Department of Veterans Affairs

Journal of Rehabilitation Research
and Development Vol. 31 No. 1, 1994
Pages 20–41

Automatic speech recognition to aid the hearing impaired: Prospects for the automatic generation of cued speech

Rosalie M. Uchanski; Lorraine A. Delhorne; Ann K. Dix; Louis D. Braida; Charlotte M. Reed; Nathaniel I. Durlach
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139

Editorial Note: *Harry Levitt, PhD, Editorial Board Member, JRRD*

This issue of *JRRD* includes this paper of special importance about a recent technological advance of great significance for the field of rehabilitation (ASR) systems. There have been many attempts over the years to develop visual displays of speech that will enable deaf and hearing people to converse conveniently with each other. The recent advances in ASR technology open up new possibilities with respect to the development of practical devices of this type. This paper by Uchanski, Braida, et al., heralds the beginning of a new era in this application of ASR technology. Unlike previous studies of this type which have focussed on narrow aspects of the problem, this study addresses technological, perceptual, and human factor issues within an integrated theoretical framework. The results of this work should pave the way to developing visual speech displays for deaf veterans and others who are deaf or have substantial hearing loss. Unfortunately, it was not possible to include this important paper in last year's special issues on Advanced Hearing Aid Technology. Because of its significance, it has been selected for recognition in this issue of the Journal to insure that it receives high visibility.

Abstract—Although great strides have been made in the development of automatic speech recognition (ASR) systems, the communication performance achievable with the output of current real-time speech recognition systems would be extremely poor relative to normal speech reception. An alternate application of ASR technology to aid the hearing impaired would derive cues from the acoustical speech signal that could be used to supplement speechreading. We report a study of highly trained receivers of Manual Cued Speech that indicates that nearly perfect reception of everyday connected speech materials can be achieved at near normal speaking rates. To understand the accuracy that might be achieved with automatically generated cues, we measured how well trained spectrogram readers and an automatic speech recognizer could assign cues for various cue systems. We then applied a recently developed model of audiovisual integration to these recognizer measurements and data on human recognition of consonant and vowel segments via speechreading to evaluate the benefit to speechreading provided by such cues. Our analysis suggests that with cues derived from current recognizers, consonant and vowel segments can be received with accuracies in excess of 80%. This level of performance is roughly equivalent to the segment reception accuracy required to account for observed levels of Manual Cued Speech reception. Current recognizers provide maximal benefit by generating only a relatively small number (three to five) of cue groups, and may not provide substantially greater aid to speechreading than simpler aids that do not incorporate discrete phonetic recognition. To provide guidance for the development of improved automatic cueing systems, we describe techniques for determining optimum cue groups for a given recognizer and speechreader, and estimate the cueing performance that might be achieved if the performance of current recognizers were improved.

Address all correspondence and requests for reprints to: L. D. Braida, 36-747 M.I.T., Cambridge, MA 02139.

Key words: *automatic speech recognition, cueing performance, improved automatic cueing systems, Manual Cued Speech.*

INTRODUCTION

Few technologies with the potential for improving communication by the deaf and hearing-impaired have advanced as rapidly during the past decade as automatic speech recognition (ASR). These advances have reduced substantially limitations associated with vocabulary size, the need to speak words in isolation, speaker dependence, and constrained grammars to achieve high levels of recognition accuracy. Research systems that recognize isolated words spoken by a single talker from a 5,000 to 20,000-word vocabulary already achieve word error rates as low as 2.9–5.4% (1). For words spoken without pauses, word recognition rates as high as 81.9% have been reported for a system with a 1,000-word vocabulary that operates in a speaker-independent mode with unconstrained grammar and 96.2% when constrained by a simple word-pair grammar (2).

Although there are many unsolved problems in developing a large-vocabulary, speaker-independent recognition system suitable for continuous-discourse and robust to environmental disturbance, the potential for using current speech recognition technology to alleviate the communication difficulties experienced by the deaf is only poorly understood. Two recent efforts illustrate a range of strategies to develop and evaluate this potential. Although these efforts use differing recognition strategies, they are perhaps more interesting because they analyze different modes in which the recognition system would be used by a hearing-impaired individual: to recognize and produce visual displays of words that could augment residual hearing; and to recognize and produce visual displays of phonetic units that could be read like text without auditory cues.

Discrete Word Recognition

This approach for utilizing ASR technology has recently been advocated by Kanevsky, et al. (3). A discrete word recognition system (4) presents visual transcriptions of words in synchrony with the acoustic signal. As configured for this task, the recognizer achieved roughly 65% accuracy on a 20,000-item vocabulary. In a formal evaluation of this system, six normal-hearing listeners were tested on 60 nonsense sentences containing five key words each. Three presentation conditions were used: H (hearing alone, in the presence of additive white noise, yielding roughly 60% correct key word scores), D (visual presentation alone, using a simulated recognition system that produced 60% correct key word transcriptions), and HD in which both the H and D displays were available. Average error rates across subjects were 43.9% (H), 44.0% (D), and 25.4% (HD). Although each of the six subjects achieved lower error rates in the HD condition than in either the H or the D conditions, for four of the subjects the gain relative to the better of the two latter conditions was small. The two subjects who achieved significant gains using the recognized words in conjunction with the acoustic signal were also the subjects who performed poorest in the H and D conditions.

Phonetic Recognition

During the past decade, the Sensory Aids Foundation (5) sponsored a preliminary evaluation of the potential of ASR systems to derive and represent continuous speech for the deaf in a stream of letter-sized phonetic units. Krasner, et al. (6) evaluated the performance of a HMM (Hidden Markov Model) phonetic recognizer operating in a speaker-dependent continuous speech mode. When trained on 30 minutes of speech data, the system was able to achieve 81% correct phone recognition with a deletion rate of 3% and an insertion rate of 13%. Slightly more than half of the substitution errors were within phone classes (e.g., fricative, plosive), with the remainder scattered across classes. Taking the performance of this recognizer as representative of current technology, Huggins, et al. (7) developed a variety of techniques for displaying the recognized phone stream (assumed to lack word boundary indications) visually so that it could be read easily. In addition, they evaluated the ability of both normal-hearing and hearing-impaired subjects to read phone representations of Harvard (8) sentences and passages under conditions that simulated both perfect transcription and a variety of recognizer error rates. For perfectly transcribed materials, one hearing-impaired subject was able to read the phonetic transcriptions of simulated rapidly spoken speech at rates in excess of 75 words per minutes (wpm) after only 4 hours of exposure and

Journal of Rehabilitation Research and Development Vol. 31 No. 1 1994

training. Five subjects were able to read this material at 25–35 wpm and two others were only able to read transcriptions of carefully pronounced material at rates of 13–22 wpm. Although the best performance achieved by these subjects is inferior to normal reading or speaking rates, 150–180 wpm (9), the results were viewed as encouraging because reading rates improved throughout the period of training and testing. One evaluation of the readability of transcriptions that included simulated recognition errors employed a single normal-hearing listener and considered error rates equal to 0, 1/3, 2/3, and 1 times a base rate (14% substitutions, 5% deletions, and 6% insertions). Reading rates declined from 69 wpm (0 errors) to 26 wpm (1/3 base error rate), 21 wpm (2/3), and 14 wpm (base error rate). Moreover, reading errors, which were nearly absent in the perfect transcriptions, increased to roughly 30% at the full error rate. In other tests, only two (of seven) hearing impaired listeners were able to read the phone sequences at rates in excess of 35 wpm with 1/3 of the base error rate; the others achieved rates of 14–18 wpm under this condition. Apparently, as soon as errors were introduced, the subjects were unable to read the phonetic streams fluently; instead the task became like solving a puzzle. Very little training could be provided under such conditions and it did not improve performance substantially. To be effective for most readers, error rates lower than 5% substitutions, 2% deletions, and 2% insertions appear to be required for this scheme.

Summary

Limitations on the performance of current automatic discrete word recognition systems and on the ability to read streams of phonetic symbols generated by real or hypothesized acoustic-phonetic recognizers suggest that communication performance with the output of current real-time speech recognition systems would be extremely poor relative to normal speech reception. An alternate approach for using ASR technology to aid the hearing impaired derives cues from the acoustic signal that are not sufficient for communication in themselves, but can be used to supplement the visual cues available from speechreading. The requirements for such an application are likely to be less demanding, because the visual speechreading signal provides some reliable cues to word boundaries and is also robust to the effects of deleterious acoustical environments. In the balance of this paper, we report on the effectiveness of Cued Speech (CS) in communication; evaluate the performance of current recognition systems as producers of cues to supplement speechreading; and estimate the segment recognition accuracy to be expected using such cues in conjunction with speechreading. We conclude by estimating the effectiveness of potential automatic cueing systems for reception of connected speech.

Effectiveness of Manual Cueing

Although many hearing impaired individuals make use of speechreading, the ability to communicate through speechreading alone is severely constrained, because many acoustic distinctions important to communication are not manifest visually. While some of the resulting ambiguity can be overcome if the receiver makes use of contextual redundancy, the concomitant demands on attention and memory are often daunting (10). Even when dealing with easy material that is spoken slowly, with repetitions, and is viewed under optimal reception conditions, speechreaders typically miss more than one-third of the words spoken. Limitations on communication via speechreading have been analyzed by studying confusions made in segment recognition tasks. Speechreaders are usually unable to distinguish between consonants differing in voicing. Roughly speaking, only 5–9 groups of consonants and 4–7 groups of vowels are resolved well by speechreaders (11,12).

Cornett (13), recognizing that the ambiguities facing the speechreader would be resolved if the talker signaled the receiver appropriately, developed Cued Speech as a supplement for speechreading that could be taught to very young deaf children and used in day-to-day communication. In Cued Speech eight visually distinct hand shapes are used to distinguish between consonants that are not well resolved via speechreading and four hand positions near the mouth provide distinctions within groups of vowels often confused in speechreading. Thus, within each group, the set of sounds are easily distinguishable through speechreading alone; sounds that are difficult to distinguish through speechreading alone are placed in different groups. In Manual Cued Speech, the consonant groups are /t, m, f/, /h, s, r/, /d, p, zh /, /ng , j, ch/, /l, sh, w/, /k, v, xh , z/, /n, b, wh/, /g, dzh ,th/ and the vowel groups are /a, uh, o/, /ae , I , uu/, /aw , eh ,

u/, and /i, er/ (**Table 1**). Diphthongs are cued as sequences of vowels.

Previous Studies

Although Manual Cued Speech has achieved some acceptance in educational settings, its effectiveness as a communication medium has not been studied extensively. Kaplan (14) evaluated the speechreading ability of 18 prelingually deaf individuals (average age of 16 yr) who had completed at least a basic course in Cued Speech, and had taken advanced subjects using Cued Speech or used Cued Speech for communication at least on a part-time basis. Test materials included nonsense C-/i/ syllables, monosyllabic words, and sets of related and unrelated sentences. Items were presented twice before subjects were required to respond, and tests were conducted under conditions of speechreading (S) and Cued Speechreading (S+C). Scores on phoneme and monosyllable tests were roughly 30% and 70% in the S and S+C conditions, respectively. In the S condition, word scores were somewhat higher for the related sentences than for the unrelated sentences (66% vs. 56%), but in the S+C condition they were roughly equal (75%). Thus, Cued Speech provided improved reception relative to speechreading on all tests, although the magnitude of the improvement was largest in the tests with low-context materials.

Ling and Clarke (15), Clarke and Ling (16), and Nicholls and Ling (17) all tested children in schools for the deaf and found substantial improvement in speech reception when Manual Cued Speech was used to supplement speechreading. In the study with the most dramatic results, Nicholls and Ling (17) evaluated 18 senior-class children who had used Manual Cued Speech for an average of 7.3 yr. Tests measured reception of syllables and words in Low-Predictability (LP) and High-Predictability (HP) sentences. They found that key-word scores for short sentence materials improved from 25.5% (LP) and 32.0% (HP) correct for speechreading alone to 96.6% (LP) and 96.2% (HP) correct when cues supplemented speechreading. Although these reports are highly encouraging, they may not document the effectiveness of the method for more difficult materials or under more realistic communication conditions. Test words were highly familiar monosyllabic nouns and were restricted to occur at the end of short (4–9 word) sentences. Scores of 42.9% (LP) and 50.0% (HP) were achieved through reception of the cues alone. Moreover, the cued and uncued test materials were not standardized for speechreading difficulty.

Table 1.
Phonetic equivalents of orthographic symbols used in the text and tables.

Orthographic representation of Phonetic Symbols.					
Consonants.			Vowels.		
Orthographic	Phonetic	Example	Orthographic	Phonetic	Example
th	θ	thin	a	ɑ	father
sh	ʃ	shin	I	ɪ	hid
xh	ð	there	uu	U	hood
ch	ʧ	chew	uh	ʌ	bud
dzh	ʤ	judge	aw	ɔ	bought
zh	ʒ	azure	ae	æ	had
wh	ʍ	where	er	ɝ	herd
ng	ŋ	sing	eh	ɛ	head

METHOD/RESULTS

Current Evaluation of Cued Speech

To obtain a clearer picture of the potential of the method, and also to obtain background data relevant to the study of artificial cueing systems, we tested experienced cue receivers on three tests of sentence reception.

Four subjects were tested (**Table 2**). All are highly experienced in the reception of Cued Speech and possess an excellent command of the English language. The subjects were 18–27 years of age and had acquired hearing losses before the age of 2 years. All have some residual hearing and wear hearing aids. Their experience with Cued Speech ranges from 8 to 25 years. For all subjects, Cued Speech was used extensively during the primary and secondary school years. During that time, transliterators were provided in their classrooms, other classmates used Cued Speech, and at least one family member used the system at home. However, at the time of this study, all subjects had graduated from high school and no longer lived at home. Their use of Cued Speech ranged from less than 1 hour to several hours per day.

Journal of Rehabilitation Research and Development Vol. 31 No. 1 1994

Table 2.
Description of the highly experienced receivers of manually Cued Speech (CS).

Subject	Age (years)	Deafness Onset (months)	Etiology	Past CS Use (years)	Current CS Use (hours/day)
S1	19	18	Unknown	13	1–2
S2	18	Birth	Unknown	15	1–8
S3	18	3	Unknown	8	1
S4	27	Birth	Rubella	25	1–2

Experiments were conducted to evaluate sentence reception through Speechreading alone (S) and Speechreading plus Manual Cued Speech (S+C). Videotaped recordings of the sentences (with no audio signal) were presented on a color monitor. Each sentence was presented only once and the subjects were given as much time as necessary to record responses. Subjects were encouraged to write down any part of the sentence that they were able to perceive and to guess if they were not sure. Correct answer feedback was not provided. S1, S2, and S3 were tested in a single 5-hour session each while S4 was tested in one 2.5-hour session.

The videotaped sentences used for the testing were recorded by two normal-hearing females (T1 and T2), both of whom are teachers of Cued Speech. Each served as a speaker/cuer and also checked the accuracy of the manual cues produced by the other. Cued and uncued sentences were recorded using the same camera viewing angle to keep the size of the speaker's face the same for both conditions. The speakers were instructed to control facial expression.

The sentence materials used for testing (**Table 3**) were CID (18), Clarke (19), and Harvard (8) sentences. The CID sentences (10 lists of 10 sentences) are typical of everyday, conversational speech. Although CID sentence length ranges from 2–13 words/sentence, each list of 10 sentences contains 50 key words which are scored for accuracy of reception. The Clarke sentences (60 lists of 10 sentences), developed to measure the intelligibility of the speech of deaf students, are also similar to conversational speech. Clarke sentences have uniform length (10 syllables) and all sentences are in statement form. The vocabulary is appropriate for children aged 7–8 years. Since a standard scoring procedure is not available for reception of these sentences, we scored each word in each sentence. The lists used in our studies contain roughly 80 words each. The Harvard sentences (72 lists of 10 sentences) are considerably more difficult than the CID or Clarke materials. Each sentence contains five key words (four monosyllabic and one disyllabic). The Harvard lists are phonetically balanced and the sentences have very low internal predictability. **Table 3** also includes measures of the average durations of the acoustic signal for the test materials used. These durations correspond to speaking rates of roughly 150 wpm for the uncued materials and 100 wpm for the cued materials.

For each subject, equal numbers of sentences were presented for the S and the S+C conditions. Three subjects (S1, S2, and S3) were each tested on a total of 100 CID, 100 Clarke, and 200 Harvard sentences. S4 was tested on 40 CID and 200 Harvard sentences. All subjects were given the same lists of sentences produced by the same speaker in the order of presentation shown in **Table 3**.

For reference purposes four subjects with normal hearing, none of whom had been trained to interpret Cued Speech, were tested on the same set of Harvard and Clarke sentences. These subjects, all of whom had normal corrected vision, were the best speechreaders screened from a group of 14 applicants. The scores of these subjects were sufficiently similar that only average scores are reported.

Strict criteria were used to score the written responses. Other than trivial spelling errors and alternates for homonyms, exact agreement was required: plural nouns were not accepted for singular nouns, verb tense had to be correct, etc. Scores are shown in **Table 3**. For the CID sentences, the

Table 3.
Reception of cued and non-cued sentences.

Sentence Lists	Speaker	Cued	Dur. (s)	Subjects S1	S2	S3	S4	Avg	4 N
CID A–E	T1	Yes	3.13	96	98	96	100	97	
CID F–J	T1	No	2.31	75	60	62	52	62	
Clarke 20–24	T2	Yes	4.50	97	99	95		97	36
Clarke 30–34	T2	No	3.11	47	43	45		45	24
Harvard 40–44	T1	Yes	4.03	72	82	68	86	78	14
Harvard 30–34	T1	No	2.83	21	23	24	16	21	16
Harvard 45–49	T1	Yes	4.15	85	93	86	91	89	22
Harvard 35–39	T1	No	2.70	34	30	33	18	29	15

Results of the sentence tests used to evaluate the visual reception of uncued and manually Cued Speech by four trained receivers of Cued Speech (S1–S4) and four subjects with normal hearing untrained in Cued Speech (4 N). Scores shown are percent correct keywords for CID and IEEE sentences, and the percent correct of all words for Clarke sentences. Only the first two lists of CID sentences (A,B and F,G) were presented to subject S4.

cue receivers' scores averaged 62% in the S condi tion and 97% in the S+C condition. Scores on Clarke sentences for these subjects averaged 45% without cues and 97% with cues. The subjects with normal hearing also achieved somewhat higher scores on the cued materials (36% vs. 24%), although the difference was much smaller than for the cue receivers. For the less contextual Harvard sentences, scores were lower in all conditions. Cue receiver scores averaged 25% (S) and 84% (S+C), while for the untrained subjects average scores were 18% for both cued and uncued materials. Overall, the cue receivers we tested appear to be very good speechreaders, achieving higher scores on all tests with uncued materials than our normal-hearing group.

These measures of the effectiveness of manually produced cues as aids to speechreading are consistent with those of Nicholls and Ling (17) in suggesting that highly trained receivers of Cued Speech can achieve nearly perfect word reception on conversational materials (CID and Clarke sentences) when cues are available. Even when there is only very restricted sentence context, as in the Harvard sentences, very large improvements can be achieved with cueing. As might have been expected and as mentioned above, the cue receivers that we tested proved to be very good speechreaders, achieving higher scores on all tests with uncued materials than our normal-hearing groups. With respect to the receptive potential for Cued Speech, the performance of the trained cue receivers is very encouraging. However the higher scores obtained with the cued materials used in this study may be partially due to differences in articulation. The cued sentences were spoken more slowly, and perhaps more clearly, than the uncued materials. This may be responsible for the slightly higher scores obtained by the untrained speechreaders on the cued Clarke materials. For Harvard sentences, however, differences in articulation between the cued and uncued materials did not affect the reception scores of our normal-hearing listeners.

CUES GENERATED BY AUTOMATIC SPEECH RECOGNIZERS

The Gallaudet-R.T.I. Automatic Cueing System

Early reports of the effectiveness of Cued Speech in classroom settings spurred attempts to develop an "Autocuer," a system that would analyze the acoustic speech signal and produce a visual display approximating Manual Cued Speech. In one implementation of this system (20), the display consists of the projected virtual image of a pair of

Journal of Rehabilitation Research and Development Vol. 31 No. 1 1994

seven-segment LED elements. Selective activation of the segments allows nine distinct symbol shapes, cueing consonant distinctions, to be created in each of four display locations, cueing vowel distinctions. The consonant groups used are /b, g, dzh/, /k, p, z/, /h, s, f, th/, /l, r,sh/, /v, t/, /m, n, ng/, /j, ch/, /d, xh, zh/, /-, w, wh/, while the vowel groups are /a, i, o/, /aw, eh, u/, /-, uh, er/, and /ae, I, uu/. Diphthongs are cued by illuminating pairs of shapes at distinct locations sequentially. These phoneme groups are not the same as those of Manual Cued Speech since it was necessary to make concessions to the difficulties of speech analysis. However, this code satisfies the criterion that sounds which are difficult to distinguish through speechreading are placed in separate cue groups.

Evaluations of the most recent implementation of this Autocuer indicate that the recognition accuracy of phonemes in isolated words is roughly 54%, with a 33% deletion rate and a 13% substitution rate.[1] With this level of recognition performance, identification scores for isolated words using the Autocuer system as an aid to speechreading were roughly 60% correct. Though not tested with sentence materials, the Autocuer's level of phone identification combined with its high deletion rate would probably not lead to good sentence reception. A major shortcoming of the recognition system used in the Autocuer appears to lie in its difficulty with proper segmentation: roughly one-third of all phonemes are undetected by this system, and consequently are unrecognized. Although a prototype of a wearable version of this Autocuer has been produced, it does not appear to be in use at the present time.

Potential Automatic Cueing Systems

Recently, Vilaclara (21) described a speech processing system designed to recognize phonemes and produce 11 cues, called "kinemes," complementary to the speechreading of French. There are six consonant kineme groups: /p, t, k/, /b, d, g/, /m, n, ng/, /f, s, sh/, /v, z, zh/, and /l, r/. Five vowel groups, appropriate for French, are also used. An evaluation of this system on sentences produced by a single speaker (450 consonant segments and 603 vowel segments) indicates that kinemes are identified with 79% accuracy. Most errors occur within consonant and vowel groups; confusions occur between consonants and vowels only roughly 5% of the time. The six consonant groups are assigned with an accuracy of 92% and the five vowel groups with an accuracy of 84%. No evaluation of this system as a speechreading supplement has been reported.

Most current speech recognizers are not designed to provide cues that aid speechreading. In order to estimate the performance that might be achieved with current and future technology, we evaluated how well cues could be derived from the speech waveform by the M.I.T. recognition system[2] (22) and by highly trained spectrogram readers.

Two different speech corpora were used to evaluate the M.I.T. automatic speech recognition system. The first corpus is derived from the TIMIT database (23). For training, 1,500 sentences from 300 speakers were used as input to the recognizer. For testing, 225 sentences from 45 speakers were used. The second corpus consisted of 100 Harvard sentences from one speaker for training, and an additional 100 Harvard sentences from the same speaker for testing. For both corpora, the test sentences were segmented prior to the automatic recognition process. Thus, the recognizer's task was solely a labelling task and hence, by experimental design, there were no insertions or deletions. The recognizer's performance is reported as confusion matrices of phonetic labels, ARTI (*A*utomatic *R*ecognizer trained on *TI*MIT materials) and ARHS (*A*utomatic *R*ecognizer trained on the *H*arvard *S*entences), with 4,229 and 3,134 sound segments, respectively.

The performance of "spectrogram readers" may provide a useful estimate of the potential performance of future automatic recognition systems. To this end, we asked three highly trained spectrogram readers (SRHT) to label spectrograms using a newly designed classification system that we expected both to provide a useful supplement for speechreading and to be easily extracted from the speech signal. This new cue system (CBG) uses six consonant classes, /p, t, k, ch/, /b, d, g, dzh/, /f, th , s, sh/, /v, xh , z, zh/, /m, n, ng , r/, and /w, wh, l, h, j/. No cue system was specified for vowels. Instead, the spectrogram readers labelled each of the

[1]R. Beadles, letter to L.D. Braida, August 31, 1989, describing the performance of the RIT Autocuer.

[2]We have been unable to obtain comparable data for other recognition systems in a form that would be useful for our purposes.

10 English monophthongs and marked, but did not identify, diphthongs.

The highly trained spectrogram readers segmented and labelled roughly 560 phonetic segments each from 30 nonsense sentences—3 male speakers, 10 different sentences per speaker (24). The readers were instructed to identify consonants as members of one of the six CBG groups, identify diphthongs as a single group, and label monophthong vowels individually with their IPA symbols (10 monophthong vowels). No time limit was imposed for assigning these labels. Thus, the readers assigned each sound segment to one of 17 groups (6 consonant groups, 1 diphthong group, 10 monophthong vowels). The performance of one expert spectrogram reader (SRVZ) was estimated from measurements made by Cole, et al. (25) in which 23 sentences (a mixture of meaningful and nonsensical, produced by two speakers) were both segmented and labelled. We mapped these labels (a total of roughly 400 segments) onto the corresponding CBG code.

Table 4 compares the performance of the spectrogram readers to that of the M.I.T. speech recognizer. Overall, SRVZ achieves the highest scores. The performance of ARHS and ARTI is roughly comparable to that of SRHT on the vowel labelling task. The recognizer studied in the VIDVOX project operating in a speaker-dependent mode achieved a score of 72% on vowels, compared to 65.2% for ARTI in a speaker-independent mode. The accuracy of Vilaclara's system in identifying the six "kineme" consonant groups (92%) is clearly superior to that of the ARTI and ARHS systems in identifying the six CBG consonant groups (78%). It is clear, however, that these automatic recognizers cannot produce reliable cues automatically. On the other hand, the inferior performance of the automatic recognizers when compared with that of the spectrogram readers for both consonants (SRVZ and SRHT) and vowels (SRVZ), suggests that future recognizers can be improved (e.g., by the incorporation of additional human speech knowledge).

INTEGRATION OF CUES WITH SPEECHREADING

The errors made by automatic speech recognizers include substitutions, deletions, and insertions. Deletions will result in the absence of cues, insertions in extraneous cues, and substitutions may result in the production of erroneous cues. Lee and Hon (26) have described a speaker-independent automatic recognition system that achieves a deletion rate of 6.6% and an insertion rate of 7.7%, close to the rates of the highly trained spectrogram readers, indicating that substantial improvement has been made relative to the Autocuer described previously. The effect of deleted, inserted, and erroneous cues on speech reception is not known. Not only are such errors likely to result in imperfect resolution of visual ambiguities, but they are likely to require the receiver to rely more heavily on context. Additionally, the time required for recognition processing

Table 4.
Performance of Human Spectrogram Readers and ASRs.

Task	Size	Recognizer			
		SRVZ	SRHT	ARHS	ARTI
Overall	17 × 17	90.2	72.2	67.9	69.2
Cons-Vow	2 × 2	99.5	94.7	92.9	94.8
Consonants	6 × 6	94.1	85.1	77.4	77.9
Vowels	11 × 11	82.9	62.0	65.4	65.2
Deletions		3.2	6.4		
Insertions		0.0	8.7		

Identification scores on four labelling tasks for the expert (SRVZ) and three highly trained spectrogram readers (SRHT), and of the M.I.T. automatic recognition system trained on Harvard sentences (ARHS) and the TIMIT database (ARTI). Also included are deletion and insertion rates for the spectrogram readers. See section on Potential Automatic Cueing Systems for details.

Journal of Rehabilitation Research and Development Vol. 31 No. 1 1994

needs to be considered. In general, one expects recognition accuracy to improve as processing time increases, but little is known about the effectiveness of cues that are delayed relative to the speechreading signal. Since such factors are not well understood at present, we restrict our attention to the problem of identifying consonant and vowel segments correctly using speechreading in combination with cues that might be generated by an imperfect or restricted recognizer.

Theory of Cue Integration for Segments

Our analysis estimates the highest identification accuracy that can be achieved by a receiver that makes optimal use of both speechreading and the imperfect cues. We assume that the cues themselves are perceived accurately and do not interfere with the reception of speechreading (i.e., the only errors expected are those associated with the production of the cues by the recognizer and the reception of segments via speechreading in the absence of cues). Moreover, we assume that, conditioned on the segment produced, the perceptual errors made in speechreading are statistically independent of those made by the recognizer. This assumption is justified in part because the recognizer is dependent on the acoustical signal rather than the visual signal.

We use a computational technique employed in the Post-Labelling Model of audiovisual integration (27). This model has been shown to underestimate audiovisual consonant reception accuracy slightly, but seems appropriate to use when discrete visual cues, rather than the continuous acoustic signal, supplement speechreading. This model predicts both the overall level of accuracy and the pattern of perceptual errors in the S+C condition based on knowledge of the error patterns characteristic of speechreading alone (S) and of the cue generator (C). We assume that the cue receiver is aware of the likelihood of occurrence of the various phones in English, as well as the likelihood that each cue and each speechread phone is perceived when each phone is spoken. The cue receiver identifies the spoken phone based on the perceived cue and the speechread phone using rules that maximize the overall likelihood of correct identification.

In the S+C condition, presentation of phone $\mathcal{P}_i$ is assumed to produce a pair of labels (S_m, C_n) corresponding respectively to the speechread and cue percepts. The collection of label pairs is assumed to be divided into mutually exclusive and collectively exhaustive sets, one set per phone. In particular, all labels belonging to the set Ω_j elicit phone response R_j so that

$$P_{SC}(R_j|\mathcal{P}_i) = \sum_{m,n\epsilon\Omega j} P_{SC}(S_m, C_n|\mathcal{P}_i). \quad [1]$$

where $P_{SC}(S_m, C_n|\mathcal{P}_i)$ is the probability that phone $\mathcal{P}_i$ gives rise to the label pair (S_m, C_n) and $P_{SC}(R_j|\mathcal{P}_i)$ is the probability that response R_j is given to this phone. As noted above, we assume that, conditioned on the spoken phone, the assignment of cues is statistically independent of the assignment of phone labels via speechreading, so that

$$P_{SC}(S_m, C_n|\mathcal{P}_i) = P_S(P_S(S_m|\mathcal{P}_i) \times P_C(C_n|\mathcal{P}_i) \quad [2]$$

There are many possible rules (or equivalently many possible choices of the Ω_j sets) that the cue receiver can use to map the label pair (S_m, C_n) into a response. When the *a priori* occurrence probabilities of the phones $\{P(\mathcal{P}_i)\}$ are known, no rule leads to greater identification accuracy (overall probability of identifying the spoken phone correctly) than the "maximum *a posteriori* probability" rule: for each (S_m, C_n), the response should be the identity of the phone for which $P(\mathcal{P}_i) \times P_{SC}(S_m, C_n|\mathcal{P}_i)$ is greatest. Let Θ_j denote the set of (m, n) for which $P(\mathcal{P}_j) \times P_{SC}(S_m, C_n|\mathcal{P}_j) > P(\mathcal{P}_i) \times P_{SC}(S_m, C_n|\mathcal{P}_i)$ for all $i \neq j$.

To make predictions for the reception of cued segments, we assume that $P_S(S_m|\mathcal{P}_i) = P_S(R_m|\mathcal{P}_i)$; that is, the speechread labels correspond to the responses in an uncued identification experiment, and that the cue confusion probabilities can be determined by measuring the performance of the recognizer used to generate the cues. Thus, the conditional response probabilities for the optimal rule are specified by

$$P_{SC}(R_j|\mathcal{P}_i) = \sum_{m,n\epsilon\Theta j} P_S(R_m|\mathcal{P}_i) \times P_C(C_n|\mathcal{P}_i), \quad [3]$$

and the overall probability of correct identification is given by

$$\sum_j P(\mathcal{P}_j) \sum_{m,n\epsilon\Theta j} P_s(R_m|\mathcal{P}_j) \times P_c(C_n|\mathcal{P}_j). \quad [4]$$

Although this theory is sufficiently general to account for situations in which the various phones have differing *a priori* probabilities of occurrence, in

the remainder of this paper we consider only the case in which these probabilities are all equal; that is, $P(\mathcal{P}_j) = 1/J$, where J is the number of phones in a given situation. This assumption is appropriate when the probabilities are unknown and is consistent with conditions in most audiovisual segment recognition tests. This assumption is not appropriate when test materials have the phonetic balance of everyday speech, such as PB words or sentences. In general, the predictions underestimate the accuracy that can be achieved when the probabilities are unequal, but known.

Human Segment Identification Studies

To estimate the expected recognition performance when speechreaders are provided with imperfect cues, we predicted confusion matrices using the maximum likelihood rule based on published data obtained in visual, auditory, and audiovisual segment recognition experiments and our own evaluations of the performance of various automatic cueing systems. For consonant segments, we used data provided by Erber (28). For vowels, we used data provided by Hack and Erber (29), Wozniak and Jackson (30), and Montgomery and Jackson (31). In the latter case, we restricted attention to the monophthongs, using the "Constant Ratio Rule" (32) to derive the confusion matrix for monophthongs from the observed confusions for monophthongs and diphthongs.

Erber (28) tested three groups (PD—profoundly deaf, SI—severely hearing impaired, and NH—normal hearing) of five children each on identification of the eight consonants /b, d, g, p, t, k, m, n/ in /a/-C-/a/ context. Materials were videotaped. Each subject was presented each consonant 45 times in each of the auditory (A), visual (V), and audiovisual (AV) conditions.

Hack and Erber (29) tested 3 groups—who obtained "good" (Group 1), "intermediate" (Group 2), and "poor" (Group 3) word recognition scores, respectively—of 6 hearing impaired children each on the identification of the 10 monophthongs /i, I, eh, ae, a, aw, uu, u, uh, er/ in /b/-V-/b/ context. Materials were videotaped. Each subject was presented each vowel five times in each of the A, V, and AV conditions.

Montgomery and Jackson (31) tested 10 normal-hearing college students on visual reception of the 15 vowels /i, I, eh, ae, a, aw, uu, u, uh, er, eI, aI, awI, au, o/ in /h/-V-/g/ context. Each vowel was spoken once by each of four female talkers. Materials were videotaped and each viewer saw each utterance of each vowel three times.

Wozniak and Jackson (30) tested 10 normal-hearing graduate students on visual reception of the 16 vowels /i, I, eI, eh, ae, a, aw, o, uu, u, uh, er, aI, au, awI, ju/ under both 0° and 90° viewing conditions. Materials were videotaped, and each viewer saw each vowel 10 times in each condition.

Predictions of Segment Reception with Cues

Predicted intelligibility scores for consonant and vowel segments computed according to the theory of cue integration for segments using data on the existing and potential cueing systems described earlier and the human perceptual data of Erber (28), Hack and Erber (29), Wozniak and Jackson (30), and Montgomery and Jackson (31) are presented in **Tables 5–11**. Observed scores for V, AV, and A reception of segments are included for reference purposes. Predicted scores for V, AV, and A reception aided by various hypothesized cueing systems are tabulated. The cueing systems include "PHONE," in which each segment has a unique cue, "MC," the cue groups of Manual Cued Speech, "AC," the cue groups of the Autocuer, and the "CBG" and "kineme" consonant groups described earlier. Predicted scores are computed for perfectly cued items (IDEAL) and cues derived by several recognition processes: SRVZ, expert spectrogram reader VZ; SRHT, three highly trained spectrogram readers; ARHS, the MIT automatic recognition system trained on Harvard sentences; ARTI, the MIT automatic recognition system trained on TIMIT sentences; and ARGV, the automatic kineme recognition system developed by Vilaclara (21). Not all combinations of cue groups and recognizers are amenable to predictions as the result of constraints on experimental design. (We also computed predictions for larger sets of consonants and vowels based on the 20-consonant matrices of Busacco (33) and the complete 15- and 16-vowel matrices of Wozniak and Jackson (30) and Montgomery and Jackson (31). These predictions are very similar to those reported in the tables.)

Certain characteristics of the predictions in **Table 5** have general properties. The IDEAL recognition system yields the highest score for any fixed

Table 5.
Visual reception of 8 consonants by hearing impaired children.

Group	Observed V	Cue	IDEAL	SRVZ	SRHT	ARHS	ARTI	ARGV
			Predicted Scores with Cues					
NH	31.6	PHONE	100			76.9	72.3	
NH	31.6	MC	92.4			66.6	59.4	
NH	31.6	AC	92.7			75.2	71.0	
NH	31.6	CBG	80.4	78.1	72.0	69.0	68.9	71.2
SI	49.2	PHONE	100			81.3	79.6	
SI	49.2	MC	96.4			75.2	68.0	
SI	49.2	AC	97.0			80.2	78.6	
SI	49.2	CBG	91.5	88.2	80.8	78.4	78.5	82.8
PD	45.2	PHONE	100			83.1	81.2	
PD	45.2	MC	98.8			77.2	69.4	
PD	45.2	AC	98.7			81.9	80.1	
PD	45.2	CBG	90.9	87.7	79.8	78.2	78.4	79.6

Predicted scores for 8 consonants by hearing impaired children using speechreading aided by speech cues. For this task the CBG and kineme consonant cue groupings are identical. Human visual data taken from Erber (28). See sections on The Gallaudet-R.T.I. Automatic Cueing System, Potential Automatic Cueing Systems, and Human Segment Identification Studies for details.

Table 6.
Audiovisual reception of 8 consonants by hearing impaired children.

Group	Observed AV	Cue	IDEAL	SRVZ	SRHT	ARHS	ARTI	ARGV
			Predicted Scores with Cues					
SI	88.0	PHONE	100			90.9	88.7	
SI	88.0	MC	99.4			90.3	88.5	
SI	88.0	AC	98.1			90.3	88.4	
SI	88.0	CBG	91.7	90.5	89.3	88.3	88.2	91.6
PD	60.5	PHONE	100			84.3	82.1	
PD	60.5	MC	99.2			79.7	73.5	
PD	60.5	AC	99.0			83.3	80.8	
PD	60.5	CBG	92.8	89.5	81.6	79.7	80.4	86.6

Predicted scores for 8 consonants by hearing impaired children using hearing and speechreading aided by speech cues. For this task the CBG and kineme consonant cue groupings are identical. Human audiovisual data taken from Erber (28). Data for the NH group, whose performance is essentially perfect without cues, are omitted. See sections on The Gallaudet-R.T.I. Automatic Cueing System, Potential Automatic Cueing Systems, and Human Segment Identification Studies for details.

set of cue groups because it provides the highest recognition rate for a given set of cues. It always yields a perfect score for the PHONE cue groups because segments can be identified without error using only the PHONE cues if they are recognized perfectly. Similarly, independent of the recognizer used, the highest predicted scores are always achieved with the use of PHONE cues. The use of a smaller number of cue groups than the number of phonetic elements (as in the AC, MC, and CBG cue

Table 7.
Auditory reception of 8 consonants by hearing impaired children.

Group	Observed A	Cue	IDEAL	SRVZ	SRHT	ARHS	ARTI	ARGV
			Predicted Scores with Cues					
SI	50.4	PHONE	100			67.4	65.6	
SI	50.4	MC	93.5			64.8	61.3	
SI	50.4	AC	75.4			62.6	58.7	
SI	50.4	CBG	57.2	57.0	55.2	52.3	52.2	64.6
PD	21.1	PHONE	100			63.7	61.4	
PD	21.1	MC	77.1			52.1	47.4	
PD	21.1	AC	67.2			53.2	49.6	
PD	21.1	CBG	44.8	45.4	42.8	39.1	40.0	47.6

Predicted scores for 8 consonants by hearing impaired children using audition aided by speech cues. For this task the CBG and kineme consonant cue groupings are identical. Human auditory data taken from Erber (28). Data for the NH group, whose performance is essentially perfect without cues, are omitted. See sections on The Gallaudet-R.T.I. Automatic Cueing System, Potential Automatic Cueing Systems, and Human Segment Identification Studies for details.

Table 8.
Visual reception of 10 monophthongs by hearing-impaired children.

Group	Observed V	Cue	IDEAL	SRVZ	SRHT	ARHS	ARTI
			Predicted Scores with Cues				
Group 1	52.3	PHONE	100	89.3	80.7	75.4	76.7
Group 1	52.3	MC	91.0	77.8	70.8	65.9	67.9
Group 1	52.3	AC	95.0	81.1	68.4	66.4	67.9
Group 2	55.7	PHONE	100	90.1	82.4	77.2	79.3
Group 2	55.7	MC	91.3	80.4	74.0	69.7	73.1
Group 2	55.7	AC	94.7	83.0	70.0	69.6	72.4
Group 3	48.0	PHONE	100	89.1	81.0	75.4	77.6
Group 3	48.0	MC	87.0	75.4	69.9	65.6	67.6
Group 3	48.0	AC	92.7	80.1	67.5	65.4	67.7

Predicted scores for 10 monophthongs by hearing impaired children using speachreading aided by speech cues. Human visual data taken from Hack and Erber (29). See sections on Potential Automatic Cueing Systems, and Human Segment Identification Studies for details.

groups) is intended to approximate the performance of the PHONE groups while reducing both display and training problems. Several predictions also generalize to all (or nearly all) of the remaining cases considered. Predicted scores for the aided AV presentation condition exceed those for the aided A and V conditions, although for the more severely impaired groups the difference is usually very small. Predicted scores for the MC, AC, and CBG cue groups typically exceed 90% for the IDEAL recognizer when used in conjunction with V or AV speechreading. This indicates that these three cue groups were well chosen as supplements to speechreading, given that the cues could be produced accurately. Predicted scores using the ARHS and ARTI recognizers are fairly similar, as one

Journal of Rehabilitation Research and Development Vol. 31 No. 1 1994

Table 9.
Audiovisual reception of 10 monophthongs by hearing-impaired children.

Group	Observed AV	Cue	IDEAL	Predicted Scores with Cues SRVZ	SRHT	ARHS	ARTI
Group 1	66.3	PHONE	100	92.9	87.6	81.5	83.9
Group 1	66.3	MC	91.7	81.7	79.1	74.5	76.4
Group 1	66.3	AC	97.3	87.0	78.5	75.3	77.3
Group 2	67.0	PHONE	100	92.3	87.2	82.8	85.0
Group 2	67.0	MC	94.0	84.4	80.0	77.5	78.9
Group 2	67.0	AC	96.7	87.5	76.0	77.0	78.6
Group 3	51.7	PHONE	100	89.6	83.2	77.2	79.2
Group 3	51.7	MC	91.3	78.7	72.6	68.5	71.0
Group 3	51.7	AC	96.3	82.9	69.9	68.3	70.4

Predicted scores of 10 monophthongs by hearing impaired children using hearing and speechreading aided by speech cues. Human audiovisual data taken from Hack and Erber (29). See sections on Potential Automatic Cueing Systems, and Human Segment Identification Studies for details.

Table 10.
Auditory reception of 10 monophthongs by hearing-impaired children.

Group	Observed A	Cue	IDEAL	Predicted Scores with Cues SRVZ	SRHT	ARHS	ARTI
Group 1	56.0	PHONE	100	88.0	77.8	76.2	74.7
Group 1	56.0	MC	87.0	75.9	69.4	67.2	64.9
Group 1	56.0	AC	92.0	80.1	70.1	68.2	67.1
Group 2	30.3	PHONE	100	82.4	70.1	66.2	66.2
Group 2	30.3	MC	70.3	62.0	55.7	51.6	52.0
Group 2	30.3	AC	74.7	64.4	56.8	54.0	53.0
Group 3	11.7	PHONE	100	78.5	66.4	62.6	63.1
Group 3	11.7	MC	60.0	51.2	44.8	42.0	42.6
Group 3	11.7	AC	60.3	51.4	44.3	43.1	41.9

Predicted scores for 10 monophthongs by hearing impaired children using audition aided by speech cues. Human auditory data taken from Hack and Erber (29). See sections on Potential Automatic Cueing Systems, and Human Segment Identification Studies for details.

might expect since the underlying recognition system was the same. Scores for SRVZ generally exceed those for SRHT and the automatic recognition systems; those for ARGV generally exceed those for ARHS and ARTI. For the MC and AC cue systems, the predicted performance in the aided V and AV conditions yields greater improvement over the unaided V and AV conditions than does the predicted aided A over unaided A. Again, this indicates that the MC and AC cue groups are well-designed as supplements to speechreading and less well-designed as aids to audition.

In general, although predicted performance for segment reception with the three automatic recog-

Table 11.
Visual reception of 10 monophthongs by normal hearing adults.

Group	Observed V	Cue	IDEAL	Predicted Scores with Cues SRVZ	SRHT	ARHS	ARTI
M. & J.	56.0	PHONE	100	88.8	81.6	76.2	78.3
M. & J.	56.0	MC	87.8	76.4	72.5	66.6	68.7
M. & J.	56.0	AC	92.0	79.8	69.5	66.6	68.5
W. & J. 1	79.6	PHONE	100	92.8	91.5	88.4	88.2
W. & J. 1	79.6	MC	97.5	89.1	88.3	85.1	86.1
W. & J. 1	79.6	AC	98.7	90.1	89.1	85.3	85.9
W. & J. 2	78.0	PHONE	100	93.0	91.0	87.9	88.3
W. & J. 2	78.0	MC	98.4	89.8	87.6	84.9	85.6
W. & J. 2	78.0	AC	99.1	90.0	88.4	84.9	85.9

Predicted scores for 10 monophthongs by normal-hearing adults using vision aided by speech cues. Human visual data taken from Montgomery and Jackson (31) and Wozniak and Jackson (30). See sections on Potential Automatic Cueing Systems, and Human Segment Identification Studies for details.

nizers (ARHS, ARTI, and ARGV) is not perfect, it appears substantially superior to that for the Autocuer. For consonants (see **Tables 5–7**), it is encouraging that in nearly all cases aided scores are usually predicted to exceed 75% correct in the V condition and 80% for the AV condition. Relative to the V condition, the use of cues is typically predicted to lead to improvements of 25 percent-age points or more. It is interesting to note that scores for the cued A condition are often predicted to exceed those for the uncued A condition by substantial amounts. This is consistent with the findings of Nicholls and Ling (17), who reported that their hearing-impaired children achieved scores for cued auditory reception of words in sentences (LP: 59.2% and HP: 68.8%) that exceeded scores for audition alone, speechreading alone, and cues alone. As in the audiovisual case, however, the improvements we predict for cued auditory reception are substantial only for the more severely impaired groups.

Most of these trends are also seen for vowels (see **Tables 8–11**), although both the aided scores and the improvements associated with the use of cues are typically predicted to be somewhat lower than for consonants. When very good reception occurs in the uncued V condition (e.g., 30), relatively little benefit to vowel identification is to be expected from the kinds of cues that can be provided by existing automatic recognizers.

Comparison of the predicted cued consonant scores in **Table 5** and **Table 6** indicates that cued speechreading is likely to be inferior to unaided audiovisual reception for the SI group, but superior for the PD group. As noted previously, cued and uncued audiovisual reception for the SI group are predicted to be roughly equivalent, whereas for the PD group cueing improves AV scores by roughly 20 percentage points. **Table 7** indicates that current recognizers are unlikely to be effective aids to impaired auditory reception of consonants. Only the PD group is expected to achieve improvements in excess of 20 percentage points. This may help explain why Kanevsky, et al. (3) typically found only minor gains when using recognized words to aid the reception of noisy speech.

Predicted cued vowel scores in **Table 8** depend very little on listener group (presumably because of similarities in speechreading abilities), greatly on the recognizer, and secondarily on the cue groups. Evidently there is an interaction between the recognizer and the groups. Scores for SRVZ are predicted to be better with the AC groups than for the MC groups, while for SRHT the reverse is true. For cued audiovisual reception of vowels, **Table 9** indicates that all three subject groups are expected to benefit substantially from cues that can be provided by

Journal of Rehabilitation Research and Development Vol. 31 No. 1 1994

existing recognizers. **Table 11** indicates that when high speechreading scores are achieved, the benefit expected from cues produced by current recognizers is small, but it is probably not negligible.

DISCUSSION

According to our experimental results, and those of Nicholls and Ling (17), speechreaders who make effective use of Manual Cued Speech are able to perceive contextual sentences with high accuracy, although when cued the sentences may be spoken somewhat more slowly and more clearly than conversational speech. These results provide additional evidence that a discrete symbolic speechreading supplement, such as that developed by Cornett (13), can provide the deaf with an effective means of face-to-face communication if the cues are produced accurately. Our theoretical analysis of the performance of spectrogram readers and automatic speech recognizers as producers of cues provides estimates of the segment identification accuracy speechreaders might achieve with present or future speech recognition systems. Accuracies of roughly 75–90% seem plausible, depending on the skill of the speechreader, the complexity of the task, and the specific recognizer and cueing system. It remains to be determined whether this level of segmental reception can support useful levels of everyday communication.

Reception of Segments, Words, and Sentences

To provide useful assistance to the hearing impaired, an autocuer must aid the reception of sentences and continuous discourse. In order to determine whether automatically generated cues can support useful levels of communication, we review some recent studies that measured the reception of both connected speech and isolated segments. In addition, we employ a mathematical model of the relationships between sentence, word, and phoneme intelligibility to estimate the level of segment reception required to achieve the performance of the experienced Cued Speech receivers we tested.

Brooks, et al. (34,35) studied the ability of one highly trained normal-hearing (but artificially deafened with masking noise) subject to recognize words, sentences, and continuous narrative using the Queens Tactile Vocoder (36) in conjunction with speechreading. When tested on 3,000 isolated common English words, the subject identified 68.7% of the words correctly. A phonemic analysis of the stimulus and response words (14 vowels, 25 consonants) indicated that the vowels were recognized with 86.9% accuracy, the consonants with 87.3% accuracy. When tested on common sentences (4–15 syllables, 5.9 words average length), the same subject identified roughly 50% of the sentences perfectly after a single presentation of the sentence. When unaided, only roughly 30% of the sentences were identified correctly. Aided, and using a "tracking procedure" (e.g., 37) on reception of dialogue typical of conversations, the subject achieved a rate of roughly 49 wpm, 61% of the rate achieved with no simulated deafness. In the unaided condition, the tracking rate was 15 wpm, only 19% of the rate achieved with no simulated deafness.

Weisenberger, et al. (38) tested three artificially deafened normal-hearing subjects on segment recognition and discourse tracking using the Queens Tactile Vocoder (36) together with speechreading (S + Q) and the Tacticon electrotactile vocoder (39) together with speechreading (S + T). Performance on the recognition of 21 consonants in C-/o/ context was roughly 72% correct in the S + Q condition and 76% correct in the S + T condition. Tracking rates, however, differed for the two aids, with the three subjects achieving 50, 64, and 70 wpm in the S + Q condition and 22, 32, and 35 wpm in the S + T condition. When unaided these subjects achieved tracking rates of 18, 18, and 25 wpm.

Reed, et al. (40,41) reported that a highly successful user of the Tadoma method achieves roughly 55% consonant recognition accuracy (24 consonants in CV context) and 57% vowel recognition accuracy (15 vowels in CVC context). The same individual recognizes roughly 50% of isolated W-22 words, and 86% of key words in the CID sentences when spoken at the approximately normal speaking rate of 4.4 syl/sec. On the more difficult Harvard sentences, this subject recognizes roughly 60% of the words at a speaking rate of 2.5 syl/sec.

The trained cue receivers we studied were able to identify key words in cued Harvard sentences with an accuracy of 84%. Although some of the errors made by the cue receivers may reflect inattention or unfamiliarity with the vocabulary used in the

sentences, the vast majority of the errors most likely result from imperfect reception of linguistic elements. On this basis, it seems likely that at least 10% of the phonetic segments are not received correctly. However, the level of phonetic accuracy needed to account for the observed levels of sentence reception is probably lower than this because of contextual redundancy. In general, the relationship between reception of segments, words, and sentences is not a simple one. Boothroyd and Nittrouer (42) provide a mathematical model of the relation between phoneme scores and scores for words and whole sentences (see **Figure 1**). While the materials used in our evaluation of Manual Cued Speech differ from those used in the Boothroyd and Nittrouer study, straightforward application of the theory to our data suggests that phoneme scores of roughly 70–80% correct may be adequate to account for the performance of trained cue receivers.[3] Although precise comparisons are difficult, such levels of phoneme reception are roughly comparable to the performance achieved in the Weisenberger, et al. study and probably superior to that obtained in the Brooks, et al. and Reed, et al. studies. All of these studies suggest that useful levels and rates of everyday communication can be achieved if phoneme reception accuracy in excess of 80% can be achieved.

It is important to note that our estimates of phoneme reception accuracy for the trained receivers of Cued Speech are derived from their performance on sentence materials rather than for isolated segments. Although we did not test these receivers on cued reception of segments, based on our experience with less well-trained receivers we suspect they would have achieved scores well above 90% correct on both consonants and vowels. Nevertheless, we believe that our estimate that only 80–90% of phonemes are received correctly by trained cue receivers in sentence context is reasonable. Preliminary analysis of the error responses made by the receivers on the cued Harvard sentences suggests that more than one-quarter of the word errors can be attributed to incorrect reception of cues for segments in words. Since our manually cued test materials were verified when the sentences were spoken,[4] our error analysis implies that the cue receivers either do not always perceive the cues accurately or do not always integrate the cues perfectly with speechreading. For our purposes, however, it is immaterial whether the cueing errors arise from correct perception of inaccurate cues or incorrect perception of accurate cues.

The main stumbling blocks faced by the designers of the Autocuer appear to result from the failure to obtain adequate performance from the recognizer used in the system. Our analysis of the recognition requirements for useful automatic cueing systems suggests that this hurdle is not a fundamental one.

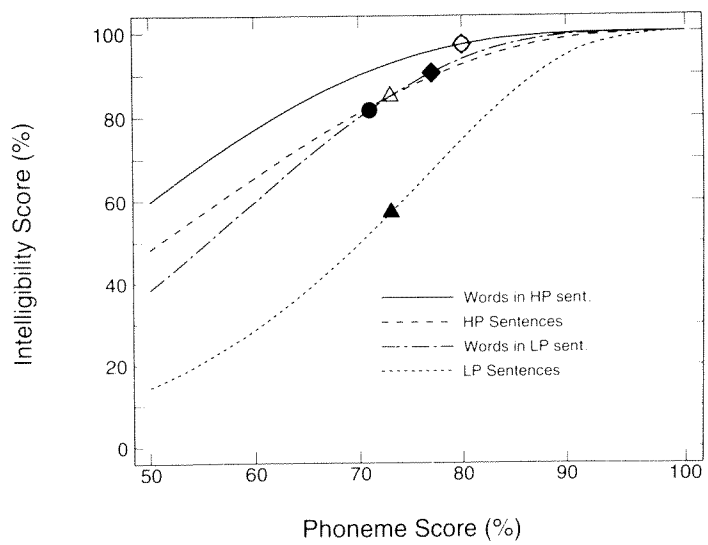

Figure 1.
Dependence of intelligibility score on phoneme intelligibility according to the analysis of Boothroyd and Nittrouer (42). The solid curve gives the word score for high predictability sentences, the dot-dash curve gives the high predictability sentence score (all words correct), the dashed curve gives the word score for low predictability sentences, and the dotted curve gives the low predictability sentence score (all words correct). To determine corresponding phoneme scores, the average Cued Speech scores for words in high-context CID (open diamond), and Clarke sentences (open circle), CID (filled diamond), and Clarke sentence scores (filled circle), the average score for words in low-context Harvard sentences (open triangle), and average Harvard sentence score (filled triangle) have been superimposed on these curves.

[3]Similar conclusions can be reached using an analysis developed recently by Rabinowitz, et al. (43).

[4]The cueing of words that were perceived incorrectly by more than one cue receiver was also checked, but no errors in cue production were found.

Journal of Rehabilitation Research and Development Vol. 31 No. 1 1994

Current recognition systems, or ones that seem plausible in the foreseeable future, should provide an adequate basis for developing effective autocuers.[5] This is not to say that the problems of developing a successful autocuer can be solved simply by waiting patiently for advances in Automatic Speech Recognition (ASR) technology. The cues derived by the recognizers must be presented so that they can be perceived accurately and integrated well with speechreading. Moreover, proficient use of Cued Speech is known to require substantial periods of training. In this respect the reception of Cued Speech is not fundamentally different from the reception of speech by other sensory substitution techniques. However, since the cues provided by automatic cueing systems are expected to be at least partially unreliable for the foreseeable future, effective strategies for dealing with such cues need to be developed. In this respect, learning to use automatic cued speech may be more difficult than learning a foreign language in that the difficulty of learning from an erratic teacher also needs to be considered.

The design of automatic cueing systems to aid the hearing impaired involves many considerations, only a few of which center on the recognizer or the cue groups. Some lessons learned from the design of Manual Cued Speech and the existing Autocuer seem pertinent. For example, the decision to use a relatively small number of cue groups and to encode consonants and vowels using perceptually distinct aspects of the display (hand shape and position) is consistent with current understanding of the characteristics of effective displays. On the other hand, the decision to restrict the Autocuer display to simple geometric shapes seems difficult to justify, particularly in view of recent advances in the technology of display graphics.

In the following sections, we suggest how the computational techniques discussed earlier can contribute to the process of designing improved autocuer systems, and consider strategies for aiding speechreading with visual displays that do not involve explicit discrete recognition of speech.

[5]If one restricts attention to situations in which speaker-dependent recognizers can be used to derive the cues, such as classroom and home environments, one's optimism about the potential of current recognition systems increases. For the same task, speaker-dependent recognizers typically achieve one-third the error rate of speaker-independent recognizers, such as those considered (44).

Improved Design of Cueing Codes and Recognizers

The cue groups used in existing cueing systems appear to have been designed primarily from the point of view of resolving speechreading ambiguities without taking into account the difficulty of assigning the specified cues automatically. The theory of cue integration that we discussed earlier provides a rational basis for the design of new cueing codes as well as the analysis of existing codes.

Consider a speech recognition system that produces an output symbol, from the set Q_k ($k=1, \ldots, K$) for each input phone. When this recognition system is incorporated into an autocuer, each recognizer output is mapped deterministically into a cue symbol belonging to the set C_n where $n=1, \ldots, N$ (we assume $N \leq K$). The cue code function ϕ specifies the mapping $C=\phi(Q)$. There are $N! \times N^K$ possible mapping functions, and an essential part of the design of an autocuer system is the specification of the optimum mapping function (i.e., the function that leads to the smallest probability of a phone recognition error when the autocuer is used in combination with speechreading).

For a fixed value of N, the optimum mapping function can be determined by estimating the probability that each cue will be assigned to each phone:

$$P_c\,(C_n \mid \mathcal{P}_i) = \sum_{k \epsilon \Phi n} P_Q\,(Q_k \mid \mathcal{P}_i). \qquad [5]$$

where Φ_n is the set of k for which $\phi(Q_k)=C_n$. Given the probabilities $P\ (C_n \mid \mathcal{P}_i)$, the computational techniques described in our theory of cue integration for segments can be used to predict the probability of error when identifying the phones using the autocuer code in conjunction with speechreading.

An example of the design of a cue system for vowels follows. Speechreading confusions are assumed to be those made when normal-hearing subjects identify monophthongs, as measured by Montgomery and Jackson (31). Two speech recognizers are considered: the IDEAL phone recognizer and a recognizer achieving the performance of SRHT. For each recognizer, each cue alphabet size, and each cue mapping function, a cue confusion matrix was computed according to Equation [5] and a cued-visual confusion matrix was then determined according to Equation [4]. The mapping function leading to the highest identification score was determined for each recognizer and each cue alphabet size. The highest identification scores are

plotted as a function of cue alphabet size in **Figure 2.**[6] The corresponding codes for the IDEAL and SRHT recognizers are shown in **Table 12.** Several mapping functions lead to the same score for the IDEAL recognizer in the case of $N=2$ cues; for $N>2$, optimal groupings are unique. However, in all cases several alternate groupings were predicted to produce close to optimal scores. Note that for all three recognizers considered, as few as 3–4 cues are sufficient to achieve near-maximum scores. The results of a similar analysis for consonants are shown in **Figure 3** and **Table 13.**

In addition to suggesting cue groups that make effective use of a given recognizer for a specified pattern of speechreading confusions, our analysis can also be used to focus efforts on improving automatic speech recognizers so that they contribute maximally to improving automatic cueing systems. Suppose we could alter the error patterns in a recognizer. Our analysis can be used to compute the resulting aided segment identification scores. For the

[6]The highest identification scores predicted when the SRVZ recognizer assigns labels are also included.

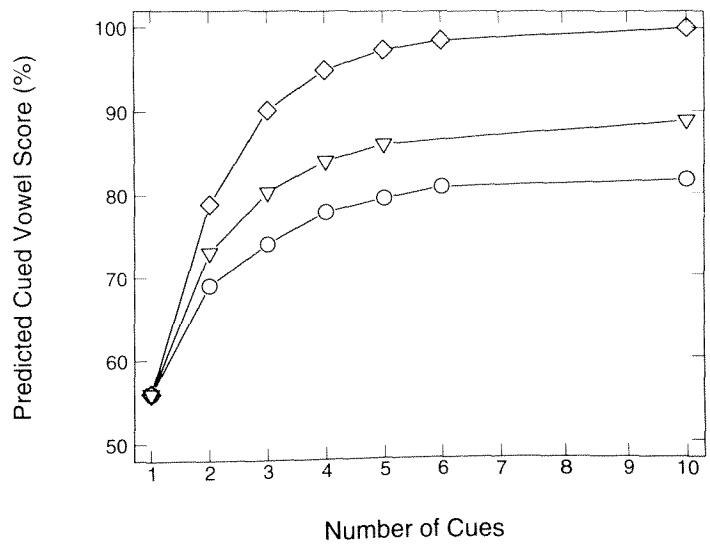

Figure 2.
Dependence of predicted cued-visual vowel identification score on number of cue groups for the 10 monophthong vowels. Human speechreading confusions were adapted from Montgomery and Jackson (31) using the Constant Ratio Rule. Diamonds are predictions for an IDEAL recognizer, triangles for a recognizer corresponding to the expert spectrogram reader (SRVZ), and circles to the three highly trained spectrogram readers (SRHT).

Table 12.
Optimum cue groups.

Cue groups for 10 monophthongs and the IDEAL recognizer		
Number of Cues	**% Correct**	**Optimum Cue Groups**
2	78.9	/I,ae,u,uh,er/ /i,eh,a,aw,uu/
3	90.2	/I,a,er/ /eh,u,uh/ /i,ae,aw,uu/
4	95.0	/ae,er/ /eh,uh/ /I,a,u/ /i,aw,uu/
5	97.4	/ae,er/ /uh/ /I,a,u/ /i,aw,uu/ /eh/
6	98.5	/ae,er/ /uh/ a,u/ /i,uu/ /I,aw/ /eh/
Cue groups for 10 monophthongs and the SRHT recognizer		
Number of Cues	**% Correct**	**Optimum Cue Groups**
2	69.1	/i,ae,a,aw,er/ /I,eh,uu,u,uh/
3	74.1	/i,a,er/ /ae,aw,uu,uh/ /I,eh,u/
4	77.9	/ae,er/ /I,uu,uh/ /eh,u/ /i,a,aw/
5	79.6	/ae,er/ /I,uu,uh/ /eh,u/ /a,aw/ /i/
6	80.9	/er/ /I,uu,uh/ /eh,u/ /a,aw/ /ae/ /i/

Optimum cue groups for aiding speechreading of 10 monophthongs by the normal-hearing adults tested by Montgomery and Jackson (31). In the upper half of the table the IDEAL recognizer is assumed; in the lower half the SRHT recognizer. See section on Improved Design of Cueing Codes and Recognizers.

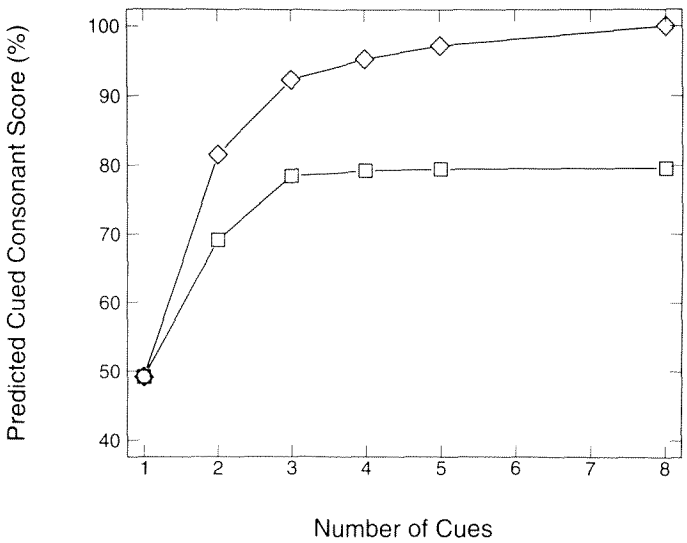

Figure 3.
Dependence of cued-visual predicted consonant identification score on number of cue groups for eight consonants. Human speechreading confusions are those of the severely impaired group of children tested by Erber (28). Diamonds are predictions for an IDEAL recognizer; squares for the ZRHS recognizer.

eight consonants used in the Erber (28) study, we computed the effects of improving the ARHS recognizer (producing PHONE labels) by artificially reducing the error rates of voicing from 15.4% to 7.7% (VO/2) and 3.9% (VO/4), of place from 29.4% to 14.7% (PL/2), and of reducing both voicing and place errors (VO/2+PL/2). As might be expected, the predictions (**Table 14**) indicate that improvements to cued visual reception result primarily from reduction of recognizer voicing errors, while improvements to cued auditory reception result primarily from reduction of recognizer place errors. However, none of the recognizer improvements considered is predicted to produce a dramatic improvement in speechreading reception, although we suspect it would be difficult to enhance the recognizers by the amounts we have considered.

Nonsymbolic Visual Speechreading Aids

In contrast to speechreading supplements based on phonetic recognition and the display of discrete symbols, visual displays that supplement speechreading without identifying discrete linguistic units have also been studied (e.g., 45,46). Recently, Ebrahimi and Kunov (47) developed and evaluated a "peripheral vision lipreading aid" that uses a 5×7 LED matrix display to present three parameters of

Table 13.
Optimum cue groups.

Cue groups for 8 consonants and the IDEAL recognizer		
Number of Cues	**% Correct**	**Optimum Cue Groups**
2	81.6	/p,d,g,m,n/ /t,k,b/
3	92.3	/b,g,n/ /d,m/ /p,t,k/
4	95.3	/k,n/ /d,m/ /b,g/ /p,t/
5	97.2	/n/ /k,m/ /b,g/ /d/ /p,t/
Cue groups for 8 consonants and the ARTI recognizer		
Number of Cues	**% Correct**	**Optimum Cue Groups**
2	69.2	/p,t,k,m,n/ /b,d,g/
3	78.5	/m,n/ /b,d,g/ /p,t,k/
4	79.3	/m,n/ /p,g/ /b,d/ /t,k/
5	79.5	/m,n/ /g/ /b,d/ /t,k/ /p/

Optimum cue groups for aiding speechreading of 8 consonants by the severely impaired children tested by Erber (28). In the upper half of the table the IDEAL recognizer is assumed; in the lower half the ARTI recognizer. See section on Improved Design of Cueing Codes and Recognizers.

Table 14.
Improving cued reception of 8 consonants.

Group	Mode	Original	VO/2	VO/4	PL/2	PL/2+VO/2
NH	V	76.9	80.6	82.9	78.6	83.9
SI	V	81.3	86.2	89.1	81.7	87.2
PD	V	83.1	87.6	90.4	83.7	89.4
SI	AV	90.9	90.9	91.6	92.5	92.7
PD	AV	84.3	88.9	91.7	84.3	90.0
SI	A	67.4	68.5	70.0	78.3	80.5
PD	A	63.7	67.2	69.0	74.4	79.7

Predicted effect of improving the ARHS recognizer on predicted cued reception using the PHONE cues. Improvements include reducing voicing errors by factors of two (VO/2) and four (VO/4), reducing place errors by a factor of two (PL/2), and reducing both voicing and place errors by factors of two (PL/2+VO/2). Human perceptual data taken from Erber (28).

the acoustic waveform: voice fundamental frequency, overall energy, and energy above 3 kHz. The aid attempts to present five distinct spatio-temporal patterns to distinguish among different consonants based on voicing and manner of articulation. During low-energy intervals no LEDs are illuminated; when periodic low-frequency components are detected, the lower two (of seven) rows are illuminated; when high-frequency components are detected, all seven rows are illuminated. The number of columns illuminated (out of five) is determined by high-frequency levels. For example, voiced plosives might be signaled by a small brief spot in the lower left corner of the display, unvoiced plosives by a momentary bright bar in the left columns, unvoiced fricatives by sustained illumination of most of the display, etc. Eight young adults (3 normal-hearing and 5 profoundly deaf) evaluated this device in a 12-consonant (VCV context) identification task. Scores improved from 41% by speechreading alone to 76% for aided speechreading. In particular, distinctions between voiced and unvoiced consonants that were made with only 55% accuracy by speechreading alone were made with greater than 88% accuracy when the aid was used. Distinctions between the five place groups also improved slightly, from 84% by speechreading to 88% with aided speechreading. Based on our analysis of the reception of Manual Cued Speech, it seems likely that if this system were equally effective for the reception of vowels, it would be a very useful aid to speechreading.

The performance achieved with the Ebrahimi and Kunov system is roughly the same as predicted using cues derived from current recognizers. For example, consonant reception predictions for Erber's SI and PD groups using the MC cues generated by the ARHS recognizer are 75.2% and 77.2%, respectively. Very similar predictions would be made if a set of 20 consonants (e.g., 33) had been considered. Given the relatively poor performance of the recognizers when generating cues for consonants, these results are not surprising. The potential advantages of discrete symbolic displays are largely vitiated when only three cue groups can be distinguished effectively. As can be seen in **Table 5**, for a cueing system to provide substantially greater consonant recognition scores, a recognizer with the performance of SRVZ would be required. In our view, the high scores achieved with the Ebrahimi and Kunov system do not argue that a relatively unprocessed continuous display, or one based on transmission of specific speech features, is superior to one based on discrete symbols. Rather, they underscore the need for more accurate speech recognition if the potential of a discrete symbol display is to be realized. Until such recognizers are available, the performance of autocuers will only be comparable to that of vastly simpler systems such as that studied by Ebrahimi and Kunov.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grants R01-DC00117 and P01-DC00361). We are grateful to Joan Besing, Robin Kittleman, Susan Russell, and Timothy Stellmach for assistance in preparing audiovisual test materials; to A.H., D.K., L.L., and S.S. for serving as receivers of Cued Speech; to Nancy Daly, Caroline Huang, and John Pitrelli for serving as spectrogram readers; to Mike Phillips and Victor Zue for providing data on the performance of the M.I.T. automatic speech recognition system; to Robert Beadles for providing unpublished data on the Autocuer; to D.A. Busacco for providing unpublished audiovisual segment recognition data; to A.W.F. Huggins and Y. Russell for providing information on the Vidvox Project; and to G. Vilaclara for providing unpublished data on his recognition system. We are also grateful to F. Destombes, P. Duchnowski, S. Goldman, D. Ling, W. Rabinowitz, and R. Rohlicek for many insights obtained through thoughtful discussion.

REFERENCES

1. Averbuch A, Bahl LR, Bakis P et al. Experiments with the TANGORA 20,000 word speech recognizer. In: Proceedings of ICASSP-87:1987:701-4.
2. Lee K-F, Hon HW, Huang MY, Huang X. Speech recognition using hidden Markov models: a CMU perspective. Speech Commun 1990:9:497-508.
3. Kanevsky D, Danis CM, Daggett G et al. A communication aid for the hearing impaired based on an automatic speech recognizer. In: Tores L, Masgrau E, Lagunas MA, eds. Signal processing V. New York: Elsevier, 1990.
4. Bahl LR, DeGennaro SV, Gopalakrishnan PS, Mercer RL. A fast approximate acoustic match for large vocabulary speech recognition. In: Proceedings of EUROSPEECH '89: 1989 Sept; Paris.
5. Russell Y. Vidvox final report. Palo Alto, CA: Sensory Aids Foundation, 1986.
6. Krasner MA, Schwartz RM, Kimball OA, Price P. Development of a Vidvox feasibility system. Cambridge, MA: BBN Laboratories, 1985 Bolt Beranek and Newman Technical Report 5906.
7. Huggins AWS, Houde RA, Colwell E. Vidvox human factors investigation. Cambridge, MA: BBN Laboratories, 1986 Bolt Beranek and Newman Technical Report 6187.
8. Institute of Electrical and Electronic Engineers. IEEE recommended practices for speech quality measurements. New York: IEEE, 1969.
9. Maclay H, Osgood CE. Hesitation phenomena in spontaneous English speech. In: Jakobovitz LA, Miron MS, eds. Readings in the psychology of language. Englewood Cliffs, NJ: Prentice Hall, 1967.
10. Douglas-Cowie E. Acquired deafness and communication. In: Cowie R, ed. Coping with acquired hearing loss. Belfast: Social Services Inspectorate DHSS, 1988:13-22.
11. Jeffers J, Barley M. Speechreading. Springfield, IL: Charles C. Thomas, 1971.
12. Jackson PL. The theoretical minimal unit for visual speech perception: visemes and coarticulation. Volta Rev 1988:90(5):99-115.
13. Cornett RO. Cued speech. Am Ann Deaf 1967:112:3-13.
14. Kaplan H. The effects of cued speech on the speechreading ability of the deaf (Dissertation). College Park, MD: University of Maryland, 1974.
15. Ling D, Clarke B. Cued speech: an evaluative study. Am Ann Deaf 1975:120:480-8.
16. Clarke B, Ling D. The effects of using cued speech: a follow up study. Volta Rev 1976:78:23-34.
17. Nicholls G, Ling D. Cued speech and the reception of spoken language. J Speech Hear Res 1982:25:262-9.
18. Davis H, Silverman SR. Hearing and deafness. New York: Holt, Rinehart and Winston, 1970.
19. Magner ME. A speech intelligibility test for deaf children. Northhampton, MA: Clarke School for the Deaf, 1972.
20. Cornett RO, Beadles R, Wilson B. Automatic cued speech. In: Proceedings of the Research Conference on Speech-Processing Aids for the Deaf. Washington, DC: Gallaudet College, 1977:224-39.
21. Vilaclara G. Speech processing to aid the profoundly deaf. J Acoust Soc Am 1988:84:S41.
22. Zue V, Glass J, Phillips M, Seneff S. Acoustic segmentation and the phonetic classification in the Summit system. In: Proceedings of ICASSP-89, 1989:534-7.
23. Zue VW, Seneff S. Transcription and Alignment of the TIMIT Database. Paper presented at the Second Symposium on Advanced Man-Machine Interface through Spoken Language, Oahu HI, Nov. 1988.
24. Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing I: intelligibility differences between clear and conversational speech. J Speech Hear Res 1985:28:96-103.
25. Cole RA, Rudnicky AI, Zue VW, Reddy DR. Speech as patterns on paper. In: Cole RA, ed. Perception and production of fluent speech. Hillsdale, NJ: Lawrence Erlbaum Assoc., 1980:3-50.
26. Lee K-F, Hon H-W. Speaker-independent phone recognition using Hidden Markov Models. IEEE Trans Acoust Speech Sig Process 1989:11:1641-8.
27. Braida LD. Crossmodal integration in the identification of consonant segments. Q J Exp Psychol 1991:43A(3):647-77.
28. Erber NP. Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing. J Speech Hear Res 1972:14:413-22.
29. Hack ZC, Erber NP. Auditory, visual, and auditory-visual perception of vowels by hearing-impaired children. J Speech Hear Res 1982:25:100-7.
30. Wozniak VD, Jackson PL. Visual vowel and diphthong perception from two horizontal viewing angles. J Speech Hear Res 1979:22:354-65.

31. Montgomery AA, Jackson PL. Physical characteristics of the lips underlying vowel lipreading performance. J Acoust Soc Am 1983:73:2134-44.
32. Clarke FR. Constant-ratio rule for confusion matrices in speech communication. J Acoust Soc Am 1957:29:715-20.
33. Busacco DA. The effect of age on the benefit derived from visual cues in auditory-visual speech recognition by the hearing impaired (Dissertation). New York: Columbia University, 1988.
34. Brooks PM, Frost BJ, Mason JL, Gibson DM. Continuing evaluation of the Queens University Tactile Vocoder. I: identification of open set words. J Rehabil Res Dev 1986:23(1):119-28.
35. Brooks PM, Frost BJ, Mason JL, Gibson DM. Continuing evaluation of the Queens University Tactile Vocoder. II: identification of open set sentences and tracking narrative. J Rehabil Res Dev 1986:23(1):129-38.
36. Brooks PM, Frost BJ. Evaluation of a tactile vocoder for word recognition. J Acoust Soc Am 1983:74:34-9.
37. De Fillippo CL, Scott BL. A method for training and evaluating the reception of ongoing speech. J Acoust Soc Am 1978:63:1186-91.
38. Weisenberger JM, Broadstone SM, Saunders FA. Evaluation of two multichannel tactile aids for the hearing impaired. J Acoust Soc Am 1989:86:1764-79.
39. Saunders FA, Franklin B. Field tests of a wearable 16-channel electrotactile sensory aid in a classroom for the deaf. J Acoust Soc Am 1985:78:S17.
40. Reed CM, Durlach NI, Braida LD, Schultz MC. Analytic study of the Tadoma method: identification of consonants and vowels by an experienced Tadoma user. J Speech Hear Res 1982:25:108-16.
41. Reed CM, Rabinowitz WM, Durlach NI, Braida LD, Conway Fithian S, Schultz MC. Research on the Tadoma method of speech communication. J Acoust Soc Am 1985:77:247-57.
42. Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. J Acoust Soc Am 1988:84:101-14.
43. Rabinowitz WM, Eddington DK, Delhorne LA, Cuneo PA. Relations among different measures of speech reception in subjects using a cochlear implant. J Acoust Soc Am 1992:4(1)1869-8.
44. Lee K-F. Automatic speech recognition. The development of the SPHINX system. Boston, MA: Kluwer Academic Publishers, 1989.
45. Upton HW. Wearable eyeglass speechreading aid. Am Ann Deaf 1968:113:222-9.
46. Gengel RW. Research with the Upton's visual speechreading aid. In: Proceedings of the Research Conference on Speech-Processing Aids for the Deaf, Washington, DC: Gallaudet College, 1977.
47. Ebrahimi D, Kunov H. Peripheral vision lipreading aid. IEEE Trans Biomed Eng 1991:38:944-52.