Computational Explorations of Speechreading

Marilyn E. Demorest
University of Maryland Baltimore County

Lynne E. Bernstein
Center for Auditory and Speech Sciences
Gallaudet University

A computational approach to describing speechreading performance is illustrated using a database obtained from 139 subjects with normal hearing who viewed videodisc recordings of the CID Everyday Sentences (Davis & Silverman, 1970) spoken by a male and a female talker. Four methods of scoring were employed: sentences correct, words correct, phoncmcs correct, and a measure of visual distance between the stimulus and response. The latter two measures were based on a sequence comparator that aligns stimulus and response phonemes to permit phonemic scoring of sentences (Bernstein, Demorest, & Eberhardt, 1991). New techniques for describing normative performance on individual sentences are presented (sentence histograms, response distributions, and a response uncertainty function), and the four measures of subjects' performance are compared. The usefulness of these descriptive methods for suggesting hypotheses about perceptual and cognitive processes in speechreading is also illustrated.

Speechreading has long been a topic of central importance in rehabilitative audiology, but it is gaining recognition as a mode of speech perception used by those with normal hearing as well. Indeed, many cognitive and speech scientists view speech perception as an inherently audiovisual process (Dodd & Campbell, 1987; Massaro, 1987). The long-term goals of our research program are to model perceptual and cognitive processes in speechreading and to characterize the nature and extent of individual differences in those processes. One underlying assumption is that detailed description of performance on different stimulus materials and careful examination of the responses of individual subjects can provide clues about these processes. A second assumption is that important information can be obtained by systematic analysis both of correct responses and of the kinds of errors that observers make.

In the first phase of this research speechreading responses to isolated sentences were obtained from individuals with normal hearing (Demorest & Bernstein, in press). Sentences were chosen as stimuli because they represent a natural unit of everyday communication and because they can be described both in terms of their visual phonetic content and in terms of units such as syllables, words, or phrases. Thus, it is possible to describe the data at several levels of analysis and thereby formulate hypotheses about the interrelationships among them.

Speechreading of isolated sentences is usually described in terms of words correct or key words correct. However, it is also apparent from inspection of responses that many incorrect words and word fragments are visually similar to portions of the stimulus sentence. If we wish to describe stimulus-response correspondences and to quantify the visual similarity of the stimulus sentence and the response, it is necessary to solve an important methodological problem: alignment of the elements of the stimulus with those of the response.

Consider the stimulus and response sentences shown in Example 1:

Stimulus: Why should I get up so early in the morning?
Response: Watch what I'm doing in the morning! (1)

Because English orthography is irregular and because the language contains homophones that are spelled differently, the sentences are first transcribed into a phonetic notation. For this purpose we have used the notational system incorporated in DECtalk (Version 2.0), a computerized text-to-speech synthesizer. The notational system is shown in Table 1, which is adapted from the *DECtalk DTC01 programmer reference manual* (Educational Services Department, Digital Equipment Corporation, 1984). In transcribed form these two sequences become:

Stimulus: wA SUd A gEt ˆp so Rli In Dx morn|G
Response: waC wxt AM duIG In Dx morn|G (2)

Three words are correct (*in the morning* or /In Dx morn|G/), but several phonemes in incorrect words appear by inspection to possibly be correct also. For example the /w/ in *why* (/wA/) may correspond perceptually to the /w/ in *watch* (/waC/) and the /t/ in *get* (/gEt/) may correspond to the /t/ in *what* (/wxt/). Moreover, it seems plausible that the /p/ in *up* (/ˆp/) might correspond to the homophenous /m/ in *I'm* (/Am/).

In order to study such stimulus-response correspondences systematically, it is necessary to have an objective procedure for aligning elements of the stimulus and response. The alignment procedure should also provide for omission (i.e., deletion) of stimulus elements and insertion of elements in the response that have no apparent correspondence with stimulus elements. Given the large number of errors that occur in speechreading, the most appropriate alignment of the stimulus and response is difficult to discern. What is needed is an algorithm (i.e., a computational procedure) for obtaining stimulus-response alignments.

Table 1

DECtalk Single-Character Notational System

Symbol	Example	Symbol	Example	Symbol	Example
a	Bob	|	kisses	S	shin
@	bat	J	gin	t	test
A	bite	k	ken	T	thin
b	bet	l	let	u	lute
c	bought	L	bottle	U	book
C	chin	m	met	^	but
d	debt	M	ransom	v	vest
D	this	n	net	w	wet
e	bake	N	button	W	bout
E	bet	o	boat	x	about
f	fin	O	boy	y	yet
g	guess	p	pet	Y	cute
G	sing	Q	Latin	z	zoo
h	head	r	red	Z	azure
i	beat	R	bird	–	(silence)
I	bit	s	sit		

Note. From *DECtalk DTC01 Programmer Reference Manual* (Table B-1) by Educational Services Department, Digital Equipment Corporation, 1984, Maynard, MA: Author. Copyright 1984 by Digital Equipment Corportation; all rights reserved. Adapted by permission.

The criteria incorporated in the algorithm should produce alignments that have a high degree of face validity and they should reflect the visual confusability of articulatory movements.

During the past few years we have developed and pilot tested a computerized sequence comparison system that produces alignments of stimulus and response phonemes for speechread sentences (Bernstein, Demorest, & Eberhardt, 1991). It is an application of algorithms presented by Kruskal and Sankoff (1983) that has been adapted to reflect phenomena in speechreading. The comparator examines all possible alignments of a phonetically transcribed stimulus sentence and a subject's transcribed response. Because it is assumed that visually similar stimulus and response elements should be aligned, the comparator selects the alignment for which the overall visual similarity of the stimulus and response is maximized. Information about visual similarity of phonemes is provided to the comparator in the form of a matrix of visual distances. The greater the distance between two phonemes, the less similar they are. The matrix of distances used by the comparator was based on multidimensional scaling of nonsense syllable confusions (for details, see Bernstein et al., 1991).

To illustrate, the comparator produced the following alignment for Example 1:

Stimulus: wA SUd A gEt ^p so R l i In Dx morn|G
Response: w a C-- – wxt Am du |G– In Dx morn|G (3)

The computational algorithm successfully aligned correct phonemes and phonemes whose visual similarity is high. In addition to aligning the phonemes mentioned above that appeared to correspond by inspection, it also aligned the /A/ in *why* with the /a/ in *watch*, the /S/ in *should* with the /C/ in *watch*, and the /s/ in *so* with the /d/ in *do*, reflecting the low visual distance between these pairs. One result that was not anticipated was the alignment of /w/ in *what* with the visually distant /g/ in *get*. However, this alignment illustrates a characteristic of the sequence comparator: The criterion for selecting an alignment does not guarantee that every phoneme is aligned only with a perceptually similar one, but rather, that across the entire sentence the visual distance is minimized.

If we accept this alignment as a description of stimulus-response correspondence, it is possible to operationally define additional measures of performance. For example, there are 12 phonemes correct, 10 phoneme substitution errors (or "confusions"), and 4 phoneme deletions. Each of these measures can be divided by the number of phonemes in the stimulus sentence or in the response to yield relative indices of performance. It is also possible to characterize the overall visual similarity of the stimulus and response using the metric in the distance matrix.

The sequence comparator has been applied to a corpus of speechread sentences obtained from observers with normal hearing. The purposes of this article are twofold: (a) to illustrate the kinds of description that are made possible by a computational approach to the study of speechreading and (b) to compare four measures of speechreading performance applied to the same data: sentences correct, words correct, phonemes correct, and overall visual distance between the stimulus and response.

METHOD

Database

The primary database for these analyses was obtained by Demorest and Bernstein (in press). It contains typewritten responses of 104 normal-hearing subjects who viewed videodisc recordings (Bernstein & Eberhardt, 1986) of the 100 CID Everyday Sentences (Davis & Silverman, 1970), spoken by a male and a female talker. Data obtained with the same experimental procedures were also available for an additional 35 subjects, most of whom had participated in laboratory studies of vibrotactile supplements to speechreading (Bernstein, Eberhardt, & Demorest, 1989; Eberhardt, Bernstein, Demorest, & Goldstein, 1990). Twenty-five of these subjects viewed the male talker only.

Transcription

Subjects' responses were reviewed for typing errors, spelling errors, and consistent use of punctuation and contractions (as reported in Demorest & Bernstein, in press). Stimulus sentences and subjects' responses were then transcribed using the text-to-speech synthesizer DECtalk (Version 2.0). The procedure was monitored by a research assistant to insure appropriate processing of word fragments, numerals, isolated punctuation, and other potential transcription problems.

Sequence Comparator

Subjects' responses were submitted to the sequence comparator described by Bernstein et al. (1991). Output of the comparator is an alignment similar to that shown in (3), but with inclusion of word-boundary markers. Deletions are represented as a stimulus element aligned with a dash in the response and insertions are represented as a response element aligned with a dash in the stimulus. Alignments were stored on the computer as text files so that their characteristics could be analyzed further. The comparator also produces a data file containing descriptive information for each sentence: the number of phonemes in the stimulus and the response, the number of phonemes correct, the number of insertions, the number of deletions, and the overall visual distance between the stimulus and the response.

Occasionally, the comparator produced more than one alignment with the same (minimum) overall visual distance. Although alternate alignments were often trivially different, resulting in identical values for performance measures, stimulus-response pairs that produced more than two alternate alignments were excluded. Of the 13,900 sentences analyzed, 632 (4.5%) had more than two alternate alignments and thus were not analyzed. When two alternate alignments were produced, one was selected randomly.

Examination of Alignments

Computer software was developed (Bernstein et al., 1991) that allows the user to search for pre-specified stimulus-response alignment patterns in the database and to tabulate the number of occurrences of each pattern. With this software it is possible to generate stimulus-response confusion matrices for selected sentences, words, phonemes, or sets of phonemes.

Scoring Methods

Sentences correct. Each sentence was scored in a dichotomous manner as correct (1) or incorrect (0). To be considered correct, all phonemes in the sentence had to be correct.

Words correct. A computer program was written to count the number of words correct in each sentence. All words were included in scoring, not just key words. No credit was given for responses that were homophones of the correct words.

Phonemes correct. The number of phonemes correct was tabulated by the sequence comparator. If the subject gave no response to a particular sentence, the comparator did not provide an alignment, but the number of phonemes correct was set to zero.

Visual distance. Each stimulus-response alignment produced by the comparator has associated with it a value for the overall visual distance between the stimulus and the response. Visual distance is the sum of the values in the distance matrix corresponding to each pair of elements in the alignment. Correct phonemes have a distance of zero, visually similar phonemes have small distances, and visually dissimilar phonemes have large distances. For example, /p/ and /m/ have a visual distance of 1, whereas /g/ and /w/ have a distance of 33. Deletions and insertions are assigned a value of 8. The alignment in (3) produces a total visual distance of 141. Although the sequence comparator provides no data when the subject gives no response, non-responses were assigned a value for visual distance equivalent to deletion of all stimulus phonemes (i.e., 8 × the number of stimulus phonemes).

Visual distance between the stimulus and response may be a useful measure of performance because it is a composite that reflects not only the correct responses made by the subject, but also the nature of the errors that are made. As noted above, errors that are visually similar to the stimulus yield small values for visual distance, whereas errors that are very different from the stimulus yield large values. Moreover, when the response bears virtually no visual similarity to the stimulus, the alignment consists of a great many deletions and insertions. Subjects with equal numbers of errors in terms of words and phonemes may differ considerably in the seriousness of their errors.

RESULTS AND DISCUSSION

Illustration: Sentence Histogram

The first type of descriptive analysis to be performed was a tabulation of the percentage correct responses for each stimulus phoneme within each of the 100 CID Everyday Sentences, based on the data for all 139 subjects. Graphs representing these data are referred to as *sentence histograms.*

To illustrate, Figure 1 is a sentence histogram for CID Sentence 1, *Walking's my favorite exercise.* The characters along the horizontal axis give the DECtalk transcription of the sentence. Data are shown separately for the male and the female talker. Sample size was reduced to 122 because 17 subjects gave no response to this sentence.

Figure 1 illustrates the potential usefulness of phonemic scoring for examining the microstructure of performance within a sentence. For example, only 7.2% of these 122 subjects responded correctly with /wcklG/, yet 75.5% correctly identified the initial consonant, /w/. This highly visible phoneme was the initial phoneme in a variety of incorrect word responses such as *what, wool, where, one,* and *watch.* This performance pattern contrasts with the results for the words /mA/ and /fevrlt/, which show a high percentage of whole-word correct responses and comparatively few additional correct phonemes. For /mA/, 42.6% of the subjects responded with the correct word and only four additional subjects (3.3%) got the phoneme /m/ correct. For the word /fevrlt/, 30.3% of the subjects were correct and additional correct phoneme responses ranged from .8% to 13.1%. Comparison of the characteristics of words that tend to be partially

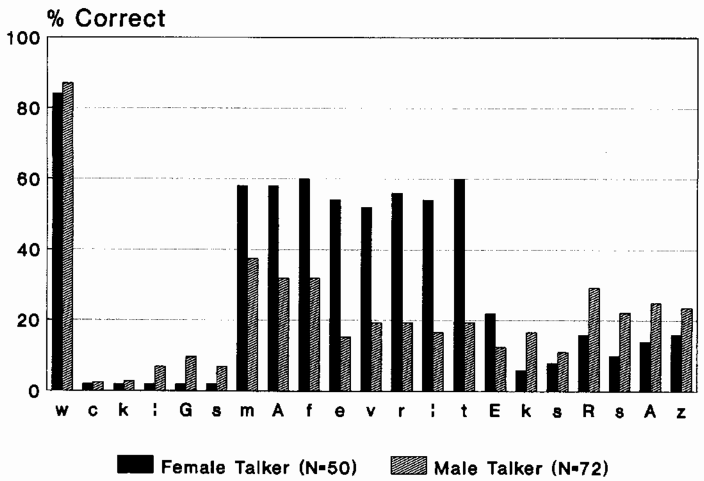

Figure 1. Sentence histogram of CID Sentence 1 for two talkers. Characters on the horizontal axis represent the DECtalk transcription of the sentence, *Walking's my favorite exercise.*

correct with those that elicit an all-or-none pattern of performance may provide insights regarding whole-word versus phonemic-level processing of the visual stimulus.

Another interesting feature of this histogram is the sharp discontinuity in performance between *walking's* and *my favorite*. Although preceding context can often constrain possible word choices, in this example the preceding context was so poorly identified that it provided little useful information for subjects to infer the topic of the sentence. The sharp rise in the percentage of correct responses must therefore reflect visual information present in the stimulus that permitted word identification independently of the preceding context. Such discontinuities can potentially be used to identify locations within sentences where speechreading is driven primarily by the visual stimulus.

One final characteristic of this sentence that is visible in the histogram is a difference between the two talkers. Although the female talker is generally more difficult to speechread than the male (Demorest & Bernstein, in press), for the words *my favorite* this difference is reversed. Such token-specific differences provide a basis for testing hypotheses about the characteristics of the visual stimulus that might account for the overall difference between the talkers. That is, characteristics that might explain the generally *higher* intelligibility of the male talker should be absent or have different values when he is *less* intelligible. Because differences between talkers are an important source of variability in speechreading performance (Demorest & Bernstein, in press), characterization and explanation of talker differences have both practical and theoretical significance.

Illustration: Response Distributions and Response Uncertainty Function

Sentence histograms provide information about correct responses but they do not reveal anything about the kinds of errors subjects make. In Figure 2 the *response distribution* for each stimulus phoneme is presented as a stacked bar graph. The graph shows the proportion of responses that were (a) correct, (b) substitution errors (i.e., phonemic confusions), and (c) deletions. The bottom portion of each bar corresponds to the combined data for the male and the female talker shown in Figure 1.

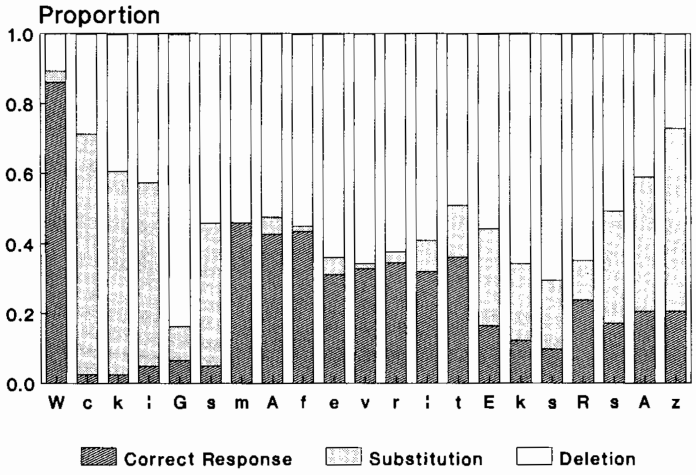

Figure 2. Response distribution for each stimulus phoneme in CID Sentence 1. Characters on the horizontal axis represent the DECtalk transcription of the sentence, *Walking's my favorite exercise.*

One notable feature of the distributions is that some phonemes elicit incorrect responses whereas others result in deletions. Consider the difference between the vowel /c/ in *walking* and the vowel /R/ in *exercise.* Incorrect word responses (see above) resulted in a large number of substitution errors for the first vowel in the sentence. However, for /R/, which has a higher proportion of correct responses, deletion is the dominant category of error.

A second way of summarizing the responses is to quantify the amount of response uncertainty for each stimulus phoneme. Response uncertainty is high when subjects make many different types of errors and it is low when there is a high percentage of correct responses and/or when errors are concentrated in a small number of categories. The uncertainty of a response distribution can be quantified (in bits) by calculating

$$-\sum_{i=1}^{k} p_i \log_2 p_i,$$

where p_i is the proportion of subjects giving response i and i is an index of summation that represents each of the $k = 47$ possible responses in turn. For example, consider the responses to the vowel /c/ in *walking*. Table 2 shows the frequencies of the different responses that occurred. The reduced vowel /x/ and deletion were the most common errors, but a variety of other vowel substitutions occurred as well. The value of $-p \log_2 p$ is shown for each response that occurred and their sum is 2.43 bits. (Although summation is across all 47 possible responses, those that did not occur have $p_i = 0$, and hence contribute nothing to the total.) Response uncertainty was calculated in this manner for each stimulus phoneme in CID Sentence 1 and the *response uncertainty function* given in

Table 2

Response Distribution and Calculation of Response Uncertainty for the Vowel /c/ in CID Sentence 1: Walking's my favorite exercise.

	Response										
	c /ɔ/	**i** /i/	**I** /ɪ/	**E** /ɛ/	**a** /a/	**^** /ʌ/	**o** /o/	**U** /ʊ/	**x** /ə/	**A** /aɪ/	**-** **Deletion**
Frequency	3	2	4	12	7	6	1	2	49	1	35
Proportion (p)	.025	.016	.033	.098	.057	.049	.008	.016	.402	.008	.287
$-p \log_2 p$	.131	.097	.162	.329	.237	.214	.057	.097	.529	.057	.517

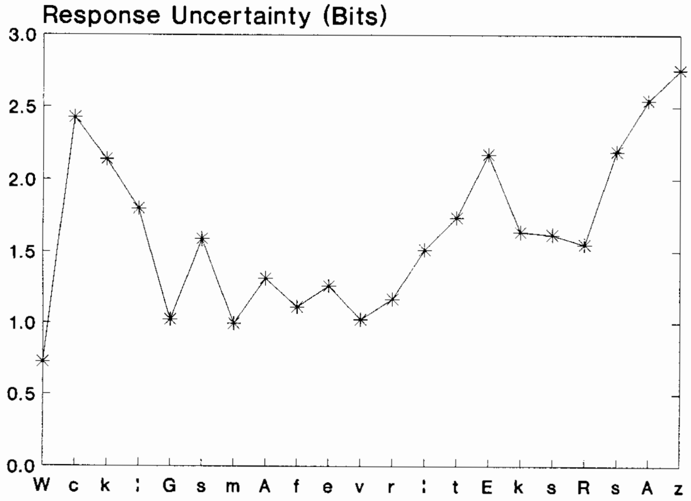

Figure 3. Response uncertainty for each stimulus phoneme in CID Sentence 1. Response uncertainty (bits) was calculated across all possible responses as $-\Sigma_i p_i \log_2 p_i$ where p_i is the proportion of subjects giving response i.

Figure 3 shows how uncertainty varies across the sentence.

For the word *walking*, the uncertainty function tracks the rise and decline in substitution errors from the beginning to the end of the word. At the beginning of the word uncertainty is low because of the high percentage of correct responses, whereas at the end of the word it is low because of the high percentage of deletions. For the phonemes of the final word, *exercise*, the percentage of correct responses is relatively constant, but uncertainty increases as the number and diversity of substitution errors increases.

The response uncertainty function provides another approach to describing performance on speechread sentences. When response uncertainty is low, it reflects unanimity among subjects and it is hypothesized that such unanimity reveals the operation of stimulus-driven and/or linguistic processes. When response uncertainty is high, subjects are making many different responses and the disparity among subjects may be attributable to idiosyncratic speechreading strategies or to other processes such as guessing. Thus systematic study of contexts with high and low response uncertainty may suggest additional hypotheses about sentence processing.

Comparison of the overall shape of response uncertainty functions from one sentence to the next is also potentially informative. For example, if speechreading were a "left-to-right" phenomenon with previous context increasingly determining responding, response uncertainty functions should generally fall from the beginning to the end of a sentence. Or, if sentences were processed in smaller units, such as phrases or clauses, the function might show discontinuities between such units. In contrast, if speechreading were exclusively determined by identification of individual phonemes, the response uncertainty function would simply track the visibility of individual phonemes.

Description of Subjects' Performance

Sentence histograms and response distributions are a method of describing normative performance on different materials. However, performance measures derived from the sequence comparator also provide information about individual differences among speechreaders. In this section two measures from the sequence comparator, phonemes correct and overall visual distance between the stimulus and response, are compared to two measures that can be obtained from conventional methods of scoring: sentences correct and words correct.

Table 3 gives descriptive statistics for the four scoring methods based on the 104 subjects reported by Demorest and Bernstein (in press). Average performance, expressed as a percentage of the maximum possible score, was 9.71% sentences correct, 20.8% words correct, and 24.9% phonemes correct. The visual distance score, which has no maximum possible value, is expressed in arbitrary units that were derived from the distance matrix used by the sequence comparator. If we divide visual distance by 2,124, the number of stimulus phonemes, the resulting value is 7.69. For comparison, note that the response in Example 1, with a total visual distance of 141 and 26 stimulus phonemes,

Table 3

Descriptive Statistics for Four Methods of Scoring Performance on the 100 CID Everyday Sentences

	Number of Sentences Correct	Total Words Correct	Total Phonemes Correct	Total Visual Distance
Mean	9.71	155.46	528.32	16336.16
SD	6.12	73.67	219.96	2601.97
Skewness	1.14	1.06	0.76	−0.02
Minimum	0	1	2	8489
Maximum	31	402	1181	25572

Note. The maximum possible words correct was 749; the maximum possible phonemes correct was 2,124.

results in an index of 5.42, a value better than the average level of performance.

The minimum and maximum scores indicate a wide range of performance for individual subjects. One subject gave only two responses, despite instructions to guess, and obtained only one word and two phonemes correct. The best performers, on the other hand, obtained 31.0% of the sentences correct, 53.7% words correct, and 55.6% phonemes correct. The increases in performance level across the three methods of scoring reflect the additional credit given for partially correct responses.

The methods of scoring that count correct responses all yield score distributions that are significantly positively skewed. The degree of skewness may be interpreted as moderate-to-severe. These measures result in many low scores but they are especially sensitive to individual differences at the upper end of the score distribution. The visual distance measure, however, is not skewed, so it is equally sensitive to individual differences throughout the performance range. This result is a reasonable one, given that visual distance provides information about differences in the seriousness of the errors that subjects make and hence provides additional information about those who make few correct responses.

Estimated test reliability. If the four performance measures are considered four methods of scoring a speechreading test, it is of interest to know whether test reliability is affected by the scoring method selected. Demorest and Bernstein (in press) have presented theoretical functions for test-score reliability, as a function of test length, for the words-correct scores summarized in Table 3. Model 1 in their graph is equivalent to a plot of coefficient alpha (internal consistency reliability). For purposes of comparison, Figure 4 shows this function for each of the scoring methods in Table 3.

Although sentences-correct scoring produces less reliable scores than the other three methods, there is little difference among words correct, phonemes correct,

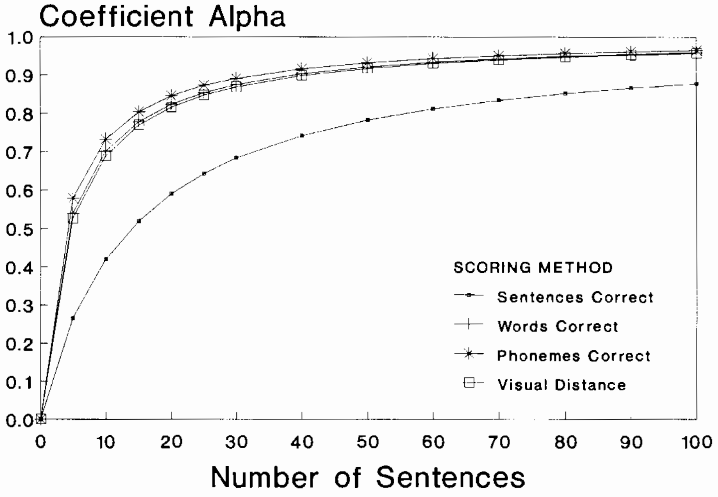

Figure 4. Internal consistency reliability as a function of test length (number of sentences) for four methods of scoring speechreading. The function for words correct is from Model 1, "Sources of variability in speechreading sentences: A generalizability analysis" by M.E. Demorest and L.E. Bernstein, in press, *Journal of Speech and Hearing Research*. Adapted by permission.

and visual distance. The poorer reliability of the sentence scores probably reflects the smaller number of items upon which they are based. That is, for 100 sentences, the words-correct score is based on 749 words and the phonemes-correct score is based on 2,124 phonemes.

Correlations among the four scoring methods. As with any set of alternate scoring methods, it is important to know whether the resulting scores are highly correlated. As shown in Table 4, the magnitude of the correlations is very high, which implies that the four methods tend to rank subjects similarly with regard to speechreading performance. The negative signs on the correlations with visual distance reflect the fact that high visual distance represents poor performance,

Table 4

Correlations Among Four Methods of Scoring Speechread Sentences

	Words Correct	Phonemes Correct	Visual Distance
Sentences Correct	.931	.882	−.854
Words Correct		.984	−.886
Phonemes Correct			−.877

Note. All correlations are significant, $p < .0005$.

whereas low visual distance indicates good performance. Because the magnitude of the correlations in Table 4 approaches that of reliability coefficients, any one of the scoring methods could probably be used as a global measure of a subject's performance on a test of speechreading sentences.

Correlations among adjusted measures of performance. The conclusion just stated is a psychometric one. It refers to the equivalence of the various scoring methods for describing individual differences using a single measure of performance. However, if we view speechreading as consisting of several interrelated tasks, such as phoneme, word, and sentence identification, it is of interest to know whether performance on these tasks is correlated. The correlations in Table 4 do not directly address this question because they are similar to part-whole correlations. Correct sentences contain correct words and correct words contain correct phonemes. In order to estimate the degree of correlation among sentence identification, word identification, and phoneme identification, per se, it is necessary to adjust these measures for the overlap that exists among them.

The correlation between sentences correct and words correct can be estimated if the words-correct score is expressed as a percentage (or proportion) of the words in *in*correct sentences. This procedure eliminates the overlap that occurs when the words in correct sentences are included in the words-correct score. Similarly, the correlation of sentences correct and phonemes correct can be estimated by calculating phonemes correct as a percentage of the phonemes in incorrect sentences. Finally, the correlation between words correct and phonemes correct can be estimated by calculating phonemes correct as a percentage of the phonemes in *in*correct words.

New, adjusted measures were obtained using the computer program that examines alignments and tabulates user-specified patterns. When words and phonemes correct were calculated using only incorrect sentences, their correlations with sentences correct were $r = .856$ and .766, respectively ($p < .0005$). Although elimination of the overlap with correct sentences reduced the correlations, the adjusted correlations are still quite high. Subjects who correctly identified words and phonemes in partially correct sentences were also, to a great degree, the subjects who were correct on a relatively large number of whole sentences.

When phonemes correct was calculated using only incorrect words, the correlation with words correct dropped to $r = .625$ ($p < .0005$). Although this correlation is high enough to suggest that word identification and phoneme identification are related processes, they are by no means equivalent. Interestingly, the correlation between this adjusted phonemes-correct score and sentences correct was only .444. Thus the number of whole sentences a subject had correct was not well predicted by phonemes correct in incorrect words.

Visual distance as a composite measure. The adjustment procedure applied to the measures of correct performance cannot be used with the visual distance measure because it is inherently a composite measure. It is sensitive to correct sentences (which have a visual distance of zero), and to correct words and phonemes, both of which contribute zeroes to visual distance. The more correct responses a subject makes, the lower visual distance is likely to be. This implies that after correct responding is taken into account, the remaining variance in the visual distance measure reflects the nature of the errors that were made. As noted above, this information tends to reveal individual differences among those whose overall performance is at the lower end of the performance range.

To describe the magnitude of this residual variance in the present database, the three measures of correct performance (i.e., sentences correct, words correct, and phonemes correct) were used as predictors in a multiple regression analysis with visual distance as the dependent variable. No adjustment was made for overlap among the measures because regression analysis adjusts for linear association among the predictors. Together the three predictors resulted in a multiple correlation of $R = .948$ ($p < .0005$), which accounts for 89.9% of the variance in the visual distance measure. The remaining 10.1% variance reflects individual differences in the kinds of errors subjects made.

CONCLUSION

This article has illustrated the application and potential usefulness of a computational approach to the study of speechreading. Sentence histograms, response distributions, and uncertainty functions are examples of the kinds of detailed description that can be derived for speechread sentences and that can generate hypotheses about the perceptual and cognitive processes that underlie performance. They represent a normative approach, which emphasizes general patterns for a group of observers. However, as the examples that were presented demonstrate, it is also informative to compare those parts of sentences where subjects tend to make the same kinds of errors with those where more idiosyncratic errors occur.

Comparisons among the various methods of scoring speechreading and describing the performance of subjects have shown how examination of individual differences can also provide valuable information for modeling speechreading. The degree of correlation among scores that represent different levels of linguistic analysis, for example, can help in identifying the component skills that comprise speechreading.

Although the database presented here consists of isolated sentences, the individual-differences approach can be applied to the broad question of what makes someone a relatively successful or unsuccessful speechreader. To answer that question, additional measures of performance on sentences are being developed and other types of materials are being explored (e.g., syllables, isolated words, topic-related sentences, and connected discourse).

A final note is that the computational techniques for studying speechreading, which have been illustrated here with data from normal-hearing subjects, can also be applied to the responses of hearing-impaired speechreaders and can be used for clinical as well as research purposes. With appropriate adjustments in the sequence comparator, they could also be used to describe auditory or audiovisual speech perception.

ACKNOWLEDGEMENTS

This research was supported by NIDCD Grant DC00695 to Gallaudet University. Portions of this research were presented at the Spring 1989 and Spring 1991 meetings of the Acoustical Society of America and at the 1991 Summer Institute of the Academy of Rehabilitative Audiology. The authors wish to thank Ms. Gale DeHaven for her invaluable assistance in data analysis.

REFERENCES

Bernstein, L.E., Demorest, M.E., & Eberhardt, S.P. (1991). *A sequence comparison system for studying sentential stimulus-response correspondences: An application for visual speech perception (lipreading).* Manuscript submitted for publication.

Bernstein, L.E., & Eberhardt, S.P. (1986). *Johns Hopkins lipreading corpus I-II: Disc I* [Videodisc]. Baltimore: The Johns Hopkins University.

Bernstein, L.E., Eberhardt, S.P., & Demorest, M.E. (1989). Single-channel vibrotactile supplements to visual perception of intonation and stress. *Journal of the Acoustical Society of America, 85,* 397-405.

Davis, H., & Silverman, S.R. (Eds.). (1970). *Hearing and deafness* (3rd ed.). New York: Holt, Rinehart and Winston.

Demorest, M.E., & Bernstein, L.E. (in press). Sources of variability in speechreading sentences: A generalizability analysis. *Journal of Speech and Hearing Research.*

Dodd, B., & Campbell, R. (Eds.). (1987). *Hearing by eye: The psychology of lip-reading.* London: Erlbaum.

Eberhardt, S.P., Bernstein, L.E., Demorest, M.E., & Goldstein, M.H., Jr. (1990). Speechreading sentences with single-channel vibrotactile presentation of voice fundamental frequency. *Journal of the Acoustical Society of America, 88,* 1274-1285.

Educational Services Department, Digital Equipment Corporation. (1984). *DECtalk DTC01 programmer reference manual.* Maynard, MA: Author.

Kruskal, J.B., & Sankoff, D. (1983). An anthology of algorithms and concepts for sequence comparison. In D. Sankoff & J.B. Kruskal (Eds.), *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison* (pp. 265-310). Reading, MA: Addison-Wesley.

Massaro, D.W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.