Ann.Bull.RILP(1988)No.22, 105-116

DESIGN OF A COMPUTER-AIDED SPEECHREADING TRAINING SYSTEM

André-Pierre Benguerel[1] and Margaret Kathleen Pichora-Fuller[2]

Introduction

Computer-Aided Speechreading Training, or **CAST**, is the natural outcome of the parallel evolution of the language laboratory, and of video and microprocessor technologies. Two versions of a **CAST** system have been designed and implemented, one for English, the other for French. The description which follows pertains more particularly to the English version, although mention of the French version will be made whenever warranted. The system to be described has been developed specifically for speechreading training of adults with acquired hearing loss, however, mutatis mutandis, similar systems could be designed for other groups of hearing-impaired individuals.

Speechreading training has been shown to be of benefit to adults with acquired hearing loss [1),13)]. Nevertheless, it is not commonly included in aural rehabilitation programs for a number of reasons:
(a) rehabilitation programs usually focus on amplification;
(b) the process of speechreading is insufficiently understood;
(c) measures of speechreading ability need to be developed to better evaluate the need for and outcome of training;
(d) methods of training are not familiar to many clinicians;
(e) there is a shortage of clinicians to provide such training.

The **CAST** system was designed for clinical use in a language laboratory situation, to increase the availability of speechreading training, and for research use, to study the process of speechreading and to further the development of measures of speechreading ability.

Language laboratory training as a supplement to face-to-face language instruction is not a new practice. It is frequently used in second-language training of adults. Learners have the advantage of training relatively independently, at their own pace and convenience. In addition, computer-aided lessons provide the learner with interactive feedback which makes the language laboratory experience more like face-to-face instruction by a teacher. The computer allows both the instructor and the researcher to record details of learner performance.

---

[1] School of Audiology and Speech Sciences, University of British Columbia, Vancouver, B.C., Canada V6T 1W5.
[2] Department of Psychology, Erindale College, University of Toronto, Mississauga, Ontario, Canada L5L 1C6.

The support of Bell Canada, the Silverman Family Foundation, and the Frank Gerstein Family Foundation is gratefully acknowledged, as well as the assistance of Mauro Cicchelli.

Computer-aided speechreading lessons have been successfully used in a language laboratory at the National Technical Institute for the Deaf (NTID) at Rochester, New York [11),12)] for over five years. The success of the NTID application encouraged us to develop a similar system for speechreading training of adults with acquired hearing loss, however, several significant changes in lesson design and materials were deemed necessary for such training to be suitable for the targeted population. The NTID lessons were based on non-interactive videotaped instructional materials for post-secondary students with severe-to-profound hearing loss. Deaf students rely more on visual speech perception than audition and have relatively poor language achievement levels, whereas adults with acquired hearing loss rely primarily on audition and have excellent language skills. The **CAST** system was developed to incorporate the goals and assumptions of face-to-face (interactive) training that had been provided to over 100 patients seen in a five-year period for aural rehabilitation. This training is tailored to the apparent strengths, weaknesses and needs of pre-retirement adults with acquired mild-to-moderate hearing loss. It aims to enhance speechreading in order to complement auditory speech reception which becomes difficult when the acoustic speech signal is degraded, background noise interferes with auditory speech reception, the rate of speech is rapid, language is complex or the topic of conversation is unfamiliar. The approach used here assumes that speechreading is a communication task which requires three skills: (1) good visual speech decoding, (2) effective use of linguistic redundancy, and (3) effective use of feedback between message sender and receiver.

Instrumental Set-up

The hardware part of the **CAST** system consists of five commercially available components:
(1) an IBM PC-AT microcomputer with 512-kb RAM, including a VT-220 keyboard, a 30-Mb hard disk, and a colour graphics card;
(2) a 14" high-quality colour TV monitor (Sony 1271Q);
(3) a high-quality Beta I video recorder (Sony SLO-325);
(4) a BCD Videolink RS 232 video controller (interface);
(5) a printer.

The software consists of:
(1) an operating system (MS-DOS, version 3.20);
(2) a set of **CAST** programs, written in C;
(3) a set of **CAST** lesson video tapes, to be described next.

Course and Lesson Structure

The structure of the course emphasizes training in visual speech decoding, since this is assumed to be the poorest of the three skills required in speechreading in those with acquired mild-to-moderate hearing loss. This group of learners is known to have excellent language skills and to function well in situations where intelligibility is sufficient. Secondary to training in visual speech decoding is its integration with the use of linguistic redundancy and sender-receiver feedback. Visual speech decoding is assumed to depend on the recognition of "visemes", units of speech characterized by visually distinct articulations. Visemes are related to groups of allophones of phonemes with a common place of articulation: for example, the allophones of /p/, /b/, and /m/ are categorized as one single viseme since they are manifested as apparently visually indistinguishable bilabial articulations. There is controversy over the number and definition of visemes, since there is variation in articulation due to coarticulatory effects, especially at fast rates of speech, and since the visibility of articulations is speaker-dependent and context-dependent [2]. Nevertheless, there is a generally accepted hierarchy of difficulty [8]. The English version of the course consists of eight lessons each emphasizing a consonantal viseme. The French version also consists of eight lessons, six of them emphasizing a consonantal viseme and two emphasizing a vowel viseme. They progress from the easiest and least controversial viseme to the most difficult and controversial, in the following sequence:

English	French
Viseme 1: /p,b,m/	Viseme 1: /p, b, m/
Viseme 2: /f,v/	Viseme 2: /f, v/
Viseme 3: /θ,ð /	Viseme 3: /t, d, n, l/
Viseme 4: /t,d,n,l/	Viseme 4: /s, z/
Viseme 5: /s,z/	Viseme 5: /ʃ, ʒ/
Viseme 6: /ʃ,ʒ,tʃ,dʒ/	Viseme 6: /k, g, r/
Viseme 7: /w,r/	Viseme U: unrounded vowels
Viseme 8: /k,g,h,j,ŋ/	Viseme V: rounded vowels

Lessons to train English vowel visemes were not developed because auditory perception of vowels is much better than that of consonants and because the definition of vowel visemes is even more controversial. For each version, two additional lessons are available which do not aim to train particular visemes but which are designed to be linguistically equivalent and suitable for pre- and post- training comparisons or for general practice by a learner who has completed the training lessons. The structure of each lesson parallels face-to-face training and has four components:

1. review of previously taught visemes;
2. training in recognition of a new viseme;
3. practice using new and old visemes by tracking discourse;
4. recapitulation.

Lesson Materials and Procedures

The review of previously taught visemes is accomplished using syllables in a visual-only presentation, each syllable consisting of a consonant viseme followed by the vowel /a/. Viseme tokens are demonstrated and identified. The learner may choose to play demonstration items any number of times.

Training in the recognition of consonant visemes is accomplished using pairs of items of the form C_1a - C_2a in a visual-only presentation, where C_1 is always a token of the target viseme and C_2 may be any consonant. The target viseme is paired once with each consonant. The order of the pairs is random. When the consonant of the second item of the pair is an instance of the target viseme, the pair should be identified as "same"; when it is not, the pair should be identified as "different". After all the pairs have been judged, pairs in error are repeated until they are judged correctly. The learner may replay pairs until satisfied that it is possible to correctly identify whether the pairs are the same or different. At the end of the task, the members of the target viseme are presented in a combined audio-visual presentation.

Practice in integrating the use of linguistic redundancy and sender-receiver feedback with visual speech decoding is accomplished using materials in which there is a high frequency of instances of the target viseme. In these viseme-specific texts, the target viseme occurs at about twice its general rate of occurrence. The use of frequency dictionaries [5),7)] proved highly valuable in the composition of these viseme-specific texts, the development of which is not without analogy with the development of phoneme-specific sentences [6)]. The samples of text were inventoried to facilitate the evaluation of the effect of lesson materials on performance.

The discourse tracking procedure [4)]

The "tracking procedure" is a procedure to train and evaluate the reception of ongoing speech. The sender reads a segment of text. The receiver reports his perception of the segment of text and is corrected by the sender until the text is repeated verbatim. Performance is measured in words per minute repeated correctly. In its original form, the discourse tracking procedure is conducted using literary passages that are not specifically designed for the training application. Often, no analysis of the linguistic characteristics of the materials is available. In the present application, the texts were inventoried in terms of lexical items which were subcategorized as types and tokens of grammatical and content words. If a lexical item recurs, the type count does not increase but the token count does. Grammatical words account for 21% of the types and 46% of the tokens in all lessons with 76% repeated tokens. Content words account for 79% of the types and 54% of the tokens with 23% repeated tokens. The inclusion of repeated tokens and strings allows the evaluation of the effect of previous practice on recognition. The text for each lesson was constructed on a particular topic, so that, within the lesson, material is semantically related. All topics are considered to be challenging but familiar to the targeted population. The topic of each lesson is known to the learner at the beginning of the practice exercise. A variety of syntactic structures and sentence lengths were included. The sample texts are presented both in paragraph and phrase format. The paragraph format is used in the introduction and conclusion of the practice part of the lesson. The phrase format is used in the body of the lesson, in a modified tracking procedure.

The face-to-face discourse tracking procedure differs from other speech reception training procedures by the fact that connected speech is used rather than lists of unrelated items of sentence length or shorter, and by the fact that the objective is correct message reception rather than a single-chance "win or lose" task. The sender re-transmits the message until the receiver responds correctly. Sender-receiver feedback consists of a modification of the transmission by the instructor/sender according to the nature of the incorrect response(s) of the learner/receiver and may include labelling of error type.

While the discourse tracking procedure does approximate natural communication better than traditional speech reception procedures, it still has limitations. The two major functions of natural communication are its "transactional" function, which is the expression of "content", and its "interactional" function, which is the expression of social relations and personal attitudes [3]. The transmission of text in the discourse tracking procedure is essentially transactional; it is specific to narrated text and consists only of the overt demonstration of verbatim accuracy of message reception, whereas the topic of the text is of minor importance. It is unlike natural dialogue because, at the outset, the entire text of the message is exactly known to the sender and is not at all known to the receiver. The novelty of the exchange stems only from the errors of the receiver. In the face-to-face discourse tracking procedure, the interactional function is present only to the extent that personal reactions to the task may be expressed. The rôle of the sender is simply to (re-)transmit the message and provide feedback contingent on the receiver's response. The face-to-face tracking procedure relies on the skill of both parties, confounding the contributions of receiver and sender.

The computer-aided procedure standardizes the rôle of the sender, completely eliminates the interactional function of communication, which is non-essential in tracking, and focuses on the rôle of the receiver in accordance with the goals of training. Receiver and sender skills are not confounded. The computer-controlled tracking procedure differs from the face-to-face tracking procedure in three major ways:
(1) sender decisions about the message transmission and feedback are pre-determined in the training algorithm and are perfectly replicable;
(2) the receiver has more choices: he can move to any phrase, or he can return to a preceding phrase to complete it;
(3) the response is typed and is displayed for the duration of the lesson on a "worksheet", on the monitor screen.
A description of each of these differences follows.

(1) Rôle of the Sender. In the **CAST** system, the sender is the computer controlled videotape. The rôle of the sender is standardized by allowing the instructor to pre-determine decisions about message (re-)transmission and feedback in the lesson algorithm. The instructor preselects three parameters of the (re-)transmission: speaking rate, modality of presentation and phrase length.

The speaking rate of phrases can be specified. The paragraph and phrase readings of the text were recorded at two speaking rates, representing slow normal speech (75 words per minute) and fast normal speech (130 words per minute). For each lesson and each learner, the instructor specifies an appropriately graded schedule for modifying the rate of presentation for up to ten trials per phrase. The same schedule of modifications applies to each phrase within a lesson.

The modality of presentation can also be specified. Both the audio and the visual signal were recorded, enabling a visual-only or a combined audio-visual presentation. Like for rate of presentation, the instructor specifies an appropriately graded schedule for modifying the modality of presentation for up to ten trials per phrase. The same schedule of modifications applies to each phrase within a lesson. The instructor may also select a presentation of the **CAST** lesson, with or without competing background noise, and with or without amplification.

Phrase length was determined prior to videotaping. Since the text was parsed only in one way, modification of the length of the (re-)transmission is more constrained in this computer-aided procedure than in the face-to-face procedure. Even though the duration of the videotaped phrase cannot be modified, in effect, the length of the message being worked on by the receiver is reduced when feedback is provided and it is increased when the receiver choses to view another phrase. It is expected that phrase length may shift the nature of the task. The reception of longer phrases is likely to be more a function of linguistic redundancy or **speechreading**. In contrast, the reception of very short phrases is likely to be more a function of visual speech decoding or **lipreading**, since less redundancy is available within the phrase. In general, more redundancy is expected to be available as more phrases are completed. The decision was made to use relatively short phrases with these hypotheses in mind and in accordance with the goal of training to increase visual speech perception ability. Phrases were recorded by one speaker at two rates for this application, however, future applications could include other word-by-word or syllable-by-syllable recordings by multiple speakers. The impact of phrase length on the particular design of the feedback algorithm must also be considered.

In both the face-to-face and the **CAST** tracking procedures, feedback is provided for verbatim accuracy and may also include labelling of error type. The other computer-aided speechreading algorithms mentioned [11),1)] provide feedback based on word-level spelling accuracy . Similarly to these, the **CAST** system includes feedback based on a word-level spelling match of response to target. Word position in the phrase is ignored if a word-level spelling match is successful. However, the goal of the **CAST** system is to train the learner in the visual decoding of speech using feedback based on viseme recognition and not just on word spelling. Therefore, feedback is also provided on the basis of the visemic match between response and target. Rules were developed to convert spelling to visemic code both for the English [10)] and the French system. If the word-level spelling match fails; the response is coded according to the spelling-viseme conversion rules and the coded response is matched to the target which has also been coded by these rules. The matching algorithm compares the response and target code strings character by character. Each viseme is coded as a number or a letter, as indicated earlier. Since in the English version, vowel visemes are not presented, a vowel or vowel cluster is coded as V; credit is given for any vowel in the correct position so long as it occurs in isolation (i.e. between word boundaries) or the adjacent consonant is correctly identified. In the French version, unrounded vowels are coded as U, rounded vowels as V. UU and VV clusters are reduced to U and V respectively. UV and VU clusters on the other hand are not reduced, thus allowing the coding of the semi-vowels /w/ and /ɥ/ as VU, and /j/ as UV.

Consonant visemes which have not yet been presented are coded simply as C. Homophenous consonant clusters (i.e. clusters constituted of consonants belonging to the same viseme) are reduced to a single instance of that viseme. Credit is given for an consonant identified as such in the correct position. In this way, feedback indicating a correct response is provided if the error is a within-viseme error for visemes which have been presented in a previous lesson or if it is a vowel for a vowel, or an untaught consonant for an untaught consonant. Matching is more demanding in later lessons as more visemes have been taught. In the **CAST** system when all visemes have been taught, a typical example might be the following:

Transaction	Presentation Mode	Item	Visemic code (not shown on screen)
Target:	videotape	Dublin	4V14V4
Response:	keyboard/screen	dumpling	4V14V8
Feedback:	screen	Dubli_	
Response:	keyboard/screen	Dublin	4V14V4
Feedback:	screen	**Dublin**	

In addition to a simple position-by-position match, a rightward search to the end of the next word takes place if the character of the response being evaluated is a trained viseme and it does not match the target character in the corresponding position. If a match is found, the search continues rightward from the position of the match. This is important because word boundaries are not necessarily apparent to the lipreader who is given a visual-only presentation. Some examples of feedback resulting from this rule are shown in the following sequences:

Transaction	Presentation Mode	Item	Visemic Code (not shown on screen)
Target:	videotape	in particular	V4#1V74V8V4V7#
Response:	keyboard/screen	aperture	V 1V74V 7#
Feedback:	screen	__ parti____r	

or

Transaction	Presentation Mode	Item	Visemic Code (not shown on screen)
Target:	videotape	post office	1V54#V2V5#
Response:	keyboard/screen	payoff	1V8V 2#
Feedback:	screen	po__ _ff___	

When the length of the phrase exceeds two words, the rightward search rule is less advantageous since the risk of skipping correctly perceived items increases. For example, with a rightward search matching rule which searches past the end of the next word, the following might occur:

	Item	Visemic Code (not shown on screen)
T:	bright colours of the new petunias	17V4#8V4V75#V2#3V#4V7#1V4V4V5#
R:	white colour from petals	7V4#8V4V7# 2 7V1#1V4V45#
F:	_right colour_ _f ___ _w p__unia_	

If the rightward search rule is constrained to stop at the end of the next word, since it is less likely that errors would extend beyond two word boundaries, the feedback gives more appropriate credit and would yield the following:

	Item	Visemic Code (not shown on screen)
T:	bright colours of the new petunias	17V4#8V4V75#V2#3V#4V7#1V4V4V5#
R:	white colour from petals	7V4#8V4V7# 27V1# 1V4V4 5#
F:	_right colour_ _f ___ ___ petun__s	

Although this matching algorithm is not always optimal, it does provide with more feedback than word-level spelling matching, which would give the receiver no credit for good guesses.

(2) Rôle of the Receiver. In the face-to-face procedure, two types of action are available to the receiver: elicit feedback by guessing the answer, or elicit a re-transmission by a response such as "what?" Depending of the response, the re-transmission may be modified at the sender's discretion. In the **CAST** system, the receiver has a choice of three types of action:

elicit feedback by typing a guess, replay the phrase, or move to another phrase with the option to return later. Speech rate and modality of presentation are modified according to the schedule pre-determined by the instructor and feedback is governed by the matching rules. There is a ceiling on the number of times the message can be re-transmitted before the answer is given, however, the receiver is free to choose any action at any time. The receiver's option to move to another phrase is approximately equivalent to the sender's decision to lengthen the phrase in the face-to-face procedure. The receiver is less passive in the **CAST** system than in the face-to-face procedure and can actively explore new strategies. The only outcome measure usually reported for the face-to-face procedure is tracking rate in words per minute. In the **CAST** system, all of the receiver actions are recorded and timed. Tracking rate is reported in words per minute, phrases per minute, and actions per phrase. Time per guess, time per phrase and time per action may also be calculated. In addition to efficiency measures, the **CAST** system provides a breakdown of the three receiver action types (play, answer (or guess), move to another phrase). A complete transcript of the responses is available, in sequential or phrase order. Strategies may be inferred from the summary of receiver actions, reported at the end of the lesson. The instructor and the learner may discuss progress and evaluate the benefit of the strategies employed.

(3) Worksheet. Two disadvantages due to the spoken nature of the face-to-face procedure are overcome by the **CAST** worksheet, which displays the receiver's response and the feedback on the monitor screen. The first disadvantage is the possible uncertainty of the sender/instructor about the response when the quality of the receiver's speech production is poor. The second disadvantage is that, since feedback is presented in the same way as the message, the receiver may have as much difficulty understanding feedback as understanding the message. In the **CAST** system, responses and feedback are always unambiguous. Throughout the lesson, completed responses and the most recent feedback remain displayed on the worksheet, providing easy reference to helpful information.

Implementation

The **CAST** software was implemented in **C**, a high-level programming language, on an IBM PC-AT computer. The C language and the IBM PC-AT computer surpass the technology used in other computer-aided speechreading training and are able to efficiently support the sophisticated design of **CAST**. While access is slower for videotape than it would be for video disc, the flexibility in creating lessons was a motivating factor in opting for videotape at this time. It should be obvious, however, that once the text of the lessons has been finalized, it could be readily transferred to video discs if economically warranted. Lesson materials were recorded in such a way as to minimize the need to search large distances on the videotape.

CAST menus and prompts are sufficient to guide basic operation of the software system and further information is available in the user's manual, however, it is recommended that new instructors be trained in a one-day workshop which would include familiarization with the goals and design of the system, hands-on experience with its implementation and operation, and a discussion of the system as a rehabilitative tool.

Since the **CAST** system is intended to be one component in a broader aural rehabilitation program, the instructor must determine if speechreading training is indicated for the hearing-impaired individual and if the individual is a good candidate for **CAST.** Those who are candidates for comparable face-to-face training and who are comfortable with the computer keyboard are suitable candidates for **CAST.** Learners can begin to use the **CAST** system in the first session. During the pre-training assessment, the instructor has an opportunity to observe the learner using the system and to identify specific difficulties in operating the **CAST** system. The results of each lesson are expected to be reviewed by the instructor with the learner so that re-assessment is ongoing as it would be in any therapy. Clinical use is considered to be viable for the targeted population.

Evaluation

The evaluation of software products for clinical use is a new but imperative concern. Such an evaluation was conducted, both from the point of view of the human-computer interface and from the clinical point of view. The results of this evaluation are reported elsewhere [9].

Further considerations

In the development of a **CAST** system, many factors have to be taken into account. Some of these factors are related to the perceptual and linguistic capabilities of the learner (e.g. the three basic skills mentioned earlier) and to the specific characteristics of his hearing-impairment. Other factors are related to the phonemic/visemic aspects of the language under consideration (e.g. the partitioning of the set of allophones into visemes, or the spelling-to-visemes conversion rules).

One interesting aspect of language which has not been discussed here sofar but which would have to be given serious thought if one were to design a **CAST** system for Japanese (or for any other language not using an alphabetic writing system) is that of the writing system used (both for the responses of the user and for displaying feedback to him). For any language with an alphabetic writing system, even when the phoneme-to-grapheme correspondence is far from being one-to-one (such as in English or French), cases where a grapheme correponds to more than one phoneme are extremely rare; in English and French, they are essentially limited to cases involving the letter "x".

In Japanese, three writing systems are available: kanji (morphemic-ideographic), kana (phonemic-syllabic), and romaji (phonemic-alphabetic). The kanji form has to be excluded for several reasons: as a means of keying in the user's response, kanjis would be impractical since they are accessible only indirectly from the keyboard (through one of the other two forms); as a means of displaying feedback, they would make it hard, not to say impossible, to express a partially correct response as a subset of strokes of a kanji. Ideally, feedback should be displayed at the phonemic or at the visemic level. This leaves us with the choice between kana and romaji. The former, in principle, has the drawback of conveying more than one viseme (in general one consonant and one vowel) per symbol, however, this drawback may not be significant in view of the fact that Japanese has an overwhelmingly -CV- syllabic structure and that Japanese vowels may all be members of the same (vowel) viseme. If this last assumption is not supported, romaji may be the preferrable form, although other factors may militate against its use, in particular its lesser familiarity to many users and its weaker relation to the way in which lexical items are represented in the mind of the speaker-listener.

Conclusion

As stated earlier, good speechreading requires three basic skills: good visual speech decoding, effective use of linguistic redundancy, and effective use of feedback between sender and receiver. Any speechreading training method, whether mechanized or not, should emphasize training in the weak area(s). These will depend on the experience of the learner. In the present case, the intended users are adults with an acquired hearing loss, but appropriate changes should permit the application of the same design principles for other hearing-impaired populations.

Another important point is that the gradation of materials should be linguistically based. This hints at the need for more information: the determination of an optimum set of visemes for a given language, their rate of occurrence, and the amount of speech information which can be transmitted in the visual modality have not yet been established.

Compared to the face-to-face discourse tracking procedure, **CAST** has several advantages: it provides standardization and allows easy measurement and analysis of speechreading errors and strategies. It promisses to reveal more about the speechreading process in general, and about the need for and effectiveness of training for specific classes of hearing-impaired learners. It allows the analysis and the systematic manipulation of connected speech subparts to assess the contribution of vision to speech perception. Hypotheses about the use of linguistic redundancy may also be tested. Only when the information bearing capacity of the visual modality is better understood and reliably quantified will evaluations of complementary channels be optimized. **CAST** should be a valuable tool for such an investigation.

REFERENCES

1) Abrahamson, J., Kopra, L.L., Kopra, M.A. & Dunlop, R.K.: Rationale for the Development of Visual Communication Training in an Adult Aural Rehabilitation Program. Paper presented at the Summer Institute of the Academy of Rehabilitative Audiology, Lake Geneva, WI, June 1986.

2) Benguerel, A.-P. and Pichora-Fuller, M.K. Coarticulation effects in lipreading. Journal of Speech and Hearing Research, 25(4), 600-607, 1982.

3) Brown, G. and Yule, G.: Discourse Analysis. Cambridge University Press: Cambridge, 1983.

4) De Filippo, C.L. and Scott, B.L.: A method for training and evaluating the reception on ongoing speech. Journal of the Acoustical Society of America, 63(4), 1186-1192, 1978.

5) Francis, W.N., and Kucera, H.: Frequency Analysis of English Usage: Lexicon and Grammar. Houghton Mifflin: Boston, 1982.

6) Huggins, A.W.F. and Nickerson, R.S.: Speech quality evaluation using "phoneme-specific" sentences. Journal of the Acoustical Society of America, 77(5), 1896-1905, 1985.

7) Juilland, A., Brodin, D., and Davidovitch, C.: Frequency Dictionary of French Words. Mouton: The Hague, 1970.

8) Pichora-Fuller, M.K.: Coarticulation and lipreading. (Master's thesis, University of British Columbia, 1980.) In: Canadian Theses, 1976/77-1979/80, 2, 50023, 1983. (Available from Canadian Theses on Microfiche Service, Collections Development Branch, National Library of Canada, Ottawa, K1A 0N4.)

9) Pichora-Fuller, M.K. and Benguerel, A-P.: Computer-Aided Speechreading Training: Development of a Rehabilitative Tool, forthcoming.

Pichora-Fuller, M.K. and Cicchelli, M.: Computer Aided Speechreading Training (CAST): Owner's Manual, 1987. (Available from M.K. Pichora-Fuller, Department of Psychology, University of Toronto, Toronto, M5S 1A5.)

11) Sims, D.G., Scott, L. and Myers, T.: Past, present, and future uses of computer-assisted training at NTID. Journal of the Academy of Rehabilitative Audiology, 15, 103-115, 1982.

12) Sims, D.G., VonFeldt, J., Dowaliby, F., Hutchinson, K. and Myers, T.: A pilot experiment in computer assisted speechreading instruction utilizing the data analysis video interactive device (DAVID). Chapter 16 in American Annals of the Deaf Symposium on Research and Utilization of Educational Media for Teaching the Deaf: Educational Technology for the '80's. American Annals of the Deaf, 124(5), 616-623, 1979.

13) Walden, B.E., Prosek, R.A., Montgomery, A.A., Scherr, C.K. and Jones, C.J.: Effects of training on the visual recognition of consonants. Journal of Speech and Hearing Research, 20(1), 130-145, 1977.