA).

Altmann, G. T. M. (Ed.) (**1990**). *Cognitive Models of Speech Processing* (MIT, Cambridge, MA).

Altmann, G., and Carter, D. (**1989**). ''Lexical stress and lexical discriminability: Stress syllables are more informative, but why?'' Comput. Speech Lang. **3**, 265–275.

Auer, Jr., E. T., Bernstein, L. E., Waldstein, R. S., and Tucker, P. E. (**1997**). ''Effects of phonetic variation and the structure of the lexicon on the uniqueness of words,'' in proceedings of the AVSP'97 workshop, (Rhodes, Greece), ESCA ISSN #1018 4554.

Bard, E. G., and Shillcock, R. C. (**1993**). ''Competitor effects during lexical access: Chasing Zipf's tail,'' in *Cognitive Models of Speech Processing: The Second Sperlonga Meeting*, edited by G. Altmann and R. Shillcock (Earlbaum, Hove, UK), pp. 163–185.

Bengeurel, A. P., and Pichora-Fuller, M. K. (**1982**). ''Coarticulation effects in lipreading,'' J. Speech Hear. Res. **25**, 600–607.

Berger, K. W. (**1972**). ''Visemes and homophenous words,'' Teacher Deaf. **70**, 396–399.

Bernstein, L. E., and Auer, E. T., Jr. (**1996**). ''Word recognition in speechreading,'' in *Speechreading by Humans and Machines: NATO ASI Series F*, edited by D. G. Stork and M. E. Hennecke (Springer-Verlag, Berlin).

Downloaded 11 Jul 2013 to 128.95.155.147. Redistribution subject to ASA license or copyright; see http://asadl.org/terms

Bernstein, L. E., Demorest, M. E., and Eberhardt, S. P. (**1994**). ''A computational approach to analyzing sentential speech perception: Phoneme-to-phoneme stimulus-response alignment,'' J. Acoust. Soc. Am. **95**, 3617–3622.

Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (**1996**). ''Speech perception without hearing,'' manuscript submitted for publication.

Carter, D. (**1987**). ''An information theoretic analysis of phonetic dictionary analysis,'' Comput. Speech Lang. **2**, 1–11.

Church, K. W. (**1987**). *Phonological Parsing in Speech Recognition* (Kluwer Academic, Boston, MA).

Davis, H., and Silverman, S. R. (**1970**). *Hearing and Deafness* (Holt, Rinehart, & Winston, Winston, NY).

Eberhardt, S. P., Bernstein, L. E., Demorest, M. E., and Goldstein, M. H. (**1990**). ''Speechreading sentences with single-channel vibrotactile presentations of voice fundamental frequency,'' J. Acoust. Soc. Am. **88**, 1274–1285.

Huttenlocher, D. P., and Zue, V. W. (**1984**). ''A model of lexical access from partial phonetic information,'' in Proceedings of ICASSP (San Diego, CA).

Jackson, P. L. (**1988**). ''The theoretical minimal unit for visual speech perception: Visemes and coarticulation,'' in *New Reflections on Speechreading*, edited by C. DeFillipo and D. Sims (A. G. Bell Association for the Deaf, Washington, DC), pp. 99–115.

Jeffers, J., and Barley, M. (**1971**). *Speechreading (Lipreading)* (Thomas, Springfield, IL).

Kricos, P. B., and Lesner, S. A. (**1982**). ''Differences in visual intelligibility across talkers,'' Volta Rev. **84**, 219–225.

Kucera, H., and Francis, W. (**1967**). *Computational Analysis of Present-Day American English* (Brown U. P., Providence, RI).

Lahiri, A., and Marslen-Wilson, W. (**1991**). ''The mental representation of lexical form: A phonological approach to the recognition lexicon,'' Cognition **38**, 245–294.

Landauer, T. K., and Streeter, L. A. (**1973**). ''The structural differences between common and rare words: Failure of equivalence assumption for theories of word recognition,'' J. Verbal Learn. Verbal Behav. **12**, 119–131.

Lesner, S. A. (**1988**). ''The talker,'' in *New Reflections on Speechreading*, edited by C. DeFillipo and D. Sims (A. G. Bell Association for the Deaf, Washington, DC), pp. 89–98.

Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (**1990**). ''Similarity neighborhoods of spoken words,'' in *Cognitive Models of Speech Processing*, edited by G. T. M. Altmann (MIT, Cambridge, MA), pp. 122–147.

McClelland, J. L., and Elman, J. L. (**1986**). ''The TRACE model of speech perception,'' Cogn. Psychol. **18**, 1–86.

Mongomery, A. A., and Jackson, P. L. (**1983**). ''The physical characteristics of lips underlying vowel lipreading performance,'' J. Acoust. Soc. Am. **73**, 2134–2144.

Nitchie, E. B. (**1916**) ''The use of homophenous words,'' Volta Rev. **18**, 85–83.

Norris, D. G. (**1994**). ''Shortlist: A connectionist model of continuous speech recognition,'' Cognition **52**, 189–234.

Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (**1984**). ''Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words.'' *Research on Spoken Language Processing PR-10* (Indiana University, Bloomington, IN).

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., and Slowiaczeck, L. M. (**1985**). ''Speech perception, word recognition, and the structure of the lexicon,'' Speech Commun. **4**, 75–95.

Scheinberg, J. C. S. (**1988**). ''An analysis of /p/, /b/, and /m/ in the speechreading signal,'' unpublished doctoral dissertation, City University of New York, New York, NY.

Seitz, P. F., Bernstein, L. E., and Auer, E. T., Jr. (**1995**). *PhLex (Phonologically Transformable Lexicon), A 35,000-word pronouncing American English lexicon on structural principles, with accompanying phonological rules and word frequencies* (Gallaudet Research Institute, Washington, DC).

SPSS (**1990**). *SPSS Reference Guide* (SPSS Inc., Chicago, IL).

Downloaded 11 Jul 2013 to 128.95.155.147. Redistribution subject to ASA license or copyright; see http://asadl.org/terms