

Technical University of Denmark



## Training of Speechreading for Severely Hearing-Impaired Persons by Human and Computer

**Bothe, Hans-Heinrich**

*Published in:*

Seventh IEEE International Conference on Advanced Learning Technologies, 2007. ICALT 2007.

*Link to article, DOI:*

[10.1109/ICALT.2007.269](https://doi.org/10.1109/ICALT.2007.269)

*Publication date:*

2007

*Document Version*

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*

Bothe, H-H. (2007). Training of Speechreading for Severely Hearing-Impaired Persons by Human and Computer. In Seventh IEEE International Conference on Advanced Learning Technologies, 2007. ICALT 2007. IEEE. DOI: 10.1109/ICALT.2007.269

## DTU Library

Technical Information Center of Denmark

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Training of Speechreading for Severely Hearing-Impaired Persons by Human and Computer

Hans-Heinrich Bothe  
Technical University of Denmark  
hhb@oersted.dtu.dk

## Abstract

*This paper describes evaluation results for a software programme that is intended to be used as a training-aid for lipreading in German. Tests were carried out in schools for hearing-impaired children in Germany which indicate that the ability to lipread increases significantly already after use of the software during a short period of time.*

## 1. Introduction

Our senses are specialized to perform certain task-related work. The cerebral cortex fuses gained information so that disorders in one modality can be compensated for by information from another one. Many hearing-impaired people utilize facial and, particularly, mouth movements during speech production as a compensation of their hearing loss. The procedure is called speechreading or, more specifically, lip-reading. Speechreading becomes possible since relevant facial movements during speech production or articulation are highly correlated with the articulatory movements and thus, with the speech signal.

## 2. Speechreading

Speechreading is essential for all those hearing-impaired people who freely want to communicate in the non-hearing-impaired society. It affects impaired or deaf children as well as adults deafened after language acquisition. It is also helpful for elderly who naturally and gradually lose their ability to hear.

Traditionally, speechreading is seen as the acquisition of a visual language, and persons affected have to follow a certain learning schedule [1], [2]. There are cultural differences in how such a schedule should look like, but two general procedures are usually present, i) a more 'mechanical' training of often used phrases and ii) a more natural and lively individual or classroom discussion [3]. Learning of speechreading

may also be combined with learning of sign language gestures or language-related sign language.



Figure 1. LIPPS user interface showing a schematic face

## 3. The training software 'LIPPS'

LIPPS is a training software for the 'mechanical' phase of speechreading on a PC, which utilizes a text-controlled computer animation of a moving face [4], [5] (see figure 1). It can be used in schools or rehabilitation centers, in special courses for elderly, or in guided self studies. It is intended to supplement and extend courses supervised by specialized and experienced teachers. The trainees can work and learn at home according to a personalized schedule designed by their teacher. Deafened trainees may use a mobile PC already after operation in the hospital, which turns out to be of high relevance since many patients are grateful to be kept busy with meaningful learning concepts.

At present, two natural face images with an animated mouth area (lips, chin, cheeks and tongue) and one cartoon face with animated mouth, tongue, and nostrils are available (see figure 1). Pedagogical concepts for the user interface of different complexity can be chosen. In general, all video clips are produced by the PC itself, which can be driven by any text input. Teachers can freely compose individual lessons for presentation and can individually configure the feedback system for the trainee (e.g., the animated face may present a 'yes' or 'no' in combination with eyebrow movements or simply show a respective text.

The computer animation is based on extensive pre-examination of video clips with ‘speaking faces’ of experienced teachers. The animation is based on morphing algorithms or, in the case of a schematic face, by line manipulation, while the motion model is driven by di-phone or di-viseme sequences.

#### 4. Evaluation of the computer animation

In the following, a study in two schools for severely hearing-impaired children in Germany is presented, which compares the training results from a human teacher with those from the computer animation. 29 children in the age between 14 and 17 years were employed. The children had to get accommodated and customized with the software during a short demonstration and test phase; thereafter, they participated in a supervised speechreading course that lasted one lecture block per day. Before and after the training course, the children were tested for their ability to speechread from LIPPS and from a human teacher. The complete evaluation period was one week.

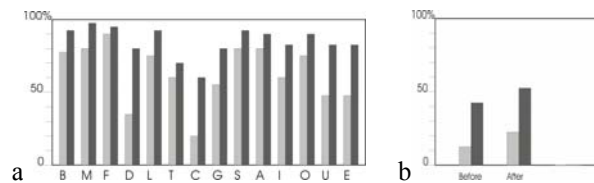
All classes used the same text corpus, consisting of 68 single words, a mixture of 28 short words (one or two syllables), 20 longer words (three to six syllables), and 20 logatomes (two to four syllables). Text material and answers of the children were transcribed into sequences of phonemes and then compared on base of the respective ‘viseme’ sequences [6]. This means that for presentation of /papa/, for example, the answer /mama/ was considered correct since /p/ and /m/ belong to the same ‘bilabial’ viseme ‘B’.

#### 5. Results

Generally, the intelligibility from a human was by approximately 30% better than from LIPPS. This holds for the tests before and after the speechreading course. Two aspects may play an essential role, i) only parts of the schematic face were moving and ii), coarticulation due to the diphone motion model does not contain any prosodic information and can still be improved.

Figure 2a shows, arranged by visemes, the overall percentage of correct answers for the LIPPS software before and after training. Bad results for the alveolar closures  $D=\{/d/, /t/\}$ , for the guttural fricatives  $C=\{/c/, /x/\}$  and the mix-up of some vowels in U and E improved significantly by the training procedure.

Figure 2b shows the hitting rates for complete words from LIPPS and a human teacher before and



**Figure 2. a. Hitting rates by viseme □ before and ■ after training, b. hitting rates for □ LIPPS and ■ teacher before and after training.**

after training. Intelligibility for LIPPS improved by approximately 10% and, more importantly, intelligibility from the human teacher improved similarly. Consequently, the test results show that training with LIPPS can improve the ability to speechread from a human speaker. This is exactly what LIPPS is intended for.

By combining (not shown) detailed results for the improvement of hitting rates arranged by visemes (e.g., with or without visible tongue movements) and results for complete words, it can further be concluded that the children put an emphasis on tongue and characteristic lip movements (e.g., lip closures or labial friction). Consequently, computer ‘speakers’ as well as humans should employ clear but yet natural tongue and lips movements when they communicate with hearing-impaired people.

#### References

- [1] Brauckmann, K., *Absehungunterricht nach dem Jenaer Verfahren*, Gustav Fischer-Verlag, 1933.
- [2] Lindner, G., *Absehen – der andere Weg zum Sprachverstehen*, Luchterhand-Verlag, 1999.
- [3] Claußen, H., S. Treue, *Das Absehen und die Aufgabe von Absehkursen. Ein geschichtlicher Überblick. Hörgeschädigtenpädagogik*, 2 (1993), pp. 84-87.
- [4] Bothe, H.-H., G. Lindner, F. Rieger, *The Development of a Computer Animation Program for the Teaching of Lip-reading*. In: Ballabio, E., I. Placencia-Porrero, R. Puig de la Bellacasa (eds.), *Technology and Informatics 9, Rehabilitation Technology: Strategies for the European Union*, IOS Press, Amsterdam, 1993.
- [5] Bothe, H.-H., *Audio to Audio-video Speech Conversion with the Help of Phonetic Knowledge Integration*, Proc. IEEE Intern. Conf. on Systems, Man, and Cybernetics (SMC’97), Orlando, 1997.
- [6] Alich, G., *Zur Erkennbarkeit von Sprachgestalten beim Absehen vom Munde*, Dissertation, Bonn, 1961.