



**KTH Computer Science  
and Communication**

**Computer-Based Speech Therapy  
Using Visual Feedback  
with Focus on Children  
with Profound Hearing Impairments**

**ANNE-MARIE ÖSTER**

**Doctoral Thesis  
Stockholm, 2006**

Akademisk avhandling som med tillstånd av Kungliga Tekniska Högskolan framlägges till offentlig granskning för avläggande av filosofie doktorsexamen i tal- och musikkommunikation med inriktning på talkommunikation, torsdagen den 15 juni kl. 10.00 i sal F2, Lindstedtsvägen 26, Kungliga Tekniska högskolan, Stockholm.

TRITA-CSC-A 2006:10

ISSN-1653-5723

ISRN-KTH/CSC/A – 06/10--SE

ISBN 91-7178-399-7

KTH School of Computer Science and Communication  
Department of Speech, Music and Hearing  
SE-100 44 STOCKHOLM, Sweden

© Anne-Marie Öster, juni 2006

Tryck: Universitetsservice US AB

## *Abstract*

This thesis presents work in the area of computer-based speech therapy using different types of visual feedback to replace the auditory feedback channel. The study includes diagnostic assessment methods prior to therapy, type of therapy design, and type of visual feedback for different users during different stages of therapy for increasing the efficiency. The thesis focuses on individual computer-based speech therapy (CBST) for profoundly hearing-impaired children as well as for computer-assisted pronunciation training (CAPT) for teaching and training the prosody of a second language. Children who are born with a profound hearing loss have no acoustic speech target to imitate and compare their own production with. Therefore, they develop no spontaneous speech but have to learn speech through vision, tactile sensation and, if possible, residual hearing. They have to rely on the limited visibility of phonetic features in learning oral speech and on orosensory-motor control in maintaining speech movements. These children constitute a heterogeneous group needing an individualized speech therapy. This is because their possibilities to communicate with speech depend not only on the amount of hearing, as measured by pure-tone audiometry, but also on the quality of the hearing sensation and the use the children through training are able to make of their functional hearing for speech. Adult second language learners, on the other hand, have difficulties in perceiving the phonetics and prosody of a second language through audition, not because of a hearing loss but because they are not able to hear new sound contrasts because of interference with their native language. The thesis presents an overview of reports made concerning speech communication and profound hearing impairment such as studies about residual hearing for speech processing, effects of speech input limitations on speech production, interaction between individual deviations and speech intelligibility, and speech assessment methods of phonetic realizations of phonological systems. Finally, through several clinical evaluation studies of three Swedish computer-based therapy systems, concerning functionality, efficiency, types of visual feedback, therapy design, and practical usability for different users, important recommendations are specified for future developments.



*till lilla Majken*



## *Acknowledgement*

I would like to thank all people that directly or indirectly have contributed to the completion of this thesis. First of all I want to express my sincere gratitude to David House, my supervisor, for his valuable support and encouragement. His gentleness and brilliant mind have stimulated me in trying to “think right and think clear”. Thanks to his guidance it has been possible for me to see the completeness in my research. I also owe my warmest and deepest gratitude to Björn Granström who always has been available to help and support me when I have needed advice.

My warmest thanks to Karl-Erik Spens for his valuable suggestions and comments on this thesis. I owe him a great debt of gratitude, for having introduced me to the Department of Speech Communication and Music Acoustics when I was a young student of phonetics, in the early seventies. Since then we have been officemates and shared many nice discussions, all laughs as well as many unforgettable moments when travelling over the world.

I warmly thank Gunnar Fant for having created the stimulating and pleasant atmosphere we all feel. I also want to thank Arne Risberg for his serious commitment, fruitful discussions, and thoughtful care of the former Hearing Technology Group, and Rolf Carlson for his kind attitude.

Especially, I want to thank my dearest friend and colleague Eva Agelfors for her deep friendship and for being able to share with her the joy and distress of daily life.

I am proud to be a member of several stimulating teams. I thank “Vingänget”: Eva Agelfors, Lennart Andersson, Mats Blomberg, Birgit Cook, Cathrin Dunger, Kjell Elenius, Inger Karlsson, Tina Magnuson, Arne Risberg, Gorda Surjadi, and Maj-Britt Wetterling for all the hilarious moments we shared together, especially in the lovely garden of Mats.

Together with “Reparationsfonden”: Eva Agelfors, Birgit Cook, Si Felicetti, Maj-Britt Wetterling, and Gunilla Öhngren Löfberg I have enjoyed luxury and sumptuousness during our memorable trips and journeys.

Finally, I want to thank the members of “Fiskmiddagsgruppen”, consisting of Eva Agelfors, Björn Granström, Peter

Nordqvist, and Karl-Erik Spens for delicious food, interesting discussions, and enjoyable company.

My sincere thanks to Kjell Elenius for his supportive attitude and for his generous invitations to “Grötgänget” when I didn’t have the time to fix my own lunch. He has been my angel of mercy during the editing of this thesis. A special thank also to Peter Nordqvist for his cheerful manners and expert help with good solutions to all my technical problems and to Inger Karlsson for her friendly drop-ins for a short chat. I express my gratitude to Rebecca Hincks, for proofreading the language of this thesis.

Thanks to Anders Askenfelt and the administrative staff Caroline Bergling, Cathrin Dunger, Markku Haapakorpi, and Niclas Horney for their resourcefulness and backup and to all my friends and colleagues at the Department of Speech, Music and Hearing for an inspiring working environment.

I also thank my colleagues at the Department of Linguistics, Stockholm University: Björn Lindblom (who led me into the world of phonetics), Bob McAllister, Olle Engstrand, Francisco Lacerda, and Ulla Sundberg for positive collaboration over the years.

Finally, I wish to thank the speech therapists of the Manilla School for the Deaf in Stockholm for valuable pedagogical collaboration and all other teachers who always made me welcome to their schools. I am indebted to Margaretha Andolf at the Language Unit at KTH, for giving me the possibility to evaluate SpeechViewer together with some of their L2 learners. Special thanks to Cecilia Melin Weissenborn for her enthusiastic work and joyful collaboration.

Enormous thanks to Ewa Bergek-Ulin and Lennart Ulin at Frölunda Data AB for combining business with pleasure and for their help to enthusiastically transform my experiences and results to be useful for speech and hearing-impaired children.

My loving thanks are due to my family, Björn, Charlotta, and Johan, who have given me so much love and happiness through the years, to Jonas my favourite son-in-law, to my dear and lovely sisters Birgitta and Monica, my nice three brothers-in-law, and last but not least, my sweet little granddaughter Majken, who has inspired me to great achievements.



## *The thesis is based on the following publications*

- Bälter, O., Engwall, O., Öster, A-M., Kjellström, H. (2005). Wizard-of-Oz Test of ARTUR - a Computer-Based Speech Training System with Articulation Correction [pdf]. In *Proceedings of the Seventh International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 36-43, October 9-12, 2005, Baltimore, MD.
- Engwall, O., Bälter, O., Öster, A-M., Kjellström, H. (2006). Designing the user interface of the computer-based speech training system ARTUR based on early user tests, to appear in *Journal of Behavioural and Information Technology*.
- Eriksson, E., Bälter, O., Engwall, O., Öster, A-M., Kjellström, H. (2005). Design Recommendations for a Computer-Based Speech Training System Based on End-User Interviews [pdf]. In *Proceedings of the Tenth International Conference on Speech and Computers*, pp. 483-486, October 17-19, 2005, Patras, Greece.
- Granström, B. & Öster, A-M. (1994a). Speech synthesis for hearing-impaired persons - in research, training and communication, *Proceedings from 2nd Int. Symposium on Speech and Hearing Sciences*, Sept. 24-25 1994, Osaka, Japan, 49-65.
- Granström, B. & Öster, A-M. (1994b). Speech synthesis for hearing-impaired persons - in research, training and communication, *STL/QPSR 2-3/94*, 93-111.
- Öster, A-M. (1985). The Use of a Synthesis-by-Rule-System in a Study of Deaf Speech. *STL-QPSR 1/1985*, 95-107.
- Öster, A-M. (1988a). Datorer i talundervisningen - Ett nytt hjälpmedel?, *Nordisk Tidskrift för Dövuundervisningen*, 1, 14-19.
- Öster, A-M. (1988b). Computer-based speech training, *Proceedings of Speech '88, 7th FASE Symposium, Edingburgh*, Book 2, 645-651.
- Öster, A-M. (1989a). Studies on phonological rules in the speech of the deaf. *STL/QPSR 1/89*, 159-162.
- Öster, A-M. (1989b). Applications and experiences of computer-based speech training. *STL/QPSR 4/89*, 37-44.

- Öster, A-M. (1989c). Applications and experiences of computer-based speech training, *Proceedings of European Conference on Speech Communication and Technology (Eurospeech)*, Paris, 714-717.
- Öster, A-M. (1990). The effects of prosodic and segmental deviations on intelligibility of deaf speech, *STL/QPSR*, 79-88.
- Öster, A-M. (1991). Phonological assessment of eleven prelingually deaf children's consonant production. *STL-QPSR* 2-3/91, 11-18.
- Öster, A-M. (1992a). *The speech of deaf children – Phonological assessment as a basis for speech training*, Thesis work for the Licentiate Philosophy degree in Phonetics, University of Stockholm, Institute of Linguistics.
- Öster, A-M. (1992b). Phonological assessment of deaf children's existing articulation skills as a basis for speech training. *Proceedings of ICSLP 92*, october 12-16, Banff, Alberta, Canada, 955-958.
- Öster, A-M. (1995a). *Principles for a complete description of the phonological system of deaf children as a basis for speech training*, Profound Deafness and Speech Comm., London: Whurr Publ. Ltd., 441-461.
- Öster, A-M. (1995b). Teaching speech skills to deaf children by computer-based speech training, *Proceedings of 18th International Congress on Education of the Deaf*, Tel-Aviv, Israel.
- Öster, A-M. (1995c). Teaching speech skills to deaf children by computer-based speech training, *STL-QPSR* 4/95, 67-75.
- Öster, A-M. (1995d). Resultat från datorbaserad röstträning med ett gravt hörselskadat förskolebarn, *Nordisk Tidskrift för Dövundervisningen*, 125-133.
- Öster, A-M. (1996). Clinical applications of computer-based speech training for children with hearing impairment. *Proceedings of ICSPL 96*, 157-160. Philadelphia, USA.
- Öster, A-M. (1997). Auditory and visual feedback in spoken L2 teaching, *Reports from the Dept of Phonetics, Umeå University, PHONUM* 4.

- Öster, A-M. (1998). Spoken L2 teaching with contrastive visual and auditory feedback, *Proc ICSLP, Sydney*.
- Öster, A-M. (1999a). Strategies and results from spoken L2 teaching with audio-visual feedback, *STL-QPSR 1-2/99*, 1-7.
- Öster, A-M., Vicsi, K., Roach, P., Kacic, Z., & Barczikay, P. (1999b). A multimedia multilingual teaching and training system for speech and hearing-impaired children – SPECO, *Proceedings of Fonetik 99*, 149-152. Gothenburg, Sweden.
- Öster, A-M. (2002a). The relationship between residual hearing and speech intelligibility - Is there a measure that could predict a prelingually profoundly deaf child's possibility to develop intelligible speech? *STL/QPSR Vol. 43*, 51-56.
- Öster, A-M., House, D., Protopapas, A., Hatzis, A. (2002b). Presentation of a new EU project for speech therapy: OLP (Ortho-Logo-Pedia), *Proceedings of Fonetik 2002, QPSR, Vol. 44*, 45-48.
- Öster, A-M., House, D., Hatzis A., Green, P. (2003). Testing a New Method for Training Fricatives using Visual Maps in the Ortho-Logo-Paedia Project (OLP), Umeå University, Department of Philosophy and Linguistics, *PHONUM 9 (2003)*, 2-X, Available online at <http://www.ling.umu.se/fonetik2003/>
- Vicsi, K., Roach, P., Öster, A., Kacic, Z., Barczikay, P., Sinka, I. (1999). SPECO – A Multimedia Multilingual Teaching and Training System for Speech Handicapped Children, *6th European Conference on Speech Communication and Technology, Eurospeech '99*, Budapest, 859-862.
- Vicsi, K., Roach, P., Öster, A., Kacic, Z., Barczikay, P., Tantos, A., Csatári, F., Bakcsi, Zs., Sfakianaki, A. (2000). A Multimedia Multilingual Teaching and Training System For Speech Handicapped Children, *International Journal of Speech Technology Vol. 3*, 289-300.

# Table of Contents

|  |            |
|--|------------|
| <b>Abstract</b> .....  | <b>i</b>   |
| <b>Acknowledgement</b> .....   | <b>v</b>   |
| <b>The thesis is based on the following publications</b> .....   | <b>vii</b> |
| <b>Table of Contents</b> .....   | <b>x</b>   |
| <b>1. Introduction</b> .....   | <b>1</b>   |
| 1.1. <i>Speech communication difficulties caused by speech perception problems</i> .....                                     | 4          |
| 1.1.1. Hearing impairment .....  | 4          |
| 1.1.2. Second language learners.....   | 5          |
| 1.2. <i>Objectives of the study</i> .....  | 6          |
| 1.3. <i>Thesis overview</i> .....  | 7          |
| 1.4. <i>Some basic concepts</i> .....  | 9          |
| 1.4.1. The term deaf .....   | 9          |
| 1.4.2. Abbreviations.....  | 9          |
| <b>2. Theoretical background and literature review</b> .....   | <b>11</b>  |
| 2.1. <i>Introduction</i> .....   | 11         |
| 2.2. <i>Speech communication and profound hearing impairment</i> .....   | 12         |
| 2.2.1. Language skills of hearing-impaired children.....   | 13         |
| 2.2.2. Classifications of hearing impairments .....  | 15         |
| 2.2.3. Speech acquisition and profound hearing impairment.....   | 20         |
| 2.2.4. The perception of speech through vision .....   | 26         |
| 2.2.5. Age of onset and speech quality .....   | 28         |
| 2.2.6. Speech quality of profoundly hearing-impaired children.....   | 29         |
| 2.2.7. Prerequisites for severely and profoundly hearing-impaired children to<br>develop intelligible speech.....            | 30         |
| 2.3. <i>Need, aim, and possibilities of speech therapy</i> .....   | 32         |
| 2.3.1. Moderate hearing-impaired children.....   | 32         |
| 2.3.2. Prelingually severe and profoundly hearing-impaired children .....  | 33         |
| 2.3.3. Adult second language learners .....  | 35         |
| 2.3.4. Speech therapy aims .....   | 35         |
| 2.3.5. General steps in an individual therapy .....  | 36         |
| <b>3. Development of speech technology in speech therapy for severely and profoundly<br/>hearing-impaired children</b> ..... | <b>37</b>  |
| 3.1. <i>Introduction</i> .....   | 37         |
| 3.2. <i>Process and product-oriented therapy systems</i> .....   | 38         |
| 3.3. <i>Tactile aids</i> .....   | 38         |

|           |   |           |
|-----------|---|-----------|
| 3.4.      | <i>Product-oriented visual aids</i> .....   | 40        |
| 3.4.1.    | Feature based visual indicators .....   | 41        |
| 3.4.2.    | Vowel and fricative displays.....   | 42        |
| 3.4.3.    | Formant displays.....   | 43        |
| 3.4.4.    | Fundamental Frequency displays.....   | 44        |
| 3.4.5.    | Computer-based speech therapy systems with visual feedback.....   | 44        |
| 3.5.      | <i>Process-oriented therapy systems with visual feedback</i> .....  | 48        |
| 3.5.1.    | Physiological devices.....  | 50        |
| 3.5.2.    | Automatic speech tutors.....  | 51        |
| 3.6.      | <i>Spoken language training for L2 speakers</i> .....   | 53        |
| <b>4.</b> | <b>Residual hearing for speech processing - methods of investigating the functional hearing for speech</b> .....  | <b>55</b> |
| 4.1.      | <i>Introduction</i> .....   | 55        |
| 4.2.      | <i>Functional hearing</i> .....   | 56        |
| 4.3.      | <i>Functional hearing and speech intelligibility</i> .....  | 57        |
| 4.3.1.    | Subjects.....   | 57        |
| 4.3.2.    | Intelligibility test.....   | 58        |
| 4.3.3.    | Listeners.....  | 58        |
| 4.4.      | <i>Results</i> .....  | 59        |
| 4.4.1.    | Speech intelligibility scores of 11 profoundly hearing-impaired children.....                                     | 59        |
| 4.4.2.    | The effect of listeners' experience .....   | 60        |
| 4.4.3.    | Relation between amount of residual hearing and speech intelligibility....  | 61        |
| 4.4.4.    | Relation between shape of audiogram and speech intelligibility .....  | 62        |
| 4.4.5.    | Relation between functional hearing and speech intelligibility .....  | 63        |
| 4.5.      | <i>Functional hearing and speech perception tests</i> .....   | 66        |
| 4.5.1.    | Introduction.....   | 66        |
| 4.5.2.    | Decisive factors for speech tests with small children .....   | 67        |
| 4.5.3.    | Test construction.....  | 68        |
| 4.5.4.    | Preliminary results .....   | 70        |
| 4.5.5.    | Conclusion .....  | 72        |
| <b>5.</b> | <b>Effects of speech input limitations on speech production</b> .....   | <b>75</b> |
| 5.1.      | <i>Deviations in the speech of moderate hearing-impaired children</i> .....                                       | 75        |
| 5.2.      | <i>Deviations in the speech of profoundly hearing-impaired children</i> .....                                     | 76        |
| 5.2.1.    | Factors that cause deviations in the speech of profoundly hearing-impaired children .....                         | 78        |
| 5.3.      | <i>Deviations in the speech of L2 learners</i> .....  | 83        |
| <b>6.</b> | <b>Interaction between individual deviations and speech intelligibility</b> .....                                 | <b>85</b> |
| 6.1.      | <i>Introduction</i> .....   | 85        |
| 6.2.      | <i>The effects of individual deviations on speech intelligibility measured by means of synthetic speech</i> ..... | 86        |
| 6.2.1.    | Introduction.....   | 86        |

|  |            |
|--|------------|
| 6.2.2. Assessment of the speech of three profoundly hearing-impaired children  | 87         |
| 6.2.3. Listening test.....   | 91         |
| 6.2.4. Results.....  | 92         |
| 6.2.5. Conclusions.....  | 95         |
| <b>7. Phonetic realizations of phonological systems .....</b>  | <b>97</b>  |
| 7.1. <i>Introduction</i> .....   | 97         |
| 7.2. <i>Speech assessment methods</i> .....  | 98         |
| 7.2.1. Phonetic error analysis.....  | 98         |
| 7.2.2. Phonological analysis .....   | 99         |
| 7.3. <i>Description of a phonological analysis of a profoundly hearing-impaired child's consonant production</i> ..... | 100        |
| 7.3.1. Step 1: Analysis of the existing articulation skills .....  | 101        |
| 7.3.2. Step 2: Assessment of the usage of the existing articulation skills through a detailed phonetic analysis .....  | 102        |
| 7.3.3. Step 3: Assessment of idiosyncratic realisations of phonological contrasts and regular error patterns .....     | 104        |
| 7.4. <i>The importance of a detailed phonetic transcription</i> .....  | 105        |
| 7.5. <i>Therapy based on existing skills</i> .....   | 106        |
| 7.6. <i>Description of a phonological analysis of a Bosnian speaker's production of Swedish vowels</i> .....           | 112        |
| 7.7. <i>Conclusions</i> .....  | 115        |
| <b>8. Design of visual feedback in Swedish computer-based therapy systems.....</b>                                     | <b>117</b> |
| 8.1. <i>Introduction</i> .....   | 117        |
| 8.2. <i>Visual feedback in speech therapy</i> .....  | 117        |
| 8.2.1. Nature of feedback.....   | 118        |
| 8.2.2. Type of feedback .....  | 118        |
| 8.3. <i>The IBM SpeechViewer</i> .....   | 120        |
| 8.3.1. Description of the system.....  | 120        |
| 8.3.2. Type of visual feedback in different exercises.....   | 121        |
| 8.4. <i>Box of Tricks</i> .....  | 126        |
| 8.4.1. Description of the system.....  | 126        |
| 8.4.2. Databases for reference speech and normal-hearing children for similarity comparisons .....                     | 128        |
| 8.4.3. Training method used in Box of Tricks .....   | 131        |
| 8.4.4. Type of visual feedback in different exercises.....   | 131        |
| 8.5. <i>OLP (Ortho-Logo-Paedia) Therapy</i> .....  | 137        |
| 8.5.1. Introduction.....   | 137        |
| 8.6. <i>System components</i> .....  | 139        |
| 8.6.1. The user interface OLPy .....   | 139        |
| 8.6.2. OPTACIA .....   | 141        |
| 8.6.3. GRIFOS .....  | 145        |

|   |            |
|---|------------|
| 8.7. <i>Conclusions</i> .....   | 146        |
| <b>9. Clinical evaluation studies of the three systems</b> .....  | <b>149</b> |
| 9.1. <i>Introduction</i> .....  | 149        |
| 9.2. <i>Clinical evaluation of SpeechViewer with profoundly hearing-impaired children</i> .....               | 149        |
| 9.3. <i>Clinical evaluation of SpeechViewer with L2 learners</i> .....  | 157        |
| 9.3.1. <i>Diagnosis of individual deviations</i> .....  | 160        |
| 9.3.2. <i>Instructions that aimed at awareness, correct realisation, and understanding</i> .....              | 164        |
| 9.3.3. <i>Further training to establish automaticity and transfer to untrained material.</i> .....            | 165        |
| 9.3.4. <i>Results of a questionnaire with the thirteen L2 speakers</i> .....                                  | 165        |
| 9.4. <i>Evaluation studies with Box of Tricks</i> .....   | 167        |
| 9.4.1. <i>Results of the Swedish questionnaire</i> .....  | 168        |
| 9.4.2. <i>Results of the Slovenian questionnaire</i> .....  | 169        |
| 9.4.3. <i>Results of the Hungarian questionnaire</i> .....  | 170        |
| 9.4.4. <i>Clinical evaluation of the Hungarian version of Box of Tricks</i> .....                             | 171        |
| 9.4.5. <i>Clinical evaluation of the Slovenian version of Box of Tricks</i> .....                             | 173        |
| 9.4.6. <i>Clinical evaluation of the Swedish version of Box of Tricks</i> .....                               | 175        |
| 9.5. <i>Clinical evaluation of the Swedish version of the OLP method with hearing-impaired children</i> ..... | 176        |
| 9.5.1. <i>Introduction</i> .....  | 176        |
| 9.5.2. <i>Method</i> .....  | 176        |
| 9.5.3. <i>Subjects</i> .....  | 177        |
| 9.5.4. <i>Treatment</i> .....   | 178        |
| 9.5.5. <i>Therapy objectives</i> .....  | 178        |
| 9.5.6. <i>Assessments</i> .....   | 181        |
| 9.5.7. <i>Results</i> .....   | 183        |
| 9.6. <i>Conclusions</i> .....   | 190        |
| <b>10. General Conclusions and Recommendations</b> .....  | <b>191</b> |
| 10.1. <i>Recommendations</i> .....  | 191        |
| 10.1.1. <i>Important demands on a visual computer-based speech therapy system</i> .....                       | 193        |
| 10.1.2. <i>Efficiency of visual feedback of prosodic parameters within spoken L2 training</i> .....           | 194        |
| 10.2. <i>Comparison of the three systems</i> .....  | 195        |
| 10.3. <i>Recommended therapy design</i> .....   | 197        |
| 10.3.1. <i>General design of computer-based speech therapy with visual feedback</i> .....                     | 197        |
| 10.3.2. <i>Structural training design</i> .....   | 198        |
| 10.4. <i>Visual feedback strategies</i> .....   | 200        |
| 10.4.1. <i>Type of visual feedback for severely and profoundly hearing-impaired children</i> .....            | 200        |
| 10.4.2. <i>Type of audio-visual feedback for L2 learners</i> .....  | 202        |

|   |            |
|---|------------|
| 10.4.3. Use of automatic speech recognition and spectral comparison of phonemes with profoundly hearing-impaired children ..... | 203        |
| 10.5 <i>Conclusions</i> .....   | 204        |
| <b>11. References</b> .....   | <b>209</b> |
| <b>12. Appendices</b> .....   | <b>227</b> |
| 12.1. <i>Appendix 1: Diacritics to assess the speech of profoundly hearing-impaired children</i> .....                          | 227        |
| 12.2. <i>Appendix 2: Swedish SAMPA symbols</i> .....  | 230        |
| 12.3. <i>Appendix 3: Swedish questionnaire for evaluation of Box of Tricks</i> .....  | 232        |



## **1. Introduction**

Technically advanced computer-based speech therapy systems with visual feedback are currently available for use as a complement to traditional methods to assist children with speech and hearing-impairments in perceiving and producing speech. Visual feedback has shown to be an efficient substitute for the missing auditory feedback in profoundly hearing-impaired children (Osberger et al., 1981; Watson, Reed, Kewley-Port, Maki, 1989; Arends et al., 1991; Yamada & Murata, 1991; Levitt, 1993; Javkin, 1994; Rooney et al., 1994; Öster 1996, 1999a; Öster et al. 1999b, 2002, 2003). It has also been shown to be a valuable supplement to audio and verbal feedback in speech training for moderately and severely hearing-impaired, normally hearing children with speech deviations as well as for second language (L2) learners.

Generally, in spoken language training with L2 learners, visual feedback is seldom used. Training is frequently carried out through traditional record-and-play-back models (Hincks, 2005) that provide audio feedback. According to Neri et al. (2002, pp. 443) there is "...a general tendency to neglect pronunciation in favour of grammar and vocabulary in research on second language acquisition." Speech-interactive and self-assessed language learning systems focus on the implementation of automatic speech recognition (ASR) providing verbal feedback (Eskenazi, 1999). ASR automatically detects segmental errors fairly well but has problems identifying prosodic errors such as deviantly produced intonation, stress patterns and fluency. However, it has been shown that the use of visual feedback has helped L2 learners in perceiving and producing especially new stress patterns and intonation of a target language (Flege, 1989; Öster 1998), a fact that supports a wider usage of visual feedback of prosodic parameters in spoken language training within second language learning.

In a computer-based speech therapy aid, visual feedback of acoustical parameters is used to give the client possibilities to evaluate his/her own speech compared to a target of accepted production. By using different amusing drawings to illustrate loudness, pitch contour, spectral distribution, etc, the child's attention can be drawn to important parameters in the speech which

can help the child to recognise whether his/her pronunciation is improving or not. Such an advanced technical aid can assist the teacher and help the client follow three very important steps in speech learning: instruction, training, and generalization. Especially in the instruction phase of speech therapy, immediate and meaningful visual feedback can help a child to be aware of the manner and place of articulation as well as of distinctive contrasts between similar speech sounds. Many distinctive features such as, for example, nasality/non nasality, are invisible through lip-reading and consequently difficult to produce correctly. By this technique it is also easier for the therapist to instruct and explain what is wrong and what is correct in the child's production. Motor learning theory in speech development also indicates that accurate feedback and repeated practice are essential to establish automaticity and linguistic use. This is the most important element in a speech therapy program but the most difficult for a therapist to carry out. The target production must be repeated and practiced in a variety of contexts. Using computer-based speech therapy with visual feedback in this situation is particularly helpful and motivating in helping the child to carry out significant amounts of additional training.

However, despite highly developed technical systems that offer flexible training the result of computer-based speech therapy (CBST), especially regarding profoundly hearing-impaired children, is not always as successful and promising as could be expected (Bench, 1992). Reasons could be that the therapy is not always based on a clinical-phonetic point of view, is not adjusted to the individual child's needs and that it doesn't pay attention to how speech sounds are realised in spoken language. The importance of such an individualised assessment was stressed by, for example, Stoel-Gammon and Dunn (1985). He stated that it is necessary to identify the unique characteristics of each child's system in order to design the most appropriate treatment plan for each child. Tanner Dyson (1988) as well as Saben and Ingham (1991) claimed that it is important to concentrate on the usage of a child's existing articulation skills, that is, the speech sounds that the child knows how to produce correctly in isolation without knowledge of their meaning and linguistic use, to identify the deviations made by the child. This is particularly important when it comes to severely and

profoundly hearing-impaired children's speech as it contains many idiosyncratic realisations of phonological contrasts. Traditionally, the speech of profoundly hearing-impaired children is assessed using a phonetic error analysis only. The articulation skill is then compared sound by sound with that of normally hearing adult speakers (Tanner Dyson, 1988) without regard to phonetic contexts or to the production's contrastive function in a specific language. Such an analysis uses a coarse phonetic transcription missing important articulatory details and provides only information on what the child is not capable of articulating. Hence, only speech sounds that the child never articulates correctly will be treated.

If instead a phonologic assessment of a very detailed phonetic analysis is used, the child's individual system can be described by identifying all segments used by the child with the help of IPA symbols and special diacritics. In this way answers could be provided to questions such as: How successfully can the child produce speech contrasts? Might a deviant production in fact be a realisation of a phonological contrast? In what ways do the phonetic elements used for contrastive function differ from those used in normal speech? Hence, basing an individual speech therapy program on a phonologically based assessment that investigates a child's unique phonological system (phonetic and phonemic inventories) as well as the linguistic use of the existing articulation skills might contribute improving, the results from computer-based speech therapy (Öster 1991, 1992a, 1992b). With the help of well understandable visual feedback strategies the child's attention can be drawn to deviated realized phonetic features and correct pronunciation can be trained with specially designed encouraging visual feedback in order to improve the intelligibility. Perhaps then successful training might come about with the help of an advanced technical computer-based speech therapy system with specially designed visual feedback.

Another problem is that the existing computer-based speech therapy systems developed for the Swedish language use different kinds of visualisations of important speech parameters, correct from the acoustic-phonetic point of view but not always useful and understandable for all kinds of users like children, adults, clients with a hearing loss, normally-hearing clients with speech defects,

and second language learners. Furthermore fundamental principles, reference speech, target spectra, methods for contrastive training, and objective evaluation techniques differ between the systems and are not appropriate to be used by all client groups. Recommendations and evaluations of what type of visual feedback strategy and therapy structure that should be appropriate to use with different user groups are seldom made. Hence a further explanation to the sometimes limited and varied progress reported from computer-based speech therapy/training could be the neglect of information about what kinds of visualisations and therapy structures would be most viable to use with individual client groups.

## **1.1. Speech communication difficulties caused by speech perception problems**

### **1.1.1. Hearing impairment**

Auditory feedback is necessary for the speech development of normally-hearing children. The child can hear his own production, compare his own production with that of others and correct his production little by little. Children with a moderate hearing loss develop spontaneous speech, but their pronunciation often suffers from distortions and lack of articulatory precision. The degree of intelligibility of their speech production is related to their hearing impairment. More impaired articulation accompanies higher hearing loss (Boothroyd, 1984; Levitt & Geffner, 1987; Ling, 1976; Smith, 1975; Markides, 1985). Considering hearing loss as a single filter (not compensating for other disordered perceptual processes), deviant articulation may be predicted from the pure tone audiogram. However, other facts have to be taken into consideration when predicting the intelligibility of the speech of moderately hearing-impaired children. The children may have been hearing-impaired at different ages, been exposed to different educational methods, have had different social and psychological backgrounds and have had different use of hearing aids.

However, for profoundly hearing-impaired children, born with hearing losses above 90 dBHL, the degree of correlation is reduced and their intelligibility can not be predicted from the pure

tone audiogram (Monsen, 1978). These children have no acoustic speech target to imitate and compare their own production with when learning speech. Therefore, they develop no spontaneous speech but must learn speech through vision, tactile sensation and, if possible, residual hearing. Other senses must replace the auditory feedback that hearing children use when they learn to speak. By relying on vision and the limited visibility of phonetic features in learning oral speech, they can establish an orosensory-motor control of their speech movements and their acoustic output. The limited speech perception potentials of these children cause them to make unavoidable phonetic deviations, which affect their ability to signal meaning differences in spoken language. The speech that these children develop varies from poor to rather good, showing typical and systematic deviations as well as unusual and non-standard features owing to the fact that their phonological systems are built up by visually detectable distinctions between different phonemes. Hence, a phonological system acquired by “hearing by eye” will differ from the norm both on the phonetic and phonological level (Dodd, 1988). The aim of speech therapy for severely and profoundly hearing-impaired children is that the children should be helped to acquire new speech patterns and develop intelligible speech for the purpose of making statements and more seldom for the purpose of communication. They should learn a “survival speech” that makes it possible for them to give and to understand simple messages in shops, in the street, in a hearing society, etc. For moderately hearing-impaired children, who rely on auditory information when learning speech, therapy is more directed towards speech correction than speech acquisition.

### **1.1.2. Second language learners**

Adults learning Swedish as a second language are another group that have difficulties in perceiving the phonetics and prosody of a second language through audition, not because of a hearing loss but because they are not able to hear new sound contrasts because of interference with their native language. Ongoing research has pointed out that this group fails to perceive difficult and new contrasts in the target language by ear alone. (Jamieson, 1995; McAllister, 1995; Flege, 1998). Age of first exposure, motivation, first

language (L1), amount of use of L1 and L2. etc. are some of the factors that affect L2 speakers' pronunciation (Piske, MacKay & Flege, 2001; Neri, Cucchiaroni, Strik & Boves, 2002). Typical pronunciation difficulties for a given target language will differ for speakers of different native languages (Dalby & Kewley-Port, 1999). Sounds that are similar but not identical cause the most serious problems both in perceiving and producing them. Flege's Speech Learning Model (SLM) is stated as follows....." L2 features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner's production of the contrast based on this feature" (McAllister et al., 2002, pp. 230). L2 learners make pronunciation errors of two types (Eskenazi, 1999). They articulate some target phonemes deviantly because they differ in number and quality from those of their mother tongue (L1) and they use prosodic parameters such as intonation, duration, and stress placement inappropriately. Speech therapy for L2 speakers aims at eliminating misarticulated speech patterns to achieve a native-like production and to improve the consistency of production.

## **1.2. Objectives of the study**

This study investigates some conceivable reasons to the sometimes limited and varied progress reported from computer-based speech therapy/training with visual feedback. Hypotheses are presented with reference to diagnostic methods, visual feedback strategies and therapy design of two client groups, namely children with severe and profound hearing losses and second language learners.

The first aim is to study the importance of an individual phonological assessment as a diagnosis to base the training on and to make the best use of visual feedback. Such an assessment investigates the linguistic usage of the client and assesses in what way the used phonetic elements differ from the phonetic elements used for contrasts in normal speech. The hypothesis is that without a phonological assessment based on a very detailed phonetic analysis it could be that therapy only results in a series of meaningless "articulation gymnastics" sessions.

The second aim is to investigate and give some recommendations of type and design of visual feedback for these

client groups that have different speech processing capabilities and rely on different senses for speech input/output. Furthermore, the intention is to give recommendations of therapy design for these client groups derived from their different aims and needs of therapy.

The third aim is to test the efficiency of visual feedback of prosodic parameters within spoken L2 training to investigate whether it is worthwhile to be used as a general tool in computer-assisted pronunciation training (CAPT) within second language learning.

### **1.3. Thesis overview**

This monograph comprises 10 chapters. Chapter 1 is the introduction, which addresses the goals of the study and reports on the speech communication difficulties of profoundly hearing-impaired children and second language learners caused by speech perception problems.

Chapter 2 contains a review of the theoretical background and literature review of work done by others on different aspects of speech communication and profound hearing impairment, such as speech acquisition, speech quality, and different classifications of hearing impairment, need, aim and possibilities of speech therapy for hearing-impaired children as well as for adult second language learners.

Chapter 3 presents the development of speech technology in speech therapy from early visual aids to more technically advanced computer-based systems using visual feedback.

Chapter 4 deals with the relation between residual hearing capabilities for speech (the functional hearing) and intelligibility of the speech of profoundly hearing-impaired children. Results of a speech intelligibility test of 11 children are reported that investigate the relationship between the children's speech intelligibility and different factors such as mean hearing loss, shape of audiograms, and functional hearing. An analytical computerized speech perception test for small children, with the effort to develop a screening method that can predict a child's ability to develop intelligible speech, is also described. The result of the test also provides an early diagnosis of difficulties to base the therapy on.

The following Chapter 5 reports on general deviations in the speech of profoundly hearing-impaired children and adult second language learners and discusses the causes why they occur.

Chapter 6 reports on a study that investigated how individual deviations interact on overall speech intelligibility. This was measured through generated synthetic speech made up from three profoundly hearing-impaired children's individual vowel, consonant and prosodic deviations that were artificially corrected towards normal speech through digital manipulation. The effects of the corrections of the different deviations on intelligibility were measured through listening tests that provided important results to be used in speech therapy.

Chapter 7 emphasizes the importance of phonologically assessing in what way profoundly hearing-impaired children's phonetic realizations of Swedish phonology differ from the normal model and in what way they, as well as adult second language learners, realize their existing articulation skills in linguistic contexts. The results of such an analysis are necessary for the planning and efficiency of an individually performed computer-based speech therapy. The description of such a phonological assessment applied to the speech production of both a profoundly hearing-impaired child and a Bosnian speaker is presented.

In the following Chapter 8 the design of the different types of visual feedback used in three Swedish computer-based speech therapy systems are exemplified, compared, and examined in relation to their potentials to be effective, attractive, easy to comprehend, user-friendly, motivational, instructive, and logical to the user.

Different clinical evaluation studies made with the three systems are reported in Chapter 9.

Finally, in Chapter 10, general conclusions of the efficiency of computer-based speech therapy using visual feedback for profoundly hearing-impaired children and audiovisual feedback for training of perception and production of the Swedish prosody within spoken L2 training are discussed and recommendations regarding therapy design and visual feedback strategies for the two groups and for the different stages during therapy are specified.



## **1.4. Some basic concepts**

### **1.4.1. The term deaf**

The term deaf in Sweden covers a number of various groups of persons, differing in etiology, degree of hearing loss and articulation skills. To be identified as either deaf or profoundly hearing-impaired is nowadays more social than biological. Deaf persons are dependent on sign language and use it in their everyday communication. Those who prefer to communicate through sign language belong to the deaf community irrespective of how great the hearing loss is (Larsson, 1997) in contrast to those who have a preference to speak, lip-read other people and use technical aids. The latter group is considered to be hearing-impaired. However, nowadays many hearing-impaired children learn sign language and are thanks to that bilingual. In many studies dealing with differences between the speech of students who are deaf and of those who are hard-of-hearing, especially outside Sweden "hard-of-hearing" typically describes children with mild and moderate losses and "deaf" usually refers to children with severe and profound hearing loss (Yoshinaga-Itano and Sedey, 2000).

Here the term profoundly hearing-impaired is used instead of deaf all through the thesis and refers to the Swedish children who join the speech clinic because they are motivated and have possibilities to develop an intelligible "survival speech" for the purpose of making statements and, more seldom, for the purpose of communication. The aim of speech therapy in Sweden is to make it possible for them to give and to understand simple messages in shops, in the street, etc. However, the term deaf is used in the literature review and in Chapter 9 by the Hungarian and Slovenian partners in their evaluation studies of Box of Tricks.

### **1.4.2. Abbreviations**

|             |   |
|-------------|---|
| <b>ASR</b>  | Automatic speech recognition  |
| <b>CALL</b> | Computer-assisted language learning, for teaching and training a second language. |

*Computer-Based Speech Therapy Using Visual Feedback*

|              |  |
|--------------|--|
| <b>CAPT</b>  | Computer-assisted pronunciation training, for teaching and training the pronunciation of a second language.  |
| <b>CBST</b>  | Computer-based speech therapy for teaching and training speech production of children with speech- and hearing impairments.  |
| <b>dBHL</b>  | Hearing loss is measured in decibels hearing level   |
| <b>L1</b>    | Mother tongue  |
| <b>L2</b>    | Second language  |
| <b>MRI</b>   | Magnetic resonance imaging   |
| <b>PTA</b>   | Pure-tone averages   |
| <b>SAMPA</b> | Speech Assessment Methods Phonetic Alphabet is a machine-readable phonetic alphabet. It was originally developed under the ESPRIT project 1541, <a href="#">SAM</a> (Speech Assessment Methods) in 1987-89, see Appendix 2 |
| <b>(sj)</b>  | Example of the orthographic representation of a speech sound   |
| <b>/ʃ/</b>   | Example of the IPA transcription of a phoneme  |
| <b>[ʃ]</b>   | Example of the IPA transcription of the articulatory aspect of a speech sound.   |

## ***2. Theoretical background and literature review***

### **2.1. Introduction**

The speech of severely and profoundly hearing-impaired children exhibits wide variations in degree of intelligibility and should be looked upon as a special idiolect that is dependent on speech reception capabilities and on other factors that these children rely upon when learning to speak. Profoundly hearing-impaired speakers often try to realize the visual representation of a phonetic contrast, which is normally signaled auditorily (Monsen, 1976). This means that a deviant production, for instance a silent articulation for /f/ with absence of air-stream but articulation present or a stop with a non-audible release may be an attempt to realize a speech sound.

Comparing the articulatory skill of profoundly hearing-impaired children with that of normal-hearing children is inappropriate as the quality of the articulation is relatively unimportant as long as the speech is intelligible. The intelligibility of the speech of profoundly hearing-impaired will depend on to what extent the phonological system and phonetic realization of the system resembles the norm of the language users in general. This can only be assessed through a phonological analysis that concentrates upon regularities in the pronunciation used in spoken language and the description of these rules. A phonological description looks for underlying rules in a child's speech, and specifies any idiosyncratic realization of phonological contrasts through an analysis of a detailed phonetic transcription.

This chapter reviews earlier studies of profoundly hearing-impaired children's speech production. The aim is to give a background to the statement that the speech of severely and profoundly hearing-impaired children is unique and to provide necessary basic issues of why an objective phonological analysis of their speech is required. The intention is also to give a background to the fact that a special visual feedback strategy and therapy design must be used during speech therapy with severely and profoundly

hearing-impaired children in contrast to what should be the most useful visual strategy in spoken language training with L2 learners.

## **2.2. Speech communication and profound hearing impairment**

Language can be written, spoken, signed and even symbolized depending on people's needs and capacities to communicate. People with severe motor disabilities use Augmentative and Alternative Communication through different kinds of symbols for social interaction and education, like the Morse code, Bliss symbols and other picture systems. Communicating through written language involves both reading and writing competences, but communicating successfully with speech requires good hearing to be able to perceive, understand and produce speech. Deaf persons who are deficient in hearing communicate through Sign language, a visual, gestural, and spatial manner that they use in their everyday communication.

Children who are born with a severe hearing loss have of course a decreased ability to learn speech through an auditory-verbal approach compared to normally hearing children. The limited auditory feedback affects not only profoundly hearing-impaired children's speech perception potentials but reduces also the opportunity of self-correction and a comparison of their own speech quality with those closest to them.

However, a hearing loss affects also certain other important factors of both the listener and the speaker that are necessary for a satisfying speech communication to be possible according to Finnerty (1996) and Loughlin (2005). These factors are:

- Auditory feedback
- Physiological control
- Knowledge of the speech code
- Identical references
- Interpretation
- Rich experience (ability of guessing)
- Awareness (interpret sounds as meaningful)
- Non-verbal skills (such as pointing, showing, gazing, head nod/shake, facial expression, eye contact, crying, imitation and vocalising)

Profoundly hearing-impaired children often suffer from disorders of physiological control like deviant respiratory patterns as well as breathiness, voice breaks, unstable pitch, nasality and vocal fry. This is due to a restricted use of the vocal apparatus and the fact that they have to learn oral language by a laborious visual imitation of speaking (Grewel, 1963). Non-visible speech elements like for instance, nasality, voicing, and fundamental frequency can be related to typical deviations in the speech of profoundly hearing-impaired children (Martony, 1971; Monsen, 1976; Öster 1992b).

The knowledge of the speech code, i.e. the knowledge of how linguistically defined units are realised in the speech act (Fant, 2001), and the possibility of identifying speech segments is of significant importance for speech communication. Severely and profoundly hearing-impaired children have incomplete knowledge of the sound system of the target language and have difficulties in interpreting sounds as meaningful, understanding the contents of a message, and using the target language in different situations. Furthermore, the speaker and the listener must have identical references to be able to understand each other.

All these important aspects of speech communication must be learnt with the help of a therapist mainly through the visual and kinaesthetic channels. It is a matter of fact that a rich linguistic environment and regular exposure to speech lead to a rich experience of the language used. This implies better possibilities of guessing when the listening condition is inopportune. According to Furth (1964) profoundly hearing-impaired children possess normal cognitive processes but suffer from general restrictions in experience, interactions, and opportunities to learn spoken language.

### **2.2.1. Language skills of hearing-impaired children**

A determining condition for a successful speech communication to take place is to know and to master the structure of the phonology, the morphology and the syntax. Limited research has been done on the language skills of hearing-impaired children or on the underlying processes (Bamford and Saunders, 1991). Normal language development has been used as a model to tell what is wrong in the language of hearing-impaired children. Longitudinal development

studies have been rare. Unfortunately, most published international studies contain data from subjects representing a whole range of hearing losses from mild to profound in degree. Background data that may affect language acquisition has not always been taken into consideration. The children may have been hearing-impaired at different ages, been exposed to different educational methods, have had different social and psychological backgrounds and have had different use of hearing aids. However, the language development of children with mild hearing losses often does not obviously differ from that of normal hearing children in its main features, but it may be delayed. A broad summary based on different studies of how the language of prelingually hearing-impaired children might differ from that of normally hearing children is given by Bamford and Saunders (1991). It was reported that hearing-impaired children use:

- reduced numbers of words (tokens)
- reduced numbers of different words (types)
- reduced numbers of grammatical correct sentences
- shorter sentences
- simpler and generally stereotyped sentence constructions  
more content words, particularly nouns and verbs
- reduced numbers of conjunctions and prepositions
- frequent errors of word orders

Borg et al. (2002) established a reference material for a language test for hearing-impaired children, "LATHIC". The test is made for children in the age range 4 - 6 years with pure tone averages of 80 dBHL or less and with spoken Swedish as the first language. The LAHTIC test battery consists of eight subtests: mental development, speech reception in noise, segmental phoneme discrimination, phonological short-term memory, phrasal prosody, phonology, speech motor functions and phoneme mobilisation. In addition the child's certainty and co-operation are judged in each test. The most important result of the analysed material was that children with hearing impairment greater than 60 dB (also unilateral) had a delayed language development compared to normal hearing children of the same age, the delay increased with larger losses, the delay decreased for older children and no differences were found between boys and girls. The total number of tested children was

almost 400, including 87 normal-hearing children who participated as a control group. 199 of the hearing-impaired children completed all tests and constituted the analysed material.

### 2.2.2. Classifications of hearing impairments

Pure tone threshold audiometry, using air and bone conduction, is a measurement of the hearing loss a child suffers from (Lidén, 1985). It measures the auditory loss at several frequencies by pure tones, generally at 125 Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 6000 Hz. Each sound frequency is sent out in the ear with the help of a headset and is gradually increased in intensity until the child can hear it. The result of pure tone audiometry indicates the dividing line between hearing and not hearing and gives a confirmation of a child's hearing sensitivity, see Figure 2.1.

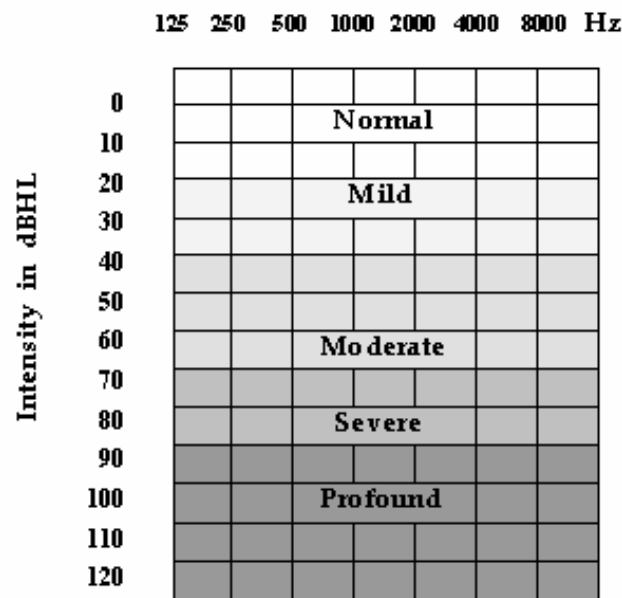


Figure 2.1 Pure tone audiogram showing a person's threshold to steady-state pure tones.

Pure tone average, PTA is a method used to calculate the results and presents a child's hearing sensitivity and hearing handicap in a practical way. It is given by averaging the air conduction thresholds

for the frequencies 500, 1000, and 2000 Hz which are considered to be the major and important frequencies for speech.

Various classifications and degrees of hearing impairment have been presented over the years. Traditionally, a classification of a hearing loss was based on how the impairment affects the ability to understand speech (Lidén, 1985). Stach (1998) illustrated the communicative effect of a hearing loss through a classification, shown in Table 2.1, based on the mean of the frequencies 500, 1000, and 2000 Hz.

Table 2.1. *Classification of degree of hearing loss and impact on communication, (Stach, B. A., 1998.)*

| <b>dBHL</b> | <b>Degree</b>     | <b>Communicative Effect</b>                               |
|-------------|-------------------|---|
| 10 to 10    | Normal            | None  |
| 10 to 25    | Minimal           | Difficulty hearing quiet speech in presence of noise      |
| 25 to 40    | Mild              | Difficulty hearing quiet or distant speech, even in quiet |
| 40 to 55    | Moderate          | Conversational speech is audible if at close distance     |
| 55 to 70    | Moderately-severe | Loud conversational speech is audible                     |
| 70 to 90    | Severe            | Conversational speech is not audible                      |
| >90         | Profound          | Loud sounds may be audible                                |

Important information about a child's possibilities to perceive the information-bearing elements in the acoustic speech signal can be provided when the child's pure tone audiogram is studied in relation to the speech banana, shown in Figure 2.2. The speech banana is taken from the measurements of Fant (1959). It is a section on the audiogram (with the shape of a banana) that covers the frequency area in which different speech sounds are produced. The limit between hearing and vibration in the low frequency range, the intensity of one's own speech and the threshold of discomfort is also indicated in Figure 2.2.



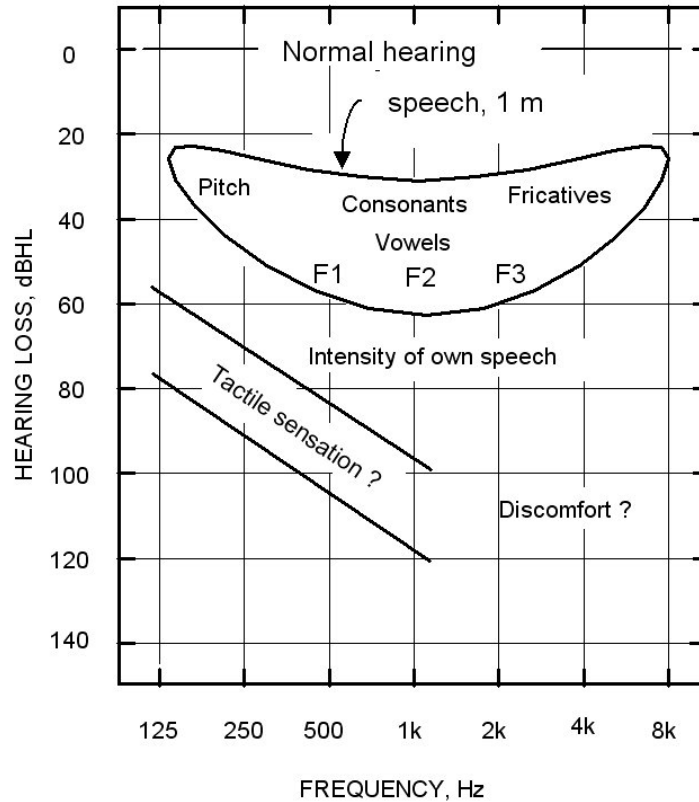


Figure 2.2. The speech banana showing the most important area for the perception of the speech sounds, Fant, 1959.

Martony and Risberg (1970) developed a method for classifying audiograms to be used when estimating hearing aid selection, methods of teaching, and results of speech training with different aids, see Figure 2.3. The method is based on degree of the hearing loss and the shape of the audiogram. By this method the audiograms are divided in a low frequency area from 125-1000 Hz and a high frequency area from 2000-8000 Hz. These two areas are then divided in 0-D for low frequencies and 0-6 for high frequencies. Consequently, by a combination of a letter and a digit, coarse information about the quality of the residual hearing and the shape of the pure tone audiogram is given.

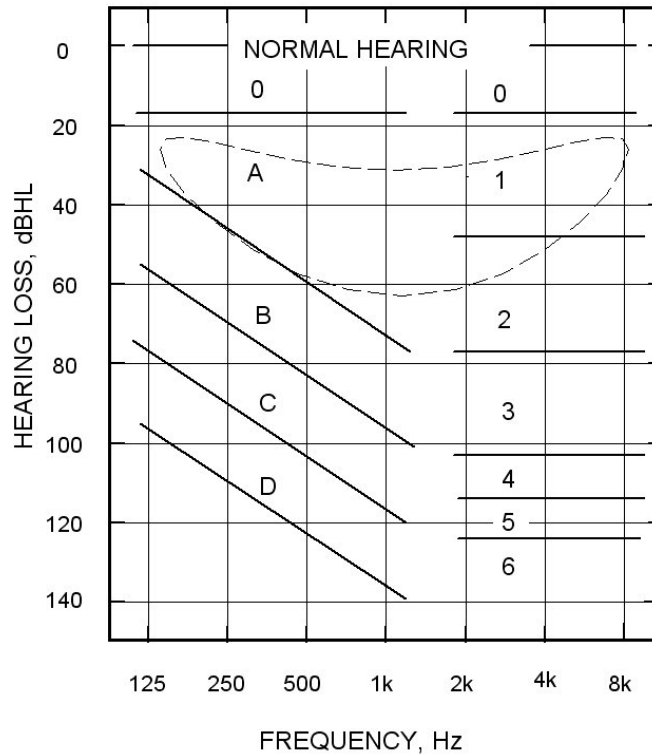


Figure 2.3. Audiogram classification according to Risberg & Martony (1970) that gives information about quality of the residual hearing and shape of the audiogram.

A classification may also be based on how the impairment affects the ability to develop speech (Risberg, 1979). It is based on the speech banana and the average thresholds (PTA) of the frequencies 500, 1000, and 2000 Hz (Table 2.2).

However, the sort of speech a hearing-impaired child develops depends not only on the amount of hearing, as measured by puretone threshold audiometry, but also on the quality of the hearing sensation and the use the child through training is able to make of his/her residual hearing. Several studies have shown that there is a close relationship between degree of hearing impairment and the intelligibility of hearing-impaired children's speech. More impaired speech accompanies higher hearing loss (Boothroyd, 1984; Levitt & Geffner, 1987; Ling, 1976; Smith, 1975; Markides, 1985).

Table 2.2. Classification of how a hearing loss affects the ability of developing speech, (Risberg, 1979).

| Threshold  | Ability to develop speech   |
|------------|---|
| < 39 dBHL  | Some delay and some deviations.   |
| 40-54 dBHL | Pronounced effect. Need for special education and hearing aids.   |
| 55-69 dBHL | Hearing aids and speech correction necessary.   |
| 70-89 dBHL | Delayed speech development. Need for speech learning and speech correction. Benefit from sign language.     |
| > 90 dBHL  | Sign language as first language. Speech development related to functional hearing. Reliance on visual cues. |

On the average, speech intelligibility decreases with increasing hearing loss up to a loss of about 90 dB. Above that the degree of correlation is reduced. However, Monsen (1978) pointed out that a good audiogram may correlate quite consistently with good speech; but, on the other hand, children with more profound hearing losses may commonly span the whole range from very intelligible to quite unintelligible speech. Hence, the intelligibility of the speech of children with pure-tone averages greater than 90 dB cannot be predicted from the degree of hearing loss, as measured by pure-tone audiometry. This was shown in a study by Adolvsson and Forsén (1968) where the intelligibility of hearing-impaired children's speech from the 8th grade of the Manilla- and Alviksschool in Stockholm was investigated, see Figure 2.4.

The intelligibility of the children's speech was rated as intelligible (3), intelligible without difficulties (2), intelligible with difficulties (1), and unintelligible (0). The figure shows that on the average, speech intelligibility decreases with increasing hearing loss (mean level at 500, 1000, and 2000 Hz) until a loss of about 90 dB. For hearing losses greater than 90 dBHL the degree of correlation is reduced. Vastly varying intelligibility scores from poor to rather good speech were shown for children with pure tone averages above 90 dBHL.

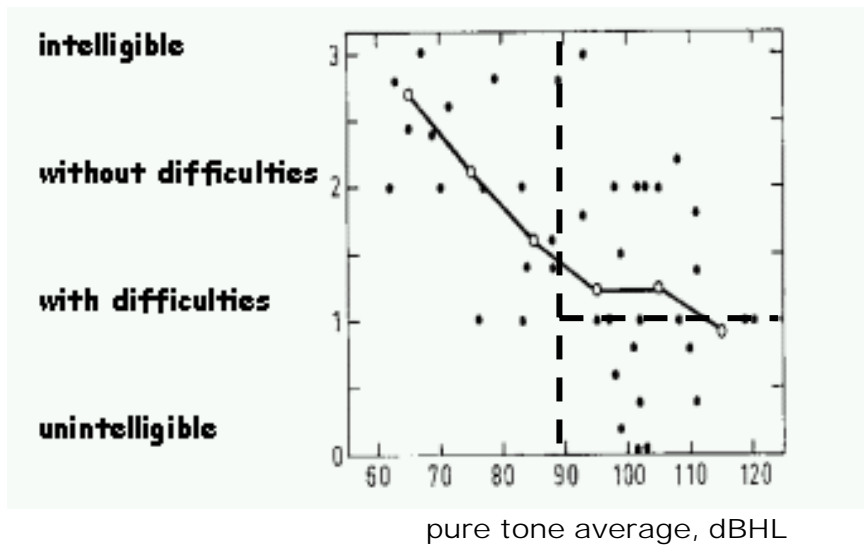


Figure 2.4. Rated intelligibility compared to hearing level of hearing-impaired children (Adolsson and Forsén, 1968).

Speech audiometry is an important qualitative complement to pure-tone audiometry. Recorded test words are transmitted to the patient with the help of a headset or a speaker. Speech audiometry can be global or analytical. The aim of a global examination is to determine the ability to understand speech in everyday use while an analytical examination determines the ability to perceive single segments and suprasegmentals (Arlinger & Hagerman, 1997).

### 2.2.3. Speech acquisition and profound hearing impairment

The hearing child perceives speech and language spontaneously through a combination of hearing and visual information, see Figure 2.5. Lip-reading is no longer seen only as a compensatory information channel to substitute for hearing. Mogford (1983), for example, has shown that visual speech information is implicated in normal processes of speech acquisition. The talking face is important from an early age in gaining and holding the attention of the listener. Wundt (1911) called the observation of the talking face "the impulse to speech". Young children observe lip-movements to find out how to articulate different speech sounds and then imitate these movements. Visual articulatory movements give concrete cues about

how to produce speech sounds, while hearing offers possibilities to control, compare and modify one's own production according to the surrounding pronunciation and linguistic usage. The hearing and sighted child will learn those sounds that have clearly visible articulation more quickly and with less errors than those with non-visible articulation. In time the hearing child is able to interpret what is being said through hearing alone, without having to watch the speaker. According to Mahshie (1996, pp. 153)...“they develop an internal mode of articulatory/phonatory outcomes or aims”.

In general the productions of sounds at the phonetic level, developed from comparison of others' and one's own articulatory patterns, precede their meaningful use at the phonological level (Ling, 1976). A phonetic description shows how a child articulates speech sounds and what speech sounds are present in his existing articulation skills without knowledge of their meaning and linguistic use. When the hearing child is becoming aware of the relationship between his own speech and those of others he gradually uses his existing articulation skills in a meaningful way at the phonological level. The auditory feedback allows a hearing child to differentiate between the sounds he makes, derive meaning from the sounds that others make, be aware of the extent to which his own speech corresponds with that of others and, what is most important, to continuously self-correct and improve his own articulation. This is illustrated in Figure 2.5 by the means of arrows that show how correct articulation is established.

Before sounds are used meaningfully (phonologically) children babble (reduplicated sequences of syllables) most of the sounds (phonetically) to reach an automatic level. Spontaneous vocalisation based on physiological conditions steadily increases and babbling becomes stabilised (Mavilya, 1972). Nearly all infants with normal hearing begin the canonical stage of babbling before 10 months of age (Roug et al, 1989). Late onset of canonical babbling may be a predictor of speech and language disorders. The phonetic repertoire will gradually and spontaneously turn into a meaningful use. Jakobson (1968) claimed that children learn contrasts, not individual sounds, in a certain order from maximal to minimal contrast.

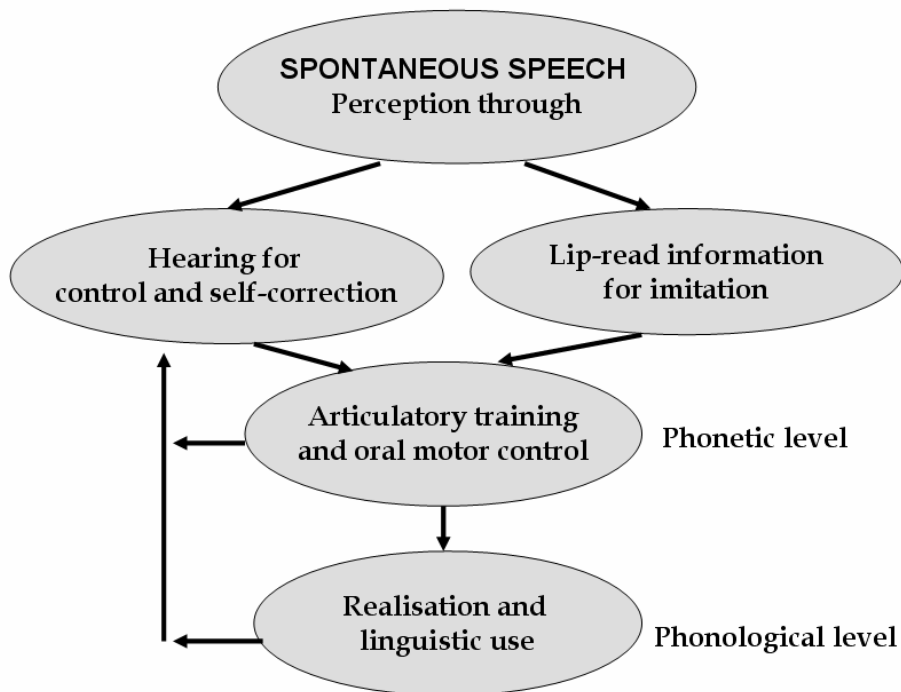


Figure 2.5. Normal-hearing children's acquisition of language and speech.

The hearing child produces and differentiates meaningful sounds by comparing his own and others' articulatory patterns through proprioception; that is audition and orosensory-motor control (Ling, 1976). At the age of 18 months children with normal hearing master the production of about 50 words. Currently, there are four schematic stages of normal vocal development that researchers agree on according to Oller (2000). Those are as follows:

- The Phonation stage (0-2 months)
- The Primitive Articulation stage (1-4 months)
- The Expansion stage (3-8 months)
- The Canonical stage (5-10 months)

David Crystal (referred in Finnerty, 1996) describes the child's vocal development in each stage to be basic biological noise, cooing, vocal play and babbling. He adds a fifth stage between 9 - 18 months that involves melodic utterances.

The blind child, however, has no access to visual information. It would then seem likely that this would affect his/her phonological development. In fact, some studies (Wills, 1981; McConachie, 1990; Preisler, 1991; Thorén, 2002) have reported that speech and language acquisition is delayed by blind children both regarding the production of the first words at about 18 months of age and of two-word-sentences at about 2 years of age.

Mills (1983) has reported that facial movements in blind children are described as "muted", and the articulation is described as less distinct. There has been little research done concerning very young children and no exact comparison has been made with sighted children at the same stage of development. Most research has been done on blind children at the age of 6 to 18 years. According to Mills (1983, pp. 156)... "blind children will follow a different and slightly slower path in earlier phonology compared to sighted children and this is attributed to the absence of lip-read information". Three visually handicapped children were studied at the age of 1-1.5 years. The children showed slower acquisition of those sounds that had a visible articulation compared to sighted children. They also made different types of errors and had a slower development. However, in the long term the lack of lip-read information was not crucial for the development of phonology. As they got older the acoustic self-correcting feedback to compare spoken sounds with identical ones produced by adults became more important. In the long term the studied children showed no sign of developing a disordered phonological system.

Göllesz (1972) investigated Hungarian vowels articulated by blind children at the age of 13 years, through electromyography and spectrography. He found more pitch modulation and less lip movements compared with normal sighted children. However, the spectrograms showed no deviations in the acoustic properties of the productions.

Imitation and self-correction are important factors in speech learning as has been stated above. Children who are born with an auditory deficit have a limited acoustic speech target to imitate and compare their own production with. Children with moderate hearing impairments acquire speech through residual hearing and vision. The speech production ability of a hearing-impaired child is to some

extent proportional to the severity of the impairment. This relationship is valid up to the point of profound hearing losses (Monsen, 1983), see section 2.2.2. However, the conventional descriptors as mild, moderate, and severe hearing thresholds cover a wide range of hearing losses and different shapes of audiograms. Therefore there is a wide variation in speech acquisition among hearing-impaired children. The use of hearing-aids from an early age and special education is of special importance to support speech acquisition. "While amplification may permit improved access to some aspects of the produced signal, amplification alone may be inadequate in many cases for successful acquisition of spoken language" (Mashie, 1996, pp. 153).

Profoundly hearing-impaired children have to mostly rely on the limited visibility of phonetic features in learning oral speech and on orosensory-motor control in maintaining speech movements. Other senses must replace the auditory feedback that hearing children use to adjust and modify their articulation when they learn to speak. These children seldom develop speech spontaneously, but their speech is traditionally developed through a structured training, using the visibility of speech articulation, reading, tactile sensations and, if possible, residual hearing. The limited speech perception potential of profoundly hearing-impaired children and their reliance on lip-read information causes them to articulate speech segments and prosodic features deviantly. Despite these limitations they may develop systematic phonological processes that however differ from those of normally hearing children like for instance, voicing: [dɔ:kə] is realised [dɔ:ga], devoicing: [rø:d] is realised [Rɔt:], fronting: [sɛ:l] is realised [fæR], backing: [jœ:ra] is realised [hy:ra] and stopping: [hɑ:v] is realised [hɑpʰ].

Several studies have shown that profoundly hearing-impaired children possess some kind of abstract and stable phonological systems (West and Weber, 1973; Oller and Kelly, 1974; Dodd, 1976, 1988; Oller and Eilers, 1981, Abberton and Fourcin, 1985; Öster 1989a, 1991, 1992a, 1992b) but these may differ from those of normally hearing speakers. Dodd (1988) has discussed the fact that the limited information available in visual aspects of speech may be used to develop a phonological system. She suggests that strong evidence is provided that lip-read and heard speech is processed in a code



insensitive to input modality. A phonological system can be derived from hearing by ear or from hearing by eye, but the resulting systems will differ in some respects.

The fact that profoundly hearing-impaired children tend to produce a speech sound in the same deviant manner, in similar contexts was confirmed in a study by Öster (1989a). A 15-year-old profoundly hearing-impaired boy, who attended the 8th grade at the School for the Deaf in Stockholm (the Manilla School), was recorded twice at an interval of three months. At both times the child read the same speech material that consisted of monosyllabic and disyllabic words as well as sentences which contained these words. The words were chosen to be common and familiar to the child. Some of the consonants did not occur in all possible Swedish positions, as can be seen in Table 2.3.

The video recordings were transcribed using the symbols of the International Phonetic Alphabet. Many peculiarities and fusions of errors occur in prelingually profoundly hearing-impaired children's speech, which made an expansion of diacritical marks necessary. Some of those which Bush et al. (1973); Grunwell (1987); Roug, Landberg, and Lundberg (1989) have developed for the transcription of babbling and phonetic development in early infancy were used (see Appendix I).

The results of a narrow phonetic analysis of the two recordings are summarized in Table 2.3 according to the position in the word and to standard phonological representation. The results show that there was a high stability in the child's consonant production and the phonological representation between the two readings. Each representation is based on at least four readings and represents words in isolation as well as words in sentences. The only exception that showed a phonologic instability between the two readings was the use of the fricatives /s/ and /ʃ/ in initial and medial position.

Table 2.3. Narrow phonetic analysis of the child's two readings (reading I and reading II) according to standard phonological representation and to position in the word.

| IPA | INITIAL                      |   | MEDIAL                     |                            | FINAL                        |                              |
|-----|------------------------------|---|----------------------------|----------------------------|------------------------------|------------------------------|
|     | I                            | II  | I                          | II                         | I                            | II                           |
| /p/ | $\underset{\downarrow}{p}$   | $\underset{\downarrow}{p}$                                | $b^h$                      | $b^h$                      |                              |                              |
| /t/ | $\underset{\downarrow}{t}$   | $\underset{\downarrow}{t}$                                | $d^h$                      | $d^h$                      | $\phi d^h$                   | $\phi d^h$                   |
| /k/ | $\underset{\downarrow}{g^h}$ | $\underset{\downarrow}{g^h}$                              | $g^h$                      | $g^h$                      | $\phi g^h$                   | $\phi g^h$                   |
| /b/ | $\underset{\downarrow}{b}$   | $\underset{\downarrow}{b}$                                |                            |                            |                              |                              |
| /d/ | $\underset{\downarrow}{d}$   | $\underset{\downarrow}{d}$                                | $d^h$                      | $d^h$                      | $\phi d$                     | $\phi d$                     |
| /g/ | $\underset{\downarrow}{g^e}$ | $\underset{\downarrow}{g^e}$                              | $g^h$                      | $g^h$                      | $\phi g$                     | $\phi g$                     |
| /f/ | $\underset{\downarrow}{f}$   | $\underset{\downarrow}{f}$                                |                            |                            |                              |                              |
| /v/ | $\underset{\downarrow}{v^e}$ | $\underset{\downarrow}{v^e}, m\underset{\downarrow}{v^e}$ |                            |                            | $\underset{\downarrow}{v^e}$ | $\underset{\downarrow}{v^e}$ |
| /s/ | - s s                        | - s   | s                          | -                          | s                            | s                            |
| /ʃ/ | s                            | s   |                            |                            |                              |                              |
| /j/ | $\underset{\downarrow}{nj}$  | $\underset{\downarrow}{j^e}$                              |                            |                            |                              |                              |
| /h/ | $\underset{\downarrow}{x}$   | $\underset{\downarrow}{x}$                                |                            |                            |                              |                              |
| l   | $\underset{\downarrow}{l^e}$ | $\underset{\downarrow}{l^e}$                              | $\underset{\downarrow}{d}$ | $\underset{\downarrow}{d}$ | $\underset{\downarrow}{l^e}$ | $\underset{\downarrow}{l^e}$ |
| /r/ |                              |   | r                          | r                          | r                            | r                            |
| /n/ | $\underset{\downarrow}{n}$   | $\underset{\downarrow}{n}$                                | $\underset{\downarrow}{n}$ | $\underset{\downarrow}{n}$ | $\underset{\downarrow}{n}$   | $\underset{\downarrow}{n}$   |
| /m/ | $\underset{\downarrow}{m}$   | $\underset{\downarrow}{m}$                                |                            |                            |                              |                              |

#### 2.2.4. The perception of speech through vision

The perception of speech through vision is difficult because many articulatory features of speech are not easily accessible from visual observation. Acoustically each speech sound is unique, but visually many sounds are hard or impossible to discriminate. Some speech sounds have almost identical visual articulatory movements while

others have non-visible articulation. Moreover, coarticulation influences the visibility of many speech movements.

Any set of speech segments that is visually contrastive from another is called a viseme (Woodward & Barber, 1960; Fisher, 1968). Confusions in both articulation and perception should therefore occur within visemes but not between them according to Martony et al., (1970). The lipreader can identify major differences in the place of articulation but has difficulties with the manner of articulation according to Risberg (1982). Woodward & Barber claimed that there are four visually contrastive units: bilabials, rounded labials, labiodentals and non-labials. However, according to Owens & Blazek (1985) the set of visemes varies from study to study due to differences in languages, talkers, stimuli, subjects' response tasks, and effects of vowel contexts. They examined the effect of vowel context on viseme identification and found that the vowel /u/ limited the number of contrastive visual units compared to /a/ and /i/.

Martony, Risberg, Agelfors, & Boberg (1970) found three visible groups for Swedish consonants according to place of articulation: bilabials /m, b, p/, labiodentals /v, f/, and 'others' /t, d, s, n, ç, h̥, l, j, r, k, g, ŋ, h/. As much as 82% of the Swedish consonants belong to the group 'others'. Concerning Swedish vowels, two groups were identified (rounded/unrounded) due to the visibility of lip rounding and jaw opening. The unrounded articulatory vowels /a/ and /i/ and the rounded /u/ were also visually contrastive. Martony (1975) emphasized this fact in a study where he showed that the distinction rounded/unrounded was consistently correctly produced by Swedish deaf children.

Erber (1974) stated that about 40 English phonemes are cut down to roughly 16 visemes in conversational speech because manners of articulation, such as voicing and nasality, are not visible. Markides (1989) stated that lip-reading gives correct identification of about 30-40% of initial consonants and only 20-30% of final consonants. According to Ewing (1941), a discrimination of 70% for consonants is required to understand speech efficiently. Consequently, prelingually and severely and profoundly hearing-impaired children, who mostly rely on lip-reading, cannot achieve this.

### **2.2.5. Age of onset and speech quality**

The age of onset seems to be very important for the kind of speech a severely and profoundly hearing-impaired person produces. Many studies have shown that speech production will differ both in severity and type due to the age of onset (Zimmerman and Rettaliata, 1981; Plant and Hammarberg, 1983; Öster 1988b; Cowie and Douglas Cowie, 1992; Fischer, 1995). If the hearing loss is acquired *prelingually*, that is before *the critical age* around the age of six (Binnie et al., 1982), before speech and language is acquired, the effects on speech production are great, both on segmental articulation and on prosody. If the hearing loss is acquired *postlingually*, when the most active period of speech and language development is ended, the loss of auditory control will not affect voice and speech immediately. Changes, when they do occur, tend to be acquired gradually. The reason for this is, according to Zimmermann & Rettaliata (1981), that the postlingually profoundly hearing-impaired speaker possesses memory traces of speech patterns and that these remain for a long time despite lack of auditory control. Post-lingual hearing impairments are far more common than pre-lingual impairments. Common treatments include hearing aids and learning lip reading.

The explanation of the fact that the age of onset has an important effect on speech quality of severely and profoundly hearing-impaired children is that children who are born with a profound hearing loss have no adequate acoustic speech target to imitate and compare their own production with. These children very rarely spontaneously develop speech but use sign language as their first language for the purpose of communication. However, prelingually and profoundly hearing-impaired children might develop intelligible speech for the purpose of making statements through a structured training using the visibility of phonetic features, reading, tactile sensations of the therapist's face, throat and expiration air and, if possible, residual hearing. Other senses such as vision and tactile stimulation must replace the auditory feedback for control and self-correction, which are important factors during speech acquisition. The child must see the words spoken in the same relationship many times, just as the hearing child has to hear them many times before understanding comes.

The fact that prelingually and profoundly hearing-impaired children are unable to hear their own spontaneous vocalisations deprives them of motivation to babble and prevents them from expanding their repertoire of speech sounds. Mavilya (1972) found both qualitative and quantitative differences in spontaneous vocalisations between three profoundly hearing-impaired infants and one normally hearing infant over a period of 3 to 7 months. Most of the vocalisations of the profoundly hearing-impaired infants were vocalic. Consonantal sounds were rare and not as well articulated as those of the normally hearing infant.

The effects of early hearing impairment on babbling were investigated by Kent et al., (1987). The study followed the phonetic development at 8, 12, and 15 months of identical twins differing in auditory function. One of the twins had a profound hearing loss bilaterally (H) and the other had normal hearing (N). The major results were that H rarely produced consonants other than alveolar and bilabial stops, that H at 24 months barely produced the range of syllables that N produced at 8 months, that H developed a restricted usage of vowel formant patterns, and that H developed an extremely variable fundamental frequency pattern and vocal fry.

Late onset, less and less babbling during the second half-year of life or atypical patterns of canonical babbling may indicate a profound hearing impairment according to Stoel-Gammon and Otomo (1986). Babies that begin canonical babbling after 10 months can be suspected to suffer from a profound hearing impairment according to Oller (2000). However, Wallace et al. (2000) reported that some of twenty deaf or hard of hearing infants with very simple babble production between 5 and 13 months developed intelligible speech, whereas some of the children with more complex early babble production were not speaking intelligibly by 5-10 years of age. The result indicates that for children with a profound hearing loss, articulation skill might be more dependent on developing an effective motor feedback acquired either visually or through touch.

### **2.2.6. Speech quality of profoundly hearing-impaired children**

Over the years, a number of qualitative and quantitative studies have documented types of segmental and prosodic deviations which are

typical of the speech of prelingually and profoundly hearing-impaired children, e.g., Ling (1976), Calvert (1961), Calvert & Silverman (1975), Hochberg et al. (1983), (see review by Gold, 1980). Prosodic aspects have not been studied as comprehensively as segmental aspects and show a much wider range of individual differences. Early studies were for the most part case studies based on collected diary accounts. Later on, more extensive studies appeared like, for example, the classic study of Hudgins & Numbers (1942), which documented the speech of 192 deaf children ranging in age from 8 to 20 years.

Vowels are extremely difficult to acquire for a profoundly hearing-impaired child because of lack of sufficient visual cues. Mosen (1976) found that the reduction of the space of vowel articulation was commonly due to a restricted range of the second formant, which is primarily responsible for the non-visible forward and backward movement of the tongue. This is in accordance with the result by Martony (1975), where spectrographic measurements of the frequency of the fundamental and the first, second, third, and fourth formants of Swedish children between 11 and 17 years with pure-tone averages (PTA) between 75-115 dBHL were analysed. He showed that jaw and lip positions generally were correct but tongue positions incorrect, which shows the impact of visibility on profoundly hearing-impaired children's speech acquisition.

### **2.2.7. Prerequisites for severely and profoundly hearing-impaired children to develop intelligible speech**

Severely and profoundly hearing-impaired children joining speech therapy should possess certain essential and important skills to develop intelligible speech. A screening and training (when possible) of the following conditions should be done before the therapy starts:

- *Case history*  
Factors like type of hearing impairment, the degree of severity, the age of onset, special education, when hearing aid was first fitted, use of hearing aid, deafness in the family, etc. have a strong influence on what kind of speech a child born with a profound hearing loss might develop.

- *Speech organ structure and function*  
An oral-peripheral examination should be done to investigate possible functional disorders that could prevent speech development (Ling, 1976).
- *Proprioception*  
Children who are born with a profound hearing loss learn speech sounds through differentiation, involving the orosensory-motor patterns that his/her speech creates within the vocal tract (Ling, 1976). Therefore screening and training of proprioception should be done to test and train his/her ability to differentiate between sounds she or he makes. This could be done through productions of vocalisations that are differentiated through proprioception of specific orosensory-motor patterns by the child. Lots of rhythm, movement, babbling and balance activities are recommended.
- *Imitation skills*  
The hearing child uses vision to find out how to articulate speech sounds and then imitates these speech movements. Audition is used as a feedback channel for control and comparison of their output with their surrounding. Profoundly hearing-impaired children use vision as a substitute for audition and are missing the option of feedback and self-control. Therefore imitation is a fundamentally important skill for profoundly hearing-impaired children to have possession of (Ling, 1976). According to Fletcher (1986, pp. 236) ...."subjects with a hearing deficit will have enhanced skill in using visual information during oral motor tasks." Testing and training can be done by for example reversing the role of teacher and child and imitating speech gestures: put out/withdraw the tongue, sticking the tongue-tip up/down, rounded/spread lips etc. Explanations of how the sound is produced and where the articulators are could be done by sign-language if needed, shown through the use of a mirror, through articulation pictures, the therapist acting as a model or through tactile and visual feedback.

- *Functional hearing*  
The quality of the child's residual hearing for speech perception and control of his own speech production, the functional hearing, might be tested and trained through speech discrimination training with spoken stimuli of phonetic contrasts using illustrations of appropriate minimal word pairs that the child has to point to.
- *Exteroception*  
Exteroception, that is the child's possibilities of deriving meaning from the sounds that others make and of interpreting sounds as meaningful, implicates the knowledge of the sound system of the target language (Ling, 1976). This awareness is necessary to develop intelligible speech and can be attained through functional hearing, visual or tactile feedback. It could be tested and trained through audio-visual stimuli of speech, speech gestures and environmental sounds.
- *Lip-reading and visual capacity*  
Lip-reading skills and good vision is of course a prerequisite for children with profound hearing losses to be able to develop intelligible speech through effective and appropriate visual strategies.
- *Sign language skills*  
A good knowledge of sign language is extremely important as the speech therapist gives instructions and explanations about articulation positions and types of visual feedback offered in computer-based speech therapy.

## **2.3. Need, aim, and possibilities of speech therapy**

### **2.3.1. Moderate hearing-impaired children**

Children with a moderate hearing loss might develop speech spontaneously but the speech contains many deviations and is often delayed. The major factor is whether they can hear their own voice



or not. A hearing loss affects a child's possibilities to perceive his or her own sounds as well as the speech of people surrounding him or her. With the help of an early hearing-aid fitting, a hearing-impaired child's possibilities to develop speech through self-correction will increase a great deal. However, the child often needs help with some aspects such as fricatives, plosives, and pitch and stress patterns. The aim of speech therapy for hearing-impaired children, born with a moderate loss, is a speech-correction of already established deviations. A hearing-impaired child must also get help in interpreting auditory sensations as meaningful and in identifying the sound system of the target language.

### **2.3.2. Prelingually severe and profoundly hearing-impaired children**

Prelingually and profoundly hearing-impaired children who more or less lack auditory feedback, have great possibilities to control and self-correct their own speech production with the help of computer-based visual feedback. This is illustrated in Figure 2.6 by means of dashed arrow lines. See Figure 2.5 for a comparison with the hearing child who acquires speech and language spontaneously through a combination of hearing and vision.

The purpose of computer-based speech training for severely and profoundly hearing-impaired children is that they should develop intelligible speech for the purpose of making statements and more seldom for the purpose of communication. They should learn a "survival speech" that makes it possible for them to give and to understand simple messages in shops, in the street, etc. Most profoundly hearing-impaired children only receive 30-60 minutes of individual training per week. No speech training or correction is practiced in the classroom during ordinary lessons. These limited opportunities for speech therapy have increased the demands for efficiency. In a speech-training program it is important to make the child aware of the manner and place of articulation as well as of contrasts between similarly produced speech sounds.

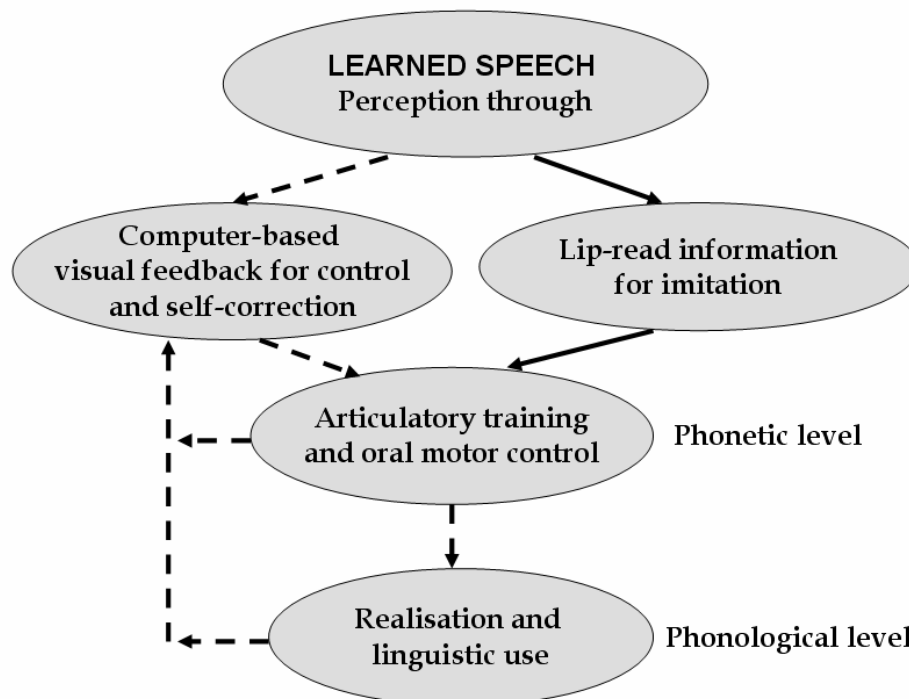


Figure 2.6. Assistance of computer-based visual feedback in profoundly hearing-impaired children's speech development.

Distinctive features like, for example, nasality and voicing are non-visible to the child through speech reading and consequently difficult to produce correctly. A computer-assisted aid has capabilities to offer a child immediate and meaningful visual feedback of these contrasts. By this technique it might also be easier for the therapist to instruct and explain what is wrong and what is correct in the child's production. The use of sign language for instruction and explanation may assist the speech therapist and help the child follow important steps during speech learning. The aim of speech therapy for profoundly hearing-impaired children is to assist them in their speech acquisition. To make their speech intelligible, to expand their existing articulation skills, and to facilitate linguistic use are the most important goals in speech therapy for profoundly hearing-impaired children.

### **2.3.3. Adult second language learners**

Speech therapy for L2 learners aims at an almost perfect pronunciation of both segmental and prosodic production of a new language. To learn a new language as an adult person is far more difficult than as a child. The brain of a child has plasticity and new networks and synapses can be developed much faster than for adults (Ellegård, 1982). For an adult person already established associations from his mother tongue must be changed. The adult person wants to use the new language straight away and learns very quickly the vocabulary of the second language. However, to learn to pronounce the words correctly requires lots of patience and a great amount of fundamental training of both perception and production of new contrasts and of contrasts that interfere with the sound system of the native language. A child learns the pronunciation of the words of the second language at the same time as he learns to utilize them and consequently gets the required fundamental training that is needed (Ellegård, 1982). Furthermore, a child has fewer demands upon the foreign language than an adult person has because the child talks about “here and now” and does not make use of the language abstractly as an adult needs to do. To sum up, children who learn a second language are able to add new patterns of behaviour to already established motor patterns while adult L2 learners have fewer possibilities to develop new connections to already established patterns.

### **2.3.4. Speech therapy aims**

Clients with deviant speech need speech therapy for different reasons. For moderate hearing-impaired children the goal is to improve articulation of especially the sibilants, to improve awareness, flow and consistency. However, the purpose of speech training for profoundly hearing-impaired children is that they should expand their systems and develop intelligible speech for the purpose of making statements and more seldom for the purpose of communication. As explained in section 1.4.1 they should learn a “survival speech” that makes it possible for them to give and to understand simple messages in shops, in the street, etc. The ambition for L2 speakers is quite the opposite. Speech therapy aims at helping them sound as native as possible.

### **2.3.5. General steps in an individual therapy**

An individual and efficient speech therapy may vary somewhat from teacher to teacher but there are certain general steps that must be followed, irrespective of age or the client's need for speech training (Risberg, 1968; Mahshie, 1995; Öster 1989b, 1989c, 1995b, 1995d, 1996). The following steps are recommended:

- *Diagnosis.* The first step of speech therapy involves an individual assessment of the deviations that should be corrected and trained in order to increase the intelligibility of the speech or, as for L2 learners, to make the speech sound as native as possible. All assessment is time-consuming and laborious. However, to prevent speech therapy from becoming a series of meaningless "articulatory gymnastic" sessions, the therapist must be prepared to spend time carrying out an appropriate diagnosis prior to therapy. Therapy and assessment are inseparable as assessments are required regularly during the training program.
- *Instructions* that aim at awareness, correct realisation, and understanding. It is important to make the learner aware of in what way his/her production deviates and show him/her how to correct it.
- *Initial training to obtain automaticity.* When the learner has acquired the target production, significant amounts of training must be invested to get him or her maintain correct production.
- *Additional and repetitive training for generalisation and transfer to linguistic use.* This is the most important element in a speech training program but the most difficult to carry out. The target production must be repeated and practiced in a variety of contexts. The ultimate goal is a system expansion that is accomplished when the learner's best production becomes his or her most common production.

### ***3. Development of speech technology in speech therapy for severely and profoundly hearing-impaired children***

#### **3.1. Introduction**

Since the early 1960's many training devices have been developed for profoundly hearing-impaired children, who learn a motor pattern for each speech sound instead of the acoustic output that normal-hearing children do through audition. Early visual aids presented acoustic information of specific speech sounds visually in real time, like isolated vowels and nasals as well as the s-sound (Pickett and Constam, 1968; Pronovost et al., 1968; Cohen, 1968; Risberg, 1968; Upton, 1968, Thomas and Snell, 1970). Many of these visual displays showed either too much or too little information, and in most cases considerable cognitive processing was needed to identify a speech sound, according to Thomas and Snell (1970). In addition, the visual feedback in these early displays was often too technical, illogical and difficult to understand for a child. Moreover, the child got no information about important improvements as the feedback mostly showed whether the pronunciation was correct or incorrect. Nowadays, efficient computer-assisted speech therapy makes use of logical and amusing visual feedback that is easy to understand for the child. It offers a child the possibility to perceive non-visible speech articulation, to imitate and compare the vocal output with that of the therapist and to display the child's smallest improvement. Beyond training on the phonetic level, extensive training on the phonological level is possible to help the child to make use of his or her oral motor pattern in speech.

In the following sections a survey is given of some essential milestones in the development of technical aids for speech therapy for severely and profoundly hearing-impaired children, from early non-computerized aids to more sophisticated computer-aided systems.

### **3.2. Process and product-oriented therapy systems**

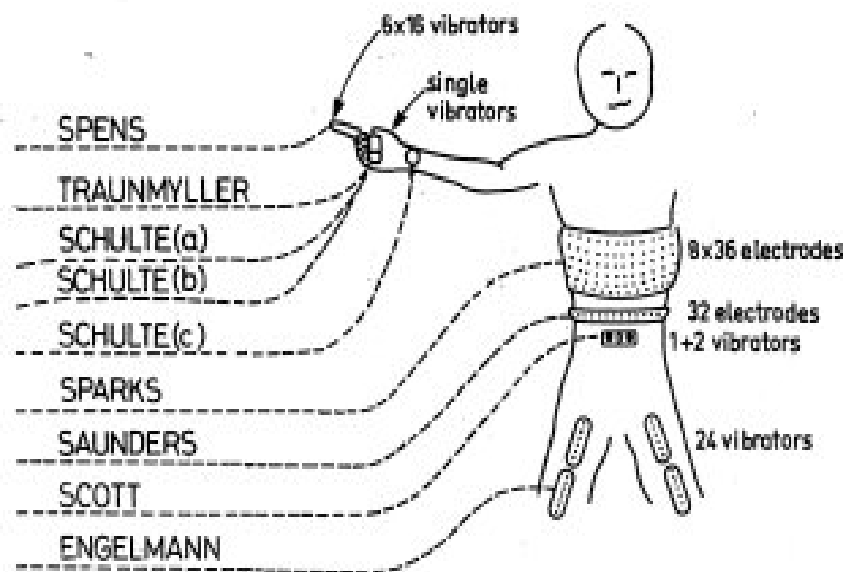
Traditional speech therapy with profoundly hearing-impaired children is based on direct imitation methods that help the children learn speech on a phonetic level by looking at the therapist's face and lips, through residual hearing or by feeling the therapist's face, throat and expiration air. The speech therapist gives instruction on how to use the speech organs while forming sounds. This is called a process-oriented approach, (Povel and Arends, 1991). Process-oriented means that such a system gives instructions on how to use the speech organs while forming various speech sounds by showing the placement of the articulators from outside and/or the inside of the mouth. A process-oriented system may replace the therapist in some stages of the training, especially in the instruction phase to guide the child how to move the articulators to improve the articulation.

In contrast, most of the existing computer-based speech therapy systems are product-oriented at present, as they give acoustic information (spectral or temporal analysis) about the speech product and the final result. These systems based on acoustic information do not give information about how the sound may be articulated. They offer an interactive parametrical feedback by showing various visual representations of acoustical parameters in real time. A target speech model is usually created to make it possible for the child to evaluate his/her own speech compared to a target of accepted production. Simple and game-like visualisations as well as "speech pictures" are also used to draw the child's attention to important parameters and help the child recognise whether his/her pronunciation is improving or not. A product-oriented computer-based speech therapy system principally requires the presence of a teacher for supervision and guidance of the adjustments of the articulators that are needed and to interact with the child by sign language if required.

### **3.3. Tactile aids**

Speech can be presented through vibrators at the fingertips and other parts of the body to indicate various elements of speech (Plant 1960;

Spens 1984; Traunmüller 1980). Many devices ranging from small portable aids to more sophisticated non-portable multichannel systems have been developed for awareness of environmental sounds as well as for speech reception. A comparison of different tactile communication aids was given by Spens (1984), see Figure 3.1.



*Figure 3.1. Different tactile aids with their respective places of stimulation and originators, Spens, 1984.*

Tactile aids can be classified in terms of how they produce touch sensations, as either vibrotactile, with signals delivered by one or more vibrators on the skin, or electrotactile, with sensations produced by electrical stimulation of the nerves that lead from the touch receptors into the skin.

Tactile aids have been shown to be good communication aids in supporting lip-reading as they provide information on rhythm, intensity and segmental boundaries (Pickett, 1980; Spens, 1984). They can help to increase the sensation of the rhythm of music and the monitoring of a hearing-impaired person's own voice for loudness and pitch.

Research has been reported by Schulte (1972) and Proctor (1995) concerning tactile aids and speech therapy. However, the use

of tactile aids has been shown to be less successful due to a major discrepancy in frequency range between the hearing sense and the tactile sense (von Békésy, 1959). The possibility to detect vibrations on the skin is highest in the area of 250 Hz and is almost none above 1000 Hz. The most sensitive area for hearing is found in higher and greater frequency range that is in the same area where most of the important parts of speech are conveyed, (see section 2.2.2, the Speech Banana). However, this was compensated for by using several vibrators each covering different ranges of the speech spectrum as can be seen in Figure 3.1. Tactaid VII that is available on the market presents sounds via seven vibrators providing information on voice-voiceless sounds, intensity, temporal cues, and the first two formants. This information is transmitted through the vibrators through place, movement, strength, and duration of vibration. (<http://www.tactaid.com/tactaid71.html>, May, 2006).

### **3.4. Product-oriented visual aids**

The visual modality has been the most preferred feedback method used in speech training with hearing-impaired children. The main motivating force for developing visual aids was a wish to display motor gestures that were non-visible and to support hearing-impaired children in developing a proprioceptive control of their speech gestures. These early visual aids were all based on some extracted acoustic information presented one by one in each display. This fact rendered the therapy nonflexible as the teacher had to change instrument according to the different training needs of the child.

Unfortunately, most of these visual displays were seldom evaluated within a pedagogical programme. Despite the fact that many of these aids have been reported to improve the speech of some children, the use of them was limited. This was probably due to the fact that the visual feedback provided by this type of visual aids was

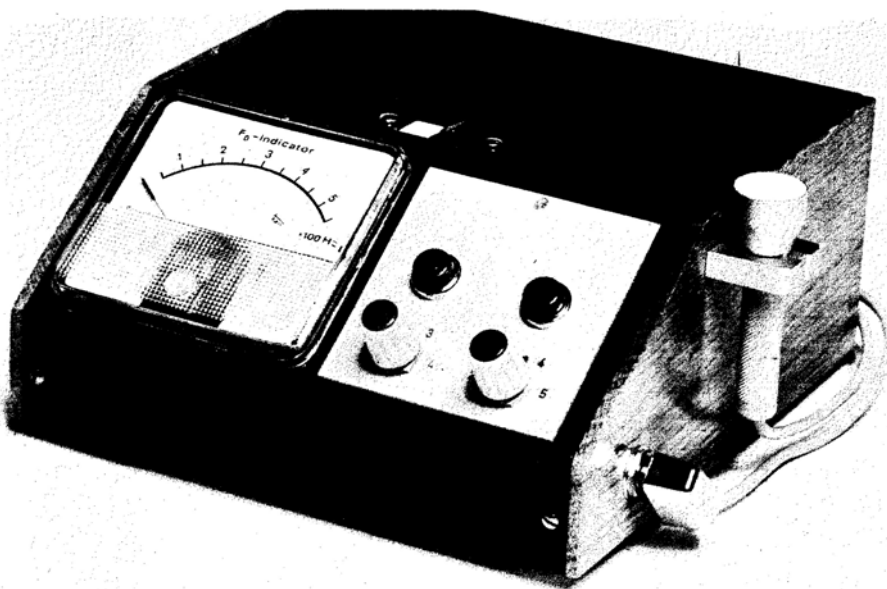
- difficult to understand
- unnatural
- delayed
- unattractive
- had no motivational impact on the children



Furthermore, due to technical problems and the lack of pedagogical manuals and evaluations, the teachers were not motivated to use them as standard procedures.

### **3.4.1. Feature based visual indicators**

Three different indicators were designed and developed during the 1960's at the Speech Transmission Laboratory at KTH, Dept of Speech, Music and Hearing. Various speech-analysing techniques were integrated in three different indicators, giving the possibilities to train s-sounds, nasalization, intonation, pitch and rhythm. A contact microphone picked up the vibration of the vocal folds and the nasal resonance. The measured frequency value of a child's production was indicated by a special lamp, showing a green light for a correct value and a red light for an incorrect value, or by the deflection of a meter needle along a frequency scale (Risberg, 1968; Martony, 1971). This type of visual feedback told the child whether the production was correct or not but gave no navigational information or explanation to the child of what to do to obtain an improved production.



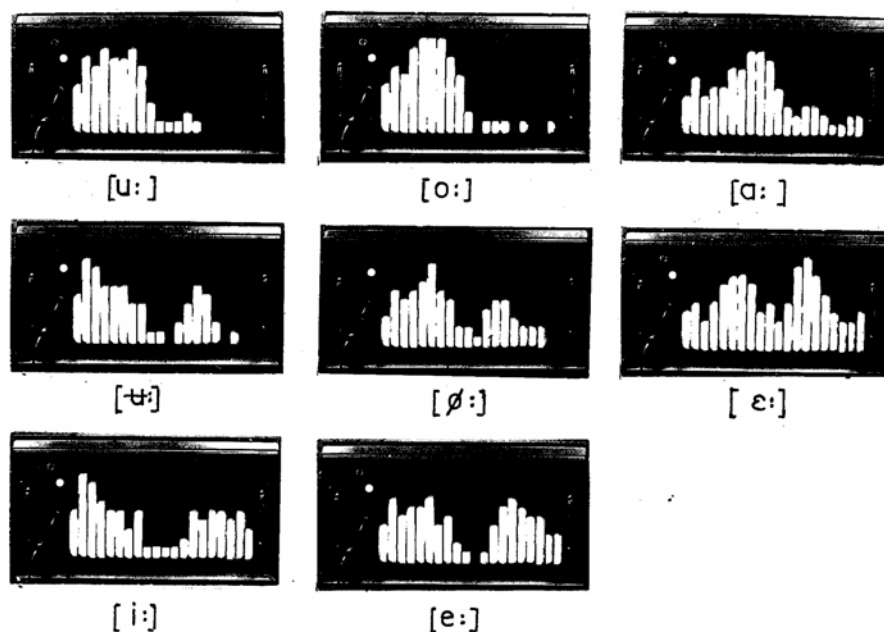
*Figure 3.2. Intonation indicator with instrument display.*

According to Risberg (1968) this type of visual aid was most useful in the instruction phase of speech therapy to describe and define the task. A promising result of voice register training was obtained by Martony (1966) with the intonation indicator (see Figure 3.2) in combination with an oscilloscope display.

### **3.4.2. Vowel and fricative displays**

A visual spectrum indicator (LUCIA) was also built by this group and was used for vowel and fricative training with hearing-impaired children at the Manilla Deaf School in Stockholm. The instrument used 20 band-pass filters in the frequency range of 200-7000 Hz.

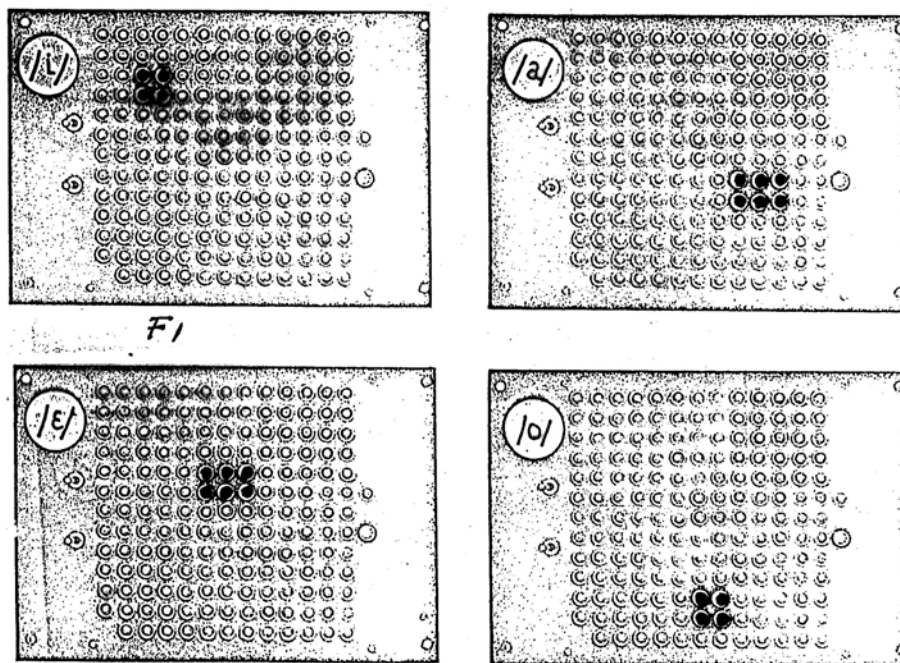
A matrix of 20 × 10 ordinary incandescent lamps gave light in 10 steps on the amplitude scale, each step about 3 dB, totally 30 dB, for showing important spectral details of long Swedish vowels, (see Figure 3.3). An innovation in this instrument was built-in memory that gave opportunities to study the attempt of the child in detail in order to give further instructions about correct behaviour.



*Figure 3.3. LUCIA with spectra of Swedish long vowels.*

### **3.4.3. Formant displays**

Displays showing the relationship between formants were developed by Thomas and Snell in 1968, Pickett and Constanam (1968) and Watanabe et al (1985). With these devices different vowels could be trained by matching visual patterns. The device by Thomas and Snell, shown in Figure 3.4, used an F1 - F2 plot in real time. The location of a vowel on a F1 - F2 plot is related to the articulatory configuration required to produce the vowel in the following way: the tongue height is related to the frequency of the first formant in such a way that F1 in a close vowel like /i/ has a lower frequency than F1 in an open vowel like (a). The position of the tongue along the vocal tract is related to the frequency of the second formant in such a way that F2 in a front vowel like (i)/ has a higher frequency than F2 in a back vowel like (o) (Miller, 1951). The expectation was that the profoundly hearing-impaired child might perceptually associate the position of different vowels on the F1 - F2 plot with correct articulatory configuration.



*Figure 3.4. Visual display patterns of sustained vowels (Thomas and Snell, 1970).*

#### **3.4.4. Fundamental Frequency displays**

Several displays were developed based on a microphone, fundamental frequency extractor and oscilloscope to improve voice pitch level and the fundamental frequency variation produced by profoundly hearing-impaired children (Plant, 1960; Martony, 1968; Boothroyd, 1970). A training system for hearing-impaired persons as well as for L2 learners was the “Laryngograph” (Fourcin & Abberton, 1971). This system was used quite successfully. The visualised signal gave a representation of the fundamental frequency and its variation shown as a frequency/time plot. Important information about prosody of a language as well as voice quality was provided.

#### **3.4.5. Computer-based speech therapy systems with visual feedback**

A significant contribution to the next generation of visual aids for speech therapy for hearing-impaired children was made by Nickerson & Stevens (1973) when they developed the first computer-based speech therapy system. The main reasons for introducing computers in speech therapy for hearing-impaired children at that time were that computers could perform complicated transformations of the speech signal to match the needs of presentation, that computers were on the whole simple to use and that they were easier to modify than machines (Nickerson et al., 1976). Thanks to the microcomputer technological development, effective visual computer-based systems have been developed and currently computer-aided speech therapy is in common use in clinics, schools and nursery schools.

The pioneering product-oriented computer-based system, made by Nickerson & Stevens in 1973, contained two motivating and playful displays. One was the “basketball game”, shown in Figure 3.5, where the aim was to get the ball into the basket avoiding the wall by controlling voice pitch. The height of the ball followed the fundamental frequency variation of the child’s voice. A happy or a sad face showed the result of the effort. This was one of the first amusing attempts to offer a child both rewarding feedback and an

exactly evaluative feedback, as is discussed in section 8.2.2 “type of feedback”.

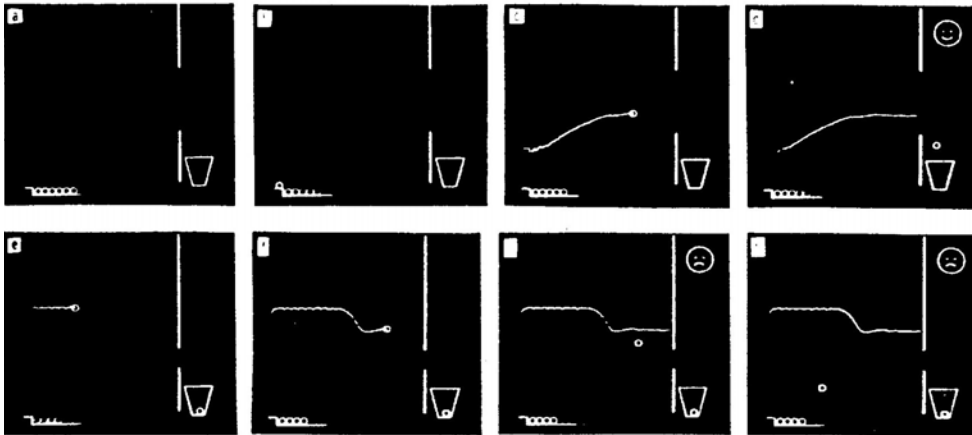


Figure 3.5. The “basketball game”. By controlling his/her vocal pitch movements the hearing-impaired child should pass a ball into the basket and avoid hitting the wall. The upper part shows a successful trial and the lower part shows a failure (Nickerson and Stevens, 1973).

The other display, shown in Figure 3.6, was the “cartoon face”, where several parameters were shown at the same time. A voiced sound was shown by an “Adam’s apple” on the throat, its height represented fundamental frequency and the loudness was shown by the size of the mouth.

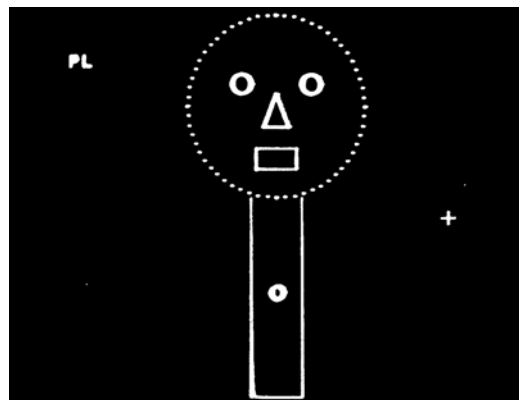
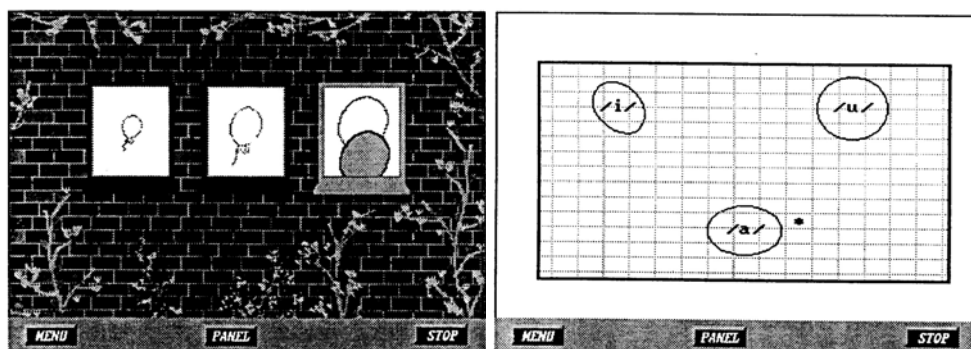


Figure 3.6. The “cartoon face” showing voicing by the “Adam’s apple”, pitch by its height and loudness by the size of the mouth (Nickerson and Steven, 1973)

Since then more advanced computer-aided speech training programs have been offered that have enhanced the possibility for speech and hearing impaired persons to improve their pronunciation (Osberger et al., 1981; Watson, Reed, Kewley-Port, Maki, 1989; Arends et al., 1991; Yamada & Murata, 1991; Levitt, 1993; Javkin, 1994; Rooney et al., 1994; Öster 1996, Öster et al. 1999b, 2002b, 2003).

Most of these systems contain a microphone, an amplifier, and a speaker connected to a sound-card that allows the user to input, store and analyse speech and then display it and play it back. The software often contains several interactive programs that have been shown to be successful in assisting speech and hearing-impaired children in achieving awareness and control over various speech attributes such as voicing, timing, pitch, and loudness as well as refining articulation and prosody.

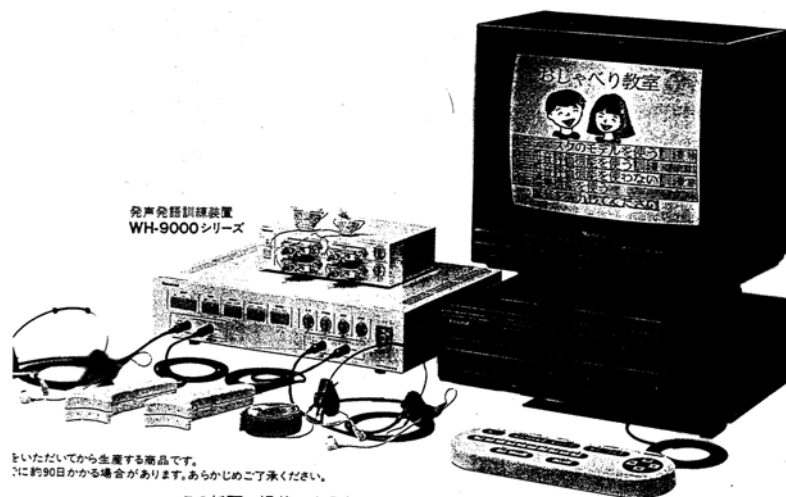
An example of these advanced computer-based systems is The Visual Speech Apparatus (VSA), developed by Povel and Arends in 1990, (Arends et al., 1991; Arends, 1993). It was designed to be used by a speech therapist who would adapt the training for each child. The program consists of attractive games which are controlled by different aspects of the child's speech. Figure 3.7 shows two selected exercises from this system. In the left panel three different ranges of loudness are shown for training of sustained voiced or unvoiced sounds. Voiced sounds are presented in red and unvoiced sounds in grey. To the right a training panel for vowel quality is presented.



*Figure 3.7. Two exercises from the Visual Speech Apparatus showing a panel for loudness training within certain ranges to the left and a vowel corrector display to the right, Arends et al., 1991.*

In this display vowels are shown as points in a panel corresponding to the first and second formant. The child's attempt is shown on the panel by a star. The intention was that the child should internalise the vowel space and achieve a correct production by changing the tongue position horizontally and vertically.

In the Matsushita speech training system (Yamada & Murata, 1991), also called the CISTA aid (Computer Integrated Speech Training Aid), as many as ten speech parameters, physiological (process-oriented) and acoustical parameters (product-oriented), are displayed simultaneously in real time using five sensors.



*Figure 3.8. The Matsushita speech training system (CISTA) showing the five sensors.*

The sensors are a nose contact microphone, an airflow sensor, a tongue palate, a neck contact microphone and an air microphone for extracting the following training parameters: nasality, contact pattern of the tongue, expiration airflow, plosiveness, fricativeness, intensity, pitch, intonation and voicing. However, such a sophisticated system with a range of training parameters, technically advanced acoustic and articulatory models, video games and graphical interfaces might most probably cause frustration as it makes great demands upon the therapist. "User-friendliness is a key factor in determining whether a particular system will actually be used in school settings" according to Youdelman (1994, pp. 77). Figure 3.8 shows the system with all sensors.

A very important contribution to the development of computer-based speech therapy with hearing-impaired children as young as four years was the approach that was implemented in the Indiana Speech Training Aid (ISTRA) by a new technology for speaker-dependent word recognition (Kewley-Port et al., 1991). The technique applied templates from a child's best production for comparison and evaluation of his/her speech quality in speech drills in game-like interfaces. The system gives feedback as a "goodness measure" based on the distance metric of the recogniser as an alternative for a human teacher. (Watson, Reed, Kewley-Port & Maki, 1989).

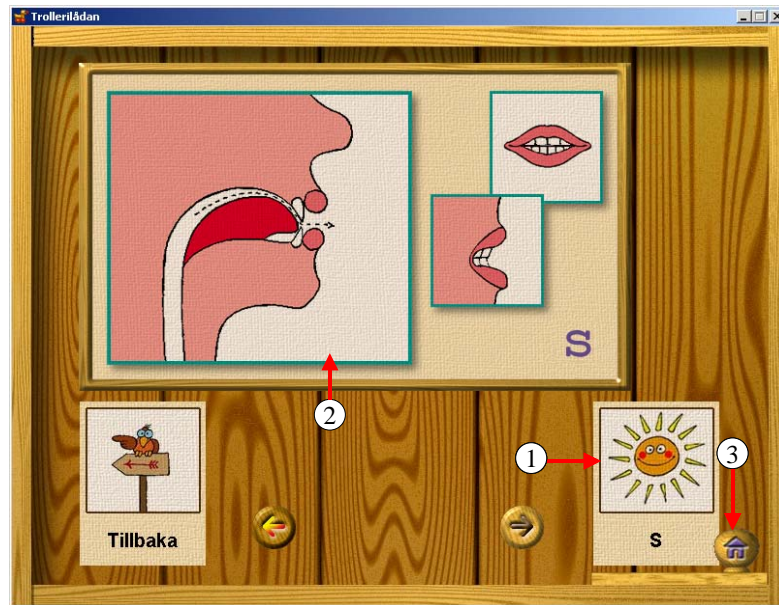
A very efficient and widely used system is the SpeechViewer (Crepy et al., 1983; Crepy et al., 1986) that was developed in a joint research project in 1979 between the IBM France Scientific Centre and professionals of speech therapy for profoundly hearing-impaired children from Institut National de Jeunes Sourds de Paris and Centre Experimental Orthophonique et Pédagogique. Today the third version is still in use and the program has been translated and adapted to many different languages. The Swedish version of SpeechViewer III (translated and adapted to Swedish by the author) is presented and discussed in Chapter 8 together with two other speech therapy programs that have been developed within two European projects with KTH as one of the partners. The two other systems are "Box of Tricks", developed in the SPECO-project: A Multimedia Multilingual Teaching and Training System for Speech Handicapped Children (Öster et al., 1999b; Vicsi et al., 1999, 2000) and the OLPy therapy system, developed in the Ortho-Logo-Paedia-project, (Öster et al., 2002b, 2003).

### **3.5. Process-oriented therapy systems with visual feedback**

Many computer-based speech therapy systems used today very often include a short section of illustrations showing cross-sections of place and manner of articulation during correct pronunciation of all speech sounds of the target language. These cross-sections can be studied and imitated by the child. Figure 3.9 shows articulation pictures of



the Swedish s-sound included in Trollerilådan, the Swedish version of Box of Tricks, developed in the SPECO-project.



*Figure 3.9. Articulation pictures of the Swedish s-sound included in Trollerilådan, the Swedish version of Box of Tricks, developed in the SPECO-project.*

One of the earliest process-oriented technical aids used with deaf children was the Danish Talemater (the Speech Mate) developed by the Institute for Electronic Systems at Aalborg University Centre (AUC) and Aalborgskolen. The aid was not computerised but used a colour-TV and a special keyboard to show cross sections of the location of the articulatory organs during various speech sounds. Visualizations of various articulatory movements were also possible to study by rapid shifts of the schematic cross sections. The Speech Mate consisted of 95 cross sections of isolated Danish vowels and consonants as well as CV-syllables. The program was later made available for work on a PC and was called "PC Talemater" (Lindberg, 1992) and was manufactured by DanVoice in Denmark.

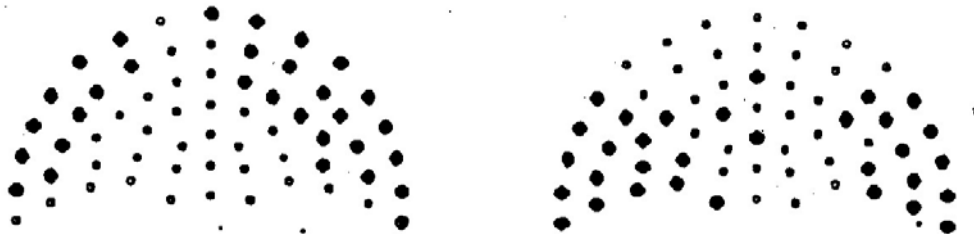
Another process-oriented computer based speech training program called SIM (Speech Illumina Mentor) was developed that provided dynamic visual information of stored images of internal and external articulators created from a series of magnetic resonance

images (MRI). Syllables could be trained through a voice recognition system in a game format (Soleymani et al., 1997).

### **3.5.1. Physiological devices**

A number of devices have been developed which provide direct feedback from physiological sensors. These devices made the child aware of articulatory gestures in the early stages of speech training. The Pneumotachograph (PTG) (Mahshie, Herbert and Hasegawa, 1984; Mahshie and Yadav, 1990) provided aerodynamic feedback obtained from the oral and nasal airflow, oral air pressure, from an electroglottograph, and accelerometer signals. Case studies showed that both significant improvements and notable carry-over (generalizations) were observed, (Mahshie, 1995).

The Electropalatograph (EPG) (Fletcher et al., 1991) used a sensor to show the tongue and palate contact during speech. An acrylic palate with embedded electrodes registered points of contact on the palate, which were displayed on a computer screen. A similar system was also included in the CISTA-aid (page 126) called the Palatograph, which provided real time dynamic tongue-palate contact by means of a tongue position sensor, see Figure 3.10.



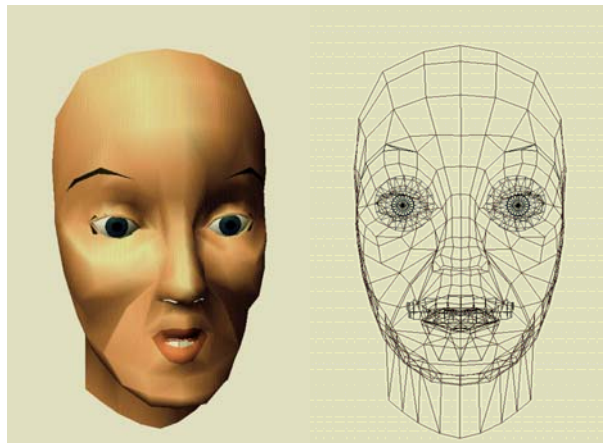
*Figure 3.10. The contact pattern for /s/ of two speakers. To the left a teacher's correct pattern is shown and to the right a child's incorrect tongue contact pattern is shown, (Yamada & Murata, 1991).*

Contacts between the tongue and inner parts of the oral cavity during the articulation of different phonemes are shown on a screen as dots, which correspond to the individual sensors. The aim of the child is to copy the correct pattern of the therapist. This enables the student to see those speech sounds that are not visible when lip-reading (Yamada & Murata, 1991; Youdelman, 1991).

### **3.5.2. Automatic speech tutors**

More recent research aims at giving knowledge of the performance and processes during articulation. A talking head has been incorporated as a virtual speech therapist, called Baldi (Massaro & Light, 2004), who is able to show the movements of his internal articulators. Baldi has been used in the early stages of speech training giving instructions to hearing-impaired children and L2 learners about how speech sounds should be produced. However, no comparison of the visual feedback with reference to the client's deviant production has been made with the help of Baldi.

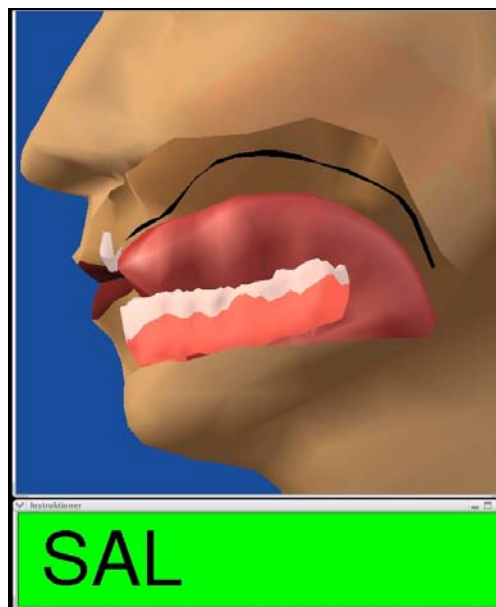
An audio-visual 3D facial model, a talking head, has been developed based on the KTH text-to-speech system (Beskow, 2003), (see Figure 3.11). Besides being used in applications like spoken human-machine dialogue systems, entertaining games, virtual reality and films, etc. a talking head might function in various communication aids for hearing-impaired users like the talking head telephone SYNFACE (Beskow et al., 2004), in language learning and pronunciation training. The advantages of using an automatic tutor in language learning are that it facilitates learning in a dialogue context and provides an always available conversational partner. Turn-taking and non-verbal feedback like encouragement, affirmation, and confirmation can be indicated as well as emphasis and focus in utterances.



*Figure 3.11. The talking face, and the underlying structure of the face-model, (Beskow, 2003).*

The assistance of an automatic tutor in pronunciation training with L2 learners and speech and hearing-impaired children could be helpful as it provides an untiring model of pronunciation, gives visual feedback for increased awareness, gives possibilities to display internal articulations, can exaggerate difficult sounds and highlight critical organs (Beskow et al., 2000).

Recently a virtual speech tutor “the ARTiculation TUtoR ARTUR” was developed at KTH in Sweden (Engwall et al., 2004; Engwall et al., 2006; Eriksson et al. 2005; Bälter et al., 2005), see Figure 3.12. This three-dimensional animated computer face with visible internal parts of the mouth shows the important differences between a correct articulation made by the three-dimensional computer face and the child’s production. It is meant to be a self-instructive device in the later stages of speech training to obtain automaticity and transfer to untrained material. With the help of computer games and the feedback from ARTUR the intention is to help the therapist with the repetitive and additional training that is necessary to establish sensory-motor associations.



*Figure 3.12. A display of the articulation tutor ARTUR giving articulatory visual feedback of the tongue position during the production of the voiceless fricative consonant /s/, (Engwall et al., 2006).*

The target production can be repeated and practiced in a variety of contexts. The tutor shows the child how to move the tongue, jaw and lips to improve his/her articulation. In this way the child is able to train alone without the guidance of a speech therapist.

### **3.6. Spoken language training for L2 speakers**

Pronunciation training has often been neglected in the teaching of Swedish as a second language. However, nowadays the importance of an acceptable pronunciation by Swedish immigrants is stressed and the uses of efficient training methods are accentuated. Traditionally spoken language has been practised in language laboratories by “repeat-after-me” methods. This type of training provides the learner with a delayed, auditory and verbal feedback from a teacher.

Speech-interactive language learning systems within the field of Computer Assisted Language Learning (CALL) focus on self-tutoring systems with automatic error detection for the correction of the articulation of L2 speakers (Eskenazi, 1996). Individual training with the guidance of a teacher is impossible due to big classes and little time. With the help of ASR techniques (Dalby and Kewley-Port, 1999), automatic pronunciation scoring, distance learning and CD-ROMs with pre-recorded utterances (LaRocca, 1994), home training is now possible.

The use of visual information as an additional feedback to the auditory information has been used in the Swedish spoken L2-training by Gårding & Bannert (1979), McAllister (1986), and Flege (1989) among others. Today two programs for Swedish are available that give possibilities to listen to pre-recorded speech in the form of dialogues. The intention is that the learner should pronounce the text and then compare a graphical curve of his pronunciation with the pre-recorded speech (Kjellin, 1997; LINGUS, Larson Education AB). However, computer-based pronunciation training with audio-visual contrastive feedback has been used very rarely although it has been shown to provide a valuable resource in spoken L2-teaching by Öster (1997, 1998).



## ***4. Residual hearing for speech processing - methods of investigating the functional hearing for speech***

### **4.1. Introduction**

How is the speech intelligibility of profoundly hearing-impaired children related to audiological data? Is there some measure of their residual hearing that could predict a child's possibility to develop intelligible speech? Studies reported in this chapter indicate that a profoundly hearing-impaired child's speech intelligibility is mostly related to the quality and the use the child has been able to make of his residual hearing, measured through the ability to recognize simple speech stimuli. The development of a computer-based analytical speech perception test for diagnostic purposes of small children from four years of age is also described in the final part of this chapter. The perception test contains easy and familiar illustrated word pairs of minimal contrasts.

As has been stated in section 2.2.2, it will be insufficient to rely on pure-tone audiometry when estimating a profoundly hearing-impaired child's possibilities of developing intelligible speech, as it gives no information about the child's speech processing capabilities. Speech is made up of complex and rapidly changing acoustic events (Ling, 1976) and the pure tone audiogram cannot indicate a profoundly hearing-impaired child's ability to distinguish between frequencies, to track formant transitions or to detect differences in intensity. The ability to perceive even simple speech material may not correlate with the ability to hear pure tones. For children with pure-tone averages worse than 90 dBHL, sound might be perceived through vibrotactile rather than auditory receptors. Vibrotactile perception is mostly limited to speech-envelope features like duration and intensity (Erber, 1974a). Auditory perception discriminates also spectral features like small differences in fundamental frequency and vowel formant patterns. The pure tone audiogram will not differentiate "vibrotactile" from "auditory"

children as it provides insufficient information about speech-processing capabilities, like the ability to perceive gap durations and small differences both in frequency and intensity.

Using the pure tone audiogram to predict a child's speech perception abilities is inappropriate because it does not reflect a child's capacity to perceive complex sounds like speech. According to Osberger (1992) individuals with similar hearing sensitivity may demonstrate very different speech reception abilities. The sort of speech a profoundly hearing-impaired child develops depends not only on the amount of hearing but also on the quality and the use the child is able to make of his/her residual hearing, the so called *functional hearing for speech*.

In order to investigate the relation between the speech intelligibility of eleven profoundly hearing-impaired children and their residual hearing capabilities a study was carried out (Öster 2002a) that contained three parts. In the first part of the study, the relation between the amount of residual hearing, measured as the better-ear average of pure-tone thresholds at 500, 1000 and 2000 Hz, and the intelligibility of their speech was investigated. In the second part, the interrelationship between the quality of residual hearing (defined as the shape of their audiogram) and the intelligibility of their speech was studied. In the third part, the relation between the degree to which the children could use their residual hearing for speech (the functional hearing), and the intelligibility of their speech was studied.

## **4.2. Functional hearing**

The term functional hearing is used to describe the degree to which a child can use his residual hearing for speech perception and control of his own speech production. The quality of the residual hearing is crucial for a child's ability to perceive speech and to develop intelligible speech and depends on many factors such as amount of hearing aid use, amount of auditory training, discrimination ability of speech features, learned ability to identify speech sounds, phonological short-term memory and speech processing capabilities. To some degree functional hearing can be trained, as some factors, for example phonetic and linguistic interpretation, might be influenced by learned ability to identify the speech sounds of the



target language. The hearing capacity on the other hand is physiological and cannot be trained. Consequently, the residual hearing capabilities for speech (the functional hearing) of a profoundly hearing-impaired child depend on the following factors:

- amount of hearing aid use
- amount of auditory training
- discrimination ability of time and frequency (hearing acuity)
- phonetic interpretation capacity (learned ability to identify speech sounds)
- speech processing capability (to combine fragmentary information to a whole meaningful message, including phonological short-term memory)
- linguistic interpretation capacity (to understand running speech and simple questions with or without speech reading)

### **4.3. Functional hearing and speech intelligibility**

#### **4.3.1. Subjects**

Eleven children from the Manilla School for the Deaf, five boys and six girls, were selected to cover the range from good to poor speakers. They were not chosen according to degree of their hearing loss. All of the subjects had been trained with traditional speech therapy. Their pure tone averages (PTA) at 0.5, 1, 2 kHz were between 90-108 dBHL in the better ear. One child was eleven years of age, while the others ranged from fifteen to seventeen years. Age, sex, mean hearing loss and shape of audiogram (cf. Figure 2.3) are shown in Table 4.1. The children were educated by sign language and their hearing losses were in the vicinity of 90 dBHL or more, areas where the hearing capabilities may rather be vibrotactile than auditory. Their attitudes towards speech therapy as well as their speech intelligibility skills varied from very poor to very good.

Table 4.1. Age, sex, pure-tone averages (PTA), and shape of audiograms for the frequencies 0.5, 1, 2 kHz in the better ear for the eleven children in the study.

| Child | Age | Sex | PTA dBHL | Shape of audiogram |
|-------|-----|-----|----------|--------------------|
| 1     | 15  | F   | 90       | B3                 |
| 2     | 16  | F   | 92       | B4                 |
| 3     | 11  | F   | 92       | B4                 |
| 4     | 16  | F   | 95       | C3                 |
| 5     | 16  | F   | 95       | C3                 |
| 6     | 15  | M   | 98       | C4                 |
| 7     | 16  | F   | 100      | C4                 |
| 8     | 16  | M   | 103      | C4                 |
| 9     | 16  | M   | 108      | C5                 |
| 10    | 16  | M   | 108      | C6                 |
| 11    | 17  | M   | 108      | C6                 |

### 4.3.2. Intelligibility test

The speech of the children was recorded on audio-tape, as they read some questions. The questions were so-called Helen-questions (Ewertsen, 1973) of the type "what colour is a lemon?", which can be answered with one word only. Five different questions read by each child were presented via headphones at a comfortable level to two groups of normally hearing persons (experienced and inexperienced in listening to the speech of profoundly hearing-impaired children) to evaluate the speech intelligibility. The listeners had the possibility of repetition. The listener's task was to write down, in Swedish orthography, the answer to the question with one word. If the answer was correct, the question is counted as being correctly understood. Only completely correctly understood questions were counted as correct.

### 4.3.3. Listeners

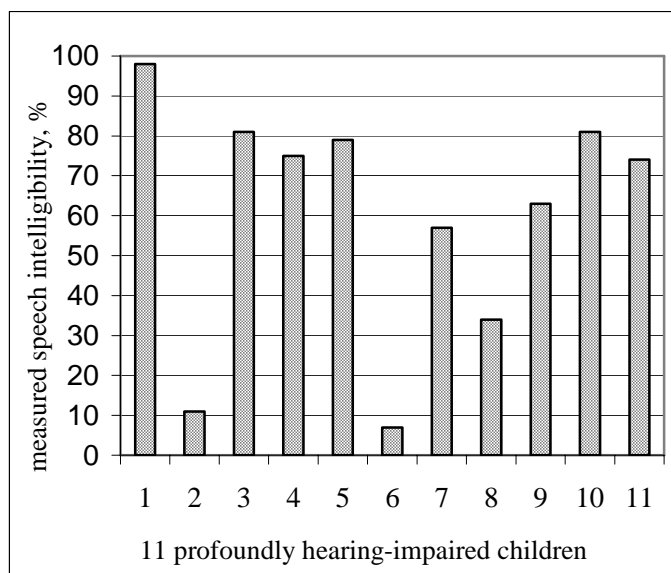
Twenty-four normally hearing persons listened to the speech-material. Sixteen of the listeners belonged to the staff of the department of Speech, Music and Hearing, KTH and were inexperienced in listening to the speech of profoundly hearing-impaired children. However, most of them were phonetically trained and had some experience of pathological speech. Eight of the

listeners were speech therapists of profoundly-hearing-impaired children and hence experienced in listening to this kind of speech.

## **4.4. Results**

### **4.4.1. Speech intelligibility scores of 11 profoundly hearing-impaired children**

Figure 4.1 shows the results of the intelligibility test averaged across all 24 normal hearing listeners.



*Figure 4.1 Speech intelligibility scores of 11 profoundly hearing-impaired children. The figure shows number of correctly perceived Helen-questions in percent by 24 normal-hearing listeners.*

The figure shows number of correctly perceived questions in percent by the listeners (each bar is based on 120 responses). The speech of three of the children was unintelligible, as the listeners could only understand 7-34% of their questions. Two of the children were semi-intelligible, 57-63% of their questions were understood. Six children were assessed to be intelligible as they were able to make as much as 74-98% of their questions understood by the listeners. Consequently, these profoundly hearing-impaired children with pure tone averages at .5, 1, and 2 kHz between 90-108 dBHL in the better ear showed

vastly varying intelligibility scores. The result indicates that the eleven children covered the range from good to poor speakers, thus confirming the selection of speakers made in this study.

#### 4.4.2. The effect of listeners' experience

Figure 4.2 shows the speech intelligibility of each child across experienced and inexperienced listeners. The average result for all questions and all speakers was 60% for all listeners, 54% for inexperienced and 65% for experienced listeners.

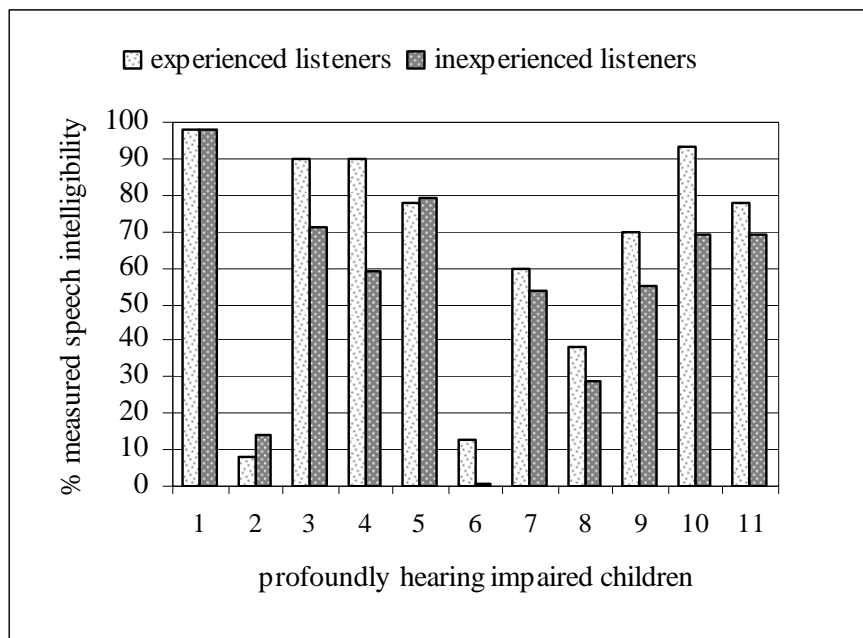


Figure 4.2. Speech intelligibility scores of 11 prelingually profoundly hearing-impaired children. The figure shows number of correctly perceived Helen-questions in percent by experienced and inexperienced listeners.

The result shows that each child's speech was understood as good, semi-intelligible and unintelligible by all listeners independent of experience. However, the experienced listeners understood each child better than the inexperienced listeners, except for child 1 and child 5, who were equally intelligible to both types of listeners.

### 4.4.3. Relation between amount of residual hearing and speech intelligibility

To investigate if the amount of the residual hearing also is predictive of the speech intelligibility of profoundly hearing-impaired children with pure-tone averages of 90-108 dBHL, the relation between the amount of residual hearing and the children's speech intelligibility was studied. The index most commonly used to indicate the amount of hearing of a hearing-impaired person is the better-ear average of pure-tone thresholds at 500, 1000 and 2000 Hz.

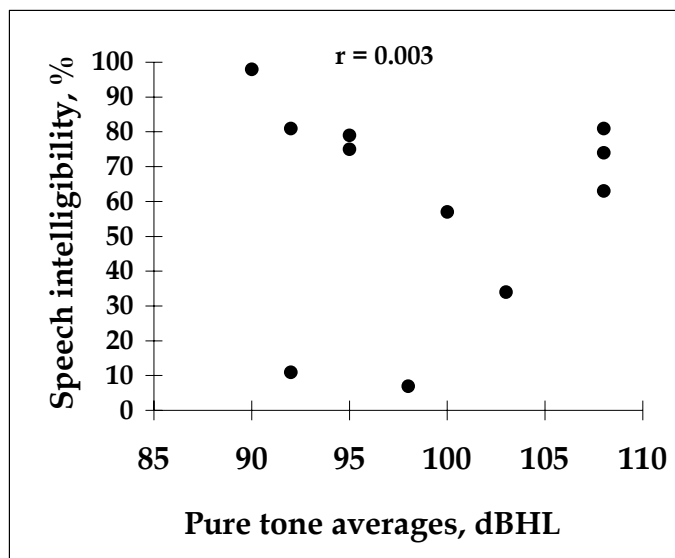


Figure 4.3. Correlation between the intelligibility of 11 profoundly hearing-impaired children's speech and their mean hearing loss for the frequencies 500, 1000, and 2000 Hz.

Figure 4.3, however, shows no correlation between speech intelligibility and PTA, which means that the intelligibility cannot be predicted from the degree of hearing loss, as measured by pure-tone audiometry. This is most likely because the pure tone audiogram gives insufficient information about speech processing capabilities. For this reason, a pure tone audiogram is not a good estimate of a profoundly hearing-impaired child's possibilities to develop intelligible speech.

#### **4.4.4. Relation between shape of audiogram and speech intelligibility**

It has been discussed whether or not the shape of the audiogram, rather than the degree of the hearing loss, may be the main factor which affects the intelligibility of hearing-impaired speakers. In general, flat audiograms are said to be associated with high speech intelligibility scores and falling audiograms with low speech intelligibility scores.

Markides (1985) investigated the relationship between shape of audiogram and rated speech intelligibility of children with similar average hearing loss of about 50 dB in their better ear. Six groups of children represented different types of audiogram shapes. However, he found no significant correlation between audiogram shape and rated speech intelligibility in that group of hearing-impaired children.

To investigate the relationship between the shape of the audiogram and the children's speech intelligibility we used the method of Risberg & Martony (1970), shown before in Figure 2.3, to classify the children's pure tone audiograms. However, the shapes of all audiograms were falling, varying among the children from B3 to C6, see Figure 4.4. Six of the children's audiograms were in the areas C4-6, which means in ranges where sound might be perceived through vibrotactile rather than auditory receptors. However, no clear relation existed between speech intelligibility and quality of their residual hearing, classified according to Risberg & Martony. Except for one child (B3 with an intelligibility score of 98%), there was no indication that a child with a classification of C4-6 should be a poorer speaker than a child with a classification of B4-C3, see Figure 4.4 (cf. Figure 4.1).

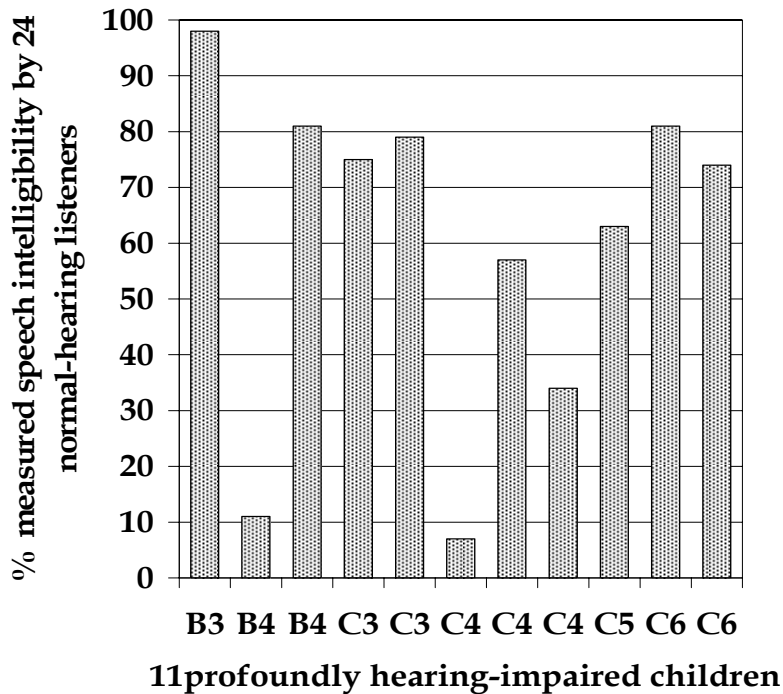


Figure 4.4. Speech intelligibility and shape of audiogram classified by the method of Risberg & Martony (1970).

#### 4.4.5. Relation between functional hearing and speech intelligibility

The functional hearing of a child, that is the degree to which he can use his residual hearing for speech perception and control of his own speech production, depends on his ability to perceive temporal and spectral information in speech, like the ability to perceive gap durations and small differences both in frequency and intensity. Speech processing capabilities are of course more appropriate to measure by means of a speech-test than by pure tones. Since the range of speech reception skills in profoundly hearing-impaired children is quite limited and since they have little experience in using speech, speech material can be difficult to utilize. Sentences might contain words that they do not know or difficult grammatical constructions. Speech tests especially designed for this group should be the best to use. However, no computerised tests have so far been developed in Swedish for use with young children who have difficulties in perceiving and producing speech. Several researchers

have shown (Cramer & Erber, 1974; Erber, 1974a; Gustafsson, 1984) that a simple spondee recognition test, for want of something better, can give valuable information about a profoundly hearing-impaired child's ability to perceive speech.

To investigate if the use of residual hearing could predict the intelligibility of profoundly hearing-impaired children's speech, the functional hearing of each child was measured by means of a speech test consisting of twelve common spondaic words (Risberg et al., 1977), familiar to the children.

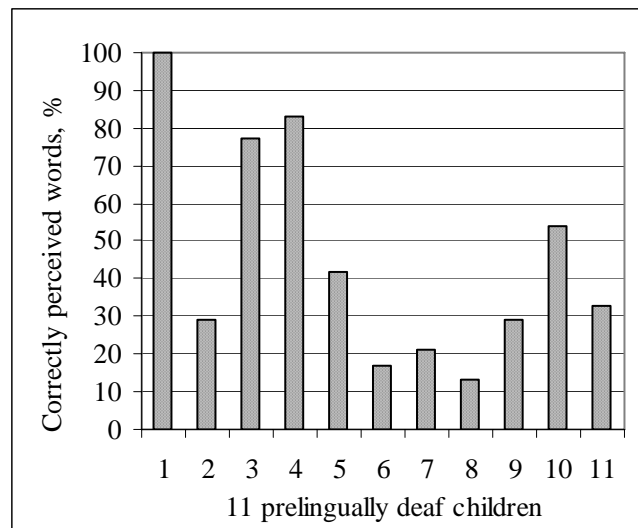


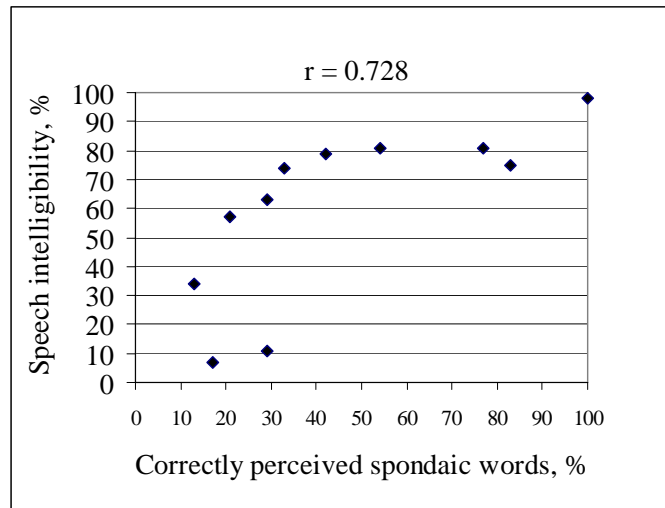
Figure 4.5. The figure shows functional hearing of each child, that is the percentage of correctly perceived spondaic words of each child. Each bar represents 24 words.

The words were presented twice (24 words) to the children via headphones at a comfortable level. The children answered by pointing to pictures that illustrated the test words. Figure 4.5 shows the functional hearing of the eleven children. Percentage of correctly perceived spondaic words is shown for each child.

The relation between measured speech intelligibility (see section 4.4.1) and functional hearing of the children is shown in Figure 4.6. A positive correlation was calculated to 0.73 between functional hearing and speech intelligibility measured for all listeners on average. The correlation between functional hearing and speech



intelligibility scores for experienced listeners was 0.74 and for inexperienced listeners 0.70. This indicates that the result of a simple speech test is a moderately good predictor of a profoundly hearing-impaired child's ability to develop intelligible speech.



*Figure 4.6. Correlation between the speech intelligibility of 11 prelingually profoundly hearing-impaired children and their possibilities to correctly perceive twenty-four spondaic words.*

The results of this study indicated that there is no clear correlation between increasing hearing loss and decreasing speech intelligibility for children with pure tone averages above 90 dBHL. As stated above it will be insufficient to rely on pure-tone audiometry when estimating a profoundly hearing-impaired child's possibilities to understand and develop intelligible speech, as it gives little reliable information about the child's speech processing capabilities. The sort of speech a profoundly hearing-impaired child develops depends not only on the amount of hearing, as measured by puretone thresholds, but also on the quality of the hearing sensation and the use the child through training has been able to make of his residual hearing.

The degree to which a profoundly hearing-impaired child can use his residual hearing for speech processing is of course only one of many factors affecting the child's speech development. Levitt (1987) has shown that there also exists a strong relationship between speech intelligibility and age at onset of hearing loss, early special education, use of hearing aid and additional handicaps.

## **4.5. Functional hearing and speech perception tests**

### **4.5.1. Introduction**

There are several speech perception tests that can be used with prelingually and profoundly hearing-impaired children and children with specific language impairment to assess their speech processing capabilities: the GASP test (Erber, 1977), the Merklein Test (Merklein, 1981), Nelli (Holmberg and Sahlén, 2000) and the Maltby Speech Perception Test (Maltby, 2000). Results from these tests provide information concerning education and habilitation that supplements the pure tone audiogram and the articulation index, because it indicates a person's ability to perceive and to discriminate between speech sounds.

Martony et al. (Martony et al., 1972; Martony, 1974; Risberg, 1976) developed an analytical rhyme speech test to be used with profoundly hearing-impaired children for testing speech discrimination ability in four frequency ranges. The test method was based on acoustic differences between contrasting phonemes and was aimed to be used in hearing aid fitting to predict the difficulties an individual child might have in learning to use acoustic information. However, the method required that the children were able to read.

As no computerized analytical speech test based on illustrations so far had been developed in Swedish for use with young children who have difficulties in perceiving and producing speech, this was done in cooperation with Risberg & Dahlqvist, (Öster 2002a). The development of the computer-based analytical speech perception test was an effort to address the need for early diagnoses and supplementary information to the pure tone audiogram about speech perception skills. More importantly, a goal of this test was to measure the potential for children to produce intelligible speech given their difficulties to perceive and produce speech. Therefore, this test was based on both acoustic and articulatory differences between contrasting phonemes. The expectation was that the result of this test might give important recommendations for individual treatment and speech-training

programs. The test seeks to evaluate the ability to perceive a range of sound contrasts used in the Swedish language. The test is tailored for measurements with small children from four years of age, who have not yet learnt to read, by using easy speech stimuli, words selected on the basis of familiarity, and pictures that represent the test items unambiguously. Profoundly hearing-impaired children with pure tone averages worse than 90 dBHL show very different abilities to learn speech and their potential to develop intelligible speech is unrelated to their pure tone audiograms. The development of this test was an effort to find a screening tool that can predict the ability to develop intelligible speech.

#### **4.5.2. Decisive factors for speech tests with small children**

The aim of an analytical speech perception test is to investigate how sensitive a child is to the differences in speech patterns that are used to define word meanings and sentence structures (Boothroyd, 1995). Consequently, it is important to use stimuli that represent those speech features that are phonologically important. Since the speech reception skills in profoundly hearing-impaired children are quite limited, and since small children in general have a restricted vocabulary and reading proficiency, the selection of the speech material was crucial. The words selected had to be familiar and meaningful to the child, be represented in pictorial form and contain the phonological contrasts of interest. However, presenting sound contrasts as nonsense syllables, so that the perception is not dependent on the child's vocabulary, was not a solution. It has been shown that nonsense syllables tend to be difficult for children to respond to and that they often substitute the nearest word they know (Maltby, 2000). Other important factors to pay attention to were:

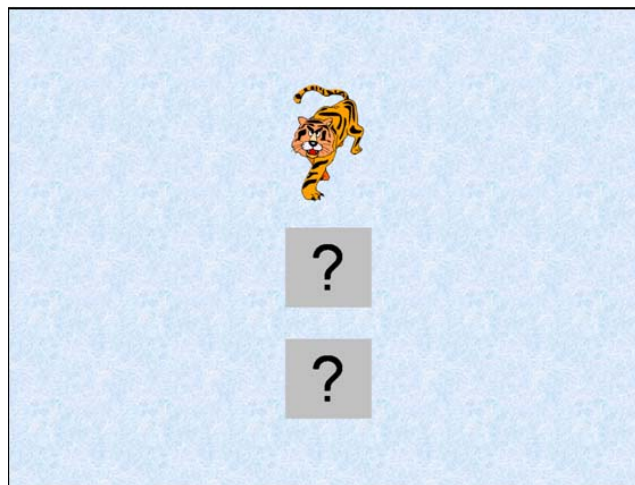
- what order of difficulty of stimulus presentation is appropriate
- what are familiar words for children at different ages and with different hearing losses
- what is the most unambiguous way to illustrate the chosen test words

Moreover, the task had to be meaningful, natural and well understood by the child; otherwise he/she will not cooperate. Finally, the test must rapidly give a reliable result, as small children do not have particularly good attention and motivation.

### **4.5.3. Test construction**

The test contains illustrations of easy and familiar words. The words contain important phonological Swedish contrasts and each contrast is tested in one of eighteen different subtests by 6 word pairs presented twice. In Table 4.2 a summary of the test shows the phonological contrasts evaluated in each subtest, an explanation of the discrimination task and one example from each subtest.

The words used were recordings of one female speaker. An illustrated word (the target) is presented to the child on a computer screen together with the female voice reading the word. The task of the child is to discriminate between two following sounds without illustrations and to decide which one is the same as the target word, see Figure 4.7. The child answers by pointing with the mouse or with his/her finger to one of the boxes on the screen.



*Figure 4.7. An example of the presentation of test stimuli on the computer screen. In this case the phonological contrast of vowel quantity tested through the words tiger-tigger [ti:ger-tiger] (tiger-begging).*

Table 4.2. The eighteen subtests included in the test.

| SUBTEST | CONTRASTING SOUNDS  | PRIMARY ACOUSTIC DIFFERENCE   | EXAMPLE                 |
|---------|---|---|-------------------------|
| 1       | Fundamental frequency                                     | Male/female/child voices  |                         |
| 2       | Number of syllables                                       | Large duration and intensity differences                            | bil/brandbil            |
| 3       | Gross discrimination of long vowels                       | F1 and F2 varying in frequency                                      | sol/sil<br>[u - i:]     |
| 4       | Discrimination of vowels                                  | Vowels differing at low frequencies                                 | fot/fat<br>[u - a:]     |
| 5       | Discrimination of vowels                                  | Vowels differing at high frequencies                                | gris/grus<br>[i - æ]    |
| 6       | Vowel quantity  | Duration of vowel and consonant                                     | tiger/tigger<br>[i - r] |
| 7       | Discrimination of voiced consonants                       | Low to high frequency voicing                                       | mus/ljus<br>[m - j]     |
| 8       | Discrimination of voiceless consonants                    | Energy varying from low to high frequencies                         | tupp/tusch<br>[p - s]   |
| 9       | Discrimination of voiced and voiceless consonants (place) | Energy varying from low to high frequencies                         | rot/fot<br>[r - f]      |
| 10      | Manner of articulation, voiced                            | Transient/continuous  | bada/lada<br>[b - l]    |
| 11      | Manner of articulation, voiceless                         | Transient/continuous  | tåg/såg<br>[t - s]      |
| 12      | Place of articulation, voiceless fricatives               | Energy varying in high frequencies                                  | sol/kjol<br>[s - ç]     |
| 13      | Place of articulation, stops                              | Energy varying  | kung/tung<br>[k - t]    |
| SUBTEST | CONTRASTING SOUNDS  | ARTICULATORY DIFFERENCE   | EXAMPLE                 |
| 14      | Voiced/voiceless consonants                               | Same place of articulation  | buss/puss<br>[b - p]    |
| 15      | l-j-r   | Same place of articulation, different manner of articulation        | ros/juice<br>[r - j]    |
| 16      | Nasal-l-j   | Different manner of articulation<br>Different place of articulation | lås/mås<br>[l - m]      |
| 17      | Front/back vowels   | Difference in F2  | för/får<br>[œ: - o:]    |
| 18      | Closed/open vowels  | Differences in F1   | får/far<br>[o: - a:]    |

The results are presented in percent correct responses on each subtest showing a profile of a child's functional hearing (see Figure 4.9) that is useful for the speech therapist for screening purposes and gives good indications of the child's difficulties in perceiving and subsequently producing the sounds of the Swedish language.

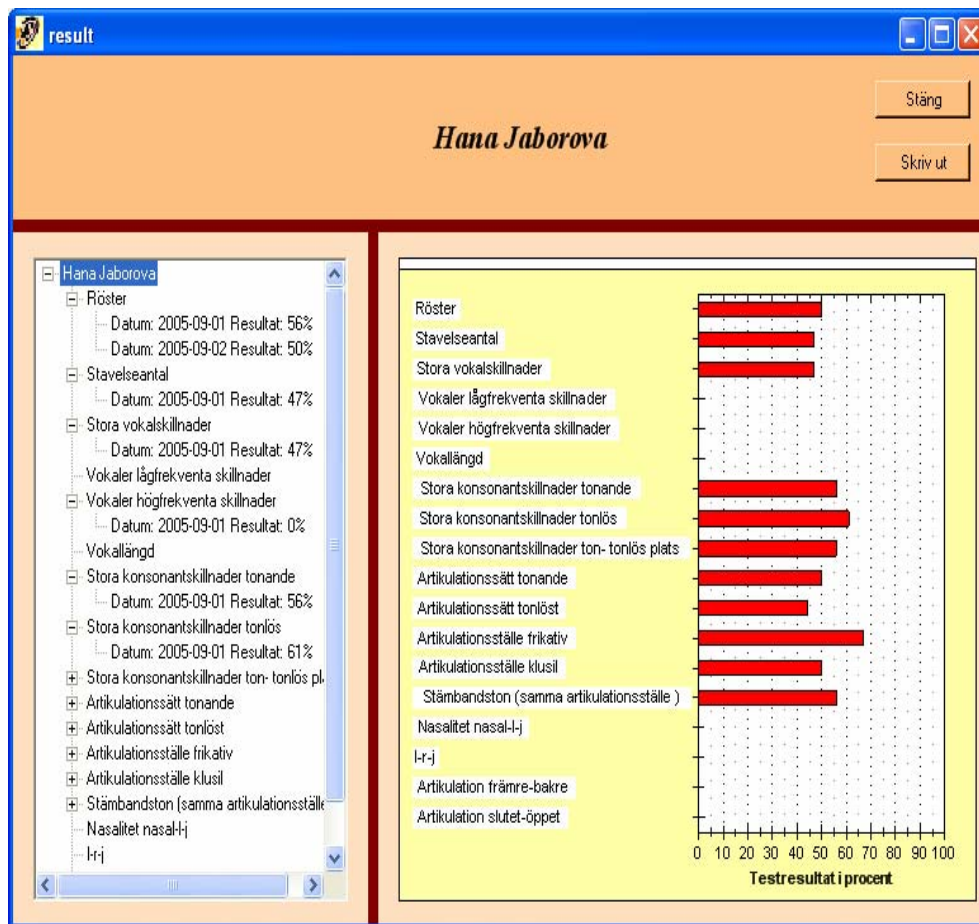


Figure 4.9 Example of a result profile for a child. Percent correct responses are shown for each subtest.

#### 4.5.4. Preliminary results

During the development of this analytical test a special reference group was established consisting of speech therapists for normal hearing children with special language impairment, moderately hearing-impaired children and profoundly hearing-impaired children. The therapists tried the different versions and gave

valuable advice continually concerning test procedure, way of response, type of interface, choice of illustrations, type of colours, and how to show the results etc. Altogether 54 children of different ages and with different types of difficulty in understanding and producing speech took part in the development and evaluation of the different versions. Eighteen of the normally hearing children with special language impairment were between 4 and 7 years of age and twelve were between 9 and 19 years of age. Nine of the children had a moderate hearing impairment and were between 4 and 6 years old and fifteen children had a profound hearing impairment and were between 6 and 12 years of age. Four of these had a cochlear implant and were between 6 and 12 years of age. Table 4.3 shows a summary of the children who tried out the program.

Table 4.3. Description of the children who participated in the development of the test. Average of pure-tone hearing threshold levels at 500, 1000 and 2000 Hz), age and number of children are shown.

| Normal-hearing children with specific language impairment |                   | Hearing-impaired children |                   |
|---|-------------------|---------------------------|-------------------|
|   |                   | < 60 dBHL                 | > 60 dBHL         |
| 4-7 years of age  | 9-19 years of age | 4-6 years of age          | 6-12 years of age |
| No. = 18  | No. = 12          | No. = 9                   | No. = 15          |

Figure 4.10 shows profiles for all 24 hearing-impaired children on some of the subtests. Black bars show mean results for the whole group and striped bars show the profile of one child with 60 dBHL pure tone averages at 500, 1000 and 2000 Hz. The result indicates that the child has greater difficulties on the whole to perceive important acoustical differences between speech sounds than the mean result of the 24 children. Many of his results of the subtests were below the result for guessing (50%). The result might be a good type of screening for what the child needs to train in the speech clinic.

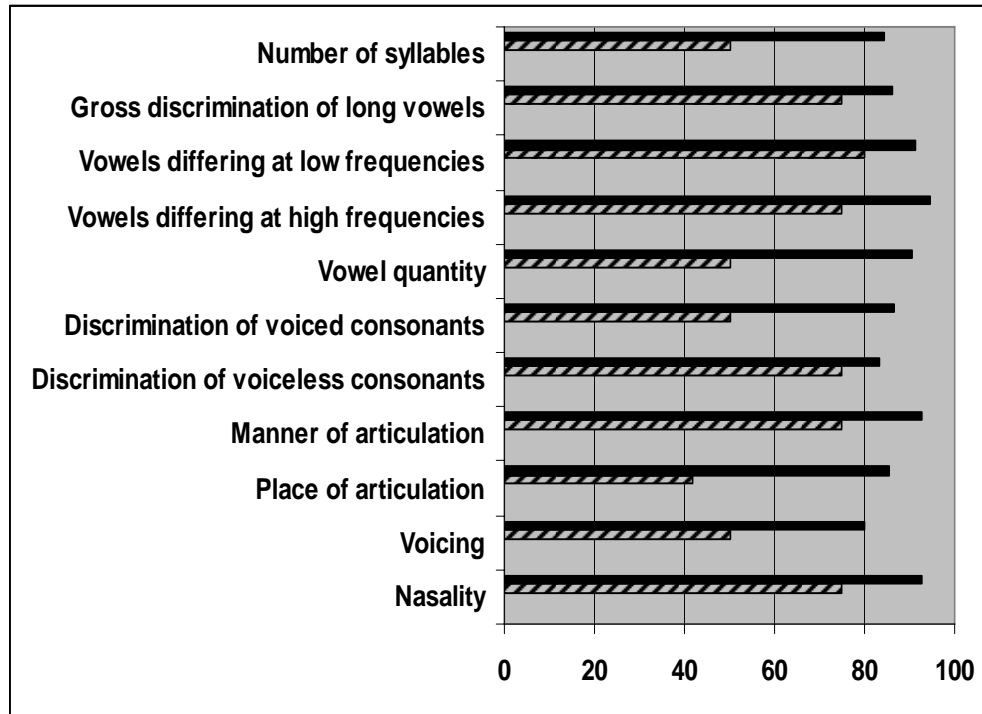


Figure 4.10. Results for the hearing-impaired children (N=24). Black bars show average results of the whole group and striped bars show the result of one child with 60 dBHL pure tone averages at 500, 1000 and 2000 Hz.

#### 4.5.5. Conclusion

The purpose of using analytical speech tests to get a measure of a listener's ability to understand or produce speech is important for many applications, e.g. in developing technical aids for transmitting speech through telephone or radio communications, in considering the acoustics of public halls, in establishing the impact of a sudden hearing loss on the ability to perceive and produce speech and in developing individual technical aids for the hearing-impaired.

The preliminary results reported here from the developmental phase of this speech perception test indicate that this type of a computerised speech test might give valuable information about which speech sound contrasts a hearing or speech disordered child has difficulties with. The child's results of the different subtests, consisting of both acoustic and articulatory differences between contrasting sounds, form a useful basis as an individual diagnosis of



the child's difficulties. This can be of great relevance for the work of the speech therapists.

The intention is that this test should be normalised to various groups of children so the result of one child could be compared to group data. This is useful supplementary information to the pure tone audiogram, especially for children with profound hearing losses. Hopefully it will meet the long-felt need for such a test for early diagnostic purposes in recommending and designing pedagogical habilitation programs for small children with difficulties in perceiving and producing speech. A training part consisting of computerized game-like exercises is now in progress. The training material will be based on the child's difficulties shown in the test result profile.



## ***5. Effects of speech input limitations on speech production***

### **5.1. Deviations in the speech of moderate hearing-impaired children**

According to the theory discussed above in section 2.2.3, speech is acquired by the developing child through the visual channel for imitation of speech movements and through the hearing channel for auditory control and self-correction. Therefore it is not surprising that the speech of hearing-impaired children often suffers from distortions and lack of articulatory precision. However, these children develop spontaneous speech and have possibilities to measure their own production attempts through self-control and self-correction depending on the severity of their hearing loss and the benefit and use of hearing aids. Often they need some help with some aspects of their own speech, as for example the pronunciation of fricatives, correct pitch, and stress patterns.

The hearing loss of these children might be considered as a single filter (not compensating for other disordered perceptual processes) and their deviant articulation may be predicted from the audiogram. If the loss is prominent in the higher frequencies for instance, fewer errors will be made on vowels than consonants. Front vowels that differ at high frequencies will then be more deviantly produced than central and back vowels that differ at low frequencies. The possibility to perceive and the ability to produce unvoiced high-frequency fricatives will be affected and the place of articulation of consonants will often be confused.

Typical speech deviations of moderate hearing-impaired children can be classified as distortions, substitutions, omissions, and insertions. A distortion is a non-standard production. A substitution is when a standard phoneme replaces another phoneme. When a deviation is defined as an omission, a speech sound is not produced at all at a place where it should be. Finally, in the case of insertions, an improper addition of a speech sound is made.

## **5.2. Deviations in the speech of profoundly hearing-impaired children**

The summary of common segmental, prosodic and voice disorders listed below is based on assessments made of the speech of Swedish prelingually and profoundly hearing-impaired children, (Öster 1992a).

### **Vowels**

- Simplified vowel systems
  - incorrect tongue position but correct lip position
- Nasalized vowels

### **Consonants**

- Substitutions of manner of articulation
  - stop consonants for fricatives and nasals
  - nasals for fricatives and laterals
  - laterals for tremulants
  - fricatives for laterals
- Substitutions of place of articulation
  - fronting of velars and palatals
  - backing of dentals and palatals
  - substitutions between sibilants
- Voicing
  - devoicing of voiced stops and fricatives
  - voicing of unvoiced stops and fricatives
- Deletions and weakening of consonants
  - deletion of laryngeal /h/ in initial position
  - absence of air-stream of fricatives
  - deletions of stops
  - non-audible release of final stops
- Insertions
  - fricatives and nasals are followed by stops produced at the same place of articulation
- Lack of co-articulation
  - two identical sounds in word-final and word-initial position are not merged into one sound but pronounced as two sounds with a pause between

### **Rhythm**

- Pausing errors
  - pauses between words and/or syllables
- Slow tempo
- No vowel quantity differences
  - preference for phonologically long vowels
- Lengthening of segments
- Incorrect stress patterns
  - monotony
- Insertions
  - a central vowel is often inserted between consonant clusters and after final consonants

### **Pitch and Voice Quality**

- Restricted frequency range
- Vowel dependent fundamental frequency
- High pitch
- Tensed voice
- Breathy voice
- Intensity variations

Some of these deviations affect the meaning of words while others affect only the naturalness of the speech. All children showed several phonological deviations, some of which were shared by all subjects. However, the children differed as to the frequency with which they applied the processes and some processes occurred in the speech of only one or two children. Many of the deviations listed above can also be found in the simplified speech of young normally hearing children. However, while these deviations disappear when the normal-hearing child matures and grows up these deviations will become fixed in the speech of profoundly hearing-impaired children if the children stay away from speech therapy.

### 5.2.1. Factors that cause deviations in the speech of profoundly hearing-impaired children

Some of the deviations found in the list above are unique for children with a profound hearing loss due to the reliance on visibility, impact of orthography, impact of different teaching methods, etiological aspects, educational aspects, and lack of physiological control.

#### Visibility

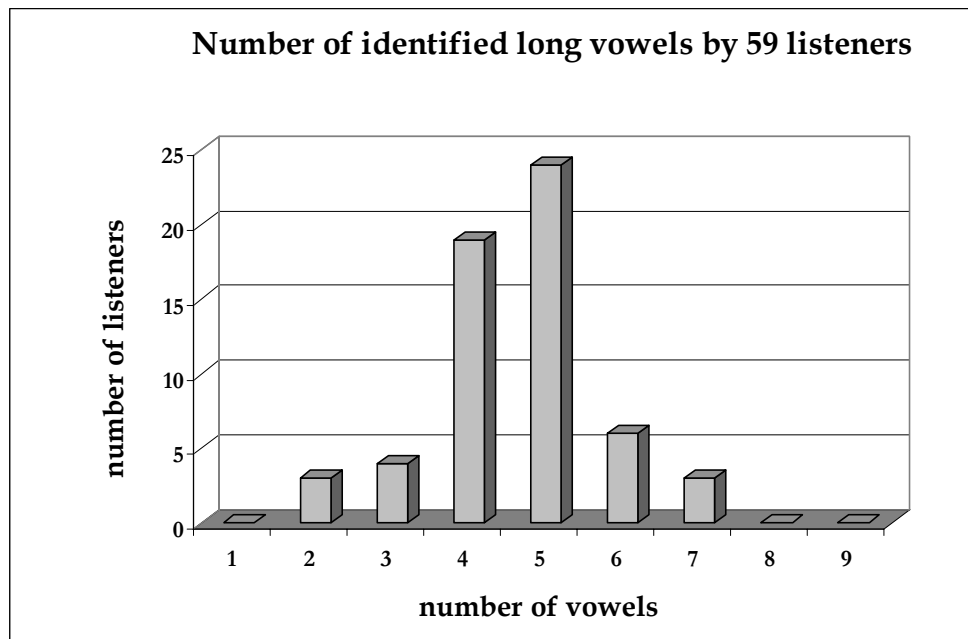
The impact of visibility on the speech of profoundly hearing-impaired children was shown by Öster (1992a), see Table 5.1. Aspects that are difficult to lip-read were related to characteristic deviations in the speech of profoundly hearing-impaired children.

Table 5.1. *Relationship between non-visible speech elements and typical deviations in the speech of prelingually and profoundly hearing-impaired children, (Öster 1992a).*

| NON-VISIBLE CUES:      | TYPICAL DEVIATIONS:        |
|------------------------|----------------------------|
| <b>Consonants:</b>     |                            |
| Place of articulation  | Fronting and backing       |
| Manner of articulation | Stopping                   |
| Voicing                | Voicing-errors             |
| Control of velum       | Nasalization               |
| <b>Vowels:</b>         |                            |
| Tongue-position        | Vowel reductions           |
| Quantity               | Preference for long vowels |
| <b>Prosody:</b>        |                            |
| F0                     | High pitch                 |
| F0-variation           | Monotony                   |
| Rhythm                 | Staccato-speech            |

As part of a course requirement, 59 students in phonetics were exposed to a listening test, that illustrates the impact of visibility on a 17-year-old prelingually and profoundly hearing-impaired boy's production of Swedish long vowels. His pure tone average was 92 dBHL. The students were requested to listen to the recorded material through a loudspeaker when the boy pronounced syllables made up

of /b/ in the connection with all Swedish long vowels. The students' task was to identify how many of the nine Swedish long vowels the boy was able to differentiate between.



*Figure 5.1. Number of long vowels identified by 59 normally hearing listeners.*

Figure 5.1 shows the result of 59 listeners' answers. It is obvious that the boy has a relatively restricted vowel system. The majority of the listeners, twenty-four persons, could identify 5 different vowels while nineteen other persons could differentiate between 4 vowels. Three persons identified 2 vowels, four persons heard 3 different vowels and six heard 6 vowels while three persons were able to identify as many as 7 different vowels.

The confusion matrix in Figure 5.2 shows that the vowels /ø:/ (57 out of 59), /ɛ:/ (55 out of 59) and /ɑ:/ (53 out of 59) were the vowels that the listeners identified the best. That means that the boy articulated these vowels very well. Most likely he controlled the rounding contrast between /ø:/ and /ɛ:/ thanks to the visibility of the difference in rounding. Moreover all these vowels are produced with a relaxed and flat tongue and a rather visible open jaw position in contrast to the vowels that the listeners had most difficulties to

identify; /i:/ (14 out of 59), /e:/(1 out of 59) and /u:/(7 out of 59), which have a non-visible closed tongue position. /ø:/ and /ɛ:/ were also those vowels that the listeners most frequently heard instead of other vowels.

|    | i: | e: | ɛ:  | ɑ: | y: | ø:  | ʌ: | u: | o: |
|----|----|----|-----|----|----|-----|----|----|----|
| i: | 14 | 44 | 1   |    |    |     |    |    |    |
| e: |    | 1  | 58  |    |    |     |    |    |    |
| ɛ: |    | 2  | 55  | 2  |    |     |    |    |    |
| ɑ: |    |    | 5   | 53 |    | 1   |    |    |    |
| y: | 3  | 12 |     |    | 29 | 2   | 11 |    | 1  |
| ø: |    |    |     |    |    | 57  | 2  |    |    |
| ʌ: |    |    |     | 1  |    | 31  | 24 | 1  | 2  |
| u: |    |    |     |    |    | 11  | 6  | 7  | 35 |
| o: |    |    |     |    |    | 12  | 6  | 14 | 27 |
|    | 17 | 59 | 119 | 56 | 29 | 114 | 49 | 22 | 65 |

Figure 5.2. Identified long vowels by 59 normally hearing listeners. Produced vowels are shown vertically and identified vowels are shown horizontally. Bottom row shows number of times each vowel was identified (grey square) or heard instead of another vowel.

### Lack of physiological control

The lack of physiological control during speech causes deviations like nasality, disordered voice production and above all disordered respiratory processes in speech, which affect the intelligibility of the speech of profoundly hearing-impaired children (Bench, 1992). Profoundly hearing-impaired children often have poor breath control during speech production. They use too much air per syllable, have a



slow speaking rate, make pauses between syllables that result in an erratic rhythm and they tend to breathe without consideration of linguistic boundaries. Their breathing often increases during speech compared to a decreasing breathing during speech by normally hearing children.

Whitehead (1983) investigated the respiratory patterns during speech of ten deaf male adults with a pure tone average loss of 105 dB. Magnetometer coils were attached to the chest, one for the abdomen and one for the rib cage, and then fed to an oscilloscope. Recordings were made when each subject read "the Rainbow Passage" (Fairbanks, 1960) at normal loudness and when the subjects talked about a topic of interest. The result showed that some deaf speakers did not inhale before speaking. As the intelligibility of their speech decreased, the volume of air in the lungs also decreased. Most of the subjects spoke on low lung volumes. Some of the subjects even initiated reading and conversation below the functional residual air capacity (FRC) without inspiration. When speaking below FRC, the speaker had to use a higher muscular pressure to achieve speech.

### **The impact of orthography**

The impact of orthography also causes deviate pronunciations of speech sounds. Each language has its own sound system and sound patterns, i.e., specific rules of how to combine phonemes to build up meaningful words and utterances. The knowledge of the phonological system of a specific language includes the knowledge of its pronunciation rules.

The new teaching situation in the Swedish Deaf schools implies that speech-training methods nowadays are based on written Swedish and the use of sign language for instruction and explanation. This means that it is extremely important that the children are well familiar with the pronunciation rules (text-to-speech rules). Insufficient knowledge of these rules causes typical deviations. Especially the various spellings of the phonemes /ʃ, ç, o:, ε:/ give rise to some deviations and the fact that two, and sometimes three, letters are pronounced as one sound in Swedish is not obvious to some children.

## **Teaching methods**

It is well known that human speakers adjust their speaking habits or speaking styles to the communicative situation. Many references to this phenomenon exist in the literature, like the "hyper vs. hypo speech" of Lindblom and Moon (1991). This "over articulation" is not merely a louder version of normal speech, but might also involve an active reorganisation of phonetic gestures.

When teaching profoundly hearing-impaired children articulation skills, the teacher often hyperarticulates to improve lip-reading. This might cause various deviations in the speech of the children as they tend to be overly sensitive to irrelevant visible variation. Examples of visible interference were found in a study by Öster (1991, 1992b). It was shown through a detailed phonological analysis that deviant phone types represented different phonemes in a profoundly hearing-impaired child's speech, despite a phonetic similarity to [b]. The child made contrasts between /p/ and /b/ in initial position through lip-protrusion instead of voicing.

## **Background variables**

Levitt (1987) has shown that the effect of background variables on the speech of hearing-impaired children is of particular interest. He emphasised the important role of special education and an early and effective intervention. He divided the background variables into the following three groups:

- *etiological*
  - age at onset of hearing loss
  - profound hearing loss in the family
  - hearing level
- *educational*
  - age at onset of special education
  - age when hearing aid was first fitted
  - use of hearing aid
  - intelligence quotient
  - reading score
  - syntactic comprehension

- *other variables*
  - other handicapping conditions
  - home language
  - parental occupation,
  - socio-economic status
  - number of siblings

### **5.3. Deviations in the speech of L2 learners**

As stated above in section 1.1.2, L2 learners articulate some target phonemes deviantly because the number and/or the quality of the speech sounds of the mother tongue (L1) differ from those of L2 (Eskenazi, 1999) and typical pronunciation difficulties for a given target language will differ for speakers of different native languages (Dalby & Kewley-Port, 1999).

For Swedish as the target language some general segmental and prosodic deviations can be summarised from studies made by Bannert (1990), McAllister (1986, 1995), Thorén (1994 ) and Öster (1998, 1999a) as follows:

**Vowels:** The Swedish language has as many as 9 phonologically long and 9 short vowels. The high number of Swedish front vowels and the fact that some of them are rounded makes it often difficult for L2 speakers to have a complete productive and perceptual mastery of all these vowels.

Moreover high and mid-high long Swedish vowels are diphthongised in open syllables. However, diphthongisation of long vowels in open syllables is seldom found in L2-speech. Some of the front vowels in Swedish are rounded. It is often difficult for L2 speakers who do not have rounded front vowels in their native language to pronounce these vowels correctly. Therefore they often substitute the Swedish front vowels (ö) [ø:] and (u) [ʉ:] with the rounded back vowels (å) [o:] and (o) [u:]. In their native language rounding is connected with a back placement of the tongue.

**Consonants:** Some of the Swedish consonants cause production and perception difficulties for L2 speakers, especially the voicing contrast between plosives, aspirated voiceless plosives, the great

number of fricatives, the nasal velar sound and the fact that dentals are retroflexed when they are preceded by /r/.

**Prosody:** Quantity is an important contrast in Swedish, which is realised as a difference in duration and, for some pairs, also with a difference in vowel quality. Furthermore, the following consonant has opposite quantity; i.e. long vowels are followed by short consonants and vice versa. It has been reported that quantity differences between phonologically long and short vowels are difficult for L2 speakers to have a complete mastery of (Bannert, 1990; McAllister, 1986; Thorén, 1994; and Öster 1998, 1999a). Common problems with Swedish prosody are to produce accent II and to produce a long consonant after a short vowel in stressed syllables.

## **6. *Interaction between individual deviations and speech intelligibility***

### **6.1. Introduction**

The effect of different phonetic deviations on the intelligibility of profoundly hearing-impaired speech has been studied using different techniques: correlational studies and qualitative studies (Hudgins and Numbers, 1942; John and Howarth, 1965; Mosen, 1983), manipulation by means of digital speech processing (Huggins, 1977; Kruger, Stromberg and Levitt, 1972; Maassen and Powel, 1984; 1985; Osberger and Levitt, 1979), or speech synthesis techniques (Bernstein, 1977; Öster 1985, 1990). With digital speech-coding techniques, only suprasegmental deviations are easily manipulated, but in the case of speech synthesis, both segmental and suprasegmental factors can be manipulated. The effect on the intelligibility of different kinds of deviations from normal production has not been decisively established. Most of the previous studies have not studied the interaction between the deviations and the individual effect of a deviation on intelligibility. Hence, the effect on intelligibility has varied depending on the contribution of deviations that has been examined. Gold (1980) claimed that,

"Whereas there is much documentation of the kinds of segmental and suprasegmental errors in the speech of the hearing impaired, there is far less evidence of the direct effects of each of these error types on overall speech intelligibility." - "Thus, although we may be able to identify those errors which occur most frequently in the speech of the profoundly hearing-impaired, we need further research to indicate how these error types interact to reduce speech intelligibility and to determine which error types should be the first to be considered when planning a training program for improved speech production in the hearing-impaired child." (p. 415.)

In this chapter a study by Öster (Öster 1985; Granström & Öster 1994a, 1994b) is described where normally hearing subjects estimated the effects of individual segmental and prosodic deviations on intelligibility in the simulated speech of three children from the Manilla School for Deaf Children in Stockholm. The deviations of each child were classified in an order of precedence according to intelligibility, giving an indication of which deviations should be corrected first, as they affected the intelligibility to the highest degree. Based on the results of the listening tests, recommendations could be given regarding an individual order for efficient correction. These findings stressed the importance of an individual assessment and an individual speech-training program for speech improvement.

## **6.2. The effects of individual deviations on speech intelligibility measured by means of synthetic speech**

### **6.2.1. Introduction**

The synthesis-by-rule system, developed at the Department of Speech Communication and Music Acoustics, KTH, (Carlson, Granström & Hunnicutt, 1982) was used in this study. By using this technique it was possible to sort out segmental and/or prosodic deviations, leaving other features unchanged to study the effect on intelligibility and the interaction between them through listening tests with normally hearing subjects. In the synthesis-by-rule system an arbitrary text is transformed into synthetic speech with the help of phonetic and phonological rules working on both the segmental and the prosodic level. On the segmental level the system is based on phonemes. The synthesis-by-rule system was judged to be useful when studying the impact of individual deviations on intelligibility, as our hypothesis was that profoundly-hearing impaired speakers have well-established speech habits, see section 2.2.3.

### **6.2.2. Assessment of the speech of three profoundly hearing-impaired children**

The children were educated mainly by sign-language. Their age at the time of recording was 11 to 15 years. Their pure tone averages were around 90 dBHL in the better ear (.5, 1, 2 kHz). It was apparent that the children had some residual hearing. Their speech was recorded when they read monosyllabic words, polysyllabic words, and a coherent text. In the words all the Swedish vowels and consonants in all phonotactically possible positions were represented. A phonetically trained listener transcribed the segmental production broadly, and the fundamental frequency and its range, pausing, speech rate, segment duration, and stress-patterns were analysed in the coherent text by means of oscillographic recordings. The assessment of the speech of the three children was made through traditional error analysis that did not pay any attention to contrastive function. The transcription was broad, and the deviations made by the children were translated into simplified phonetic rules. In this study, the intention was not to investigate what the speech of the children did express, but rather to gain knowledge of the relative effect on intelligibility of a specific phonetic deviation. Below, a systematic description of the three children's segmental and prosodic deviations is presented which formed the basis of the programmed phonetic rules that generated the synthetic speech.

#### **Consonant deviations**

The plosives of child A were all voiced except /g/ which was unvoiced. All the /s/ and /ʃ/ sounds coincided with the pronunciation of [ç]. There were no nasals and /m/ and /n/ were produced at the right place but in a wrong manner. The retroflexed consonants were not articulated as one sound but as two. Child B had plosives that were voiced in initial position and unvoiced and aspirated in final position, except for /g/, which was nasalized in medial and final position, see Table 6.1. The sibilants were very indistinct and preceded or followed by /k/.

Table 6.1. Description of the realized consonant systems of three profoundly hearing-impaired children in initial, medial and final positions. Shaded areas pertain to phonotactically impossible positions.

**Child A**

**Target**

|                         |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |
|-------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|
| p                       | t | k | b | d | g | f | v | s | ʃ | ç | j | h | m | n | ŋ | l | r  |
| <b>Initial position</b> |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |
| b                       | d | k | b | d | k | f | f | ç | ç | ç | d | h | b | d |   | l | ʀç |
| <b>Medial position</b>  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |
| b                       | d | k | b | d | k | f | f | ç | ç |   | d |   | b | d | ç | l | ʀç |
| <b>Final position</b>   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |
| b                       | d | k | b | d | k | f | f | ç | ç |   | d |   | b | d | ç | l | ʀç |

**Child B**

**Target**

|                         |    |    |    |    |    |    |    |    |    |    |   |   |   |    |    |    |   |
|-------------------------|----|----|----|----|----|----|----|----|----|----|---|---|---|----|----|----|---|
| p                       | t  | k  | b  | d  | g  | f  | v  | s  | ʃ  | ç  | j | h | m | n  | ŋ  | l  | r |
| <b>Initial position</b> |    |    |    |    |    |    |    |    |    |    |   |   |   |    |    |    |   |
| b                       | g  | g  | b  | g  | g  | f  | f  | ʃk | ʃk | kʃ | g | h | b | g  |    | g  | ʀ |
| <b>Medial position</b>  |    |    |    |    |    |    |    |    |    |    |   |   |   |    |    |    |   |
| p                       | t  | k  | p  | t  | ŋg | f  | f  | ʃk | ʃk |    | g |   | m | ŋg | ŋg | g  | ʀ |
| <b>Final position</b>   |    |    |    |    |    |    |    |    |    |    |   |   |   |    |    |    |   |
| ph                      | th | kh | ph | th | ŋg | fh | fh | ʃk | ʃk |    | g |   | m | ŋg | ŋg | lə | ʀ |

**Child C**

**Target**

|                         |   |   |   |   |   |   |   |    |   |   |   |   |    |    |   |    |   |
|-------------------------|---|---|---|---|---|---|---|----|---|---|---|---|----|----|---|----|---|
| p                       | t | k | b | d | g | f | v | s  | ʃ | ç | j | h | m  | n  | ŋ | l  | r |
| <b>Initial position</b> |   |   |   |   |   |   |   |    |   |   |   |   |    |    |   |    |   |
| p                       | t | k | p | t | k | f | f | ts | ç | ç | ç | h | mb | nd |   | l  | r |
| <b>Medial position</b>  |   |   |   |   |   |   |   |    |   |   |   |   |    |    |   |    |   |
| p                       | t | k | p | t | k | f | f | t  | ç |   | ç |   | mb | nd | ŋ | l  | r |
| <b>Final position</b>   |   |   |   |   |   |   |   |    |   |   |   |   |    |    |   |    |   |
| p                       | t | k | p | t | k | f | f | ç  | ç |   | ç |   | mb | nd | ŋ | lɛ | r |

The dental sounds /j/ and /l/ were often realized as a voiced velar sound. Many of the consonants were produced far back in the mouth. The retroflexed dental consonants were articulated as a sequence of /r/ and a dental, in accordance with the orthographic representation. The plosives of child C were unvoiced and some



fricatives sounded like [ç]. Like child B this child produced some consonants differently depending on the position in the word. Most of the nasals were followed by a plosive produced at the same place. Retroflexed consonant clusters were produced as separate sounds.

### **Vowel deviations**

Tables 6.2 - 6.4 give a detailed description of the children's vowel systems.

Table 6.2-6.4. *Description of the realised vowel system of three profoundly hearing-impaired children.*

#### *Child A*

| <b>Key word</b> | <b>Target</b> | <b>Realised as</b> |
|-----------------|---------------|--------------------|
| Sil (strainer)  | i:            | ɛ                  |
| Sill (herring)  | ɪ             | ɛ                  |
| Vet (know)      | e:            | ɛ                  |
| Vett (sense)    | ɛ             | ɛ                  |
| Säl (seal)      | ɛ:            | ɛ                  |
| Säll (blissful) | ɛ             | ɛ                  |
| Här (here)      | æ:            | æ                  |
| Herr (Mr)       | æ             | æ                  |
| Syl (pricker)   | y:            | œ                  |
| Syll (sill)     | ɣ             | θ                  |
| Föl (foal)      | ø:            | œ                  |
| Föll (fell)     | ø             | œ                  |
| För (prow)      | œ:            | œ                  |
| Förr (before)   | œ             | œ                  |
| Hus (house)     | ʌ:            | œ                  |
| Hund (dog)      | ə             | ə                  |
| Rot (root)      | u:            | ɔj                 |
| Rott (rowed)    | ʊ             | ɔj                 |
| Gå (go)         | o:            | o:                 |
| Gått (gone)     | ɔ             | ɔj                 |
| Hat (hate)      | ɑ:            | a:                 |
| Hatt (hat)      | a             | a                  |

*Computer-Based Speech Therapy Using Visual Feedback*

*Child B*

| Key word        | Target | Realised as |
|-----------------|--------|-------------|
| Sil (strainer)  | i:     | ə           |
| Sill (herring)  | ɪ      | ə           |
| Vet (know)      | e:     | e:          |
| Vett (sense)    | ɛ      | ɛ           |
| Säl (seal)      | ɛ:     | ɛ:          |
| Säll (blissful) | ɛ      | ɛ           |
| Här (here)      | æ:     | æ:          |
| Herr (Mr)       | æ      | æ           |
| Syl (pricker)   | y:     | ʏ:          |
| Syll (sill)     | ʏ      | ʏ           |
| Föl (foal)      | ø:     | œ:          |
| Föll (fell)     | ø      | œ:          |
| För (prow)      | œ:     | œ:          |
| Förr (before)   | œ      | œ           |
| Hus (house)     | ʉ:     | o:          |
| Hund (dog)      | ø      | ø           |
| Rot (root)      | u:     | ø:          |
| Rott (rowed)    | ʊ      | o           |
| Gå (go)         | o:     | ø:          |
| Gått (gone)     | ɔ      | ø           |
| Hat (hate)      | ɑ:     | ɑ:          |
| Hatt (hat)      | a      | a:          |

*Child C*

| Key word        | Target | Realised as |
|-----------------|--------|-------------|
| Sil (strainer)  | i:     | i:          |
| Sill (herring)  | ɪ      | ɪ:          |
| Vet (know)      | e:     | e:          |
| Vett (sense)    | ɛ      | ɛ:          |
| Säl (seal)      | ɛ:     | æ:          |
| Säll (blissful) | ɛ      | ɛ:          |
| Här (here)      | æ:     | æ:          |
| Herr (Mr)       | æ      | æ:          |
| Syl (pricker)   | y:     | i:          |
| Syll (sill)     | ʏ      | ʏ:          |
| Föl (foal)      | ø:     | ø:          |
| Föll (fell)     | ø      | œ:          |
| För (prow)      | œ:     | œ:          |
| Förr (before)   | œ      | œ:          |
| Hus (house)     | ʉ:     | ʉ:w         |
| Hund (dog)      | ø      | ʏ:          |
| Rot (root)      | u:     | əu:w        |
| Rott (rowed)    | ʊ      | œ:          |
| Gå (go)         | o:     | əo:         |
| Gått (gone)     | ɔ      | œ:          |
| Hat (hate)      | ɑ:     | a:          |
| Hatt (hat)      | a      | a           |

Child A reduced the Swedish vowels to six short and two long vowels. All vowels were nasalized. All close vowels were pronounced half-opened. Child B had nine long and seven short vowels. Deviations were found in both quantity and quality. All back vowels were realized as central [ə]. The child's lip positions were correct, while the tongue positions sometimes are incorrect. This shows the impact of lip-reading on the speech of severely hearing impaired children. Child C had as many as twelve long vowels and

only one short vowel. Unrounded vowels were nasalized and some rounded vowels were diphthongized.

### **Prosodic deviations**

Child A had a monotonous rhythm due to incorrect pausing between syllables. Her fundamental frequency was extremely high with a lowering at the end of every vowel. The child also had extended segment duration. Child B had a normal speaking rate. However, the speech was not fluent, due to the fact that he breathed after every second word. He emphasised the beginning and the end of every sentence. The most interesting deviation was his vowel-dependent fundamental frequency variation because of an excessive articulatory tension. Child C had a normal fundamental frequency but a remarkably slow tempo. This child made a pause after every word and extended the last syllable in every phrase. Furthermore, he extended the occlusion phase in the production of the plosives /p, t, k/. The prosodic deviations of each child formed the basis for the rules used to generate the prosody of the simulated speech.

### **6.2.3. Listening test**

Based on the above assessments, simplified phonetic rules were constructed that generated the synthetic speech for listening tests. The speech of each child was represented in 32 sentences. Every sentence contained four key words. Every deviation or combination of deviations was presented in groups of four sentences (16 keywords). Table 6.5 shows the different combinations of deviations used in the listening test, corresponding to corrections of the different deviations.

The listeners were 21 students at the Department of Linguistics, Stockholm University. They listened through earphones to one child at a time with a short interval between each child. Every sentence was presented twice. The listeners wrote down what they understood. The test started with eight synthesized sentences with no deviations to familiarize the listeners with synthetic speech.

Table 6.5. *Combination of deviations used to simulate the speech of three profoundly hearing-impaired children and the corresponding corrections of deviations measured through a listening test.*

| <b>Type of deviations</b>               | <b>Corrections of deviations</b> |
|---|----------------------------------|
| No deviations                           | All deviations corrected         |
| Vowel deviations                        | Consonants and prosody           |
| Consonant deviations                    | Vowels and prosody               |
| Prosodic deviations                     | Vowels and consonants            |
| Vowel & consonant deviations            | Prosody                          |
| Vowel & prosodic deviations             | Consonants                       |
| Consonant & prosodic deviations         | Vowels                           |
| Vowel & consonant & prosodic deviations | No corrections                   |

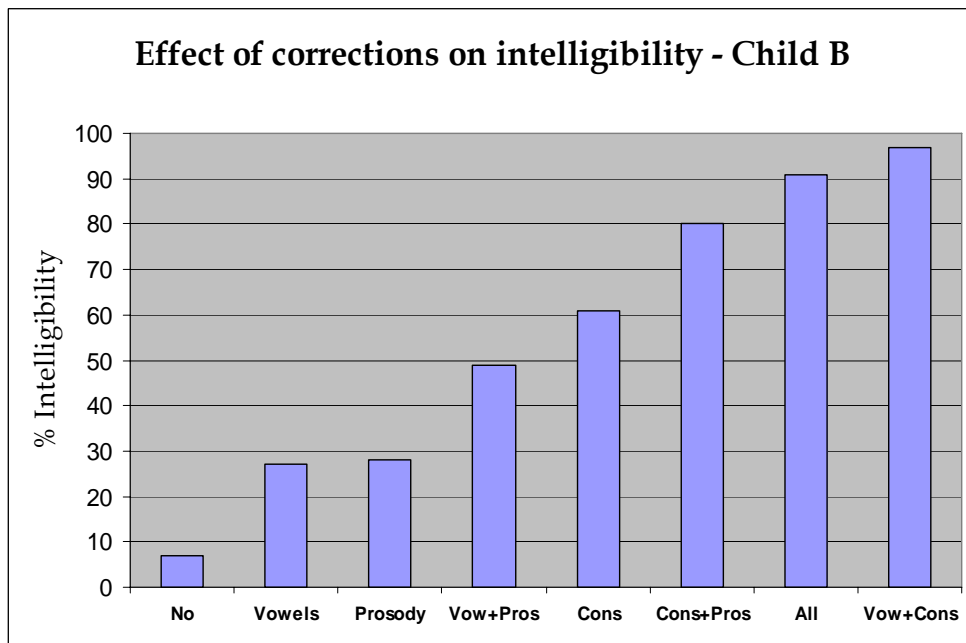
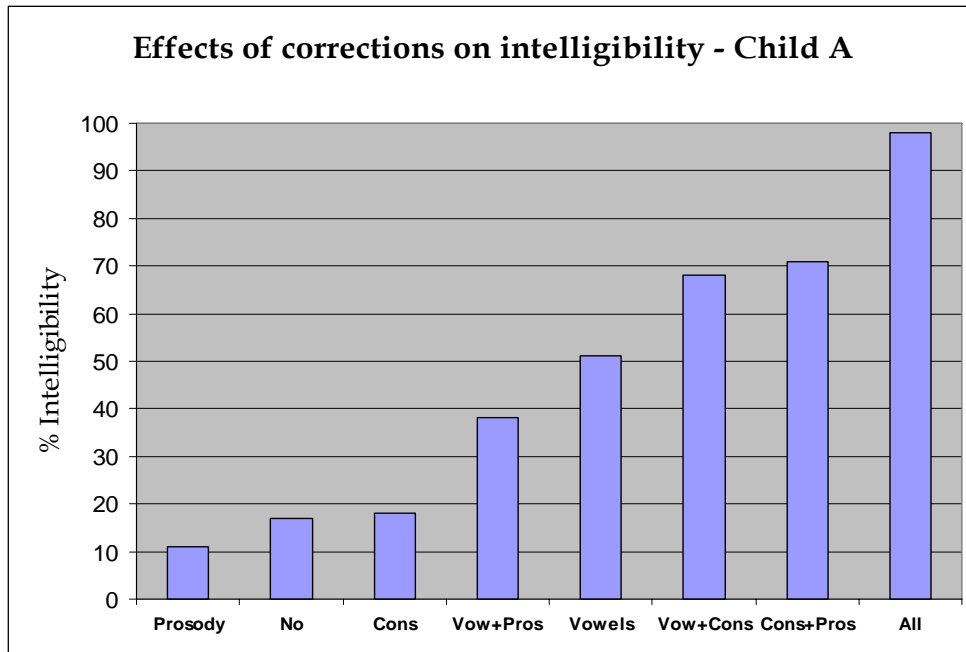
#### **6.2.4. Results**

Figures 6.1-6.3 show the result of the listening test of the simulated speech of the three children. The deviations are arranged according to increasing speech intelligibility. The results are classified in three groups of intelligibility:

- Unintelligible speech = 0-25% correctly identified key words
- Semi-intelligible speech = 25-50% correctly identified key words
- Intelligible speech = 50-100% correctly identified key words

Based on these results some recommendations could be given regarding the speech training of the children. The speech of child A was unintelligible as the intelligibility was 17%. The vowels should be corrected first, increasing the intelligibility to 51%. After that, her consonants should be trained. Without consonantal deviations, the intelligibility would increase further to 66%. Finally, the prosody deviations should be corrected. The simulated speech of child B was also unintelligible with only 7% intelligibility. The consonant deviations had a seriously deteriorating effect on the intelligibility. If these are corrected, the intelligibility will increase to 61%. Thereafter, the vowels should be trained. With the vowels corrected the intelligibility will further increase to 97% and the speech will be completely intelligible.

*Interaction between individual deviations and speech intelligibility*



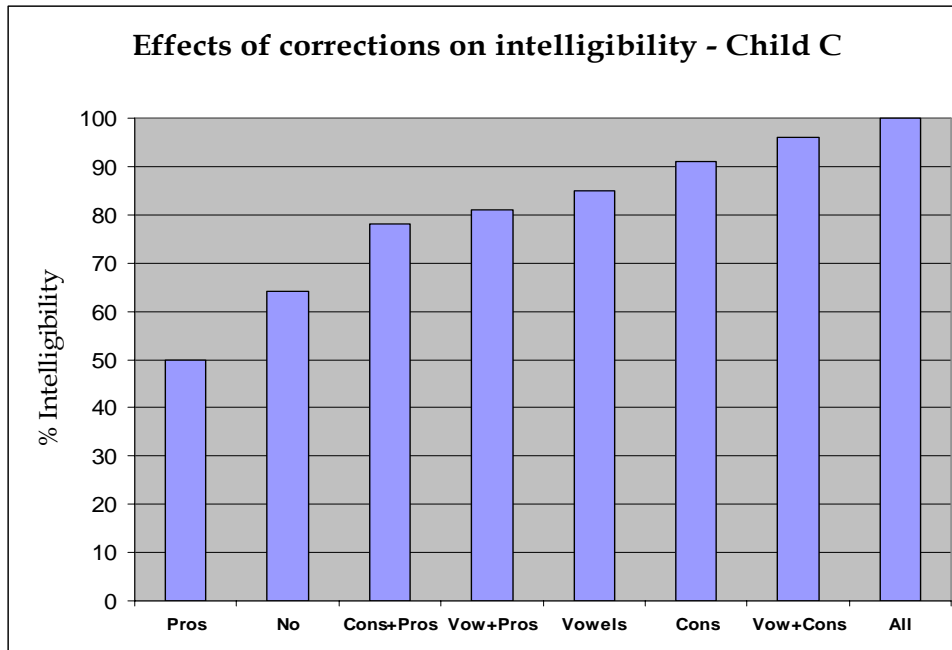


Figure 6.1-6.3. Intelligibility of synthesized versions of the speech of the hearing-impaired children. Corrections of vowel (VOW), consonant (CONS) and prosodic (PROS) deviations have been simulated.

Finally, the prosodic deviations should be corrected, but not motivated primarily by an expected increase in intelligibility. However, the simulated speech of child C was intelligible (64%). If the consonantal deviations were corrected first, the intelligibility would be as high as 91%. Then a correction of the vowel deviations would increase the intelligibility to 96%.

An interesting result of this study was that the effect of the prosodic deviations on intelligibility was negligible, perhaps due to the short sentences. For child A and child C the intelligibility even decreased when only the prosodic deviations were corrected. This was probably owing to the slow speaking rate and the frequent pausing that gave the listener more time to segment and interpret the meaning of the sentences. Thus, a slow tempo impacts the intelligibility positively on the condition that the speech contains grave segmental deviations. It makes the speech unnatural but still not unintelligible.

### **6.2.5. Conclusions**

One conclusion drawn from this study is that a correction of errors in the segmental articulation of speech is probably necessary before a significant improvement in intelligibility can be achieved. An improvement in the prosodic elements alone does not result in an immediate gain in intelligibility - it may instead improve the naturalness and overall quality of the speech. This had already been shown by Osberger and Levitt (1979) and later by Maassen and Powel (1985) in studies where temporal deviations in sentences spoken by profoundly hearing-impaired children were artificially corrected towards normal speech through digital manipulation. The samples were then subjected to intelligibility tests. It was observed that segmental deviations, as opposed to prosodic deviations, were primarily responsible for reducing the intelligibility of the speech of profoundly hearing-impaired children. Only when the segmental articulation was satisfactory, could improvement in intelligibility be obtained from the correction of prosodic deviations.

Another important conclusion is that each child's speech production is unique. Therefore, it is extremely important that an individual diagnosis is made prior to therapy to find those deviations that impact on the intelligibility of the speech. A complete diagnostic assessment includes both an analysis of the existing articulation skills as well as an analysis of how the articulation is realized in linguistic use. Because of the fact that profoundly hearing-impaired children depend on lip-reading when learning speech it is also important to investigate how the interference of visibility affects their possibilities to realize phonological contrasts in the Swedish language. Methods to analyze deviant speech to assess what the speech expresses are discussed in the next chapter.





## ***7. Phonetic realizations of phonological systems***

### **7.1. Introduction**

Several studies of the phonological systems of hard-of-hearing children (West & Weber, 1973; Abberton, Fourcin, Hazan, 1985; Oller & Kelly, 1974; Dodd, 1988; Oller & Eilers, 1981; Öster 1991, 1992a, 1992b, 1995c) have shown that, although profoundly hearing-impaired children make deviations in production, they possess some kind of abstract and stable phonological system. Studies of the phonological systems of children with profound hearing losses have been inspired by Clinical Phonology (Hodson, 1980; Shriberg & Kwiatkowski, 1980; Grunwell, 1987; Ingram, 1989) that was developed to analyse phonological processes in disordered speech through descriptions of contrasts and processes. According to Grunwell (1987), there are five major clinical assessment procedures that are based on a phonological process analysis:

- Phonological Process Analysis (PPA; Weiner, 1979)
- Natural Process Analysis (NPA; Shriberg and Kwiatkowski, 1980)
- Assessment of Phonological Processes (APP; Hodson, 1980)
- Procedures for the Phonological Analysis of Children's Language (PPACL; Ingram, 1981)
- Phonological Assessment of Child Speech (PACS; Grunwell, 1985)

Linguistic theories such as Taxonomic Phonemics and Generative Phonology, Jakobson's Child Phonology theories (1968) and Stampe's Natural Phonology (1979) have contributed to the development of Clinical Phonology. Taxonomic Phonemics, developed in the 1940's and 1950's, classifies contrasting sound units or phonemes which cause a difference in meaning. Generative Phonology, described by Chomsky and Halle in the 1960's uses distinctive features and formal rules to describe the sound patterns of a language. Jakobson (1968)

claimed that children learn contrasts, not individual sounds, in a certain order from maximal to minimal contrast. Stampe (1979) stressed that small children have a tendency to simplify adult speech by innate rules or processes.

## **7.2. Speech assessment methods**

All assessment is time-consuming and laborious. However, to prevent speech therapy from becoming a series of meaningless "articulatory gymnastics" sessions, the therapist must be prepared to spend time carrying out an appropriate diagnosis prior to therapy. Therapy and assessment are inseparable, as assessments are required regularly during the training program.

A deviation may be looked upon as a phonetic deviation or a phonological deviation. A phonetic deviation affects the naturalness of the speech and could be the result of an incorrect phonation or an articulatory movement that has no effect on the speaker's ability to signal meaning differences. The presence of a phonetic deviation is defined as a phonological deviation if it changes the meaning of the word. Phonological deviations affect the intelligibility of a child's speech and cause meaning differences in spoken language. In speech therapy, it is most important to first concentrate on a special treatment of phonological deviations in order to improve intelligibility and then focus on phonetic deviations to improve the naturalness of the speech.

### **7.2.1. Phonetic error analysis**

As has been stated in the Introduction, deviant speech is traditionally assessed using phonetic error analyses only. A phonetic error analysis is proper to use with less deviant speech, as for instance for assessment of the speech of moderate hearing-impaired children, as the articulation skill is compared with that of normally hearing speakers without paying attention to phonetic contexts or to the production's contrastive function in a specific language. Such an analysis uses a coarse phonetic transcription, missing important articulatory details, and provides only information on what the client is not capable of articulating. Hence, only speech sounds that a client never articulates correctly will be treated. A phonetic error analysis

takes for granted that the sounds that the child articulates correctly in isolation, the existing articulation skills, also are used correctly. Obviously, this approach provides insufficient and misleading information on which to base an effective speech therapy program with more deviant speech, as it pays no attention to either the usage of the existing articulation skills or to whether a deviant articulation might signal a "correct" contrast. Profoundly hearing-impaired speakers are dependent on visual cues and often try to realise the visual representation of a phonetic contrast that is signalled auditorily. This means that a deviant production may be an attempt to realise a speech sound contrast (Monsen, 1976) as discussed in sections 2.2.3 and 7.3.3.

In a study by Öster (1989a) it was shown that profoundly hearing-impaired children understand the phonological contrasts between visually similar consonants but have difficulties in realising them correctly. Systematic deviations from normal were found in the speech of a 15-year-old, prelingually and profoundly hearing-impaired boy, educated using sign language, when he produced Swedish stops. His pure tone average in the better ear was 108 dBHL. A systematically deviant contrast was found in initial position between voiced and unvoiced bilabial stops and in final position between voiced and unvoiced dental and velar stops. The child contrasted unvoiced stops from their voiced cognates in initial position by lip protrusion and in final position by adding a neutral vowel instead of a voicing contrast.

### **7.2.2. Phonological analysis**

Nowadays, the assessment of exceptionally deviant speech, such as the speech of profoundly hearing-impaired children and some types of L2-speech, is more concentrated on deeper aspects of speech production like information about the linguistic use of the existing articulation skills in all possible positions and in what way the child's phonetic realisations differ from the normal model. This is done through phonological analyses.

According to Grunwell (1987) children's speech problems are largely confined in effect to the patterns of consonantal usage. This is very obvious for profoundly hearing-impaired children who develop systematic contrasts for consonants, which however differ from those

of normally hearing speakers, principally due to an increased reliance on visual cues. To identify the phonetic deviations that affect a child's ability to signal meaning differences in spoken language, a very detailed transcription must be used and an expansion of IPA symbols and diacritics is necessary (see Appendix I). As concerns vowels, deviant speech most often involves systematical substitutions, which is why a coarse transcription provides sufficient information.

If speech therapy is made without awareness of the child's phonological system, it may destroy already built-up couplings between abstract entities and articulation. This might even result in a decreased intelligibility of the speech after training owing to a reduced number of contrasts. A speech therapy program based on a phonological assessment will be more directed towards training of distinctive features and of the correct production of contrasts between often visually similar consonants, rather than towards a correct pronunciation of a specific sound. When a deviantly produced distinctive feature is corrected, the articulation of all speech sounds that contain this feature will improve.

In the following sections descriptions of phonological analyses are shown of a profoundly hearing-impaired child's consonant production and of an L2-speaker's (Bosnian speaker) vowel production.

### **7.3. Description of a phonological analysis of a profoundly hearing-impaired child's consonant production**

Öster (1995a) describes a phonologically based assessment method on which to base speech therapy. The analysis outlines a child's unique phonological system (phonetic and phonological inventories) and expresses what the child's speech expresses. It is done in three steps, which are described below. First the individual child's existing articulation skills are assessed. After that the child's usage is analysed. Finally the idiosyncratic realisations of phonological contrasts are analysed through the detailed phonetic transcription. The outcome of such a complete assessment provides significant and valuable information to base the therapy on.

As an example of this method, the video-recorded speech of a fifteen-year-old child, educated using sign language, will be phonologically analysed below (Öster 1995a). His pure-tone-average (at .5, 1 and 2 kHz) in the better ear was 108 dBHL. Obviously, speech information was not perceived at all without a hearing aid. The child was video-recorded saying 58 polysyllabic Swedish words. Within the constraints of the Swedish phonotactic system, these words sampled each Swedish consonant at least twice in the initial, medial, and final positions. Consequently, each phoneme was sampled at least 6 times. The video-recorded speech was transcribed in detail using IPA symbols and some of the specific diacritics used by Bush et al. (1973); Grunwell (1987) and Roug, Landberg, and Lundberg (1989) developed to transcribe babbling and phonetic development in early infancy, see Appendix 1.

### **7.3.1. Step 1: Analysis of the existing articulation skills**

Many children articulate several speech sounds correctly in isolation but have problems in producing them contrastively in various contexts. Therefore it is important to concentrate upon those speech sounds which the child can produce in isolation or in syllables, but which are not correctly realised in a linguistic context. These speech sounds should be established in all contexts before attempts are made to introduce the speech sounds, which are not yet within the child's productive inventory. By assessing the existing articulation skills, it is also possible to exclude a motor disorder as a cause of a phonological deviation.

Table 7.1 shows the consonants that are present (the existing articulation skills) and the consonants that are absent in the child's inventory (the blank cells). This is done without regard for the accuracy of the child's productions. Although the child has an articulatory knowledge of eight of the eighteen Swedish consonants, some deviations occurred in different positions, due to limited information of phonetic features, limited knowledge of the rules of pronunciation, etc. The table shows that he produced only five of his eight consonants correctly in the initial position, six of them in the medial position and five of them in the final position. The consonants which caused deviations were /g, ç, m/.

Table 7.1. The table shows the existing articulation skills of a profoundly hearing-impaired child (in bold), the word position where the consonant was correctly articulated and the word position where a deviation occurred, (Öster 1995a).

| IPA      | Existing articulation skills | Used correctly |        |       |
|----------|------------------------------|----------------|--------|-------|
|          |                              | Initial        | Medial | Final |
| p        |                              |                |        |       |
| t        |                              |                |        |       |
| <b>k</b> | <b>k</b>                     | yes            | yes    | yes   |
| <b>b</b> | <b>b</b>                     | yes            | yes    | yes   |
| <b>d</b> | <b>d</b>                     | yes            | yes    | yes   |
| <b>g</b> | <b>g</b>                     | no             | no     | no    |
| <b>f</b> | <b>f</b>                     | yes            | yes    | yes   |
| v        |                              |                |        |       |
| s        |                              |                |        |       |
| ʃ        |                              |                |        |       |
| <b>ç</b> | <b>ç</b>                     | no             |        |       |
| j        |                              |                |        |       |
| h        |                              |                |        |       |
| <b>m</b> | <b>m</b>                     | no             | yes    | no    |
| n        |                              |                |        |       |
| ŋ        |                              |                |        |       |
| l        |                              |                |        |       |
| <b>r</b> | <b>r</b>                     | yes            | yes    | yes   |

### 7.3.2. Step 2: Assessment of the usage of the existing articulation skills through a detailed phonetic analysis

After this, the usage of the consonants that the child knows how to articulate is assessed, because the consonants that the child can articulate in isolation are not always correctly realized in different linguistic contexts. The child needs to learn to use his existing articulation skills appropriately in all word positions before new speech sounds can be taught. The assessment is done by analysing

the detailed phonetic transcription made. The transcriber must try to capture all the relevant visual details, either at the time of recording or through the use of video, in order to complete the picture of the child's contrastive system, according to Summerfield (1979). In Table 7.2 the existing articulation skills are shown in bold style /k, b, d, g, f, ç, m, r/ and the deviations made for those consonants in initial, medial and final word-position.

Table 7.2. *Phonetic consonant deviations made in various word positions in a profoundly hearing-impaired child's speech. The existing articulation skills are shown with arrows (Öster 1995a). Shaded cells = phonotactically not applicable.*

|   | IPA | Initial        | Medial         | Final                 |
|---|-----|----------------|----------------|-----------------------|
|   | p   | b              | b              | b <sup>ə</sup>        |
|   | t   | d              | d              | d <sup>ɫ</sup>        |
| → | k   |                |                |                       |
| → | b   | <u>b</u>       | <u>b</u>       | b <sup>ɫ</sup>        |
| → | d   |                |                |                       |
| → | g   | k              | k              | k                     |
| → | f   |                |                |                       |
|   | v   | fK             | fK             | fK                    |
|   | s   | ç              | ç              | ç                     |
|   | ʃ   | ç <sup>k</sup> | ç              | ç <sup>ə</sup>        |
| → | ç   | k              |                |                       |
|   | j   | d <sup>F</sup> | d <sup>F</sup> | d <sup>F</sup>        |
|   | h   | ∅              |                |                       |
| → | m   | <u>m</u>       | m <sup>b</sup> | <u>m</u> <sup>ə</sup> |
|   | n   | <u>n</u>       | <u>n</u>       | <u>n</u>              |
|   | ŋ   |                | g              | g <sup>ə</sup>        |
| → | l   | d              | d              | d                     |
| → | r   |                |                |                       |

Blank cells indicate that the consonant was pronounced correctly. Hence only the consonants /k, d, f, r/ were correctly produced in all word positions. It is also interesting to study how the child used his

existing articulation skills for those consonants that he had not yet learned. Special attention should be paid to the consonants /g, ʒ, m/. The child has an articulatory knowledge of these sounds, but his productions always (except /m/ in medial position) involve substitutions of other phonemes when he used them in a linguistic surrounding. It can be seen that [g] is used for /ŋ/ but when asked to produce /g/ it is realized as [k].

Similarly [ç] is substituted for /s/ and /ʃ/ but when the child is asked to produce /ç/ he substitutes it with a [k]. The nasal /m/ is produced almost correctly in the medial position but in the initial and final position it is phonetically similar to /b/.

### **7.3.3. Step 3: Assessment of idiosyncratic realisations of phonological contrasts and regular error patterns**

Finally by studying Table 7.2 closely, important information is offered about substitutions and articulatory details used to signal differences between consonants. Through this analysis it is possible to determine whether a deviant pronunciation is, in fact, an attempt by the child to realize a phonological contrast and to find the child's regular error patterns (phonological rules). Some of the deviant phone types represent different phonemes despite the phonetic similarity. For example, many of the deviant phone types are similar to /b/ and /d/. It can be assumed that the child, despite the phonetic similarity to [b], makes contrasts between /p/, /b/ and /m/ in the initial position through a voiced non-aspirated /p/, lip protrusion for /b/ and nasal air emission for /m/. It is also very likely that the child, despite the phonetic similarity to [d], contrasts /t, d, j, n, l/ in the final position, through a non-audible release for /t/, retroflexion and frication for /j/, nasal air emission for /n/ and retroflexion for /l/. Obviously, the child understands the phonological contrasts between these visually similar consonants, but has difficulties in realising them correctly.

Many of the deviant phonological processes found in the speech of normally hearing children (Nettelbladt, 1983), as for instance simplifications like "stopping" of fricatives, can also be



found in the speech of prelingually and profoundly hearing-impaired children. Normally hearing children produce these simplifications during a short period of their development. However, profoundly hearing-impaired children often preserve these processes together with other deviant processes that are due to the special condition under which they learn to speak, (Oller and Kelly, 1974; Dodd, 1974).

## **7.4. The importance of a detailed phonetic transcription**

The outcome of this phonological analysis through a detailed transcription has provided significant and valuable information that could not have been obtained if only a traditional phonetic error analysis and a coarse transcription had been used. By studying the detailed transcription, revealing important articulatory details, it is possible to determine if the child's deviant pronunciations are, in fact, attempts to realize phonological contrasts due to visible interference.

Figure 7.1 shows the different outcomes from a coarse and a detailed phonetic transcription of the child's consonant production. A coarse phonetic transcription would have missed the important articulatory details that the child used, trying to realise the contrasts discussed above.

The result of the coarse phonetic transcription would have been that /p, m/ were substituted by /b/ and /t, j, n, l/ by /d/, i.e., that the child was missing /p, m, t, j, n and l/ in his phonetic inventory. Articulatory training of each of these consonants in isolation might dissolve the unique phonological system of the child. Instead, the speech therapy must be directed to deal with all these contrasting consonants simultaneously.

Visual feedback has shown to be excellent for developing contrasts between visually similar consonants in the speech of profoundly hearing-impaired children. In a study by Öster (1989b, 1989c) improvements in producing durational contrasts between long and short vowels and voicing contrasts between velar stops in medial and final word positions were found after some training with a computer-based speech therapy program with visual feedback.

This type of training made the children aware of their deviant way of expressing contrasts and helped them train correct contrasts through the immediate and meaningful visual feedback.

| <u>Coarse transcription</u> | <u>Detailed transcription</u>    |
|-----------------------------|----------------------------------|
| [p]                         | [b]                              |
| [b]                         | [ <u>b</u> ]    initial position |
| [m]                         | [ <u>b̃</u> ]                    |
| <hr/>                       |                                  |
| [t]                         | [d <sup>̣</sup> ]                |
| [d]                         | [d]                              |
| [j]                         | [d F]    final position          |
| [n]                         | [ <u>d̃</u> ]                    |
| [l]                         | [d ]                             |

Figure 7.1. The outcome of a coarse and a detailed phonetic transcription of a profoundly hearing-impaired child's phonetic deviations, (Öster 1995a).

## 7.5. Therapy based on existing skills

As has been shown above, a phonological assessment provides information on existing skills, i.e. the amount and type of correctly articulated consonants and vowels regardless of position, how these are used and if a deviant production in fact may be an attempt to realize a phonological contrast.

The fact that this kind of assessment shows what the child's speech expresses means that it provides suitable information to base the speech therapy on. To expand the usage of existing articulation skills of a child before new speech sounds are taught has shown to be an effective and fruitful therapy method in order to extend the child's inventory and improve the intelligibility of his/her speech.

To get a general view of the average phonetic and phonological competence of a group of 11 profoundly hearing-

impaired children a study was made by Öster (1991). The phonetic analysis also showed how visibility affects the acquisition of profoundly hearing-impaired children's segmental production and which vowels and consonants they learn first. Eleven prelingually profoundly hearing-impaired children, educated by sign language, participated in this study. One child was eleven years of age, while the others ranged from fifteen to seventeen years. Their pure tone averages (0.5, 1 and 2 kHz) were between 90-108 dBHL in the better ear. The children were video-recorded when they read a list of familiar words also illustrated by pictures. The word list contained at least two presentations of all Swedish consonants and vowels in initial, medial and final position, if phonotactically possible. A narrow phonetic transcription was made from the video-recorded speech by the author using IPA symbols and additional diacritics, see Appendix 1.

Figure 7.2 shows the number of children who, at least once in the material, articulated each Swedish consonant correctly regardless of position and without any reference to contrastive function. The children, as a group, articulated 70% of the Swedish consonants correctly. The consonants which most of the children articulated correctly were those that are visually contrastive and easy to lip-read, such as bilabial and dental stops, labiodental fricatives and the lateral consonant. The children also produced the unvoiced stops, /p, t, k/, much better than the voiced stops, /b, d, g/.

This result shows that profoundly hearing-impaired children first learn to pronounce those consonants with visible speech organ gestures through imitation. However, those distinctive features like, for example, nasality/non nasality and voiced/unvoiced that are invisible through lip-reading can very well be taught to the children by a meaningful graphical visual feedback, for instance through different colours; green for unvoiced, red for voiced and blue for nasals.

Correctly articulated Swedish consonants

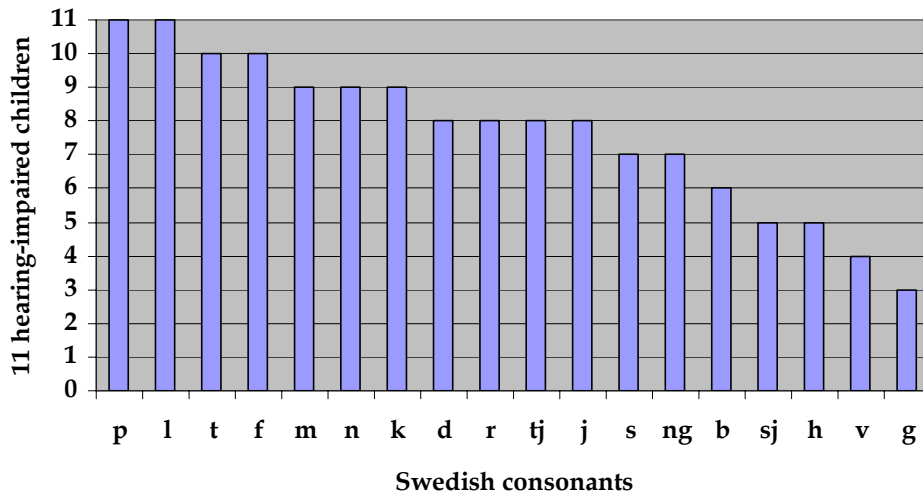


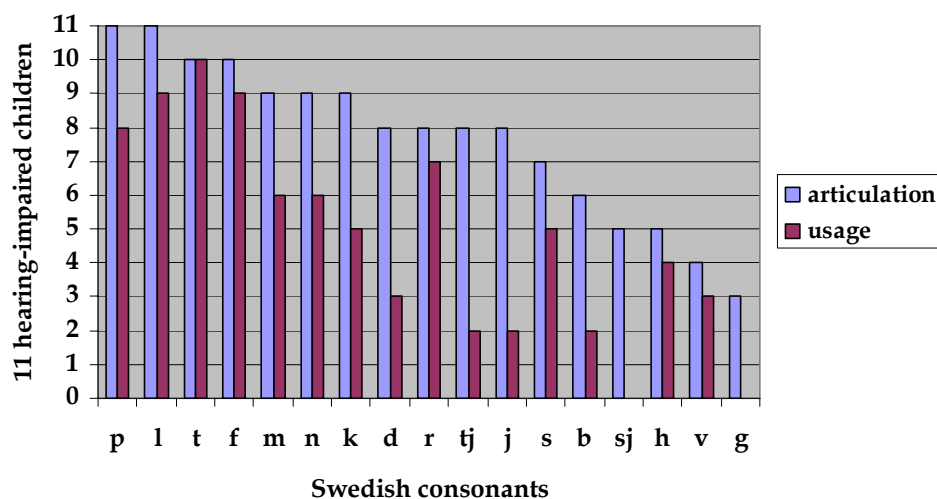
Figure 7.2. Existing articulation skills of the Swedish consonants by eleven profoundly hearing-impaired children of grade 10, (Öster 1991).

Figures 7.3-7.5 show the correct usage (black bars) in different positions of those consonants which the children knew the articulation of (grey bars). The children, as a group, correctly articulated 70% of the Swedish consonants, but could only make correct use of 43% of them in initial position, 50% of them in medial position and 50% of them in final position. This indicates a discrepancy between correct articulation regardless of position and correct articulation with reference to contrastive function and phonetic contexts. A large difference in heights between the two bars, representing correct articulation (phonetically) and correct use (phonologically) indicates that this particular consonant was difficult for profoundly hearing-impaired children to realize correctly in this position. By the means of a phonological assessment of a detailed phonetic transcription, as in Table 7.2, it was possible to examine in what way a consonant was realized in various word positions and what constitutes the discrepancy between correct articulation and the effect of phonetic context.

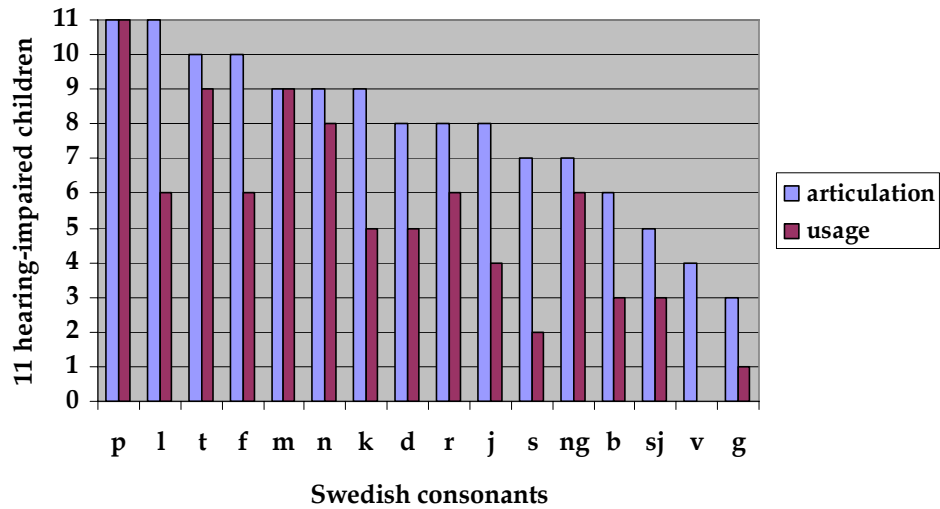
Only a few consonants were controlled as well phonologically as phonetically like /t/ in initial position, /p/ and /m/ in medial position and /ŋ/ in final position. Others varied in correct use according to word position. For example, five children could articulate /fj/ in isolation, but none could articulate it correctly in the initial position, only three in medial position and four in final position. Five of seven children were able to use /s/ correctly in the initial position, five of seven in medial position and four of seven in final position. None of the four children who articulated /v/ correctly in isolation pronounced it correctly in medial and final position. However, three of the children pronounced /v/ correctly in initial position. Some of the consonants, /j, d, k, b/, were equally difficult to use and pronounce correctly in all word positions.

The same discrepancy between correct articulation of vowel quality and their representation in spoken language can be seen in Figure 7.6. Profoundly hearing-impaired speakers have a preference for long vowel duration, which according to Oller (1981) is explained by the fact that they depend heavily upon vision and that vision simply does not operate in as rapid a time frame as audition. In this study, the main interest was to find out which vowels are generally acquired first of all through their visual accessibility; the quantity was of less importance than the quality.

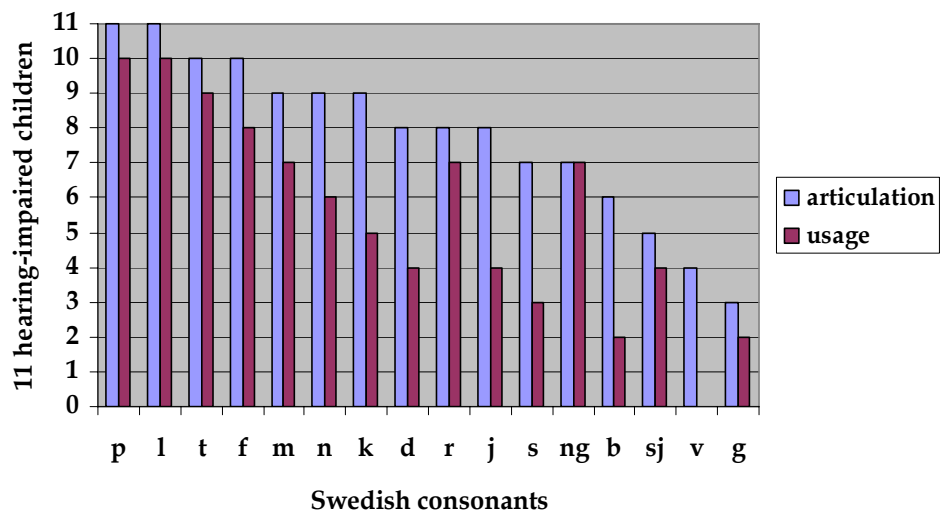
### Initial position



### Medial position



### Final position



Figures 7.3-7.5. Grey bars show the number of children who articulated each consonant correctly. Black bars show the number of children who made correct use of their articulation in initial, medial and final word-positions.

The result shows that all children controlled the quality of the short front and open vowel /a/, which means that this vowel could be seen as the first acquired vowel by prelingually and profoundly hearing-impaired children. Almost all children knew the production and the rounding contrast between /ø:/(ö:) and /ɛ:/(ä:) thanks to visibility.

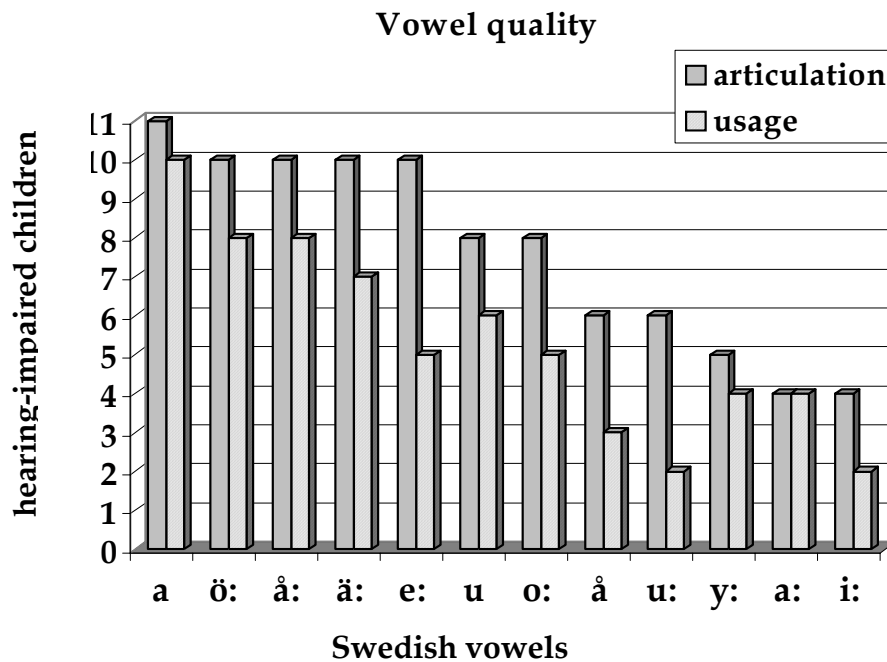


Figure 7.6. Grey bars show the number of children who articulated each vowel correctly regardless of phonetic context. Light-colored bars show the number of children who made correct use of their articulation in spoken language. The vowels are represented by orthographic symbols.

Also the quality of the back close-mid long vowel /o:/(å:) that is pronounced with a relaxed and flat tongue and a rather visible open jaw position was well articulated by ten of eleven children. However, the peripheral vowels /a:/(a:), /i:/(i:), and /u:/(o:) with the most distinct articulatory positions were those vowels that the children had less articulatory control of. Many children made several confusions and substitutions when they tried to use the vowel quality correctly in spoken language. Especially the quality of /e:/

(e:) and /u:/ (u:) and to some extent the quality of /ɛ:/ (ä:), /u:/ (o:) and /ɔ/ (å) seems to be difficult to use for many children. These vowels were substituted by other vowels.

The first step of a therapy based on existing articulation skills (those consonants and vowels that the child knows the articulation of regardless of position) should be to train correct pronunciation of these, in all possible contexts in spoken language, in order to level out the discrepancy between correct articulation and correct usage.

Individual differences will determine the particular training material and order selected for training, since the training is based on existing skills. The treatment is influenced by the Sensory-Motoric Approach devised by McDonald (1964). Bisyllabic drills are used, beginning with small changes of tongue movement between the production of the vowel and the consonant followed by progressively larger changes of movement. Game-like strategies with visual feedback have shown to be helpful and motivating in increasing the children's accuracy of articulation, repetition of syllables with different articulatory positions, coordination of articulators, breath control, and rhythm and stress placement.

## **7.6. Description of a phonological analysis of a Bosnian speaker's production of Swedish vowels**

As has been said before in section 1.1.2, L2 speakers articulate some target phonemes deviantly because they differ in number and quality from those of their mother tongue (L1). Many vowels are similar by sight and cause systematic vowel substitutions in their speech. In this case a coarse transcription of the pronunciation of vowels provides sufficient information to base the therapy on. Table 7.3 shows a phonological analysis based on a coarse transcription of a Bosnian speaker's production of Swedish vowels.

The Bosnian speaker trained together with twelve adult immigrants with a computer-based speech training system with audio-visual feedback of both perception and production of spoken Swedish. Six of them originated from Bosnia, the others from Cuba, Peru, Saudi Arabia and Russia. On an average they participated in a



total of six half-hour training sessions. The training was carried out twice a week. Their speech was analysed phonetically and phonologically to get an individual diagnosis of their deviations, their existing articulation skills as well as the linguistic representation.

In the first column "Articulates", all the client's different productions made of each of the Swedish vowels are shown. The more alternatives there are for each vowel the more uncertain the client was about how to articulate the vowel. The result is similar to a phonetic error-analysis, as it shows what the client is not capable of producing correctly. The symbol (: ) after a vowel means that the client sometimes produced the vowel as a short and sometimes as a long variant.

The next step is to find those vowels that the client knows how to produce correctly, i.e. the existing articulation skills. This is compiled in the next column "Existing articulation skills". It can be seen that the Bosnian speaker knew the quality of 9 of the 22 Swedish vowels.

The last column reveals how the client substituted the vowels that he could articulate for other vowels that he could not articulate. In other words, the last column "Represents" shows what his speech expressed and the representation of his existing articulation skills in spoken language. For instance, it can be seen that he used the rounded back vowel /ɔ(:)/ for all rounded back vowels and also for the rounded front vowel /ø(:)/.

A close analysis of his articulation and usage of Swedish vowels gives important information about the client's difficulties and what causes them. This can be of great use in the therapy. It can be seen that he has difficulties in producing the difference in rounding between /I, i: and Y, y:/. Actually, he is not able to produce any of the rounded front vowels /y:, Y, u:, ø, ø:, ø/ but substitutes these with the closest back vowel. This is due to the fact that front rounded vowels are missing in his native language. He has learnt to withdraw his tongue together with lip rounding, as only back vowels are present in his native language.

Another explanation for some of his deviations is the fact that he has problems with the orthography. The vowels /ɛ:-o:/ (ä-å) are confused as well as the vowels /u:-ø:/ (o-ö) and /ɑ:-ɔ:/ (a-å). This is

most likely due to a lack of knowledge of the Swedish pronunciation rules.

Table 7.3. A phonological analysis of a Bosnian speaker's production of Swedish vowels.

| IPA | Articulates | Existing articulatory skill | Represents        |
|-----|-------------|-----------------------------|-------------------|
| i:  | i: ɪ        | i:                          | i: y: ʏ           |
| ɪ   | ɪ           | ɪ                           | ɪ ʏ i:            |
| e:  | e:          | e:                          | e:                |
| ə   |             |                             |                   |
| ɛ:  | ɛ           | ɛ:                          | ɛ o:              |
| ɛ   | ɛ:          | ɛ                           | ɛ:                |
| æ:  |             |                             |                   |
| æ   |             |                             |                   |
| a   | a(:)        | a(:)                        | a ɑ:              |
| ɑ:  | ɔ a(:)      | —                           |                   |
| y:  | i:          | —                           |                   |
| ʏ   | i: ɪ        | —                           |                   |
| u:  | u:          | —                           |                   |
| ʊ   | ʊ           | —                           |                   |
| ø:  | ɔ(:) u:     | —                           |                   |
| ø   | ɔ ʊ         | —                           |                   |
| œ:  |             |                             |                   |
| œ   |             |                             |                   |
| u:  | ʊ ɔ(:) u:   | u:                          | ʉ: ø: u:          |
| ʊ   | ɔ           | ʊ                           | ø u: ø ɔ          |
| o:  | ɔ(:) ɛ:     | —                           |                   |
| ɔ   | ʊ ɔ         | ɔ(:)                        | ɑ: u: ø ø: ʊ o: ɔ |

Such an assessment should be the basis for speech therapy of L2 speakers. It is important to make the client aware of his difficulties

and the cause of them. His usage of his existing articulation skills should be trained before new vowels are taught.

## **7.7. Conclusions**

The outcome of the phonological analyses described in this chapter reveals that there are both similarities and differences between hearing-impaired children and second language learners when learning Swedish and that a phonological analysis provides important information which is of great value both in CBST and CAPT. In both cases the phonological analyses showed what the profoundly hearing-impaired child and the Bosnian speaker were not capable of producing correctly and how they substituted the phonemes that they could articulate for other phonemes that they could not articulate in a linguistic context. However, for the profoundly hearing-impaired child, due to interference from the visual modality, the use of a fine phonetic transcription is necessary to reveal details of the idiosyncratic realizations of normal phonological contrasts.



## ***8. Design of visual feedback in Swedish computer-based therapy systems***

### **8.1. Introduction**

Today, computer-based speech therapy systems with visual feedback have been given a positive response, and they are commonly used by speech therapists and teachers in Sweden (Öster, 1995b, 1996). Computer-based speech therapy with visual feedback has turned out to be a valuable and efficient complement to traditional speech therapy for all types of hearing-impaired children and is meant to be supervised by a teacher.

In this chapter my research dealing with design, functionality and type of visual feedback of three different Swedish systems will be reviewed. The systems will be related to the different needs, aims and possibilities of computer-based speech training with visual feedback of moderate, severely and profoundly hearing-impaired children and L2 learners. This is of importance for how the type and pattern of the external visual feedback ought to be shaped and what type of pedagogical applications should be used to fit each group.

Two of the systems are commercial products and one is a research prototype in progress. All of them are product-oriented systems.

### **8.2. Visual feedback in speech therapy**

Visual information as an additional source of feedback can be provided by traditional training (by the teacher alone without the assistance of computer-based systems) or by a computer-based system.

The visual feedback provided by traditional therapy is given by the teacher and can be summarized as follows:

- external
- delayed
- verbal
- subjective

The visual feedback provided by a computer-based speech therapy aid alone is quite opposite. It is:

- internal
- in real time
- non verbal
- objective

A combination of these two possibilities of feedback, meaning that a teacher uses a computer-based system with visual feedback as a complement to the traditional training, ought to give the best result during treatment. In this way the teacher has all possibilities to assist the child in developing awareness of his own production, comparing his production to a target and following his own progress.

### **8.2.1. Nature of feedback**

The nature of feedback of the product-oriented computer-based speech therapy systems is parametrical as it gives acoustic information (spectral or temporal analysis) about the speech product in real time. Within motor learning theory, this feedback is called Knowledge of Result (KR) and is an essential component of learning a motor behaviour (Mahshie, 1995, 1996). The feedback of process-oriented systems, on the other hand, gives instruction on how to move the articulators to reach the target production. Within motor learning theory this feedback is called Knowledge of Performance (KP) and according to Mahshie (1995) this feedback is most useful in the instruction phase to get the child aware of his deviant production and understand how a correct realisation should be produced.

### **8.2.2. Type of feedback**

The following types of visual feedback will be exemplified in detail in the following sections where each therapy system is described. They will also be discussed according to clinical usability.

#### ➤ **Animated graphics**

- for basic *awareness*, which illustrates selected dimensions of speech such as pitch, loudness, timing, and the presence or absence of voicing, to establish a relationship between the

graphics and vocal aspects. This can be a balloon, for instance, that gets larger in relation to the loudness of the child's vocalization.

- for correct *control* of the voice, pitch, or intonation
- **Evaluative feedback** provides an indication of success or failure or provides a measure of "goodness". It can be either:
  - *exact* and tells whether the production was correct or incorrect
  - *acceptable* and shows that the child is "on the way" and has made some improvements. This can be shown through a digit from 1 to 5 according to how correct the production was.
- **Navigational feedback** informs the child how to reach the production and is instructional and informative. This can be shown by
  - *comparable*, often through a comparison of the child's production with a correct production of a reference model, see Figure 8.5, or
  - *criterion-referenced* based on visual targets of the child's production that are compared to the norm and to their own improving productions, as in the visual maps of OPTACIA, see Figure 8.18, or
  - *articulation pictures* showing correct position of the vocal organs during articulation of individual speech sounds.
- **Rewarding feedback** should stimulate and motivate the child to go on with the training. The training must be pleasant and amusing. The rewarding feedback can reward a production that is
  - *hitting the mark* and consists of both visual and auditory effects such as a duck that moves to the right and opens its mouth to show that it is happy with the articulation. If not, the duck will remain standing with its mouth closed. It can also be
  - *encouraging* as it reinforces any improvement through for instance a flower in a pot that develops from nothing to a

sprout, to a few leaves, to a blooming flower according to the correctness of the pronunciation.

- **Recognition** used together with different graphics helps the child with the repetitive and additional training that is needed to stabilise a recently learned pronunciation.
  - *phoneme-based comparison* by comparing spectra with models of the child's best productions
  - *ASR* (automatic speech recognition) of words with targets of the child's best productions.
- Finally, **optional auditory feedback** can be used with hearing-impaired children with moderate hearing losses and with L2 learners through the playback of stored recordings and through rewarding sound effects.

## **8.3. The IBM SpeechViewer**

### **8.3.1. Description of the system**

The original SpeechViewer program was developed 15 years ago (Crepny et al. 1983) and is now available in twelve different languages. It has been translated and adapted to Swedish in three different versions by Öster (1988a, 1989b, 1996; Lotsson, 2001). The system consists of a PC with a colour display. A microphone, an amplifier, and a speaker connected to a M-ACPA card allow the user to input, store and analyse speech and then display it and play it back.

The software contains 13 interactive programs, shown in Figure 8.1, aimed at assisting a child in achieving awareness and control over various speech attributes such as voicing, timing, pitch and loudness as well as refined articulation and prosody. Feedback is given immediately through a variety of graphical designs and game-like strategies synchronised with optional auditory playback. The programs are grouped into three sections for different areas of use: basic awareness, skill building and phonology exercises, and speech patterning modules.





Figure 8.1. The thirteen different programs included in SpeechViewer III.

### 8.3.2. Type of visual feedback in different exercises

The displays for **basic awareness** illustrate selected dimensions of speech such as pitch, loudness, duration, voicing and breath control for phonation by playful and easy exercises. These exercises are meant to be used by young profoundly hearing-impaired children to develop fine control of their voices. All the exercises are easy to understand and provide visual feedback in the form of **animated graphics** for basic *awareness* according to the child's production. In the "Sound" exercise, something moves on the screen showing that a sound is produced. A helicopter flies when the child changes the pitch up and down in the exercise "Pitch range". In the exercise "Voicing," voiced sounds turn a figure red and unvoiced sounds turn it green. As an example, Figure 8.2 illustrates how to train correct loudness in the exercise "Loudness range".

## Computer-Based Speech Therapy Using Visual Feedback

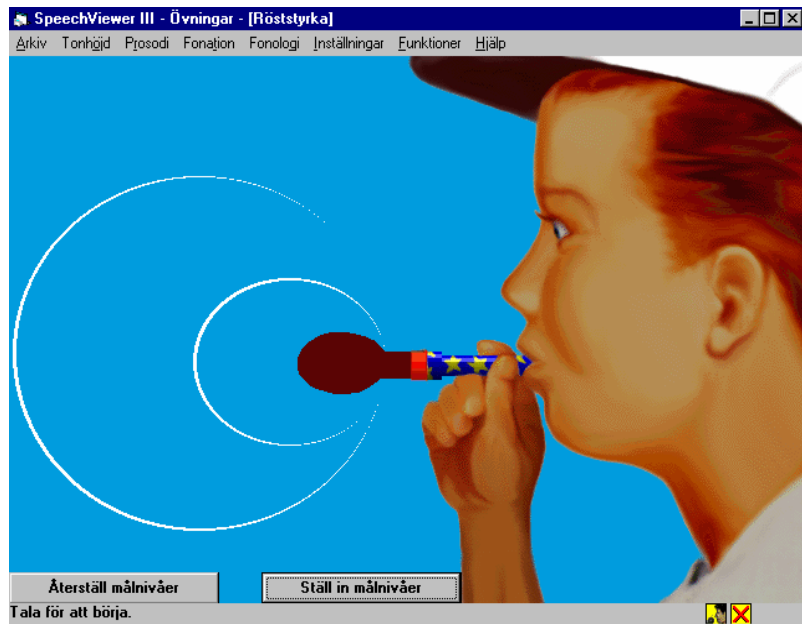


Figure 8.2. A screen of the exercise “Loudness range”.

The child can see the balloon grow as he makes the sound louder. It is also possible to set two different targets for matching low and high volume.

The **skill building modules** provide a game-like strategy to strengthen ability in pitch, timing, voicing, and speech sound production. In the “Pitch control” module for example, the teacher is able to arrange a “steplechase course” consisting of a number of targets and obstacles placed over the screen. The child has to move a figure through the course, hitting targets and avoiding obstacles by controlling his pitch, see Figure 8.3. **Animated graphics** are used for *control* of the voice and pitch. An *exact, evaluative feedback* is given after complete success. A **rewarding** visual and optional **auditory feedback**, given when the child *hits the mark*, should stimulate the child to go on with the training.

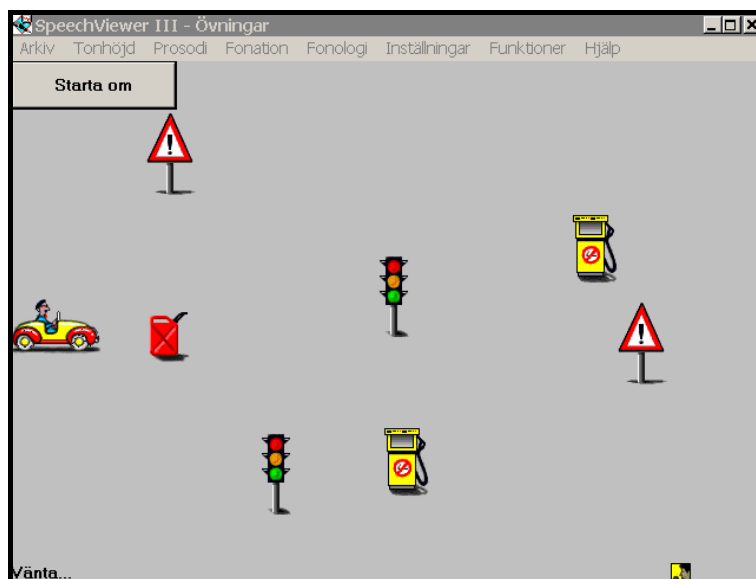


Figure 8.3. Example of an exercise for pitch control. The goal is to hit the targets (petrol can and petrol pumps) and avoid the obstacles (traffic signs) by controlling the car by raising and lowering pitch.

**The phonology exercises** aim at establishing a consistent and intelligible pronunciation of the Swedish phonemes. Through a *phoneme-based comparison*, that compares spectra with models of the child's best productions, the computer evaluates the accuracy of the child's production. These exercises are to be used in the phase of speech therapy where the child is "on the way" to acquiring a correct production of a particular speech sound. The child's own "best productions" should be recorded and stored as a target and used as models in exercises for accuracy, matching words containing the target phoneme and contrasting. The goal is accomplished when the child's best production becomes his or her most common production. With regard to inter-speaker variability problems, dialects, and deviant speech due to hearing disorders this strategy, to use the child's own best productions as models, is also used with the CISTA-Aid training (Youdelman, 1994) as well as with the ISTR-Aid (Kewley-Port & Watson, 1995). In previous versions of SpeechViewer, standard models, consisting of normally hearing children and adults, were constructed as models for training in the instruction phase. This approach was of course not satisfactory and was a common concern of many users of the earlier versions

(Mahshie, 1998). In the exercises for “Accuracy”, “Chaining” and “Contrasting” the sustained phonation of one to four phonemes can be trained or matched in words. Figure 8.4 shows an example of the Two-Phoneme Contrast exercise, where the task is to maneuver the jeep through an obstacle track by pronouncing /s/ to turn to the left and by pronouncing (tj) [ç] to turn the jeep to the right.

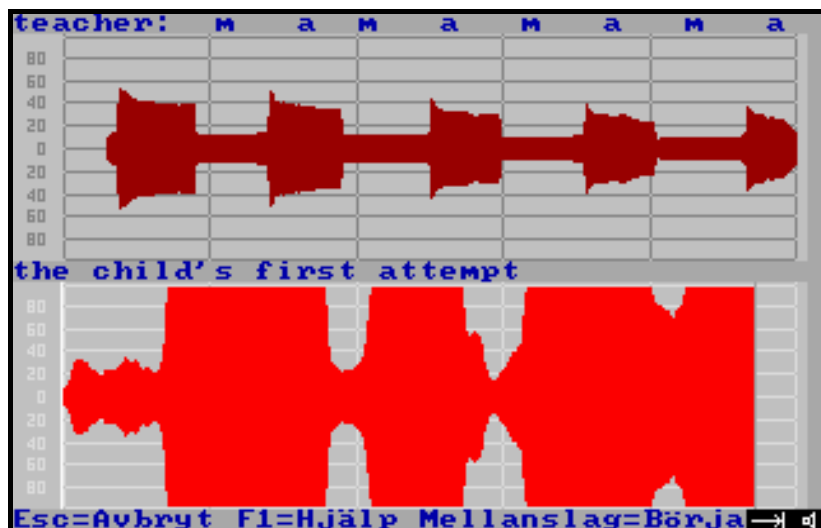


Figure 8.4. Display of the phonology exercise “Contrasting”. The task is to manoeuvre the jeep along the track avoiding the animals. When pronouncing /s/ the jeep goes to the left and by pronouncing /tj/ [ç] the jeep goes to the right.

All these exercises use **animated graphics** for *control* and give an **evaluative exact** feedback as well as a **rewarding** visual and optional **auditory feedback** when the child *hits the mark* to stimulate the client to go on with further repetitive and additional training.

**The speech patterning programs** display the speech signal as oscillograms, spectra or spectrograms for analysis and training of refined articulation and prosody. The Speech Patterning Module “Pitch and Loudness” displays F0 and/or intensity in real time. A split screen provides a comparison of a learner’s utterance with a model of the teacher’s. The comparison of the two visual patterns gives the learner possibilities to discriminate between any important differences of the distinctive features in the Swedish language. This

possibility of comparison of both the visual pattern and the auditory feedback has been shown to be very effective in the instruction phase for both L2 learners and hearing-impaired children (Öster 1995b, 1995c, 1998, 1999a). This *comparable navigational feedback* is logical and shown in real time and makes the clients understand in what way their speech differs from the teacher's. Figure 8.5 shows a training session of repeated syllable with this module. In the upper portion, the therapist's pattern of repeated production of the syllable /ma/ is shown. In the lower portion, a profoundly hearing-impaired child's deviantly produced nasal in the repeated production of the syllable is shown.



*Figure 8.5. Display of loudness over time in "Pitch and Loudness". In the upper portion, the therapist's pattern of repeated production of the syllable /ma/ is shown. In the lower portion, a profoundly hearing-impaired child's deviantly produced nasal in the repeated production of the syllable is shown.*

Throughout the system the **clinical management** allows the creation of sustained phoneme models, speech models, and client profiles as well as the reporting and management of client data. Statistics of each activity are saved to follow the improvements of the child.

## **8.4. Box of Tricks**

### **8.4.1. Description of the system**

This teaching and training system for speech- and hearing-impaired children aged 4-10 was developed in the EU project "SPECO" within the INCO-Copernicus program between 1998 and 2001 (Vicsi et al. 2000). Five partners were involved in the project: the Technical University of Budapest, Hungary, the University of Reading, United Kingdom, the University of Maribor, Slovenia, the Robot Control Software, Hungary and KTH, Sweden. The system was developed for four languages: Hungarian (Varázsdoboz), English (Box of Tricks), Swedish (Trollerilådan), and Slovenian (ARTI).

The system is product-oriented and offers parametrical acoustic feedback of the speech in real time. The system displays important articulation features graphically. These are called speech pictures and are presented in a clear and entertaining way. The speech pictures make it easy for a child to distinguish between an incorrect articulation and the correct articulation of a reference model. The children are able to learn a correct pronunciation by looking at their own speech pictures and comparing them with the speech picture of the correct pronunciation. The structure of the system is presented in Figure 8.6.

The system comprises a general language-independent measuring tool and a database editor. The separation of the complex sounds into their component frequencies is done in critical filter bands, from 80 Hz to 8 kHz. 20 critical bands are used. The database editor makes it possible to construct different modules and vocabularies for other languages. Two language-dependent speech and picture databases, one for teaching and training vowels (the "Vowel Support") and the other for fricatives and affricates (the "Fricative Support"), are constructed for each of the four languages. All training words are illustrated for small children who have not yet started reading.

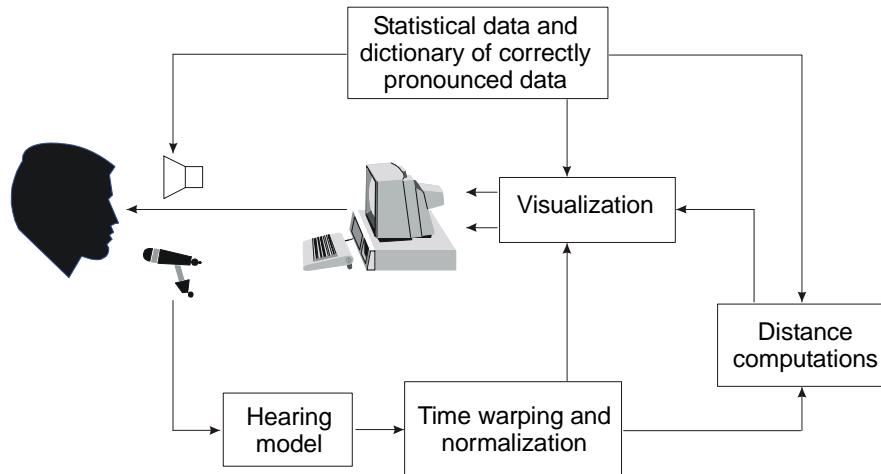


Figure 8.6. The structure of Box of Tricks

Figure 8.7 shows the main menu, consisting of the microphone adjustment possibility, the sound preparation exercises, the vowel training vocabulary, the fricative training vocabulary and the intonation exercises. Figure 8.8 shows an example of illustrated words for training, containing the vowel /ä/ [æ, ε] in the middle position.

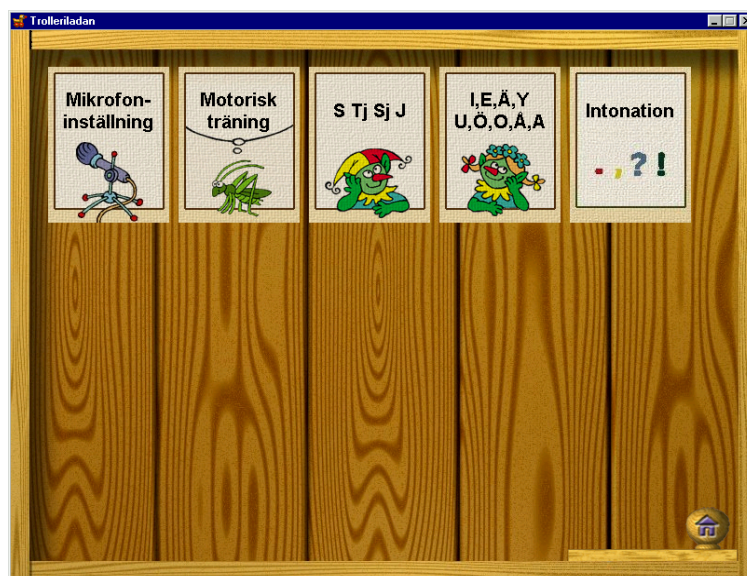


Fig.8.7. Main menu of the Swedish version of Box of Tricks.





Fig.8.8. Choice of illustrated words for training of the Swedish vowel (ä)

#### 8.4.2. Databases for reference speech and normal-hearing children for similarity comparisons

Recording and storing a child's best pronunciation, as the therapist has to do when working with SpeechViewer, has been shown to place too great a demand on some therapists' experience of basic computer knowledge as well as articulatory and acoustic phonetics (Öster 1996). Therefore the basic idea when developing Box of Tricks was to incorporate a reference speaker as a model for visual and auditory training to make this system more user-friendly. The reference speaker was a ten-year-old Swedish girl with clear and good pronunciation of standard Swedish. The system consists of two types of databases, one for the reference speech that makes comparisons with a correct model possible, and the other called "the child database" that was used to determine models and limit values for the phoneme-based recognition of vowels and fricatives and for background pictures that represent the lower and the upper limit of the spectral envelope for an acceptable pronunciation, see section 8.4.4.



### **Database for reference speech**

The Swedish system contains stored reference speech examples recorded as the girl read 17 isolated phonemes, 174 speech units, 350 words, 43 minimal word pairs and 127 sentences. Each utterance was recorded three times using a special editor incorporated in the system. The best utterance was saved and chosen to appear as the reference model in the exercises.

The system presents acoustic parameters of the reference speech in a way that is understandable and interesting for young children, also correct from the acoustic-phonetic point of view. Amusing illustrations, called "Speech Pictures", emphasize either important speech parameters to make them understandable or show the reference speaker's pronunciation of phonemes, words and sentences on the upper half on the screen for comparison with the child's production on the lower part of the screen. Articulation pictures, spectrograms and target spectra are also used to make the visual feedback intelligible. All training phonemes and words are illustrated for small children who have not yet learned to read. All reference speech training material is segmented, and time-warping algorithms are used to present the reference speech and the child's speech immediately below each other, even if the speaking rates are different. The child's pronunciation of a vowel and a fricative consonant in isolation, in a syllable, in a word and in a sentence is compared according to similarity with the reference speaker's pronunciation. Different types of feedback are used to motivate and stimulate the child to go on with the training.

### **Child database**

A Swedish database was collected to measure the acceptable range of speech parameter variation of Swedish children around six years of age with intelligible speech. The text material contained all Swedish sibilants and /j/, all long vowels, and the three short vowels (u, a, å) (differing in quality between long and short variants) in isolation, syllables and short series of syllables, words and sentences. In words, all examined speech sounds occurred in all possible positions. One, two and three syllabic words were included.

31 children between 6 and 9 years of age were recorded as they read the text material. Table 8.1 shows age and gender

distribution of the children. All children belonged to a local primary school in the south of Stockholm. All children had normal hearing and a distinct and good pronunciation. The group was assessed by the author to be a good sample of Swedish children who talk properly for their age.

Table 8.1. *Age and gender distribution of speakers.*

| <b>AGE</b> | <b>BOYS</b> | <b>GIRLS</b> |
|------------|-------------|--------------|
| <b>9</b>   | <b>4</b>    | <b>4</b>     |
| <b>8</b>   | <b>4</b>    | <b>3</b>     |
| <b>7</b>   | <b>7</b>    | <b>5</b>     |
| <b>6</b>   | <b>2</b>    | <b>2</b>     |

The recordings took place in a quiet room in the school building. The recordings were made on a Casio digital tape recorder DA-7 and a Monacor ECM-100 electret microphone that was fitted in a stand. The person who recorded the speech of the children was sitting beside each child and took care that the children did not change their position or that they did not touch the microphone or the stand. Children who could not read fluently repeated by ear. The total recording time per speaker was about 15 minutes.

The recordings were classified by the author as good, acceptable, and unacceptable productions. The averaged energy values and spectrum lines of the good and acceptable productions were used to determine allowed spectral deviations for the vowels and fricative background pictures. During the training the actual spectrum lines must fall within the two spectrum lines in the speech picture, see Figure 8.10.

### **8.4.3. Training method used in Box of Tricks**

The system is based on important steps of traditional speech therapy (Öster et al., 1999b, Vicsi et al., 2000). These steps are:

- Sound preparation:
  - Loudness
  - Pitch
  - Rhythm
  - Spectrum
  - Voicing
- Sound development:
  - articulation pictures
  - isolated phonemes
  - syllables with the target phoneme in all positions
  - repetition of syllables
- Sound sequences in words, minimal word-pairs and sentences:
  - fricatives can be practised at the beginning, in the middle or at the end of a word
  - vowels can be practised in monosyllabic or polysyllabic words
- Intonation exercises
- A user manager gives the possibility to
  - set the children's data
  - describe their speech problems
  - document their improvements.

### **8.4.4. Type of visual feedback in different exercises**

#### **Sound preparation**

In the sound preparation exercises, the child has the possibility to train different speech parameters like loudness, rhythm, sustained sound, voicing, and pitch. The aim is to get the child used to paying attention to the screen and be aware of important speech parameters. The child can see how a figure moves further up when the loudness increases, or keep the pitch or loudness steady for a short time to move a butterfly over a worm or for a longer time over a flower; the

neck of a duck moves up and down according to the child's pitch and so on. All the exercises are easy to understand and provide visual feedback for *awareness* in the form of **animated graphics** according to the child's production.

Figure 8.9 shows an example of pitch training with one of the sound preparation exercises. The neck of the duck moves up and down according to the child's pitch. The pair of cherries and the bee above them provide two targets for practice.



Figure 8.9. A display of the Pitch module of the Sound preparation exercises.

## Sound development

Among the sound development programs there is a possibility to study *articulation pictures* giving **navigational feedback** of correct position of the vocal organs during articulation of all phonemes that can be trained. Vowels and fricatives can be trained in isolation or in all positions in a syllable and short units. Frequency spectrums are shown on the screen, which are calculated and displayed every 20ms. In the "Isolation" exercise the form of the spectrum depends on the phoneme that is being trained. When correct pronunciation is attained, the spectrum falls inside a path that represents the lower and the upper limit of the spectral envelope for an acceptable pronunciation, based on the child database described in section 8.4.2. Figure 8.10 shows examples of speech pictures representing target spectra of accepted isolated production of (ä) [æ:] and (u) [u:]. The

display to the left shows a successful pronunciation of the vowel (ä) [æ:] as the spectrum falls inside the path. However, in the display to the right the active spectrum does not match the stored model of (u) [u:].

The automatic feedback, here shown as a flourishing flower, is based on a distance calculation between the spectral components of the stored spectrum and the active spectrum and gives the child a **rewarding feedback** that is *encouraging*. It can also be varied to display a digit, showing the outcome from 1-5, where 5 is the best.



*Figure 8.10. Two examples of training results with the sound development exercise Isolation. The automatic feedback, in the form of a flower, is one of several variants of the encouraging evaluative visual feedback in the system, (see text for details).*

This type of feedback is an **acceptable evaluative feedback**. It ranks the performance of the child and shows when the child is on a winning streak. The use of automatic feedback provides opportunities for children to practise with the system alone or at home with parents. By using background pictures for comparison with a correct model this module also supplies a **comparable navigational feedback**.

The calculated average spectrum patterns of sibilants based on the child databases for each of the four languages are presented in Figure 8.11.

Computer-Based Speech Therapy Using Visual Feedback

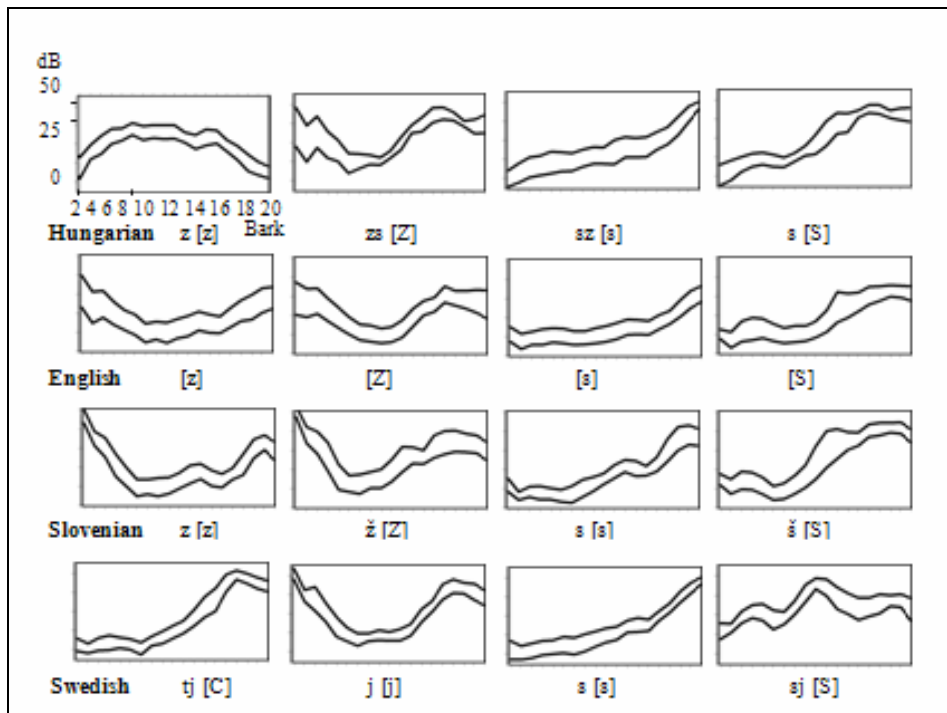


Figure 8.11. Average spectra of sibilants used for the four languages represented in Sampa symbols, see Appendix 2..

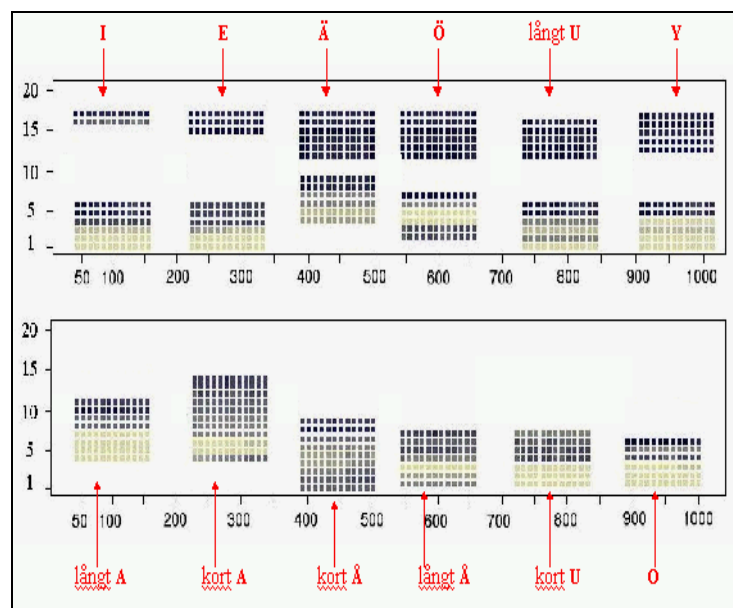
The background pictures in the “Syllable-training” exercise in Figure 8.12 emphasize the energy measured in each frequency band of a spectrogram. The child has to focus on the segmented part containing the background picture illustrating the target phoneme for training, in this example /s/ as in (sun).



Figure 8.12. Two examples of training results with the sound development exercise Syllable-training.

The task for the child in the case of the sibilants is to cover the clouds, see Figure 8.12, in the lower part of the display with dots representing spectral energy, but not to cover the other parts of the symbolic background picture. In Figure 8.12 two displays of the training of an /s/ in the syllable (is) are shown. In the display to the left, a sibilant is produced in which the noise frequency is too low, which can be seen in the lower portion of the screen. The upper portion of the screen shows the model of the reference speaker. The background picture of the sun is covered with dots and the result is consequently 1 point. In the display to the right, a correct pronunciation is shown. The child's production in the lower part of the display looks very similar to the correct model in the upper part of the display and the automatic feedback shows 5 points.

In Figure 8.13 spectrograms of the nine long and three short Swedish vowels that can be trained are shown.



*Figure 8.13. Spectrograms of the Swedish vowels used in Box of Tricks.*

The spectrograms form the basis of the symbolic background pictures shown in Figure 8.14 that are used as targets in the vowel support menu. The task for the child is to cover the boxes representing the lowest formants, and not to cover the symbolic background picture representing the vowel.



Computer-Based Speech Therapy Using Visual Feedback

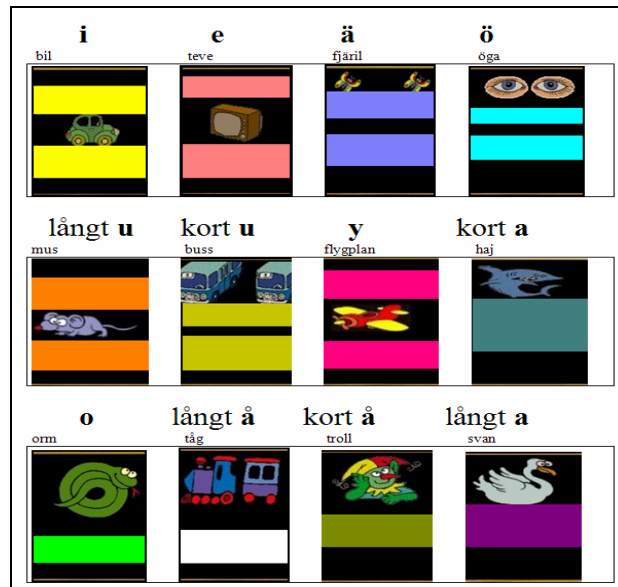


Figure 8.14. The symbolic background pictures of the Swedish vowels used in Box of Tricks.

Figure 8.15 shows an example of how the symbolic vowel speech picture for /i/ is used in a word training layout with a **rewarding feedback** that is *encouraging* (a moving duck that opens his mouth).

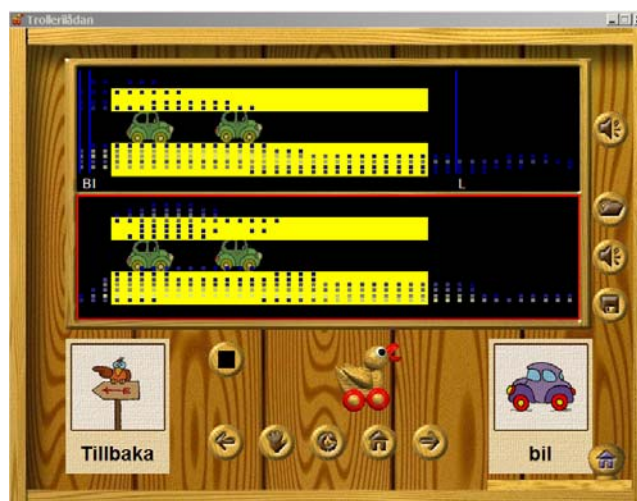


Figure 8.15. A good result of a child training the vowel /i/ in a single word. The feedback consists of an encouraging rewarding feedback as well as a comparable navigational feedback (comparison with a correct model).



## **Sound sequences**

When it comes to training the phonemes in words, minimal word-pairs and sentences, all phonemes can be trained in different positions and contexts. Fricatives are presented in CV, VCV, VC and VC-VC-VC positions and connected with all vowels. In the Fricative Support, all phonemes are presented in initial, medial and final position, and in the Vowel Support, all phonemes occur in one syllable and polysyllable words. Minimal word-pairs are used to train differences between two phonemes in similar words by visual speech pictures. Sentences containing all phonemes, from simple to more complex are also available for training.

In all these exercises an *acceptable evaluative* feedback is used. It ranks the performance of the child and shows when the child is on the winning streak. It can also be *encouraging rewarding* as it reinforces any improvement. The use of an automatic feedback provides opportunities for children to practise with the system alone or at home with parents. By using background pictures for comparison with a correct model a *comparable navigational feedback* is also used. The exercises work together with **recognition** through a *phoneme-based comparison* of the child's production with the pronunciation of the reference speaker.

## **8.5. OLP (Ortho-Logo-Paedia) Therapy**

### **8.5.1. Introduction**

The OLP-therapy method was developed within an EU Quality of Life and Management of Living Resources project, coordinated by the Institute for Language and Speech Processing, Athens, Greece, with participation from France (Arches), Greece (Altec S.A. and Logos Centre for Speech-Voice Pathology), Spain (Universidad Politécnica de Madrid), Sweden (KTH), and the UK (Sheffield University and Barnsley District General Hospital). Three basic types of pathologies were addressed in the OLP project. These were clients with dysarthria (English language), cleft palate (Greek language) and children with hearing impairments (Swedish language). This organization provided the project with a natural division of client

groups into languages and pathologies, thus enabling each clinical partner to concentrate on one particular pathology.

The project aimed at improving the quality of life of persons with articulatory impairments by applying a new technological aid to support (not replace) traditional speech therapy at the level of articulation, and making it available over the Internet. The therapy method integrates automatic speech recognition technology based on the child's best productions, and the therapy is tailored specifically for each child. OLP follows standard therapy design through therapy schedules of levels of therapy, which are specified based on knowledge of therapy design needed for each target client group. For Swedish prelingually and severely hearing-impaired children, the levels are specified in isolated vowels, syllable, repetition of same syllables, repetition of different syllables, monosyllabic words, polysyllabic words, and short phrases. Different types of exercises with a choice of technological tools are available at the various levels. Table 8.2 shows the available tools that are integrated in the OLP system and the tasks that are executed through them.

Table 8.2. *Function of available software tools and their use in different tasks. (Protopapas, A., 2004, User's Manual OLP document QL1971-ILS-IN-C-097-a3).*

| <b>Tool</b>       | <b>Functions</b>  | <b>Tasks</b>  |
|-------------------|---|---|
| <b>Recorder</b>   | Records sound from the microphone saving the recordings as audio files  | Record words for assessment. Record single sounds to train phonetic maps. Record words to train recogniser.   |
| <b>Trainer</b>    | Creates a speech recogniser based on a set of recorded sounds   | Train syllable recogniser.<br>Train word recogniser.  |
| <b>Recogniser</b> | Matches sounds from the microphone to a set of "known" words learned previously during "training" and displays the result graphically in the context of a "game". | Evaluate the production of a syllable and a word using graphical displays. Evaluate production of one command word from a word set relative to confusable others. |

|                                      |   |  |
|--------------------------------------|---|--|
| <b>STAPtk</b>                        | Support the user in recording sounds, labelling them, designing, creating, and using phonetic maps. | Label recorded sounds. Design a phonetic map display. Train a phonetic map. Display degree of match between spoken sound and a set of targets on a phonetic map. |
| <b>Pitch display</b>                 | Displays graphically the pitch of a voice detected in the microphone.                               | Evaluate ability of the client to produce a rising pitch contour. Evaluate ability of the client to produce a steady pitch.                                      |
| <b>Loudness display</b>              | Displays graphically the loudness of whatever sound is picked up at the microphone.                 | Evaluate ability to sustain phonation and to produce rising loudness contour.  |
| <b>Recogniser configuration tool</b> | Supports the user in setting options for using the recogniser tool.                                 |  |
| <b>Sound selector</b>                | Displays and selects from available recordings according to criteria specified by the user.         | Transfer recordings made with the OLPy Recorder to be used with phonetic maps (STAPtk).  |

## **8.6. System components**

The OLP therapy system consists of four components: the user interface (OLPy) containing graphical interfaces for training with automatic word recognition, OPTACIA that uses the software component (STAPtk) for creating and using phonetic visual maps, the automatic speech recogniser (GRIFOS) and the functionality of remote administration (TELEMACHOS).

### **8.6.1. The user interface OLPy**

OLPy is the part of the OLP system that keeps track of clients, schedules, tasks and exercises. It communicates with the system database where everything is stored, and organizes the appropriate tools when tasks are to be carried out. There is a structural design in

OLPy that requires the therapist to follow certain steps to create and configure a schedule consisting of different tasks of interest for the child to train. Figure 8.16 shows the main OLPy window showing the therapist's and child's names, and a description of the tasks that are involved in added exercises.

Exercises must be configured by the therapist who sets up the parametrically adjustable features of the individual tasks that belong to an exercise.

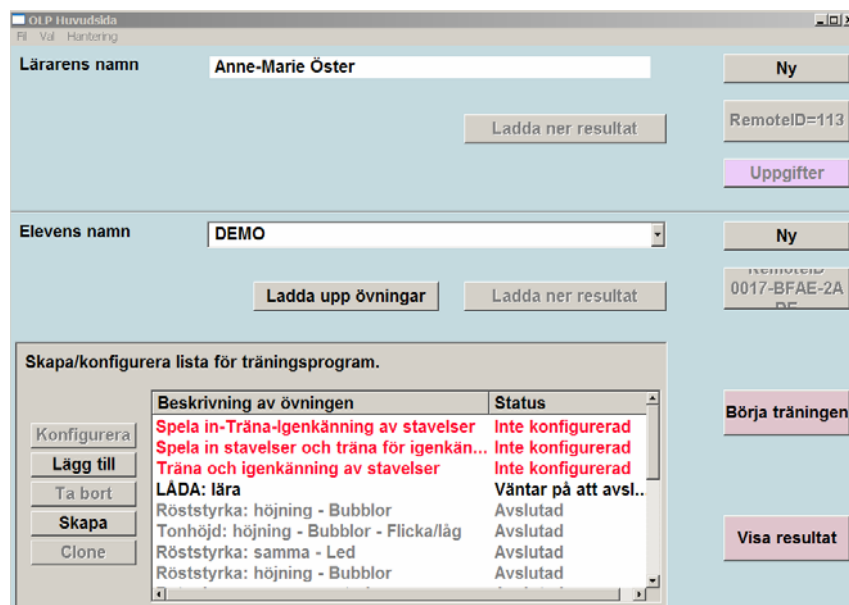
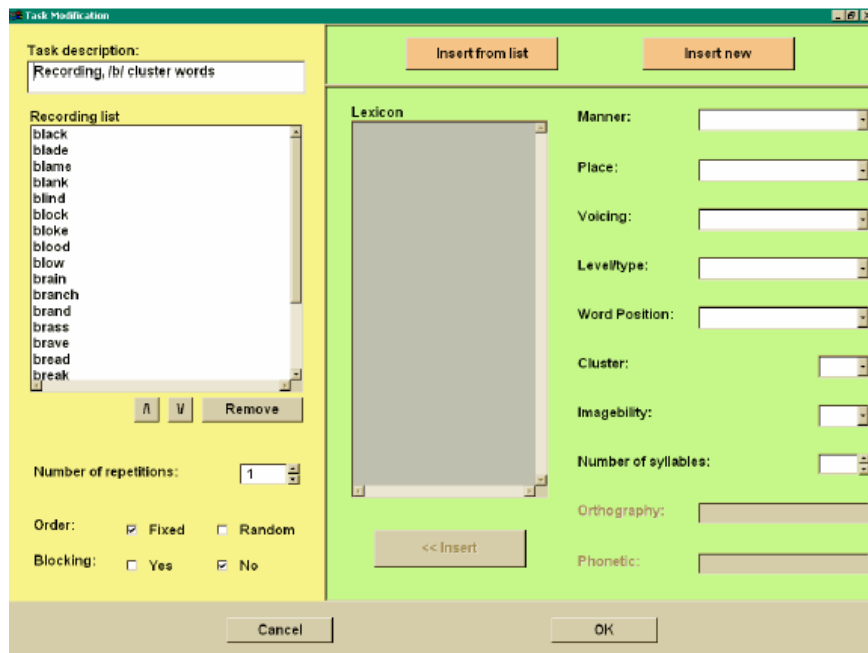


Figure 8.16. The main OLPy window showing the child's name and schedule.

A library that supports therapy design for the specified client group is included in the system with the option of inserting new words. This wordlist contains all Swedish vowels in mono- and polysyllabic words (at least three examples of each) as well as all Swedish consonants in all possible positions (initial, medial, final) followed by both rounded and unrounded vowels if applicable. Since all tasks are performed by several different software tools there are many configuration windows in OLPy. Figure 8.17 shows one of the configuration windows for a recorder task using the recorder tool.



*Figure 8.17. Recorder configuration window with existing library to the left and the option of inserting new words to the right.*

## **8.6.2. OPTACIA**

OPTACIA utilizes the software STAPtk to create and use phonetic visual maps and was developed from the Optical Logo-Therapy OLT, (Hatzis, 1999; Hatzis et al., 1999; Hatzis & Green, 2001). Optacia visualises phonetic contrasts between sounds and provides real time audio-visual feedback through a tailored acoustic-to-articulation kinematic mapping in 2D. OPTACIA is based on three basic, well-founded treatment principles: visuomotor tracking, visual contrast feedback, and visual reinforcement. Visuomotor tracking (Ziegler, Vogel, Teiwes, and Ahrndt, 1997) is a special case of biofeedback where some dynamic physical measure of performance is portrayed visually in real time. Visual contrast feedback is the important contrast between the correct and the misarticulated produced sound pattern that gives the children opportunities to become aware of differences in various articulatory configurations (Öster 1996). Visual reinforcement can support the child to increase the rate of response.

## Computer-Based Speech Therapy Using Visual Feedback

The aim of the therapy using OPTACIA is to strengthen, establish, and hopefully maintain correct articulation. The therapist has to record the sounds of interest to create the target for training, manually label the utterances, train the map, design and save the map, or select a predefined one. The articulator configuration corresponds to map position and articulator movement corresponds to map trajectory. In this way an individual child is provided with real time visual feedback about her/his speech. It is possible to re-use data collected during therapy sessions to re-train the map.

Figure 8.18 shows a designed map during training of the front rounded Swedish vowel [ɥ:] (u) and the back rounded Swedish vowel [u:] (o). The targets are shown by squares and the child's productions by circles. It can be seen in the figure that the child is making progress to the targets. This type of feedback is *critereon-referenced navigational*.

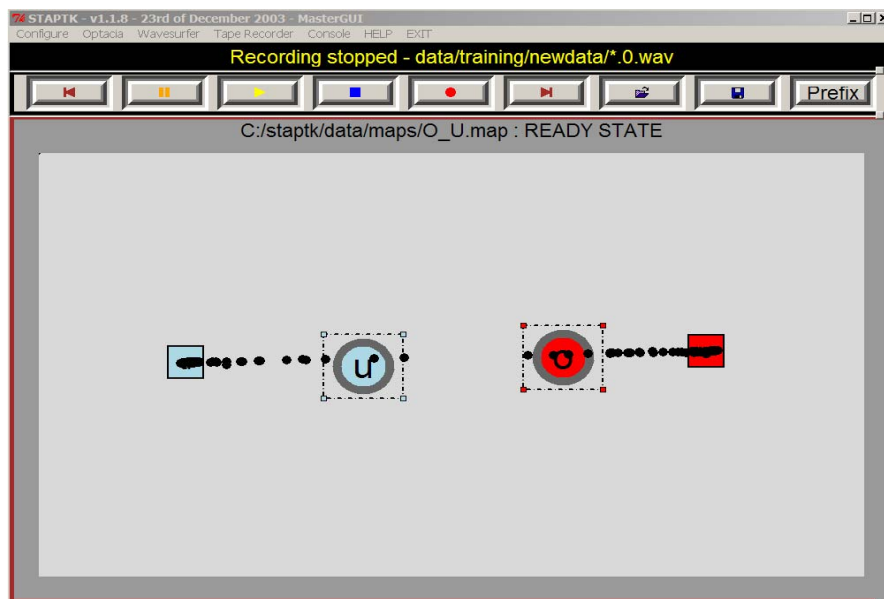
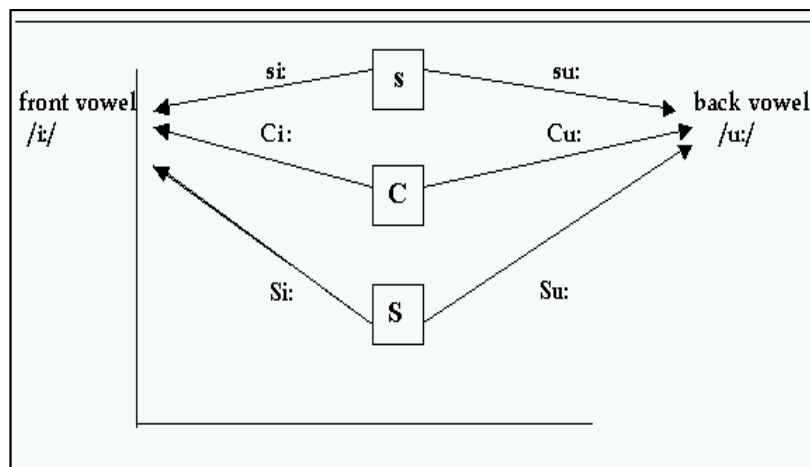


Figure 8.18. A visual map during training designed to correspond to correct production of the front rounded Swedish vowel [ɥ:] (u) and the back rounded Swedish vowel [u:] (o). The targets are shown by the small squares and the child's productions by the circles.

The technique was tested with the speech of three hearing-impaired subjects for Swedish fricatives (Öster et al., 2003). The results

demonstrated the utility of the mapping techniques for visually portraying correct and deviant pronunciations and indicated the potential use in actual clinical practice with hearing-impaired children. To produce the map, it was necessary to have training data, in the form of recorded speech time-annotated wave-files with target labels of those speech sounds, which are to appear on the map. The design of the Swedish sibilant fricatives map with acoustic and phonetic information of the target sounds is shown in Figure 8.19. The phonetic symbols are presented in the Speech Assessment Methods Phonetic Alphabet (SAMPA), which constitutes the best robust international collaborative basis for a standard machine-readable encoding of phonetic notation (see Appendix 2 for more details).

The acoustic difference between the three Swedish sibilants is represented by the vertical axis, which shows frequency range. By the positions of the fixed points along the axis the relationship between the acoustics and the articulation of each Swedish sibilant fricative is demonstrated. Articulatory targets for training each sibilant together with either a front spread [i:] or a back rounded [u:] in syllables are also inserted in the map, providing visual paths during training.



*Figure 8.19. Design of the map for Swedish training of the sibilant fricatives in Sampa symbols.*

The data files were recorded and manually labelled in the Wavesurfer application, <http://www.speech.kth.se/wavesurfer/> (Sjölander & Beskov, 2000). The recorded speech of three profoundly

hearing-impaired children served as test material for the map. The children (one girl and two boys) were 16 years of age and belonged to a school for the deaf in Stockholm. The children were recorded when they repeated CVC combinations. All the produced combinations of (sis) [si:s] and (sos) [su:s] were selected and played back to the map. The results of two of the children (one girl and one boy) can be seen in Figure 8.20.

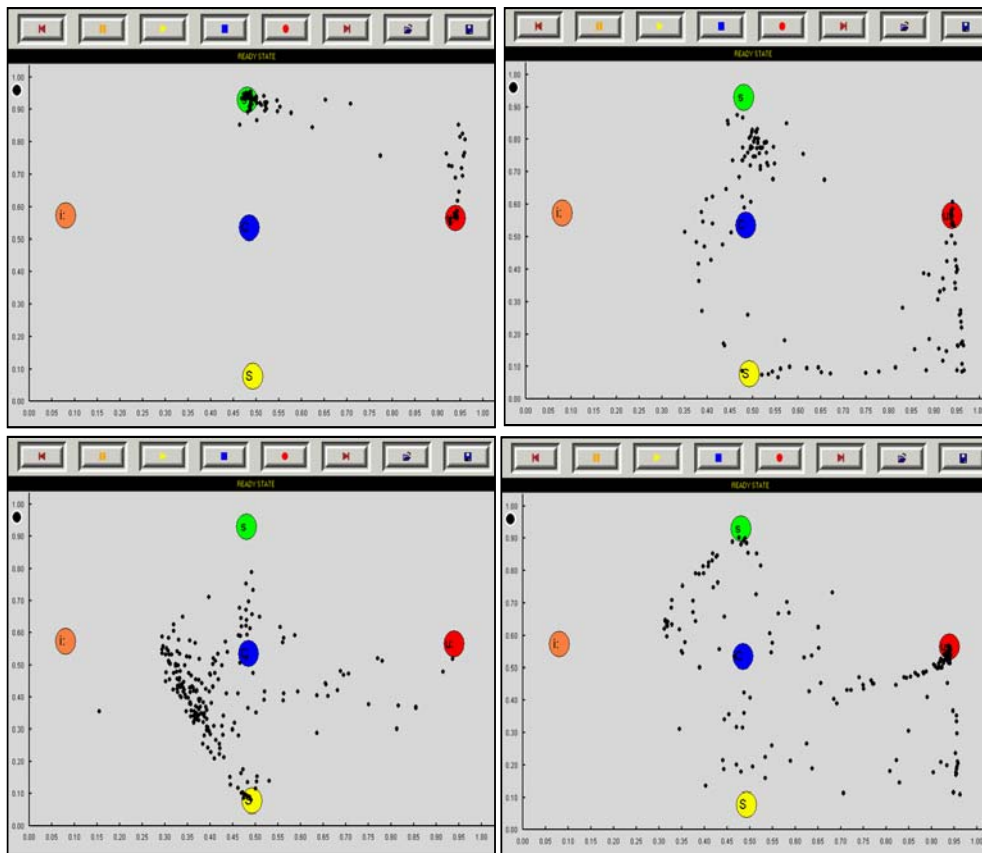


Figure 8.20. The result of two hearing-impaired children's repetitions of (sis) and (sos) played back to the map. The left panels are (sis) and the right panels are (sos). The upper panels are from the girl and the lower ones from the boy.

The upper left panel shows the girl's production of /sis/. In this production, she is quite successful at producing the fricative, but the vowel maps onto the [u:] instead of the intended [i:]. This could be the result of an incorrect [i:] production which includes lip-rounding



and strong nasalization. In the upper right panel, the same girl has intended a pronunciation of /sus/ but here the map indicates the instability of her fricative. Her /u/ vowel is correctly pronounced. In the lower left panel, the result of the boy's /sis/ pronunciation is shown. He has more difficulty with the fricative /s/ than the girl, but he comes closer to an acceptable pronunciation of the /i/ vowel.

Finally, in the lower right panel, the result of the boy's /sus/ is shown. In this attempt, his first /s/ is quite successful, as is his vowel. It is the /s/ in final position which he has difficulty with. The results for the other boy were not as clear. As he had low intelligibility and a very indistinct articulation, the map showed no targets or trajectories.

### **8.6.3. GRIFOS**

GRIFOS is a speaker-dependent, small-vocabulary, automatic speech recognition system (based on HMM models) used in OPTACIA sessions and in the graphical user interfaces of OLPy. The acceptability of a child's speech productions is evaluated to a trained target during speech therapy. The **recogniser** is used together with **animated graphics** for an amusing *exact evaluative feedback* as well as a **rewarding feedback**. The rewarding feedback can either be *hitting the mark* as it rewards a correct produced word or syllable and considers all productions that differ from the target as incorrect, or be *encouraging* and reward a produced word or syllable according to the degree to which the child's production matches the target by presenting the progress visually. Figure 8.21 shows an example of one of the different graphical interfaces, "the Diver" that is used for an abundant and meaningful training. A production matching the target model will cause the diver to swim up to the rock and pick a jellyfish. When the predetermined number of hits is reached, the diver swims out of the screen. A target bar is shown at the bottom of the interface. When pressing the button with the little head the child's target model is replayed. This model is made up of the child's best productions. A colour bar shows the number of attempts configured in the exercise.



Figure 8.21. The graphical user interface “the Diver” used together with word recognition. A production of the word “apple” is recognised as correct and the diver picks a jellyfish.

## 8.7. Conclusions

This chapter has reviewed many different types of visual feedback that are used in computer-based speech therapy today to motivate and assist speech and hearing-impaired children during speech training. The systems are product-oriented as they give parametrical feedback by showing visual representations of acoustical parameters in real time. The feedback can be of different kinds and purposes: for basic awareness and control, for evaluation, for navigation and guidance, for rewards, for repetitive and additional training, and for auditory playback and sound effects. The exercises consist of animated graphics that change in colour and form, game-like exercises, comparable models of spectra, spectrograms, and oscillograms, target speech pictures, and visual maps.

All systems are technically well developed and contain several programs to choose from, as well as many efficient types of visual feedback. However, some of the programs are more useful for children and hearing-impaired persons than for adults and hearing persons and vice versa. Moreover, some of the visual feedback types are more efficient during certain stages of the therapy than others.

*Design of visual feedback in Swedish computer-based therapy systems*

These facts are brought up in the clinical evaluations of the systems reported in the following chapter. Recommended visual feedback strategies and therapy design for different users are discussed in Chapter 10.



## ***9. Clinical evaluation studies of the three systems***

### **9.1. Introduction**

This chapter reports on several evaluations of the three systems carried out within national projects as well as within EU projects. The SpeechViewer was evaluated for Swedish with hearing-impaired children and L2 learners through case studies. Evaluations of Box of Tricks were also made by the Hungarian and Slovenian partner in almost the same manner but the groups of children who participated in the evaluation differed between the partners. The OLP therapy and OPTACIA were evaluated by the Swedish, Greek, and British partners through an AB/BA design. The Swedish subjects consisted of two groups of hearing-impaired children.

### **9.2. Clinical evaluation of SpeechViewer with profoundly hearing-impaired children**

In several pre-schools and schools for deaf and hard-of-hearing children in Sweden, the SpeechViewer has become a standard and valuable complement to the regular speech training activities. This is in part due to an ongoing collaboration between the therapists and the department of Speech, Music and Hearing at KTH.

The positive effect of a systematic training of prosodic contrasts was shown in a study by Öster (1989a, 1989b, 1990). Two prelingually profoundly hearing-impaired children with difficulties in producing certain phonological contrasts in Swedish were trained systematically with the system during eight weeks in order to evaluate its efficiency.

One of the children (child I) was a 15-year-old boy, with some residual hearing in the low frequencies. He had difficulties in producing quantity differences between phonologically long and short vowels. This is an important contrast in Swedish, which is

realised as a difference in duration and, for some pairs, also with a difference in vowel quality. Furthermore, the following consonant has opposite quantity, i.e., long vowels are followed by short consonants and vice versa. The child controlled only the phonological contrast between long and short (o) [u: - ʊ] before training. His realisations of short (å) [ɔ] and short (u) [ø] were always produced long. In some cases he did not control the vowel quality. It was not the intention to train vowel quality in this study, but in some cases the pronunciation improved or became more stable after the durational training. Child II was a 13-years-old girl, who had difficulties in producing distinctions between voiced and voiceless velar stops. Her hearing threshold was within the region of vibration.

The children were video recorded before and after training when they read minimal word pairs containing the phonological contrasts of vowel-length and voicing. Short sentences in which the target words were included were also recorded to study the pronunciation of the words in isolation compared to the pronunciation of the words in running speech. Untrained word pairs that contained the same contrasts were also recorded to study the generalisation effect. Narrow phonetic transcriptions of the video recordings were made by the author using the symbols of the International Phonetic Alphabet and some diacritical marks (see Appendix 1) that have been developed for the transcription of babbling and phonetic development in early infancy (Bush et al., 1973; Grunwell, 1987; Roug, Landberg, and Lundberg, 1989).

The children were trained with the speech patterning program "Pitch and Loudness" by their therapists for about ten minutes twice a week during eight weeks. This is a graphical presentation of the speech signal where the voiced/voiceless contrast is clearly indicated by different colours. Voiced sounds are red and voiceless sounds are green. Discrimination between long and short vowels is visible through the differences in duration of the red colour that indicates voiced vowels. Otherwise all vowels look the same. In Figure 9.1 the display of the speech-patterning program "Pitch and Loudness" is shown.

In the upper portion the teacher's pattern of the word pairs (haka) [ɑ:] (chin) and (hacka) [a] (chop) is shown and in the lower portion the child's production before training is shown. It can be

seen that the child pronounced the long quantity for both long and short /a/.

It was easy to explain to the children what was deviant in their production by comparing their speech with the speech of the therapist on the split screen display. Figures 9.2 and 9.3 show the subjective assessment (done by the author) of correctly produced quantity of long and short vowels by Child I before and after training.

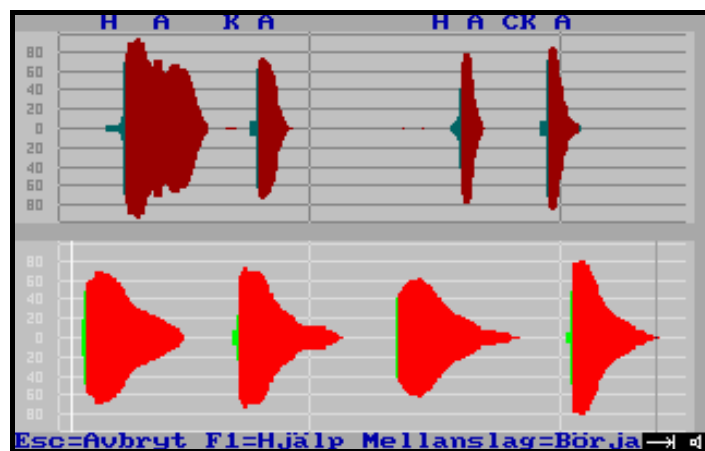


Figure 9.1. Display of the speech-patterning program "Pitch and Loudness". In the upper portion the teacher's pattern of the word pairs (haka) [haka] (chin) and (hacka) [haka] (chop) is shown and in the lower portion the child's production before training is shown. It can be seen that the child only mastered the long quantity of the vowel /a/.

The result shows that, in spite of the limited amount of data, the child learned to produce the short versions correctly after training. Short vowels improved more than long vowels. An improvement in the production of durational contrast between vowels after systematic training was found in all vowels except for /y/. An improvement was also found in running speech and in untrained words, similar to that found in trained words.

Child II also improved in producing voicing contrasts with the help of the SpeechViewer training. Before training, (g) was pronounced without voicing in medial position. In final position, (g) was omitted.

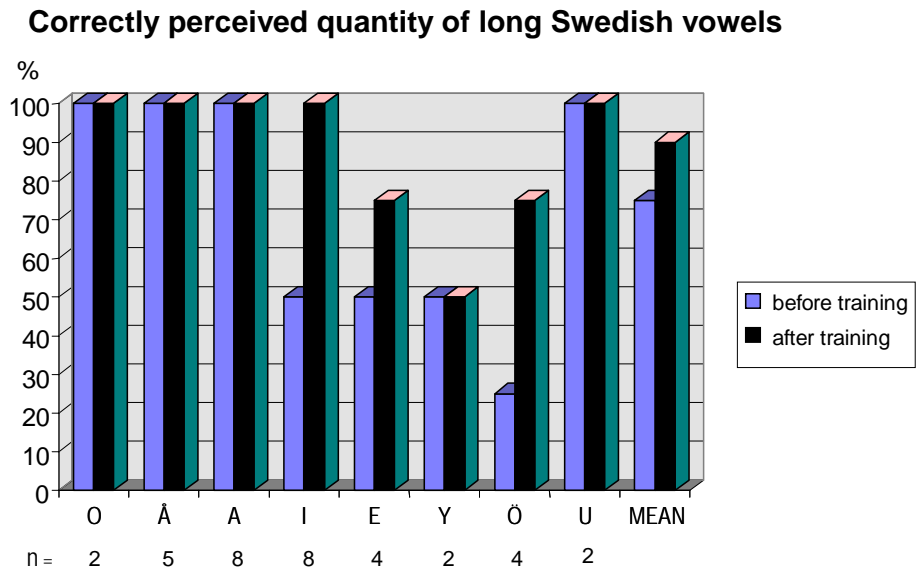


Figure 9.2. Subjective evaluation of correct duration of Swedish long vowels before and after training. The vowels are represented by orthographic symbols.

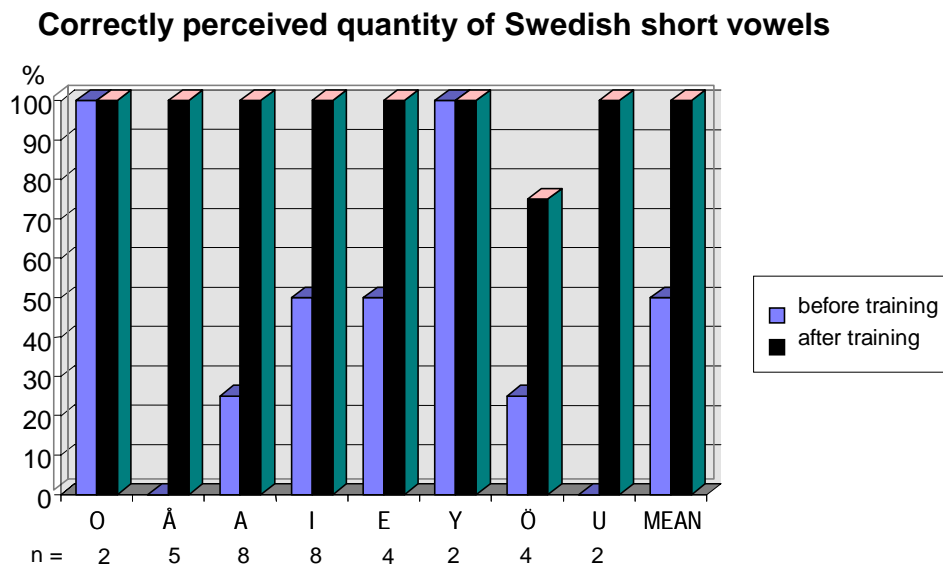


Figure 9.3. Subjective evaluation of correct duration of Swedish short vowels before and after training. The vowels are represented by orthographic symbols.



However, after training, (g) was pronounced as a voiced velar stop in both medial and final positions. This child also produced the differences in voicing between velar stops when she read the words in isolation as well as when she read the words in running speech. Similar to Child I, the improvement in trained words was also found in untrained words.

SpeechViewer was also introduced in the pre-school training of the Danderyd Hospital in Stockholm. The ambition was to take care of early skills during the sensitive period for learning and individually train such abilities as respiration, loudness, pitch and voice quality before the children join the speech clinic of a school for deaf children. Results were reported (Öster 1995c, 1995d 1996) of a five-year old prelingually deaf boy (D), who used sign language for communication. His pure-tone averages were 78 dB in the right ear and 102 dB in the left ear. His phonation was too high and monotonous around 700 Hz, which he could not perceive or control by himself. This was very disturbing for those closest to him and he was constantly reduced to silence and faced with irritation. His teacher wanted him to learn and establish a natural pitch with a voice he could make use of. During training D wore his hearing aids. By way of introduction it was important to get D to understand that what happened on the screen was dependent on his phonation and pitch. We started the training with one of the simple pitch awareness modules to increase awareness of vocal pitch and quantify his pitch range. A helicopter changes position vertically as pitch variations occur and two white marks indicate minimum and maximum pitch attained, see Figure 9.4. D's pitch was high but his range was small: 688-756 Hz.

D observed the helicopter on the highest floor but he was unable to lower it. The feedback made him aware of his vocal pitch but it did not show him how to lower it. Instead we used the "Pitch and Loudness" patterning program, where time is represented on the horizontal axis and pitch along a vertical frequency scale in Hz. The split screen in Figure 9.5 displays the teacher's input of a sustained /ɑ:/ with a natural pitch in the upper portion and D's phonation during 3.8 sec. with a mean of 604 Hz in the lower portion of the screen.



Figure 9.4. Display of the module "Pitch" where D's pitch and range are shown.

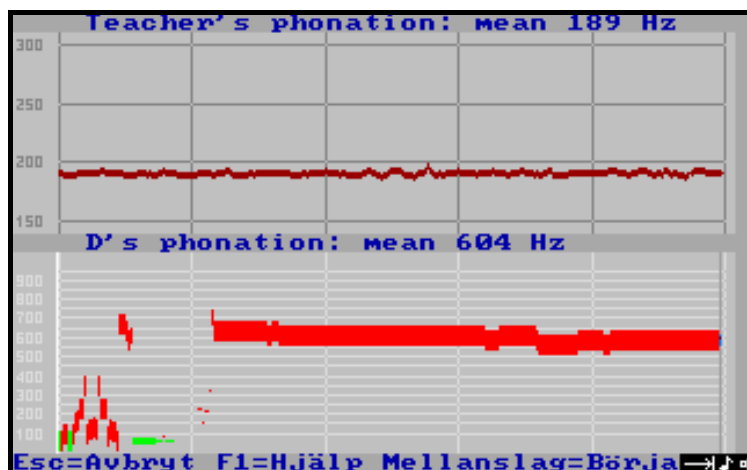


Figure 9.5. Display of the speech patterning program "Pitch and Loudness". In the upper portion the teacher's phonation during 4.4 seconds is shown. In the lower portion D's sustained spontaneous phonation during 3.8 seconds is shown.

The next step was to get D to vary his pitch. He touched the teacher's larynx and observed her pitch variation in the upper part of the screen many times over. In the lower part of the screen in Figure 9.6 the positive result of D is shown during 5 seconds.

The third step was to get D to lower his pitch. Figure 9.7 shows, in the upper part of the screen, the pattern of the teacher when she repeatedly lowered her pitch at the same time as D

touched her larynx. The lower part shows that during the first 3 seconds D's pitch was very high but suddenly it dropped to 327 Hz. He was stunned and went back to the high and varied pitch pattern to get control over his voice. From now on D varied between high and low pitch and at a given sign by the teacher he immediately lowered his pitch.

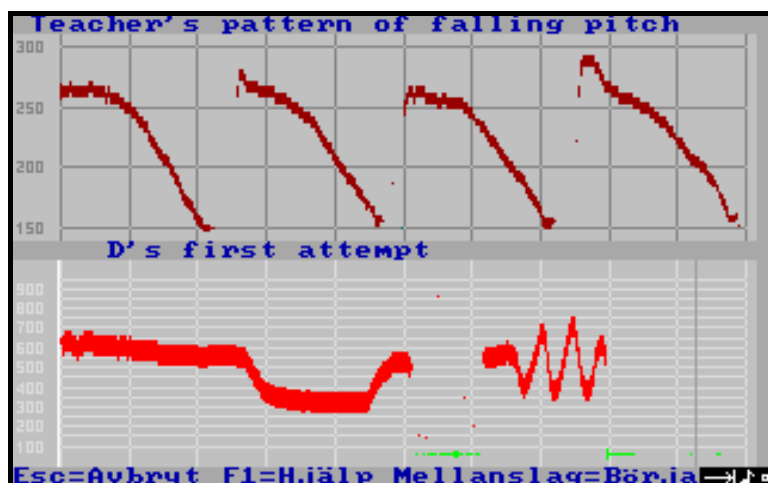
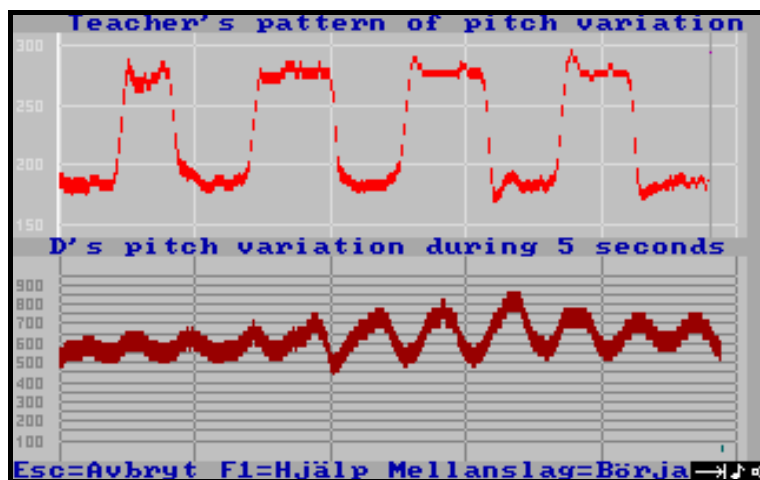


Figure 9.8 shows a typical training session of repeated phonation with a natural pitch. To vary the training and strengthen his control of pitch, D also tried skill-building programs like the one in Figure 9.9, where he had to produce correct pitch variations to control the vertical movements and sustained voicing to control the horizontal movement of an object toward targets arranged in a curve.

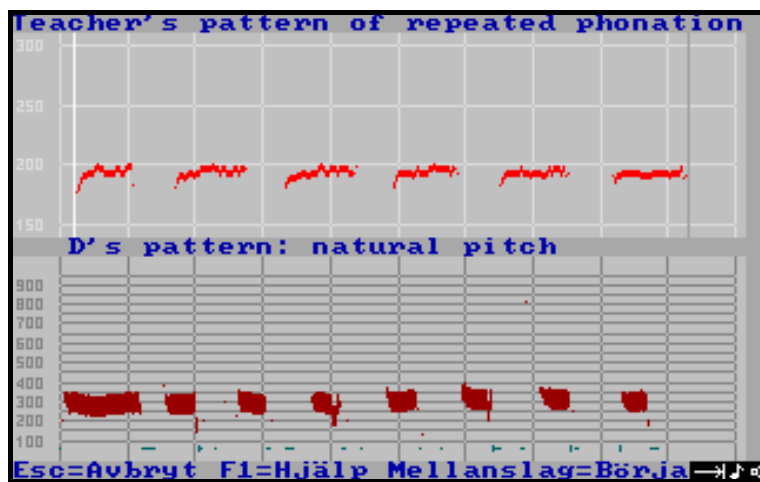


Figure 9.8. Display of a typical training session of sustained and repeated phonation with a natural pitch.



Figure 9.9. Display of the skill-building program "Pitch". D. monitored the figure with his voice towards the gold pieces by varying his pitch between 275 and 325 Hz.

Figure 9.10 shows that D learned and established a lower pitch more natural for his age, and that he got awareness and control over his voice. The upper portion of the screen shows the spontaneous phonation two weeks after D finished training. The average pitch was 266 Hz. The lower portion of the screen shows his spontaneous phonation eight weeks after training. The average pitch was 263 Hz and his voice was a soft, nice voice of a five-year old little boy.

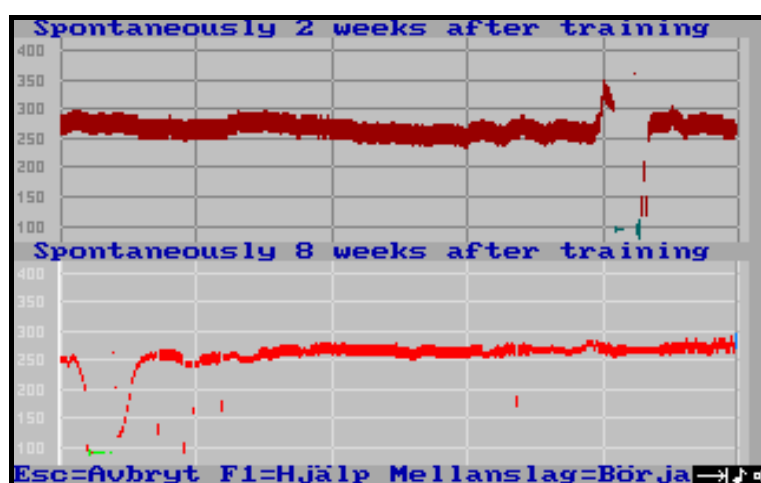


Figure 9.10. The upper portion shows D's spontaneous phonation two weeks after finishing his training and the lower portion shows his spontaneous phonation eight weeks after finishing training.

### 9.3. Clinical evaluation of SpeechViewer with L2 learners

In three studies by Öster (1997, 1998, 1999a) strategies and results from a project together with the Unit for Languages and Educational Research and Development at KTH were reported where this new L2-teaching strategy was used with 13 adult international engineers (9 males and 4 females) learning technical Swedish. The learners participated in a total of six half-hour training sessions twice a week, training both perception and production of spoken Swedish. All of them but one had an academic degree in engineering from a university in their home country. Their ages varied from 25 to 46 years. Six of them originated from Bosnia, the others from Cuba, Peru, Saudi Arabia and Russia. The Speech Patterning Module "Pitch

and Loudness” was used and some of the learners also used “Skill building modules.

Before training, the speakers were recorded when they read a text of 17 sentences and 110 isolated words. This diagnostic speech material contained all Swedish vowels (long and short) and all Swedish consonants in initial, medial and final positions. Two trained phoneticians transcribed the recorded words and an assessment of each L2-speaker’s individual segmental deviations was made. The recorded sentences were evaluated by the phoneticians according to the speakers’ accuracy in producing stress patterns, intonation, word accent and rhythm. The segmental and suprasegmental deviations found formed the basis for each speaker’s individual speech material in the pronunciation training that followed.

By means of the audio-visual feedback given in the module “Pitch and Loudness”, efficient training of both perception and production of spoken Swedish was viable. It was easy to make the learners aware of in what way their production deviated and show them how to correct specific deviations. In the following two figures (9.11 and 9.12) the visual information that this module provides can be studied.

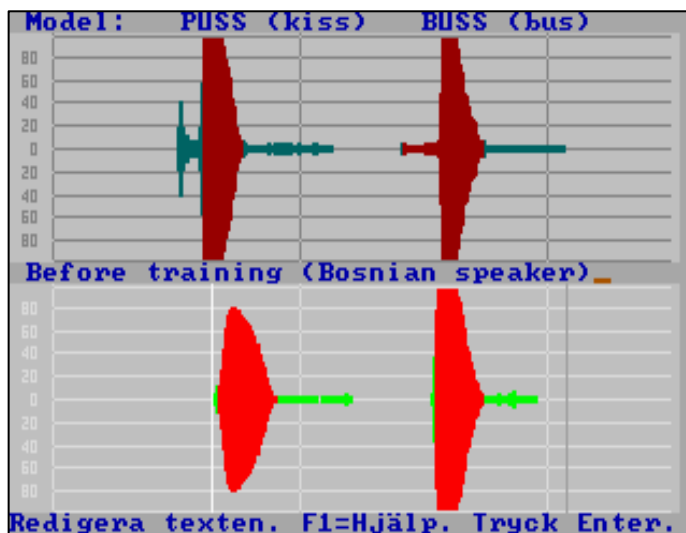


Figure 9.11. The upper portion shows the teacher’s model of a correctly produced voicing contrast between the initial plosives /p/ and /b/. The lower portion shows a deviant production of a Bosnian speaker.

The program displays F0 and intensity in real time. The split screen provides a comparison of a learner's utterance with a model of the teacher. In Figure 9.11, the voiced and voiceless contrast is clearly indicated by different colours in the loudness application of the module "Pitch and Loudness". Voiced sounds are red (dark) and voiceless sounds are green (light). In the upper portion, the teacher's correct production of the word pairs /puss/ (kiss) and /buss/ (bus) is shown and in the lower portion a Bosnian speaker's production is shown. It can be seen that the Bosnian speaker has difficulties in producing the phonetic contrast between the initial unvoiced aspirated plosive and the voiced plosives in Swedish.

Figure 9.12 shows a training session of the intonation pattern of the Swedish sentence "Karin sjunger i sängen" (Karin sings in bed) in the application where fundamental frequency and intensity are displayed in real time. Stressed syllables are visible on the teacher's screen through pitch modulation, longer duration and higher intensity (thickness of the line).

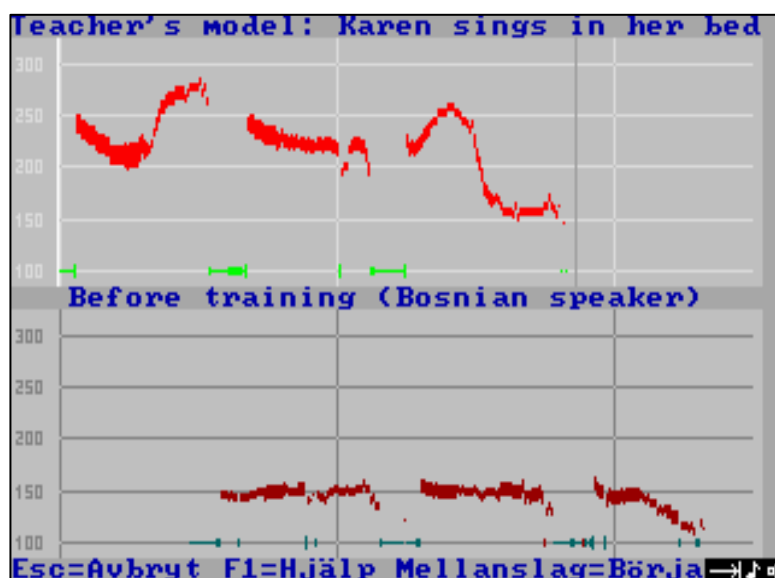


Figure 9.12. In the upper portion, the teacher's correct intonation pattern of the phrase "Karin sjunger i sängen" (Karin sings in bed) is shown. In the lower portion, a Bosnian speaker's intonation pattern is shown.

On the lower screen, a Bosnian speaker's quite monotonous production is shown. The contrastiveness of the visual patterns gives

the learner possibilities to discriminate between various distinctive features that underlie phonological contrasts in Swedish.

While the specific structure of spoken L2-teaching may vary somewhat from teacher to teacher there are certain general steps to be followed to obtain efficient training. The following steps, earlier discussed in section 2.3.5 are recommended:

- Diagnosis of individual deviations
- Instructions that aim at awareness and correct realisation
- Initial training to obtain automaticity
- Additional and repetitive training for generalisation and transfer to linguistic use

### **9.3.1. Diagnosis of individual deviations**

The first step involves an assessment of individual deviations that should be trained. Examples of individual results of a phonological assessment (Öster 1997) of segmental deviations by a group of six Bosnian and four Spanish (Cuba, Peru) speakers are shown. Deviantly produced Swedish vowel quality is shown in vowel diagrams and deviantly pronounced consonants are inserted in a table that shows applicable positions. A summary of general prosodic deviations is also shown.

#### **Vowels**

Displaying vowel deviations in diagrams was a good pedagogical aid to describe the role of the tongue in vowel production. The Swedish language has 9 phonologically long and 9 short vowels. As previously stated, quantity is an important contrast in Swedish, which is realised as a difference in duration and, for some pairs, also with a difference in vowel quality. Furthermore, the following consonant has opposite quantity; i.e. long vowels are followed by short consonants and vice versa. High and mid-high long Swedish vowels are diphthongised in open syllables. Noteworthy is also the number of Swedish front vowels and the fact that some of them are rounded. It is often difficult for L2 speakers to have a complete productive and perceptual mastery of these vowels (Bannert, 1990). The diagrams in Figures 9.13 and 9.14 show typical vowel confusions



and difficulties that the L2 learners had depending on their mother tongue.

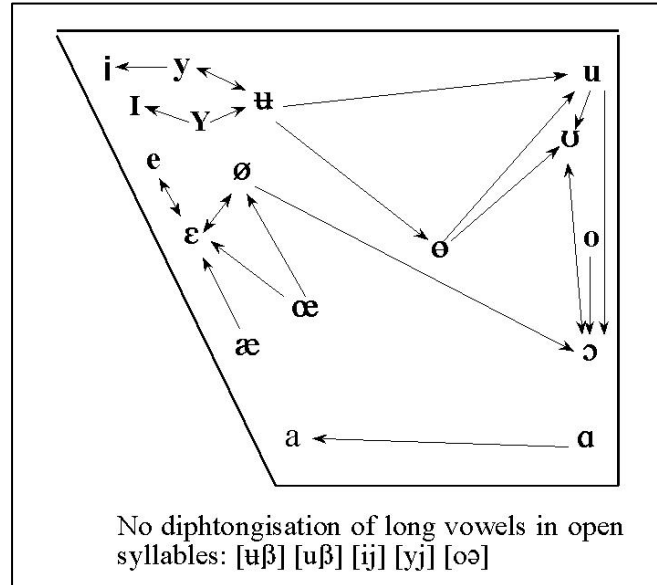


Figure 9.13. Deviantly produced Swedish vowel-quality by six Bosnian speakers. Lines and arrows indicate typical confusions.

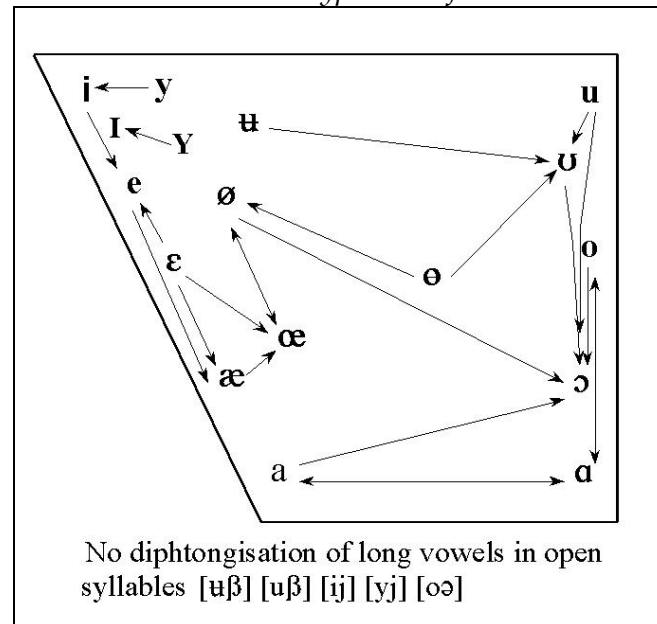


Figure 9.14. Deviantly produced Swedish vowel-quality by four Spanish speakers. Lines and arrows indicate typical confusions

## Consonants

Deviantly produced consonants are inserted in tables that show all possible positions.

Figures 9.15 and 9.16 show the deviations made for those consonants in initial, medial and final wordposition (cf. section 7.3.2). Blank cells indicate that the consonant was pronounced correctly. Shaded cells indicate positions where the consonant was not phonotactically possible. The consonant deviations that were trained depended on the mother tongue of the L2 speakers. Some of the Swedish consonants caused production and perception difficulties for the L2 speakers, especially the voicing contrast between plosives, aspirated voiceless plosives, the great number of fricatives, the nasal velar sound and the fact that dentals are retroflexed when they are preceded by /r/.

| IPA | initial | medial | final |
|-----|---------|--------|-------|
| p   | b       | b      | b     |
| t   | d       | d      | d     |
| k   | g       | g      | g     |
| b   |         |        | p     |
| d   | t       | t      | t     |
| g   |         | k      | k Ø   |
| f   |         |        |       |
| v   |         |        |       |
| s   |         |        |       |
| ʃ   | ç fj sj | çj     |       |
| ç   | ʃ       |        |       |
| j   |         |        |       |
| h   | ç χ     |        |       |
| m   |         |        |       |
| n   |         |        |       |
| ŋ   |         | ŋg     | ŋg    |
| l   |         | ɫ      | ɫ     |
| r   |         |        |       |

Figure 9.15. Deviantly produced Swedish consonants by six Bosnian speakers.

| IPA | initial | medial | final |
|-----|---------|--------|-------|
| p   | b       |        | b     |
| t   | d       | d      |       |
| k   |         |        |       |
| b   |         |        | p Ø   |
| d   |         |        | t Ø   |
| g   |         | k y    | k Ø   |
| f   |         |        |       |
| v   |         |        | f     |
| s   | əs      |        |       |
| ʃ   | f x     | x      |       |
| ç   | ʃ       |        |       |
| j   |         |        |       |
| h   | Ø       |        |       |
| m   |         |        |       |
| n   |         |        |       |
| ŋ   |         | ŋg     | ŋg    |
| l   |         |        |       |
| r   |         |        |       |

Fig. 9.16. Deviantly produced Swedish consonants by four Spanish speakers.

### Prosody

Prosodic deviations were evaluated and summarized by two phonetically trained persons when listening to the recorded text. The most common prosodic deviations made by both groups of L2 speakers were:

- Perceiving and producing accent II
- Producing a long consonant after a short vowel in stressed syllables
- Producing quantity differences between phonologically long and short vowels.

### 9.3.2. Instructions that aimed at awareness, correct realisation, and understanding

It was important to make the learner aware of in what way his/her production deviated and show him/her how to correct these deviations. Instruction was based on this possibility of comparison in the module "Pitch and Loudness". The most instructive speech material consisted of minimal word pairs that contained the contrast that the learner produced deviantly. When studying the following figures it is obvious that this contrastive training using audio-visual feedback had a certain effect on the L2 speakers' productions. Figures 9.17 and 9.18 show positive results after six half-hour training sessions twice a week of the same material as shown in figs 9.11 and 9.12 (same speakers).

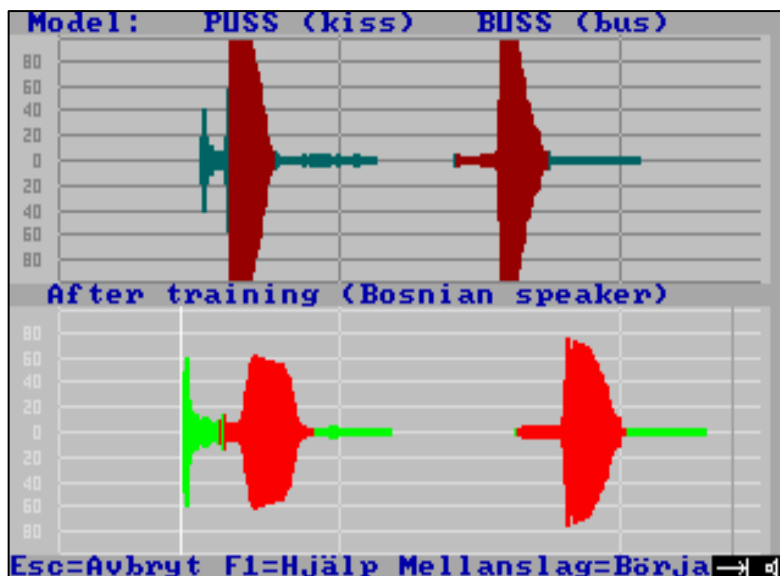


Figure 9.17. The upper portion shows the teacher's model of a correctly produced voicing contrast between the initial plosives /p/ and /b/. The lower portion shows a correct production by a Bosnian speaker after some training (cf. Figure 9.11).

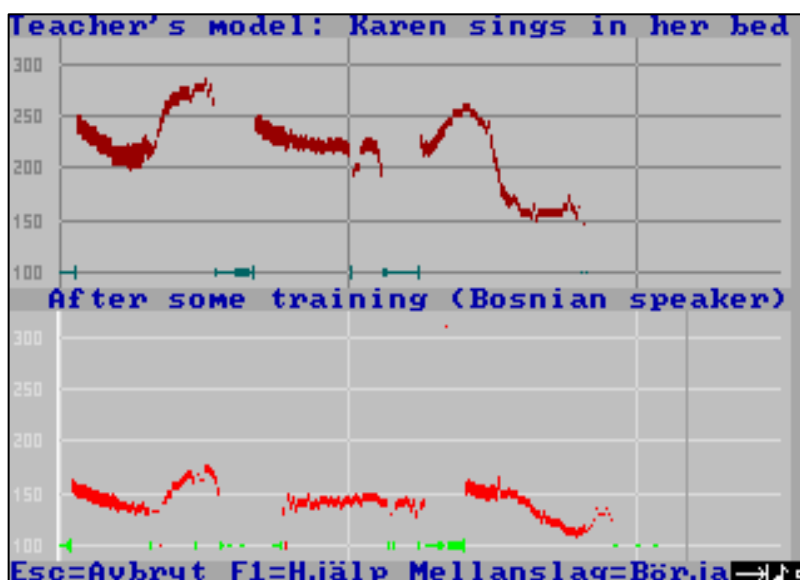


Figure 9.18. The upper panel shows the teacher's correct intonation pattern of the phrase "Karin sjunger i sängen" (Karen sings in her bed). In the lower panel, a Bosnian speaker's improvement after training is shown (cf. Figure 9.12).

### 9.3.3. Further training to establish automaticity and transfer to untrained material

Skill building modules were also used which provide a game-like strategy to strengthen ability in refined articulation. Phonemes produced by the learner were matched against models by comparing target spectra. To avoid the problem of interspeaker variability the L2-speaker's own "best production" was stored as a target.

### 9.3.4. Results of a questionnaire with the thirteen L2 speakers

After the training period the L2 speakers were asked some questions concerning this new L2 teaching strategy. In the following figure some of their answers are summarised (Figure 9.19).

*Computer-Based Speech Therapy Using Visual Feedback*

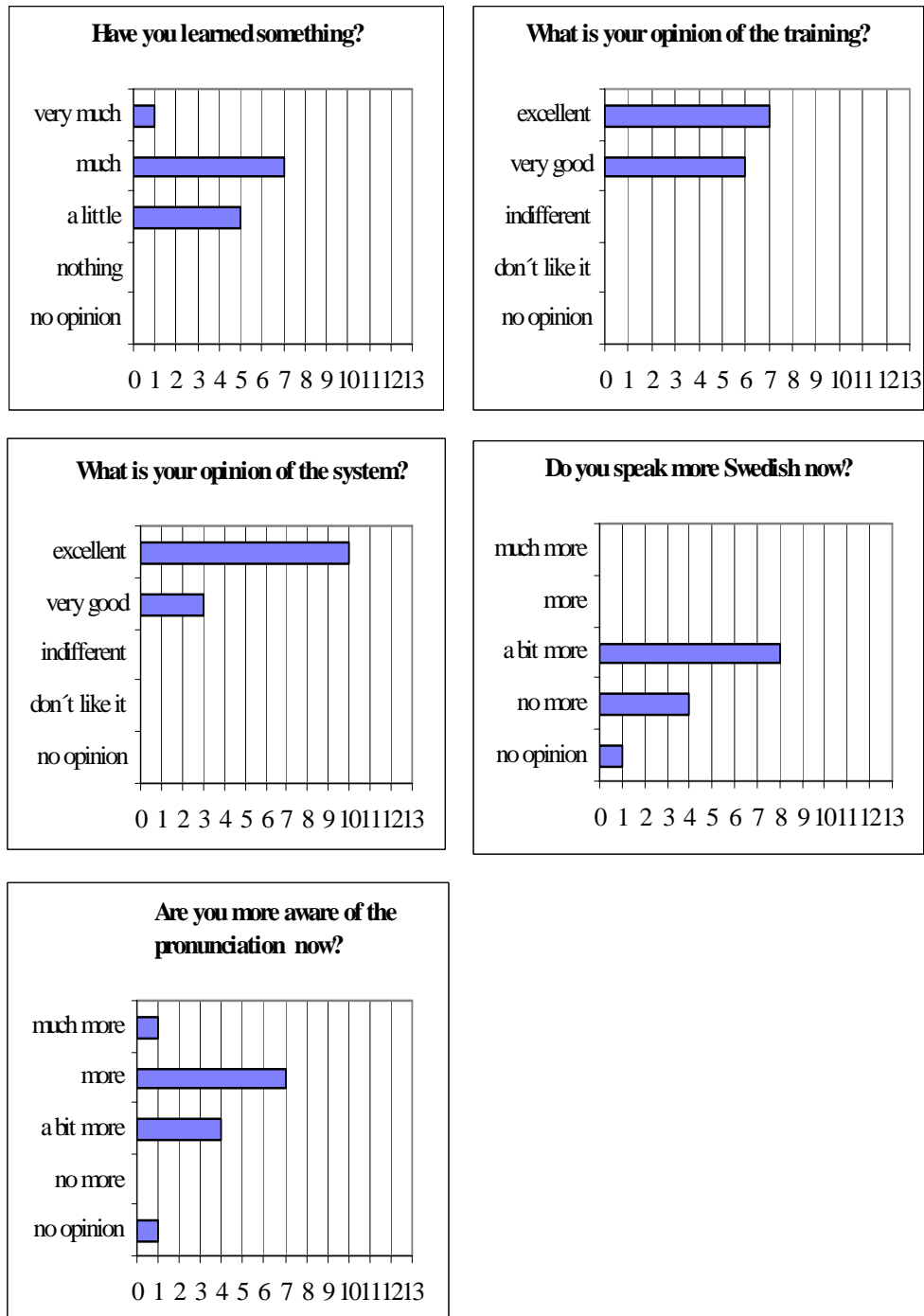


Figure 9.19. Results from a questionnaire with thirteen L2 speakers after some experience of audio-visual speech training provided with the SpeechViewer.

Below some important comments made by the L2 speakers after training with SpeechViewer are summarised:

- "I feel more confident now"
- "I realise the importance of a good pronunciation"
- "Now I immediately notice by myself when I pronounce something deviantly"
- "I am listening to other people more than before, trying to imitate their pronunciation"
- "Nowadays I am aware of my pronunciation"
- "I think that I talk more like a Swede now"
- "I talk more slowly now"
- "I want to train a lot more. My Swedish is not automatised"
- "It seems as if I have lost my identity"

#### **9.4. Evaluation studies with Box of Tricks**

This system was developed for speech- and hearing-impaired children aged 4-10 within the EU project "SPECO" between 1998 and 2001. The system was developed for four languages: Hungarian (Varázsdoboz), English (Box of Tricks), Swedish (Trollerilådan), and Slovenian (ARTI), see section 8.4.1. The main objective of the project was to develop a user-friendly system through the combination of prerecorded training vocabularies, comparable targets of a reference speaker, symbolic speech pictures for each phoneme, and illustrated training words.

During the development of the system, all partners were keeping close contacts with speech therapists from different educational fields, asking them repeatedly for their opinion. When the program and the User Manual were finished it was given to the therapists for evaluation. Two types of evaluations were made. One was based on the therapists' answers from a questionnaire that was constructed and used for all four languages. The therapists answered the questionnaire after having worked with the program from three months to half a year. Another clinical evaluation was made by the Slovenian and the Hungarian partner.

### **9.4.1. Results of the Swedish questionnaire**

The children who were trained with the program were between 4 and 12 years of age and consisted of three groups: hearing-impaired pre-school children, profoundly hearing-impaired children from a school for the deaf in Stockholm, and normal-hearing children with central language impairment. The evaluation was based on the answers of eight speech therapists who worked with the system during three months. The therapists had among other things experienced that:

- “The program supplies a long-felt need. There are so many computer-based language-training programs but so few speech-training programs on the market”
- “The sound preparation exercises attracted the children the most”
- “The children were very curious and wanted to see all the illustrations before starting therapy”
- “The auditory feedback of their own pronouncing stimulated them a lot”
- “Even if the children did not understand the acoustic pattern they got an understandable feedback by comparing their pattern with that of the reference speaker”
- “In the beginning some children had problems in focussing on the most important things”
- “Most children were very interested and co-operated”
- “Every time they came to school they wanted to work with the program. The program never failed. It was technically extremely stable”
- “Both the visual and auditory feedback worked without any problem. The program seems to be very well thought-out and carefully planned”
- “Using symbolic cards illustrating each speech-sound was a good idea”
- “Some parts of the program depending on the maturity and ability of the child”
- “Nice colours, and the illustrations are beautiful and clear”



- “If the child is motivated the program seems to be very effective”
- “In the beginning the program was difficult to grasp”
- “The program is very big and sometimes it was difficult to find the way to the accurate exercise”
- “The more I work with the system the more I appreciate it”
- “I especially valued the fixed vocabularies and the reference speaker”

For a full report about the Swedish questionnaire see Appendix 3.

#### **9.4.2. Results of the Slovenian questionnaire**

Five therapists answered the Slovenian version of the questionnaire. All therapists were speech and language pathologists. Three of them had used the program for three months, one for five months and one for seven months. Sixteen children between the ages of 5 and 15 worked with the system. Five of them were speech-impaired, three were hearing-impaired, three were deaf, and five children had a cochlear implant.

In general, the therapists were very satisfied with this method of speech therapy. Their opinion was that the period of evaluation was too short to make an evaluation that would give more than indicative conclusions. However, the method was considered to be very useful, especially as a supplement to traditional methods. It was a much appreciated modification in their work, pleasant for the children and useful for therapy of children with speech and/or hearing disorders. From a pedagogical point of view the system was highly rated. The user manual offered the therapists enough help and information. They were all of the opinion that there was no need for additional exercises. The children who had used the program showed a lot of interest and loved to work with it. They all understood the logic of the speech pictures and the background pictures and found all illustrations to be very amusing and colourful. In general, the opinion of the therapists was that the program stimulated the children to work for a longer time with each exercise and also motivated them to reiteration.

### **9.4.3. Results of the Hungarian questionnaire**

The survey was concluded in "Török Béla" Hard of Hearing School in Budapest. Severely hearing-impaired children, hearing-impaired children, deaf children (sign language), children with cochlear implant and normal-hearing children with difficulties in understanding and learning speech participated in the evaluation. A summary of the therapists' answers follows below:

- 1) Children can use this multimodal system by themselves very easily and they use it with pleasure, which is a very important factor from the point of view of efficiency.
- 2) The visual feedback helps children to see whether their pronunciation is correct or not, and how far it is from the correct one. They are not in need of relying only on the teacher's opinion. In particular, this is very important in case of speech handicapped with hearing loss.
- 3) In general, sounds were formed sooner than in control groups where they did not use any computer-based teaching system. It was found that a consistently shorter time was required for improving a speech sound than was the case with corresponding children of similar mental ability and impairment level who had been instructed by the traditional method. However, it is difficult to express the results in quantitative data because the result depended on many other factors (for example, one highly important factor was how much additional help the child received at home).
- 4) In those cases where these sounds were very resistant to the traditional therapy, the new method helped to repair these sounds.
- 5) The system is a useful tool for teachers in the individual linguistic training. It gives possibility to train in small groups too and pupils can use the system themselves and practice alone. Of course, especially for young children the visual tool itself does not substitute the work of the speech therapist. Firstly, this tool is a good aid. It helps the work of the therapist and gives a variety to teaching. Secondly, at the automation-phase or in case of older

children, the visual tool itself gives a good possibility to practice alone.

#### **9.4.4. Clinical evaluation of the Hungarian version of Box of Tricks**

The aim of the clinical evaluation of the Hungarian version, called the SPECO-method or *Vara'zsdoboz*, was to study whether the system was effective, if the intelligibility improved after training, and whether the children maintained a good pronunciation after therapy (Vicsi, 2001). The result of speech therapy with the Hungarian version of the system and with traditional therapy (without computer-based visual feedback) was compared using listening tests. Forty children between six and eight years of age were selected and grouped according to their degree of hearing impairment. Eight children had normal hearing, eight children had a mild hearing impairment, eight children had a moderate hearing impairment, eight children had a severe hearing impairment, and eight children (two from each group) constituted the control group (traditional therapy). All children except the control group constituted the test group. All children were recorded when they read eighteen words before therapy, immediately after therapy, and five months after therapy.

Thirty persons, inexperienced with the speech of hearing-impaired children, participated in a listening test where the same word recorded at three different occasions was presented from each speaker. The words were presented as word-pairs before and immediately after therapy, before and five months after therapy, and immediately after therapy and five months after therapy. The listeners were asked to decide which word out of two was the most intelligible.

The results of the listening tests are presented in Figures 9.20-9.22. The effect of therapy is evident. Most of the time the listeners found the words after therapy to be more intelligible than the words before therapy. The intelligibility was higher for those groups who had been trained with the Box of Tricks than for the control group who was only trained with traditional methods.

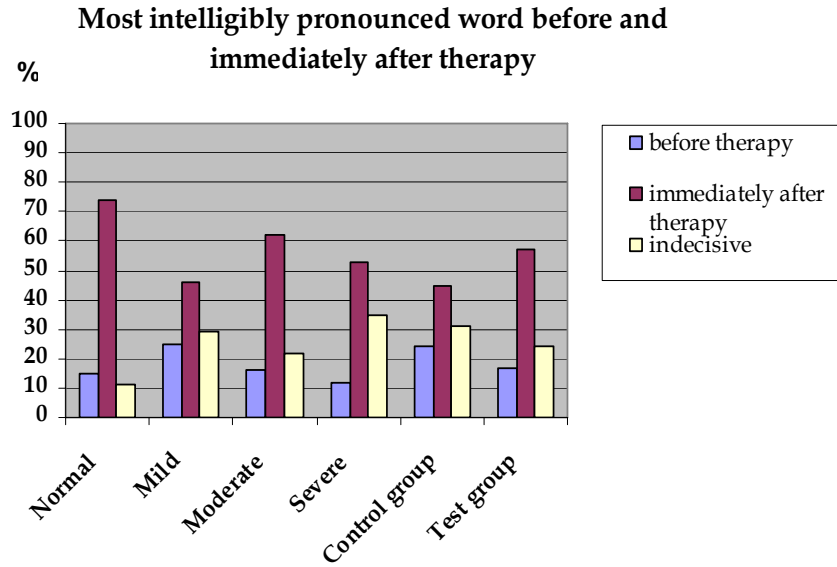


Figure 9.20. Effects of therapy with four groups of children. Results from a listening test by 30 inexperienced listeners, who judged words pronounced by the different groups of children before therapy and immediately after therapy (see text for more details).

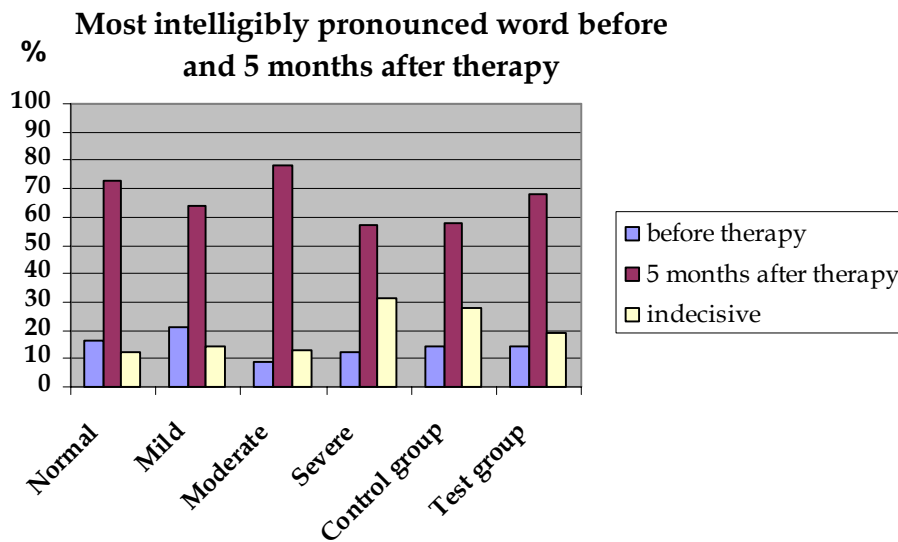


Figure 9.21. Results from a listening test by 30 inexperienced listeners, who judged words pronounced by four different groups of children before therapy and 5 months after therapy.

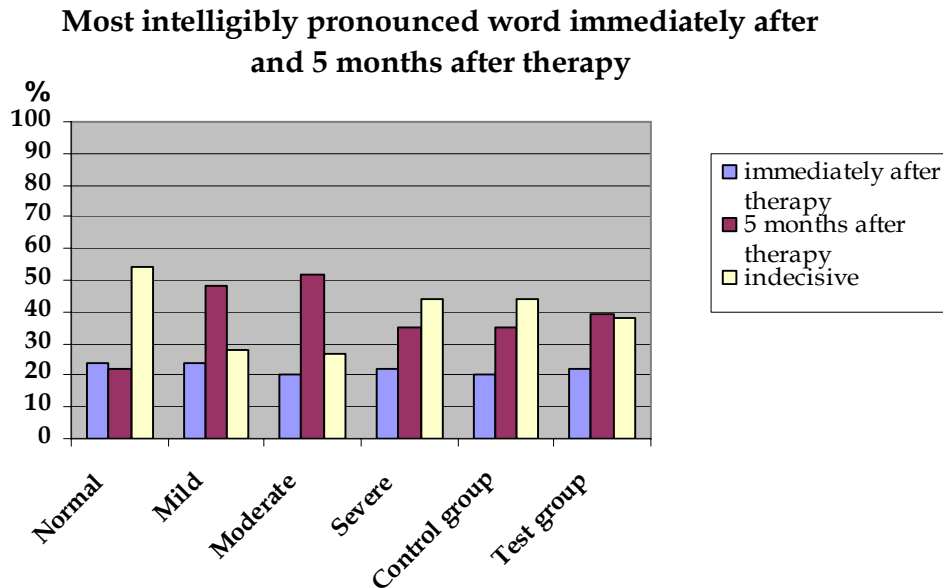


Figure 9.22. Results from a listening test by 30 inexperienced listeners who judged words pronounced by four different groups of children immediately after therapy and 5 months after therapy.

When comparing the groups with each other it can be seen that the normally hearing children improved more than the other groups immediately after training with the system.

The eight children with a moderate hearing impairment improved more than the other groups of hearing-impaired children after therapy. It is also obvious that the intelligible pronunciation obtained after therapy was established and maintained five months after therapy.

#### **9.4.5. Clinical evaluation of the Slovenian version of Box of Tricks**

The Slovenian version called “ARTI” was evaluated in the Speech and Hearing Centre in Maribor, Slovenia. Thirty-one children between the ages of 5 and 15 years used the system for three months. The children were divided into four groups: 11 children were speech-impaired, 6 were hearing-impaired, 5 were deaf using sign-language,

and 9 children had a cochlear implant. Each group was further divided into a control (15 children) and a test group (16 children). The children in the test group were trained with the system and the children in the control group were trained with traditional methods, without computer-based therapy with visual feedback.

A listening test was performed to evaluate the effect of speech therapy with ARTI. Recordings were made when each child read eighteen words before and after therapy. The listeners consisted of two types of listeners, naive and experienced listeners of the speech of hearing-impaired children. Eight naive listeners and seven speech therapists from the Speech and Hearing Centre in Maribor listened to the utterances and had to decide which pronunciation was the better of the same two words. The order of utterances for each word was randomly selected and the listeners were not aware which pronunciation was recorded before therapy and which was recorded after therapy.

The result of the listening test for all listeners' estimated improvements of the test groups, relative the improvement of the control group, showed that the deaf children improved to a higher degree, 18,65%, in intelligibility after therapy with the system than the hearing-impaired children, 8,25%, did, see Figure 9.23. The children with cochlear implant improved only with 0,1%, while the speech-impaired children did not improve in intelligibility after therapy with the system. They were judged to be less intelligible, -4,5 % after therapy with the system compared to the control group that only had traditional therapy without computer-based training with visual feedback.

Figure 9.23 shows the result of the listening test of all listeners', naive listeners', and speech therapists' choice of the best pronounced word, before or after ARTI-therapy, for the four different groups of children, compared to the result of the control group. The result of the listening test with experienced speech therapists shows that all hearing-impaired children who used the system improved in intelligibility compared to the children who were trained with traditional methods. The improvement was biggest for the deaf children using sign language. However, the results of the listening test with naive listeners shows that the children with cochlear implant did not improve after training with

the system when compared to the control group. The results can be looked upon as indicative because the training phase was rather short and the children would probably have needed more time to adapt to the new method.

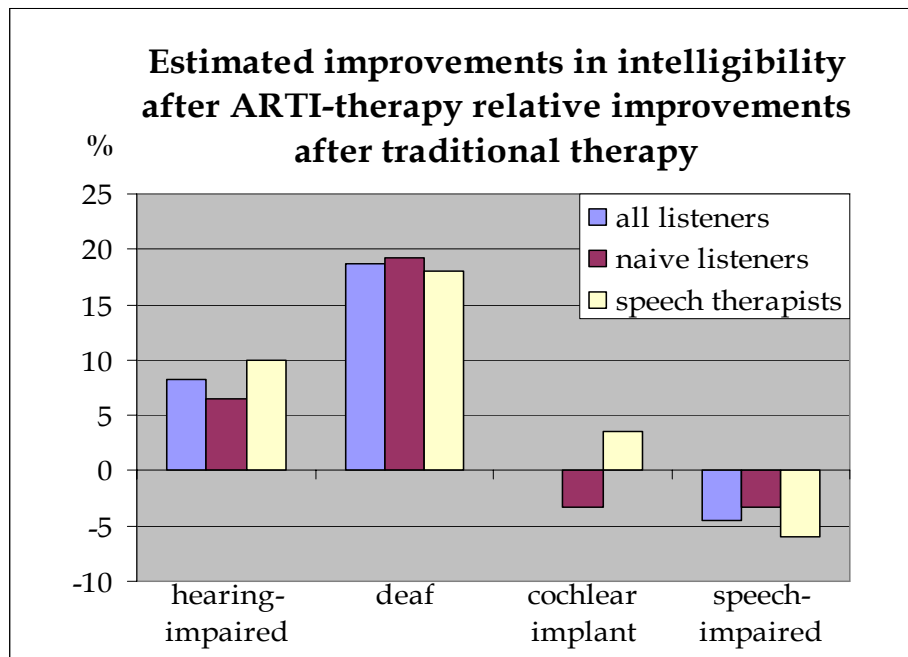


Figure 9.23. Result of the listening test by all listeners', naive listeners' and speech therapists' estimation of best pronounced word, before and after ARTI-therapy, with four different groups of children compared to the result of the test group.

#### 9.4.6. Clinical evaluation of the Swedish version of Box of Tricks

A controlled clinical evaluation of the Swedish version has not yet been performed. The system is available as commercial software in Sweden and includes a training course on how to use the system in the most effective way. Thanks to these courses continuous and resumed contacts with the users have shown that the therapists as well as the children appreciate the system as it is user-friendly, amusing and helps the children in their speech acquisition.

## **9.5. Clinical evaluation of the Swedish version of the OLP method with hearing-impaired children**

### **9.5.1. Introduction**

In Sweden, therapy with the OLP method was compared to traditional therapy for children with hearing impairment. The evaluation gave details about the effectiveness, feasibility and accessibility of the OLP therapy method in the treatment of articulation disorders by hearing-impaired children. The OLP therapy method was designed to increase intelligibility and accuracy of articulation by providing rewarding, navigational, and encouraging visual feedback.

### **9.5.2. Method**

Two speech therapists from the Manilla Deaf School in Stockholm were enrolled in the program during the iterative development of the OLP therapy method. They practised selecting and running the exercises via the OLPy user interface, recorded their own speech and set up exercises for practise, tried to understand how to use the system with hearing-impaired children, reported problems and gave suggestions for improvement. When the OLP software was sufficiently developed, the OLP-prototype was clinically evaluated by the speech therapists with some of their children. Three important questions were provided in order to evaluate the OLP therapy method with hearing-impaired children:

- Is the OLP method of therapy effective in increasing the intelligibility of the speech of:
  - prelingually hearing-impaired children between 8 and 14 years with a moderate to severe hearing loss ?
  - prelingually hearing-impaired children between 8 and 14 years of age with a severe to profound hearing loss?
- What degree of transfer is there of speech skills to untrained speech material?
  - Does the OLP therapy method function in a real environment?



### 9.5.3. Subjects

Eight children were recruited for the evaluation. The children ranged between 8 -14 years of age and their pure tone averages (PTA) in best ear were between 55–108 dBHL. Age, sex, and hearing data for the six children who finalized the evaluation are given in Table 9.1 and Table 9.2. All children used sign language for communication and fulfilled the specified inclusion/exclusion criteria. This implied that they all had good vision, cognition and reading age of over 8 years.

Table 9.1. Age, sex, and hearing loss in the better ear for three children with a moderate to severe hearing impairment who participated in the clinical evaluation.

**Group B(OLP) A: Moderate to severe hearing impairment**

|       |     |     | Hearing Loss in the better ear, dB |     |     |      |      |      |           |
|-------|-----|-----|------------------------------------|-----|-----|------|------|------|-----------|
| Child | Age | Sex | 125                                | 250 | 500 | 1000 | 2000 | 4000 | PTA, dBHL |
| 1     | 9   | F   |                                    | 20  | 45  | 40   | 80   | 95   | 55        |
| 2     | 11  | F   | 50                                 | 60  | 70  | 70   | 70   | 60   | 70        |
| 3     | 11  | M   | 55                                 | 65  | 65  | 80   | 85   | 115  | 77        |

Table 9.2. Age, sex, and hearing loss in the better ear for three children with a severe to profound hearing impairment who participated in the clinical evaluation.

**Group A B(OLP): Severe to profound hearing impairment**

|       |     |     | Hearing Loss in the better ear, dB |     |     |      |      |      |           |
|-------|-----|-----|------------------------------------|-----|-----|------|------|------|-----------|
| Child | Age | Sex | 125                                | 250 | 500 | 1000 | 2000 | 4000 | PTA, dBHL |
| 4     | 14  | M   | 70                                 | 55  | 70  | 70   | 105  | 105  | 81        |
| 5     | 10  | M   | 80                                 | 90  | 85  | 105  | >    | >    | 103       |
| 6     | 8   | F   |                                    | 85  | 90  | 100  | >    | >    | 108       |

All children had sufficiently well developed proprioception, imitation and exteroception skills to be able to benefit by speech training and support of technical devices. However, only six children finished the evaluation. Two children interrupted the trial because of various reasons.

### 9.5.4. Treatment

It was decided to evaluate the OLP therapy method through an AB/BA design in order to control for carry-over effects of one treatment condition to another. 'A' represented traditional therapy where explanations of manner and place of articulation were done by sign language if needed and shown through tactile and visual feedback by the use of a mirror, articulation pictures, or the therapist as a model. B represented therapy with the OLP method. Three children with severe to profound hearing impairment received treatment A then B (OLP) and three children with moderate to severe hearing impairment received B (OLP) then A which is shown in Table 9.3. All children trained two times a week for about 20 minutes each time during two periods of five weeks each. Each child required treatment for specific sounds with specific therapy objectives.

Table 9.3. AB/BA design of the clinical trial. Traditional versus computer assisted treatment with two groups of hearing-impaired children. Each square represents one week.

|                 |   |   |   |   |   |   |   |   |   |   |   |   |   |
|-----------------|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GroupAB         | X | A | A | A | A | A | X | B | B | B | B | B | X |
| GroupBA         | X | B | B | B | B | B | X | A | A | A | A | A | X |
| Assessments = X |   |   |   |   |   |   |   |   |   |   |   |   |   |

Before therapy started and after each therapy phase all children were recorded when they read two lists of words consisting of treatment and non-treatment words. These recordings were assessed to investigate therapy effects.

### 9.5.5. Therapy objectives

Before any treatment the target therapy objectives were selected for each child, based on the children's need for articulation training and according to the therapists' familiarity with each child's speech

deviations. The objectives were the same for both A/B treatments. All children required treatment for specific sounds. The range of different therapy aims was small because targets of for example changing pitch or decreasing volume were not important for these children to train because they all had a rather good control of these parameters. A list of treatment words containing the therapy objectives was constructed for each child. This was used in all therapy sessions. A list of non-treatment words with the same structure as the treatment words was also created for each child. These words were not used during therapy. All children were recorded when they read all words of the two lists before 1) any training, 2) after OLP training and 3) after traditional training to investigate therapy effects.

### **Child 1 (PTA, better ear 55 dBHL)**

This girl was 9 years of age and had difficulties with /s/, /l/ and consonant clusters. She also omitted final consonants in words. It was decided that her treatment objectives with work with Optacia and Griphos should be syllables and words containing the following vowels and consonants:

#### **Visual maps (Optacia)**

- Syllables: LA, LI, LO, SA, SI, SO
- Long Swedish vowels

#### **Automatic recognition with graphical interfaces (Grifos)**

- Syllables
- Single words with /s/ in initial, medial, final word-position and the cluster [st]
- Single words with /l/ in initial, medial, final word-position and the cluster [bl]

### **Child 2 (PTA, better ear 70 dBHL)**

This eleven-year-old girl's most serious problem was the pronunciation of the Swedish sibilants and the orthographic letter (x) [ks]. Her treatment objectives with work with Optacia and Griphos were decided to be isolated sibilants and words containing following consonants:

**Visual maps (Optacia)**

- Isolated sibilants: /s, ç, fj/

**Automatic recognition with graphical interfaces (Grifos)**

- Single words with /s, ç/
- Single words with /fj/ in initial, medial, final word-position
- Cognates
- Single words with (x) [ks]

**Child 3 (PTA, better ear 77 dBHL)**

Child 3 was an eleven-year-old boy who had great difficulties with the /f/ and /v/ sounds. He always substituted /f/ with /ç/ and /v/ with /d/. He was extremely interested in sports and his treatment words were decided to be sports words containing the following consonants:

**Visual maps (Optacia)**

- Isolated /s, ç/
- Syllables: /f, ç/ + long vowels
- Syllables: /v, d/ + long vowels

**Automatic recognition with graphical interfaces (Grifos)**

- Single words with /f, v/ in initial, medial, final word-position

**Child 4 (PTA, better ear 81 dBHL)**

This boy was fourteen years old and his greatest pronunciation problems were the Swedish sibilants and the consonants /k/ and /ŋ/. His objectives were set to:

**Visual maps (Optacia)**

- Isolated sibilants
- Syllables: sibilants together with [u:]

**Automatic recognition with graphical interfaces (Grifos)**

- Single words with /s, ç, fj/ and /k, ŋ/ in initial, medial, final word-position

### **Child 5 (PTA, better ear 103 dBHL)**

This 10-year-old boy had disordered pronunciation of all sibilants and substituted /m/ with /b/. He also confused the production of the vowels /u:, ʌ:, o:/. His training material was as follows:

#### **Visual maps (Optacia)**

- Isolated sibilants and sibilants together with /u:, ʌ:, o:/
- Isolated vowels /u:, ʌ:, o:/

#### **Automatic recognition with graphical interfaces (Grifos)**

- Single words with /s, ç, ʃ/ in initial, medial, final position
- Single words with /m/ and /b/

### **Child 6 (PTA, better ear 108 dBHL)**

The last child was an 8-year-old girl with similar difficulties. All sibilants and consonant-clusters were difficult for her to pronounce intelligible. Her objectives were:

#### **Visual maps (Optacia)**

- Isolated sibilants

#### **Automatic recognition with graphical interfaces (Grifos)**

- Single words with /s, ç, ʃ/ in initial, medial, final position
- Single words with consonant-combinations [ls], [sl], and [mp]

## **9.5.6. Assessments**

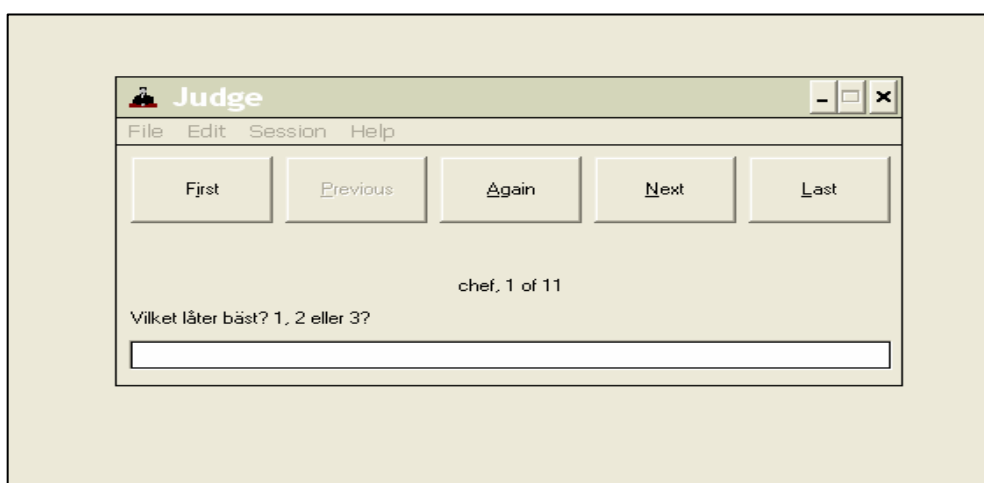
Three different assessments were performed to study the effects of treatment A and/or B, whether treatment A and/or B had a positive impact on untrained material and to obtain the therapists' opinion of possibilities and limitations of the OLP therapy method.

### **Evaluation of the effects of AB and BA therapy by using a listening test**

The evaluation compared the results of the speech training with the help of OLP therapy with traditional methods for therapy, by using a listening test. Ten naive adult listeners listened to the recorded speech samples of the treatment word list produced before and after each treatment phase see Table 9.3, and selected the best pronunciation out of three of the same word for every child. The

order of utterances for each word was played randomly so the listeners were not aware which pronunciation was recorded before or after any therapy.

The listening-test was administered by the program “Judge” (Granqvist, 1998). The subjects compared the stimuli with one another and rated them by entering text. A screen of the program is shown in Figure 9.24.



*Figure 9.24. Screen of the program “Judge”. The subject has to decide which pronunciation out of three that sounds best? 1, 2, or 3?*

### **Evaluation of transfer to untrained material by using a listening test**

In order to assess whether effects of treatment A or B had been generalised to untrained material, the listeners also listened to and rated the recorded speech of the separate list of non-treatment words with the same structure as those used during treatment. The words were recorded in the same way before start and end of each treatment phase and the procedure of the listening test was the same as reported above.

### **Questionnaire**

An evaluation was also carried out based on a questionnaire (see Table 9.4) that the two speech therapists answered after finishing the OLP therapy method containing the following questions:

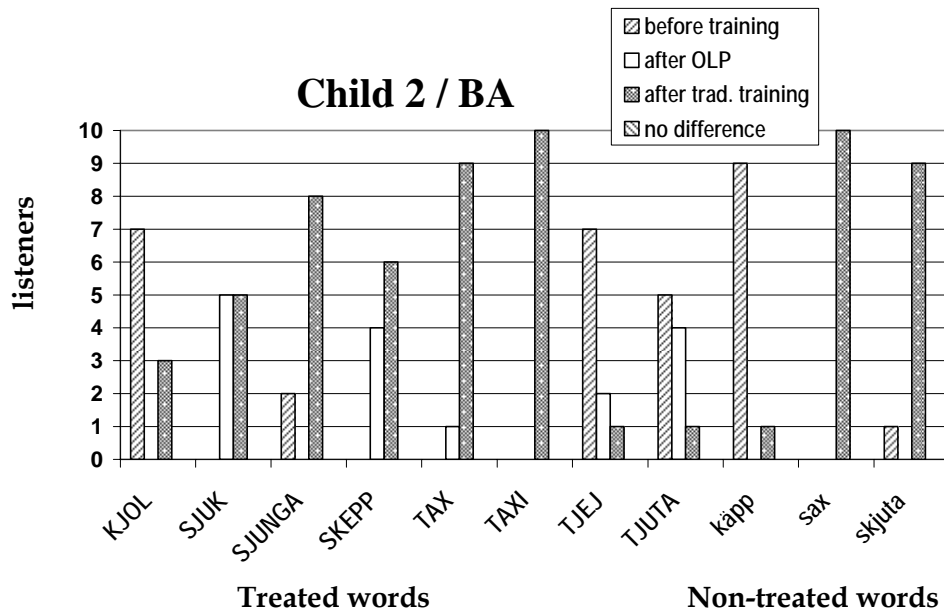
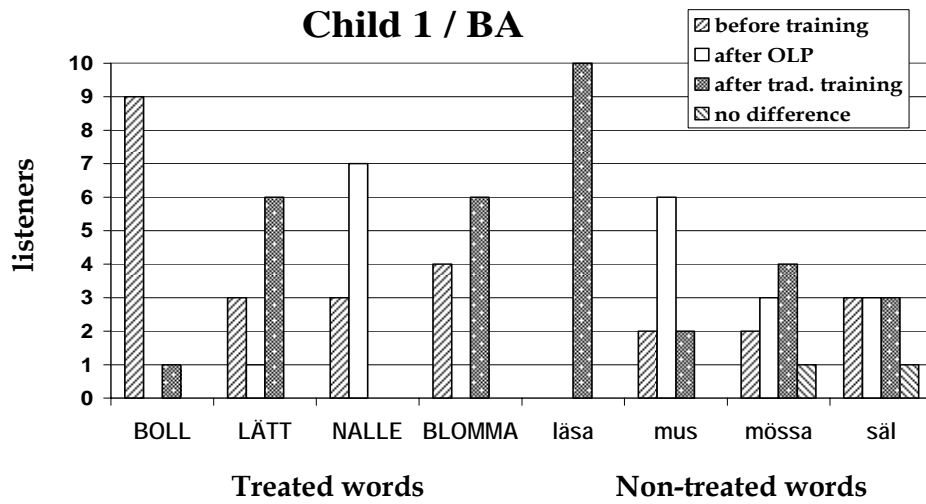
Table 9.4. *Questions that were answered by the two therapists after OLP therapy.*

1. Do you think that experience with computers is necessary to be able to use this system?
2. Do you think that phonetic knowledge is necessary to work with this system?
3. How did the system meet with your expectations?
4. Was the system easy to handle?
5. Did you consider the training as meaningful?
6. Did you like the graphics?
7. How was the system from a pedagogical point of view?
8. Was the system reliable?
9. How was the interaction with the children?
10. Did the children learn anything?
11. Did the children understand the feedback from the visual maps?
12. Did the children understand the feedback from the recognition?
13. Should any element have had more or less training opportunities?
14. Was the child motivated to train with the system?
15. Did you miss the possibilities to train something? If yes, what?
16. Additional information:
17. Try to estimate how much time you have spent for each session.

### **9.5.7. Results**

#### **Evaluation of the effects of treated and non-treated words**

Figures 9.25-9.30 show the results of the selections made by the subjects of the six children's best pronunciation out of three for both treated and non-treated words. Words written in capital letters indicate treated words and words written in small letters represent non-treated words.





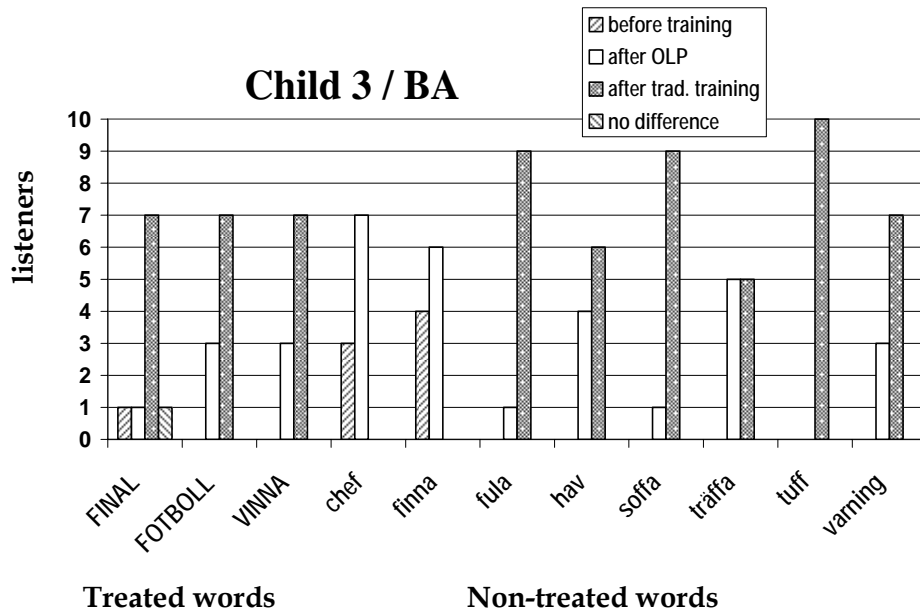
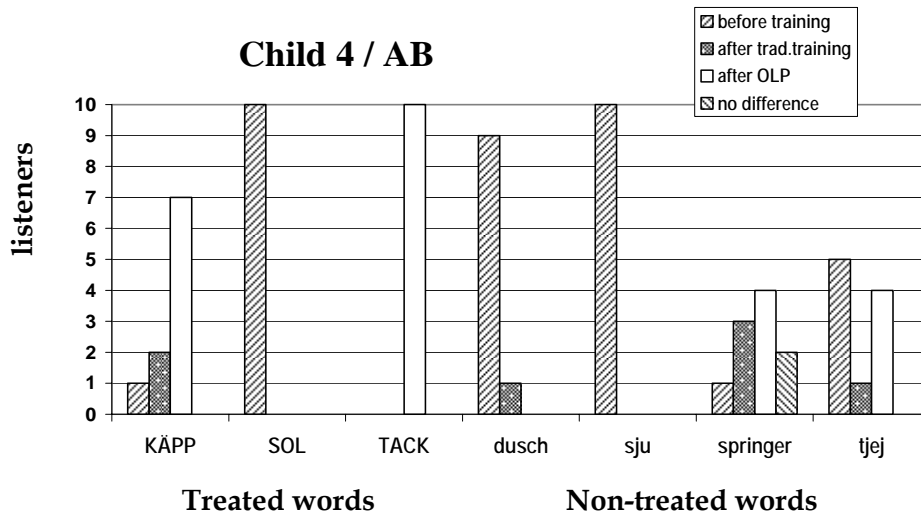


Figure 9.25-9.27. Results of ten listeners' ratings of the best pronunciation out of three of treated and non-treated words produced by three children with moderate hearing impairment from Group B(OLP) A. Words written in capital letters indicate treated words and words written in small letters represent non-treated words.



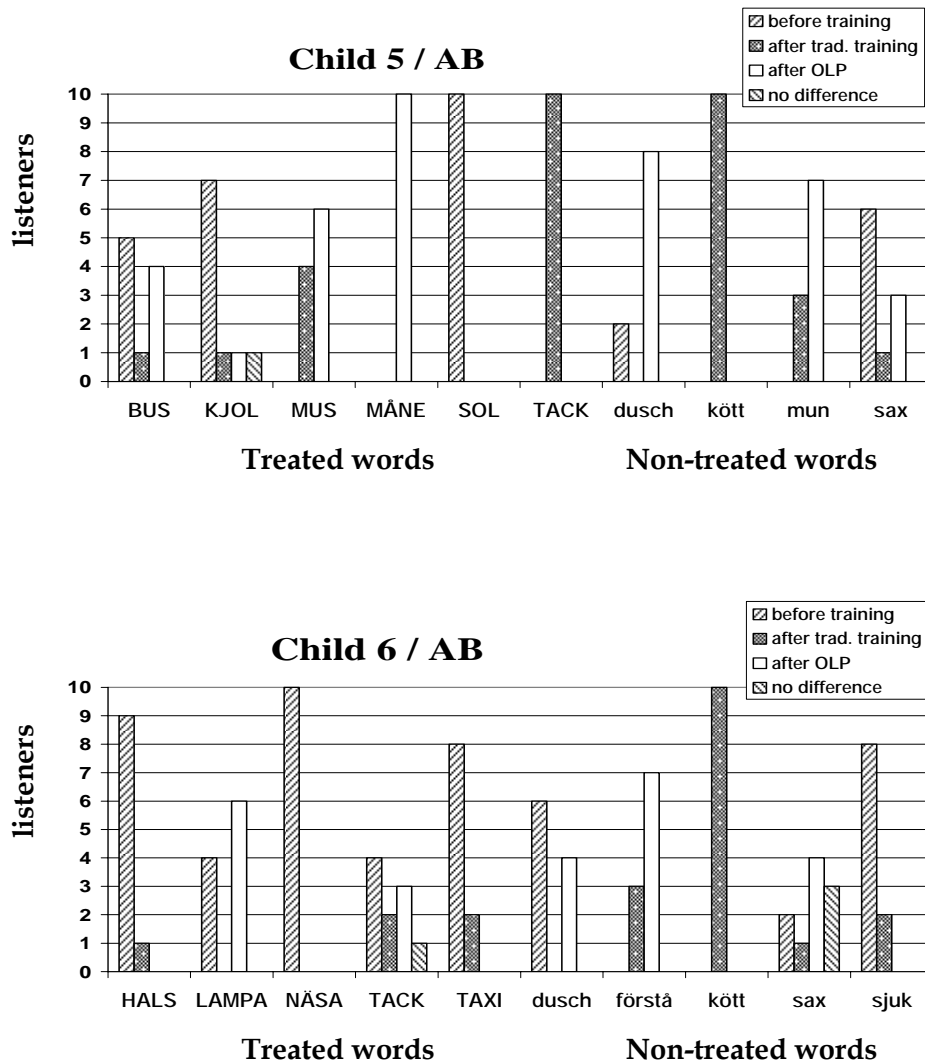


Figure 9.28-9.30. Results of ten listener's ratings of the best pronunciation out of three of treated and non-treated words produced by three children with profound hearing impairment from Group A B(OLP). Words written in capital letters indicate treated words and words written in small letters represent non-treated words.

The results show that most of the children improved their articulation after training with the OLP therapy method as well as with the traditional method. This was highly characteristic for motivated children with some functional hearing. Not all objectives

were reached but in all cases except for child 6 some of the targets were achieved.

Children with a moderate to severe hearing loss gained more knowledge and became more skilled in pronouncing their target therapy objectives than the children who had a severe to profound hearing impairment. To some extent this might depend on the therapy order but most likely this was due to the fact that these children had better speech processing capabilities and were more motivated to speech therapy than the children of Group AB. All children except one showed transfer of speech skills to untrained words of the same structure as the treatment words after having finished the trial.

The most noticeable impact of the OLP therapy method was that this type of training seemed to start a process of awareness, understanding and development of the children's speech production if this type of therapy preceded traditional training, which was the case for the children of group BA. The listeners chose most of these children's pronunciations after traditional training as the best ones even if the OLP therapy method had paved the way.

### **Questionnaire**

The observations and opinions of the two therapists after finishing the OLP therapy could be outlined in possibilities and limitations of the OLP therapy method. The therapists were happy with the evaluation and the possibility to connect with research by experimenting with a modern speech training aid under development. They thought that this method gave possibilities of co-operation as some children wanted to participate in the management of the system. The child's motivation to achieve correct articulation was enhanced with the system, and a correct pronunciation was easier to be established and kept in mind. They also thought that the OLP system could be excellent to be used by motivated hearing-impaired adults outside school to give them possibilities to train speech on their own. The most positive thing with the system, according to the therapists, was that it helped to get a child to repeat the same word/sound many, many times, which is impossible by traditional training. To sum up, the therapists agreed that the Optacia served the purpose for speech therapy with moderately, severely, and profoundly hearing-impaired children. The possibility

in OPTACIA to easily change the layout or to visualize articulatory movements rather than sounds, e.g. tongue retraction, nasality, lip rounding, etc, seems to be unique and promising.

In the instruction phase OPTACIA offered navigational feedback in the instruction phase through the visual maps that was instructive and easy for the children to understand. The children became aware of how their articulation differed from correct behaviour and understood how correct realisation should be reached. With the help of the graphical interfaces of OLPy, intensive training on existing skills to establish automaticity was possible. It was obvious that the extensive training that the OLP therapy method offered also transferred skills to untrained words, as can be seen in Figures 9.25-9.30.

However, the therapists also provided information about some weak points of the system and some important negative aspects after finishing the OLP therapy. Many of these drawbacks referred to the

- lack of user-friendliness
- the type of feedback that the system offered
- the use of automatic recognition
- the reliability of the system.

They claimed that the OLP therapy method required too much time to learn to run on one's own, necessitated very good knowledge of computer techniques, and contained a complicated procedure to reach an exercise. They maintained that the feedback from the graphics too seldom gave the child any encouragement, that the child felt unsure when there was no feedback after a good try, and that there were no possibilities of a contrastive visual training by visually comparing the child's production with a correct model. It was also found that some moving parts of the graphics were too slow-moving. This caused some children to lose interest. Some graphics needed more obviously contrasting colours and some needed a more obvious starter as the children complained that it was not apparent to them when to start their production.

However, there was a general opinion that the visual maps of OPTACIA gave the children navigational and meaningful visual feedback and the use of them during training was a contributing cause of the positive result of the OLP therapy method. However, the

therapists stated that they would like predefined phonetic maps to be integrated in the system to make it more user-friendly.

The therapists were skeptical about using automatic speech recognition with profoundly hearing-impaired children because there was a risk that the children were training against incorrect models due to the great variability and peculiarities in their speech. Some children were also too emotional, required a lot of support, were unmotivated, and were too immature to work with automatic speech recognition as it offered limited navigational feedback. The therapists also experienced that the recognizer was unreliable and gave inconsistent feedback.

## **Discussion**

Three important questions were provided in order to evaluate the OLP therapy method for hearing-impaired children. The questions were: if the OLP method of therapy was effective in increasing the intelligibility of the speech of prelingually hearing-impaired children with moderate to severe hearing losses and with severe to profound hearing losses, if there was any transfer of speech skills to untrained speech material and if the OLP therapy method functioned in a real environment.

All children were motivated and were happy to join the trial. They liked the graphics, considered the training as meaningful and came punctually in a good mood to the speech clinic. The result showed that the system was most useful for the children who had some residual hearing. All children of Group BA improved their production and became more aware of how to articulate their objectives more intelligibly after the OLP treatment. The OLP treatment also had a positive effect on the traditional treatment. By starting with OLP therapy it seemed as if a process of awareness and understanding was commenced. The navigational feedback of the visual maps was easy to understand and gave the children good instructions how to move their articulators in a correct manner. Despite the small number of children tested, the therapist thought that OPTACIA had potentials to be a useful, uncomplicated and quick method if predefined visual maps were available.

To sum up, it could be said that OPTACIA as a whole tended to function in a real environment but the graphical interfaces

together with word recognition was too complicated to use with severely and profoundly hearing-impaired children as it was unstable, unreliable and very time consuming to use in an exercise.

## **9.6. Conclusions**

When introducing new technology as well as new pedagogical and phonetic methods for speech therapy it is extremely important to investigate the effects, efficiency, efforts, and benefits involved. However, this can be especially difficult to carry out especially with profoundly hearing-impaired children, who constitute a very heterogeneous group. Therefore, mostly case studies are performed which follow a child for a longer period of time. Also difficult to perform are evaluations that comprise many schoolchildren, who should have the same type of therapy for the same period of time, due to holidays, absence from school, and the necessity for the children to leave the classroom and miss important classes. For that reason statistical methods are seldom used in these kinds of assessments.

## ***10. General Conclusions and Recommendations***

### **10.1. Recommendations**

The expected effect of speech therapy in general is:

- to establish automaticity
- to expand the speaker's phonological system
- to make the learner's best production be his/her most common production
- to transfer skills to untrained situations and linguistic use

Evaluations and experiences from visual computer-based speech therapy have shown that this type of therapy is a valuable and effective expansion in the speech clinic as it assists the therapist in the expected effects described above as well as in many problematic aspects that are included in a speech therapy program. Computer-based speech training with visual feedback stimulates, and motivates the children through amusing and variable visual feedback and seems to start a process of awareness, understanding and development of their speech production. However, it is important to point out that even the best computer program could never replace the therapist but only assist and facilitate his or her work. Computer-aided speech training is a supplement to traditional methods and has a pedagogical value for the therapist who has a good knowledge of articulatory and acoustic phonetics as well as of the computer technique.

Results have shown that the visual feedback provided by these systems helps severely and profoundly hearing-impaired children and L2 speakers to understand what is wrong and what is correct in their production. It offers meaningful feedback of distinctive contrasts that are not visible via speech-reading and consequently difficult to learn to produce correctly. Especially appreciated is the very objective evaluation of a child's speech that is provided by a visual computer-based system. The speech therapist often has a difficult role of encouraging and motivating the child at

the same time as she/he must criticise and evaluate the child's attempts. Visual speech training systems give the therapist and the child better possibilities to cooperate.

Moreover, game-like strategies to strengthen ability in refined articulation offer a child extensive training on existing skills and help the therapist in the repetitive and additional training phase. Phonemes produced by the children are matched against models by comparing target spectra. To avoid the problem of interspeaker variability, the children's own "best production" is stored as a target.

Another general advantage that has been found is that it gives the speech therapist increased flexibility in training since the task levels and performance criteria offer many choices. Much of traditional training is combined in one piece of equipment, built to display several speech parameters or features, which makes the selection of the most suitable training easy. Furthermore, efforts and improvements in training can easily be registered and documented since the information on the screen can be printed out or saved in files.

One explanation of the positive training results in the clinical evaluations of the three described therapy systems is the very instructive and pedagogical information that these systems provide in the form of a meaningful, motivational, easily comprehensible, objective visual feedback that is shown without delay. It gets the children to understand what is wrong and what is correct in their production, especially by means of a comparable navigational visual feedback, that simultaneously shows the correct model of the teacher and the deviant production of the learner on a split screen. It has been shown to be easy for the therapist to show and instruct the child about "new" speech sounds and the distinctive acoustic and articulatory cues that distinguish Swedish phonemes from each other. Training with minimal word pairs, which only differ with respect to one opposition, has proved to be an appropriate and efficient training material, especially in the instruction phase.

Assessing the speech of profoundly hearing-impaired children phonologically prior to therapy has shown to be extremely important since valuable pedagogical information about systematical phonological deviations of the existing articulation skills due to visible interference can be derived. By this means, further



development of deviant processes can be avoided during the speech acquisition of profoundly hearing-impaired children. A phonological analysis is also of importance when assessing the representation of L2 learners' existing articulation skills in spoken language.

There are some evident risks of computer-based speech training that are worth consideration. To some extent new training techniques need to be developed. There is also always a risk that therapy will be adjusted to the system and not to the child's needs.

It should also be pointed out that a computer-based therapy system must be user-friendly in order to be used by a therapist. If a program is too time-consuming to set up, too hard to survey or too demanding, this might act as a deterrent to using it, no matter how well developed the rest of the system might be. Besides, detailed manuals should be elaborated for all programs.

A long-standing wish is that training courses in computer-based speech training and training in performing phonological analyses should be an integrated part of the linguistic schooling of speech therapists. This has also reference to the teachers of second language learning when assessing the representation of L2 learners' existing articulation skills in spoken language.

### **10.1.1. Important demands on a visual computer-based speech therapy system**

Results and experiences from computer-based visual speech therapy for profoundly hearing-impaired children of various ages have shown that in order to be efficient and enhance the possibility for a child to develop intelligible speech, a visual speech training aid has a number of important requirements as follows:

- Clear instructions and pedagogical manuals must be created and made available for use with different groups of children and clients.
- The visual feedback of the child's voice and articulation should be shown immediately and without delay.
- The system must be acceptable to the therapist as well as to the child, which means that the system must be attractive,

interesting, easily comprehensible, easy to handle, and motivating.

- The visual pattern must be natural, logical, and easily understandable. This means that the training parameters as, e.g., pitch should be displayed vertically as pitch variations occur, intensity should be shown through the size of an object that becomes larger as a sound becomes louder and smaller as the sound becomes softer, intonation and stress through a continuous red curve, duration could be shown horizontally and voicing through a relationship between voicing and the change of a colour.
- The system should provide a contrastive training, that is, the correct model of the therapist and the deviant production of the child are shown simultaneously and compared with each other.
- The system should provide a flexible, individual, and structural speech and voice training and give an objective evaluation of the child's training results.

### **10.1.2. Efficiency of visual feedback of prosodic parameters within spoken L2 training**

L2 speakers' ability to discriminate between and to produce Swedish speech sounds, stress patterns, intonation, word accent and rhythm improved with computer-based speech training with audiovisual feedback. Especially prosodic features improved due to the comparable navigational feedback that made them understand in what way their speech differed. Some of the Swedish phonetic contrasts were more easily learned than others. A general opinion of the L2 speakers was that they became more aware of the Swedish pronunciation and spoke more Swedish after training. However, most of them wanted the training period to be extended for better long-term results. The possibility of home training was also a general desire. This work shows that it would be meaningful and advisable to use computer-based speech training with audiovisual feedback for training of both perception and production of Swedish prosody within spoken L2 training (CAPT).

## **10.2. Comparison of the three systems**

Figure 10.1 summarises a comparison of the three product-oriented computer-based speech therapy systems with visual and optional auditory feedback that are reviewed in this thesis. All systems are product-oriented, but the sounds produced by the speaker are in different ways visualized on the computer screens after processing. All systems have optional auditory feedback, game-like exercises, and a user management. Dissimilarities and the most distinguishing qualities of each system are shown in the figure.

SpeechViewer is a very powerful system for real time speech therapy. The thirteen different programs can be used with both children and adults, and normally hearing persons as well as hearing-impaired persons. Some of the programs, aimed at therapy with small children, are very easy to use by the therapist while others are technically very sophisticated and somewhat complicated to adjust to the client. There is no integrated vocabulary for training or structural training design in the system, a fact that requires good knowledge of phonetics and traditional clinical therapy to use the system in the most effective way. Some resourceful programs, such as "Pitch and Loudness" and "Phonology", are extremely useful also for L2 speakers but they are dependent on the therapist's proficiency to make the best use of them.

Box of Tricks, on the other hand, is a much more user-friendly system and is useful for speech and hearing-impaired children below 12 years of age. Prerecorded vocabularies, comparable targets of a 10-year-old reference speaker, symbolic speech pictures for each phoneme, and illustrated words for children who have not yet learned to read make the system easy to handle but exclude the possibility to set up special individual training exercises.

The OLP therapy method for distance learning, on the other hand, contains a library of training words to choose from and a structural training method but no reference speaker. The system uses ASR with best productions that makes the system complicated and time-consuming to handle. The evaluation showed that this method was not suitable to use with hearing-impaired children (see section 10.3.3) but is more suitable to use with adults suffering from functional articulation deviations.

*Computer-Based Speech Therapy Using Visual Feedback*

| <b>Box of Tricks</b>  | <b>SPV3</b>  | <b>OLP therapy</b>                                     |
|---|--|--|
| - Prerecorded vocabularies                                      | - No vocabulary  | - Library  |
| - Reference speaker   | - No reference speaker   | - No reference speaker                                 |
| - Speech pictures   | - Curves, oscillograms and spectra   | - Visual maps  |
| - Animated graphics: awareness                                  | - Animated graphics: awareness, control  | - Animated graphics: recognition                       |
| - Symbolic pictures   |  |  |
| - Illustrations of words  |  |  |
| - Easy to handle  | - Needs good knowledge of articulatory and acoustic phonetics and computer technique | - Complicated to set up exercises                      |
|   |  | - Needs good knowledge of computer technique           |
|   |  | - Time-consuming                                       |
| - Phoneme-based spectral comparison                             | - Phoneme-based spectral comparison  | - ASR of syllables and words                           |
| - Follows steps in traditional therapy based on existing skills | - Dependent on the therapist   | - Follows steps in traditional therapy                 |
| - Contrastive training without exception                        | - Contrastive training in 2/13 programs  | - No contrastive training                              |
| - Targets based on a child-data-base and a reference speaker    | - Targets based on the client's best productions                                     | - Targets mostly based on the child's best productions |
| - Animated graphics: awareness                                  | - Animated graphics: control, awareness  |  |
| - Evaluative: acceptable  | - Evaluative: exactly  | - Evaluative: exactly (OLPy)                           |
| - Rewarding: encouraging  | - Rewarding: hitting the mark  | - Rewarding: hitting the mark(OLPy)                    |
| - Navigational: comparable articulation pictures                | - Navigational: comparable   | - Navigational (OPTACIA): criterion-referenced         |
| - Complement the therapist                                      | - Complement the therapist   | - Might replace the therapist                          |
| - Real-time therapy   | - Real-time therapy  | - Distance-learning                                    |

*Figure 10.1. Comparison of three computer-based speech therapy systems with visual feedback*

However, the integrated program OPTACIA, which works with phonetic visual maps, was shown to be cost-effective and helpful to use with hearing-impaired children.

Suggestions are given in section 10.4.1 for the most efficient use of the different types of visual feedback found in the three systems during different phases of speech therapy.

### **10.3. Recommended therapy design**

#### **10.3.1. General design of computer-based speech therapy with visual feedback**

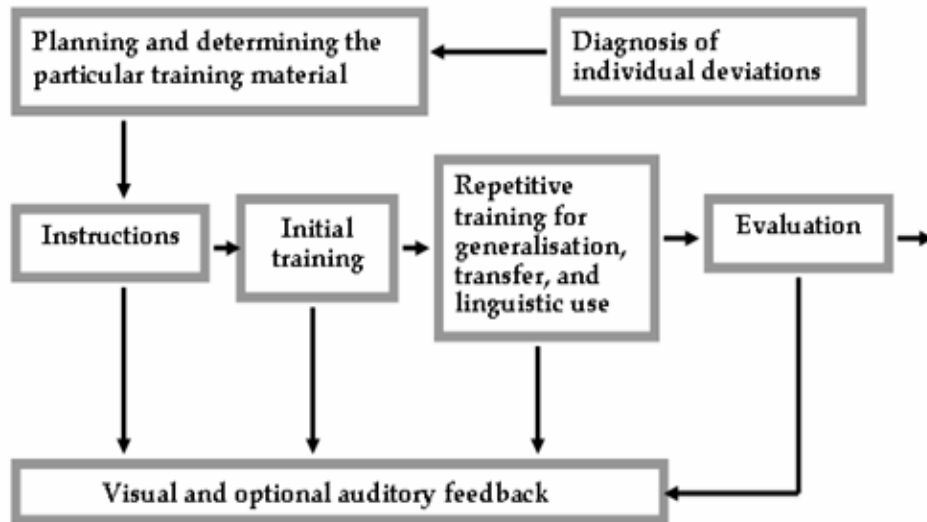
The following recommendation of a general design of computer-based speech therapy with visual and optional auditory feedback, shown in Figure 10.2, is a result from experiences of pedagogical aspects, phonetic and phonological methods, and speech technology concerning the three existing systems described in this thesis.

First the individual deviations that have the most impact on the intelligibility must be diagnosed. After that speech material must be introduced, for instance minimal word pairs, that gives good visual instruction to make the child aware of his/her deviant production, in what way it deviates, and how to produce it correctly. The next phase is a training phase aimed at changing a deviant realization, strengthening a successful production and establishing a more correct and intelligible pronunciation. To transfer skills to untrained material and linguistic use, a phase of repetitive and additional training is necessary. The target production must be repeated and practised in a variety of contexts frequently recurring.

Guidance, reinforcement and assessment should be made at every stage of this learning process through specially designed visual and optional auditory feedback. Figure 10.2 displays a diagram illustrating these principles.

A training session should not exceed 30 minutes to be efficient. A complete training-period should contain about 10 sessions before the final evaluation. The evaluation should be done by the therapist, by for example comparing a special recorded text material that the child reads before and after the training period.

When it comes to small children who cannot yet read, the evaluation must be based on an illustrative material.



*Figure 10.2. Components and information flow in computer-based speech training with audio-visual feedback*

### **10.3.2. Structural training design**

A structural therapy and assessment procedure to be used with children who have some verbal skills was developed by Öster (1996) to help the therapists in their work for the following two reasons. One reason was that SpeechViewer is a tool for speech training that does not contain any speech-training material and is not structured in different levels of therapy but is dependent on the therapist's power of invention. It consists of different programs for different types of training. Because of this the therapists working with the program must have good knowledge of both phonetics and computer techniques to utilise the system in the best way. The other reason was that therapy and assessment are inseparable procedures, as assessment is required regularly during training. To assist the therapists both in the assessment procedure as well as in the structural training, the procedure was elaborated in the form of a protocol.

Training and assessment should be done on six levels, following steps of traditional speech therapy, starting with motor training on isolated speech sounds for training of basic skills like

respiration, intensity, pitch, phonation and articulation. Then follow syllables where the consonant is chosen on the basis of visibility, repeated syllables with different stress patterns, repeated alternated syllables for automatization of prosody, words containing the speech sounds of current interest for relevant training, and finally a short phrase containing the topical words.

All the time the child's mastery of breath control, intensity, pitch, duration, and voice quality on all levels should be listed in the protocol by the therapist, according to statistics provided by the system and according to the therapist's opinion. This structural training design in six levels is also added in Box of Tricks as well as in the OLP therapy method, see list below. Characteristic of such training is that it is the most effective one as it is based on existing skills and expands the existing articulation skills.

**Respiration**

- Intensity
- Pitch
- Phonation
- Articulation

**Vowels**

- Isolation
- Syllables
- Repeated syllables with different stress patterns
- Words (one-two-polysyllables)
- Minimal word-pairs
- Short phrases

**Consonants**

- Isolation
- Syllables and Sound sequences (initial, medial, final)
- Repeated syllables
- Words (one-two-polysyllables)
- Minimal word-pairs
- Short phrases

Training exercises are designed on the different levels so that training on the segmental level will be expanded to the word level with the same phonemes. Training on the word level is then

expanded to the phrase and sentence level resulting in an expansion of the training of the same phonemes and words.

## **10.4. Visual feedback strategies**

Experiences from these three systems have shown that in order to make the therapy as effective as possible, it is important that certain types of visual feedback should be connected to specific phases in the training process of different user groups.

### **10.4.1. Type of visual feedback for severely and profoundly hearing-impaired children**

An applicable strategy for severely and profoundly hearing-impaired children is recommended in Figure 10.3. Children with or without hearing impairments with other types of articulatory impairments will also benefit from computer-based speech therapy with this type of well-developed visual feedback strategies. During the instruction phase, animated graphics and a comparable navigational feedback is appropriate to use in order to make the child aware of his/her deviant production, in what way it deviates, and how to produce it correctly. A criterion-referenced navigational feedback, based on visual targets of the child's production that are compared to the norm and to their own productions, is also proper to be used in the instruction phase as well as in the training phase to let the child see the improvements. A navigational feedback that is dependant on correct control of the voice, pitch or intonation is informative in the training phase, as well as an evaluative feedback that immediately rewards the child if the production is acceptable and more correct than before. The reward is aimed to be encouraging as it reinforces the slightest improvement. Training based on a phoneme-based comparison of a stored model of the child's so-far best production can stimulate the child in various amusing game-like layouts. As the child's production is improving, new models must of course be stored as targets in a simple and quick way.



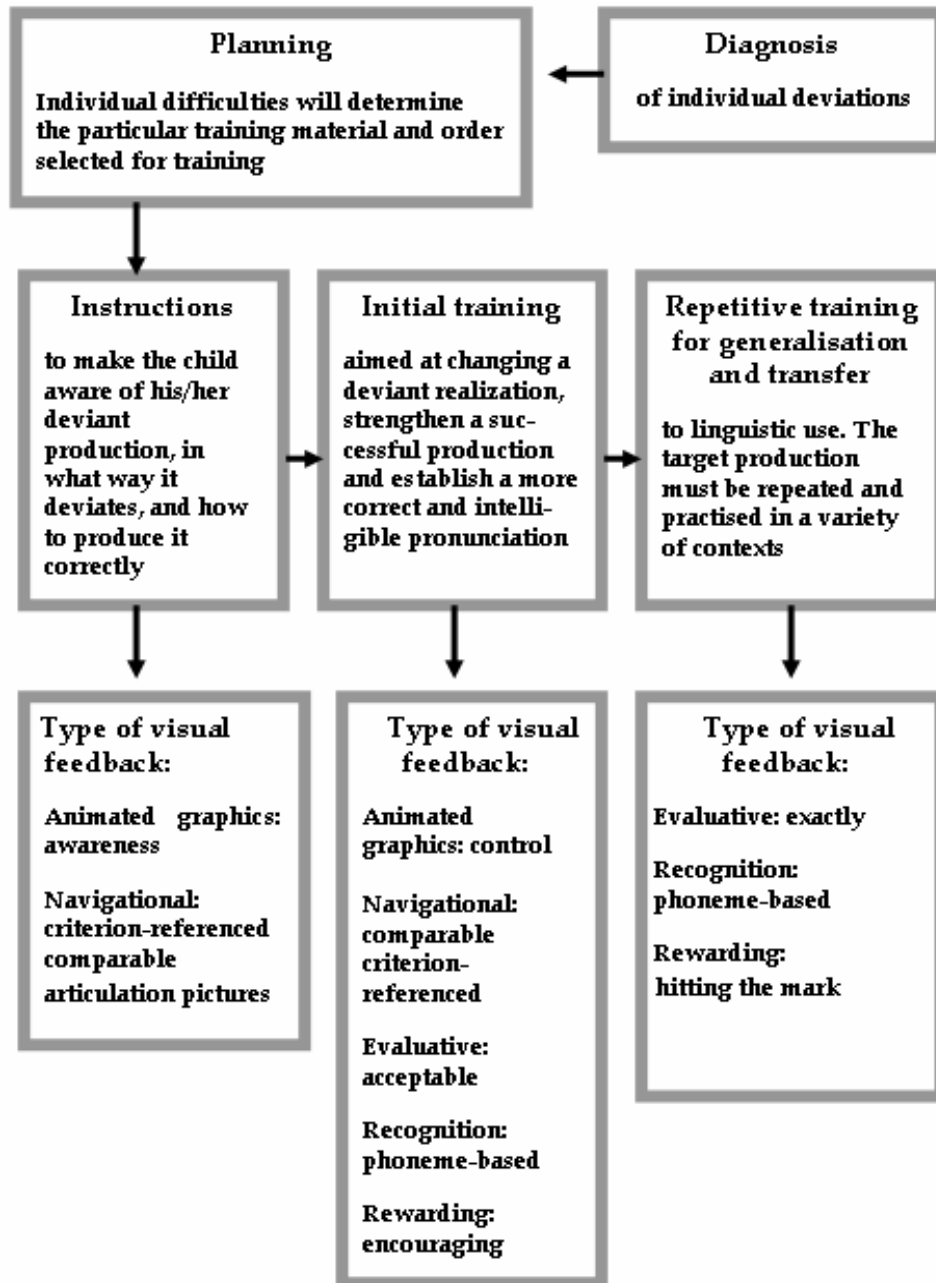


Figure 10.3. Therapy design and visual feedback strategies for severely and profoundly hearing-impaired children.

During the last phase, where lots of repetitive and additional training is needed to automate correct production for transferring skill to

linguistic use, the evaluative feedback should only reward the child when the production is correct and hitting the mark. Lots of phoneme-based recognition exercises that are amusing, motivating, and challenging are needed to intensify the training.

### 10.4.2. Type of audio-visual feedback for L2 learners

Audio-visual feedback of prosodic parameters such as intonation contours, stress patterns and fluency has been shown to be efficient for L2 learners and supports a wider usage within CAPT. Figure 10.4 illustrates relevant parts in this approach. The training phases are identical as can be seen in the figure.

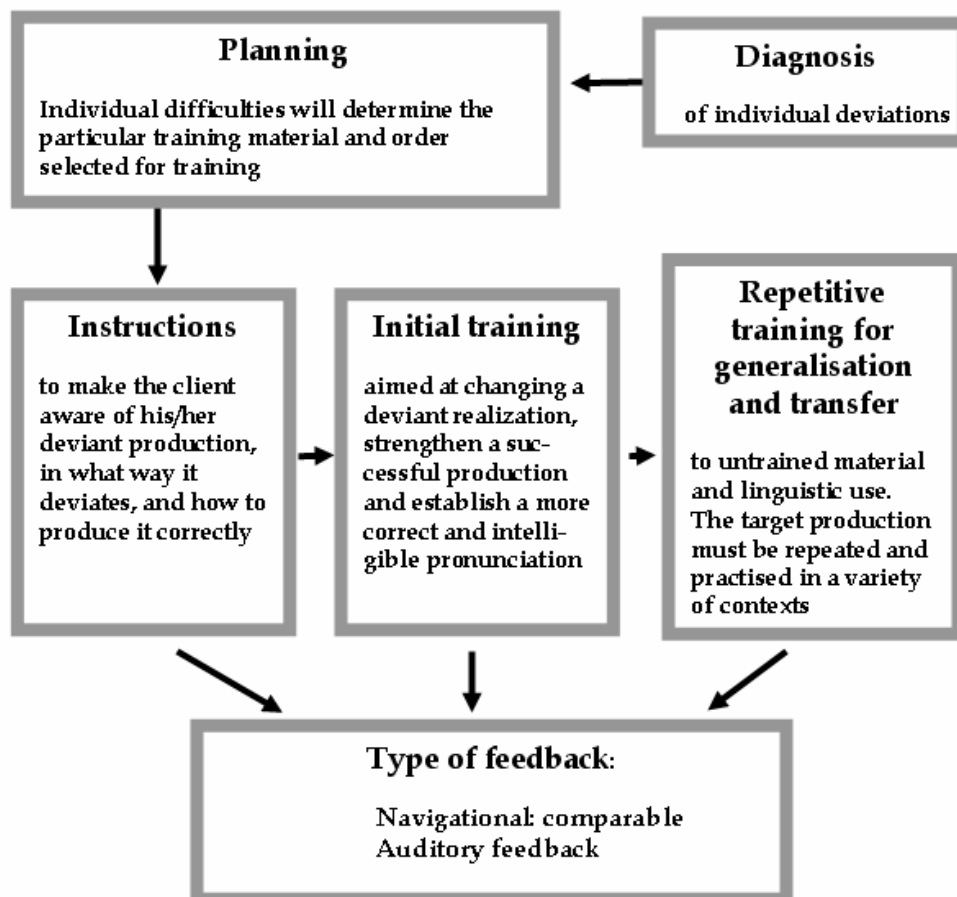


Fig. 10.4 Feedback strategies for L2 learners.

In all phases (instruction, training and repetitive and additional training for transfer) a comparable navigational visual feedback accompanied by an auditory feedback for perceptual training is recommended. The auditory and visual similarity of the teacher's model and the client's production is enough to stimulate the client to continue the training in a meaningful way.

There is substantial research concerning the use of automatic word recognition in CAPT as it automatically detects segmental errors fairly well. However, there are still problems with identifying prosodic deviations, a fact that supports a wider usage of this type of comparable navigational audio-visual feedback of prosodic parameters in CAPT.

### **10.4.3. Use of automatic speech recognition and spectral comparison of phonemes with profoundly hearing-impaired children**

The use of automatic word recognition and spectral comparison of phonemes has been evaluated with profoundly hearing-impaired children in this study. It was a general opinion that the feedback from the automatic word recognition in the OLP method was improper for use with profoundly hearing-impaired children for three reasons. In the first place, the children felt unsure when there was no encouraging rewarding feedback after a good try. Moreover there was no navigational feedback during training as well as no possibility of a contrastive training that visually compares the child's effort with a correct model. These three types of visual feedback have been shown to be the most central ones to stimulate and inform severely and profoundly hearing-impaired children in their speech acquisition.

The normal goal of ASR is to classify all utterances correctly, even if they are not pronounced accurately. However, in pronunciation teaching with severely and profoundly hearing-impaired children the system must be able to do more than distinguish between good and poor pronunciations. Intelligibility is in focus and the quality of the child's speech determines whether most people can understand them easily or with difficulty. Therefore such a system must provide an acceptable evaluative feedback that

provides a measure of goodness and shows that the child is doing very well and has made some improvements. Thus the two systems, ASR and pronunciation teaching systems have different aims.

The other reason is that severely disordered speech produced by profoundly hearing-impaired children often contains temporal, phonatory and aerodynamic deviations that correlate poorly with ASR accuracy. Instead the segment-based speech comparison used in SpeechViewer and Box of Tricks that compares a recorded phoneme inside a word with a stored production behaved extremely well. This particular method is known as speaker-dependent, in which a particular utterance is compared to a stored target of a child's best production. Phoneme-based speech recognition has shown to be useful in programs for repetitive and additional training that aim at strengthening a successful production and get the child's best production to be his/her most common production in all possible surroundings.

## **10.5 Conclusions**

In this thesis, work carried out over many years is presented dealing with severely and profoundly hearing-impaired children's possibilities to acquire speech skills through computer-based speech therapy with visual feedback. This technique is also tested and evaluated in the pronunciation training with adult second language learners. Studies that investigate the effects of speech input limitations on speech production, the interaction of individual deviations on speech intelligibility, phonetic realizations of phonological systems and the use of visual feedback in computer-based speech therapy reflect the problems and possibilities that must be considered when teaching severely and profoundly hearing-impaired children speech. Hypotheses are presented that a specialized individual diagnostical method must be carried out before therapy to base the therapy on. Different visual feedback strategies and therapy design for each of these two client groups: children with severe and profound hearing losses and adult second language learners, should be utilized to get the clients to benefit from computerized speech training.

It is well documented that the speech intelligibility of severely and profoundly hearing-impaired children varies a great deal and

cannot be predicted from the degree of hearing loss, as measured by pure-tone audiometry. The sort of speech a profoundly hearing-impaired child might develop depends more on the functional hearing for speech, that is, residual hearing for speech, amount of hearing aid use, amount of auditory training, discrimination ability of speech features, learned ability to identify speech sounds, phonological short-term memory and speech processing capabilities. A method for testing the functional hearing for speech through an analytical speech perception test that gives supplementary information to the pure tone audiogram is described and tested.

An important part of the thesis discusses different relevant aspects of the speech of severely and profoundly hearing-impaired children, as for instance age of onset of hearing loss and dependence on visibility of speech sounds, and reviews determining factors and prerequisites for severely and profoundly hearing-impaired children to benefit from speech therapy and develop intelligible speech. The objective is to give necessary basic details of why a special visual feedback strategy and therapy design must be used during speech therapy for severely and profoundly hearing-impaired children, in contrast to what should be the most useful in spoken language training with adult second language learners. The hypothesis that the use of audio-visual comparable feedback helps L2 learners in perceiving and producing especially prosodic parameters is tested in the thesis. The clinical evaluation shows promising results that indicate a wider usage of audio-visual feedback in CAPT systems.

The hypothesis that involves a diagnosis method in the form of an individual phonological assessment based on a detailed phonetic analysis, states that this kind of analysis gives information about how the articulation is realized linguistically, and reveals deviantly realized phonological contrasts. Part of the work deals with the development of a method for such a phonological assessment which investigates what profoundly hearing-impaired children's speech does express and how speech sounds are realized by L2 learners. The method is tested and is shown to assist the therapist in deriving significant and individual information about the clients' speech productions as it gives possibilities to outline a constructive and individualized speech therapy with well-designed visual feedback. The phonological assessment provides information

about deviations in different linguistic contexts of those speech sounds that the clients know the articulation of and the way these differ from the normal model. In addition, the method also provides information about the speech-sounds that they cannot yet produce. A phonological assessment of a profoundly hearing-impaired child's speech also shows whether a deviant pronunciation in fact is a realization of signaling a contrast of meaning and if so in what way the phonetic element used differs from the normal model.

Three Swedish computer-based speech therapy systems with visual feedback are presented in the thesis. The systems are assessed with reference to functionality, speech visualisations, and types of visual feedback, therapy design and practical usability. Fundamental principles of the three systems and their different solutions for working with reference speech, target spectra, contrastive training, objective evaluation, understandable visualisations and clinical management are discussed. Two of the systems are available as commercial products: the IBM SpeechViewer III and Box of Tricks, called Trollerilådan in Sweden. The original SpeechViewer program was developed 15 years ago and is available in twenty different languages. Box of Tricks is a result of the completed SPECO-Project, funded by the EU INCO-COPERNICUS program and is developed in four languages: English (Box of Tricks), Hungarian (Varázsdoboz), Slovenian (ARTI) and Swedish (Trollerilådan). The third system, OLP-therapy, is a result of the recently finished EU Quality of Life and Management of Living Resources project, with participation from Greece, France, Spain, Sweden, and the UK. All the systems are developed for children with speech and/or hearing impairments who have difficulties in understanding and producing spoken language. They all use visual information as an alternative feedback but present it in different ways. All systems use different kinds of visualisations and speech pictures of important speech parameters, correct from the acoustic-phonetic point of view, designed to be understandable and interesting for young children. The SpeechViewer and Box of Tricks are designed to supplement speech therapy while the most innovative aspect of the OLP-therapy system is a distance learning application entirely built on distributed (Internet) technology.

## *General Conclusions and Recommendations*

Several clinical evaluation studies with both severely and profoundly hearing-impaired children as well as with adult second language learners investigate the effectiveness of the systems, the degree of transfer and generalization, and the overall functionalities. The findings from the evaluations showed positive results in many ways, as for instance assisting the therapists to instruct and explain what is wrong and what is correct articulation through comparable visual feedback, in helping the client with significant amounts of additional training, in making the clients aware of nonvisible manner and place of articulation through an immediate and meaningful visual feedback and to get a more positive cooperation in the speech clinic. The specified recommendations in the end of the thesis about how to utilize different types of visual feedback during different phases of therapy as well as the fact that a structured therapy design should be adjusted for different client groups are important to take into consideration for future developments of computer-based speech therapy systems with visual feedback.





## 11. References

- Abberton, E., Fourcin, A., Hazan, V. (1985). Phonological competence with profound hearing loss, *Paper presented at the Int. Congress on Education of the Deaf*, Manchester.
- Adolvsson, K., Forsén, H. (1968). *Samband mellan talförståelse och talkvalitet*, Examensarbete vid Lärarhögskolan i Stockholm (Manillaseminariet).Handledare Arne Risberg, KTH.
- Amcoff, S. (1973). Relationer mellan språkliga uttrycksformer. En undersökning av elever i specialskolan för hörselskadade. *Pedagogisk forskning*, Uppsala, No.1.
- Arends, N., Povel, D. J., Michielsen, S., Claassen, J., Feiter, I. (1991). An Evaluation of the Visual Speech Apparatus (VSA), *Speech Communication*, 10, 405-414.
- Arends, N. (1993). *The visual speech apparatus*, IvD/RES/9303/01, Instituut voor Doven, Sint Michielgestel, The Netherlands.
- Arlinger, S., & Hagerman, B. (1997). *The Swedish approach to speech audiometry. Speech Audiometry*, 2:a uppl. (red. M. Martin): Whurr Publishers, London.
- Bamford, J., Saunders, E. (1991). *Hearing Impairment, Auditory perception and Language Disability, Studies in Disorders of Communication*, Whurr Publ., London Jersey City, ISBN 1-870332-01-6.
- Bannert, R. (1990). *På väg mot svenskt uttal*. Lund: Studentlitteratur.
- von Békésy, G. V. (1959). Similarities between hearing and skin sensations, *Psychol. Rev.*, 66, 1-22.
- Bench, R. J. (1992). *Communication skills in hearing-impaired children*, Whurr Publishers Ltd, ISBN 1-56593-075-4, London N1 2UN, England.
- Bernstein, J. (1977). Intelligibility and simulated deaf-like segmental and timing errors. *Record IEEE Int. Conf. Acoust. Speech and Signal Processing*, Hartford.
- Beskow, J., Granström, B., House, D., Lundeberg, M. (2000). Experiments with verbal and visual conversational signals for an automatic language tutor. *Proc of InSTIL 2000*, Dundee.

- Beskow, J. (2003). *Talking Heads. Models and Applications for Multimodal Speech Synthesis*, Ph.D Thesis, KTH, Sweden. ISBN 91-7283-536-2.
- Beskow, J., Karlsson, I., Kewley, J. and Salvi, G. (2004). SYNFACE - A Talking Head Telephone for the Hearing-impaired. In K Miesenberger, J Klaus, W Zagler, D Burger eds, *Computers helping people with special needs*, 1178-1186.
- Binnie, C.A., Daniloff, R.G. and Buckingham, H.W. (1982). Phonetic disintegration in a five-year-old following sudden hearing loss, *J. Speech and Hearing Disorders*, Vol 47, p 181-189.
- Boothroyd, A. (1970). Concept and control of fundamental voice frequency in the deaf - an experiment using a visible pitch display, Paper presented at the *International Congress of Education of the Deaf, Stockholm, Sweden*.
- Boothroyd, A. (1984). Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *Journal of Speech and Hearing Research*, 27, 128-134.
- Boothroyd, A. (1995). *Speech perception tests and hearing-impaired children*, in *Profound Deafness and Speech Communication*, G. Plant and K-E Spens, London, Whurr Publisher Ltd, pp.345-371.
- Borg, E., Risberg, A., McAllister, B., Undemar, B. M., Edquist, G., Reinholdson, A-C., Wiking-Jonsson, A., Willstedt-Svensson, U. (2002). Language development in hearing-impaired children. Establishment of a reference material for a 'Language test for hearing-impaired children', LATHIC. *Int. J. Pediatr. Otorhinolaryngol.* 1;65(1):15-26.
- Bush, C.N., Edwards, M.L., Luckau, J.M., Stoel, C.M., Macken, M.A., Petersen, J.D. (1973). On specifying a system for transcribing consonants in child language: A working paper with examples from American English and Mexican Spanish. *Report, Dept. of Linguistics, Stanford University*.
- Bälter, O., Engwall O., Öster, A-M., Kjellström, H. (2005). Wizard-of-Oz Test of ARTUR - a Computer-Based Speech Training System with Articulation Correction [pdf]. In *Proceedings of the Seventh International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 36-43, October 9-12, 2005, Baltimore, MD.

- Calvert, D. (1961). *Some Acoustic Characteristics of the Speech of Profoundly Deaf Individuals*. Ph.D. thesis, Stanford University, Palo Alto, CA.
- Calvert, D.R. and Silverman, S.R. (1975). *Speech and Deafness*. Washington: Alexander Graham Bell Association for the Deaf.
- Carlson, R., Granström, B., Hunnicutt, S. (1982). A multi-language text-to-speech module, *Proc. ICAASP 1982*, 3, 1604-1607..
- Cohen, M. L. (1968). The ADL Sustained Phoneme Analyzer, *American Annals of the Deaf*, 113, 168-177.
- Cowie, R. and Douglas Cowie, E. (1992). *Postlingual acquired deafness: Speech deterioration and the wider consequences*, Berlin, ISBN 3110125757, Publisher Mouton de Gruyter.
- Cramer, K.D., and Erber, N.P. (1974). A spondee recognition test for young hearing-impaired children. *Journal of Speech and Hearing Disorder*, 39, 304-311.
- Crepy, H., Denoix, B., Destombes, F., Rouquie, G., Tubach, J-P. (1983). Speech Processing on a Personal Computer to Help Deaf Children, *World Computer Congress*, 669-671.
- Crepy, H., Destombes, F., El Breze, M., Rouquie, G. (1986). *Speech training on a personal computer*, Int. Congress on Acoustics, Toronto.
- Dalby, J., Kewley-Port, D. (1999). Explicit pronunciation training using automatic recognition technology, *Calico Journal*, 16, 3, 425-445.
- Dodd, B. (1974). *The acquisition of phonological skills in normal, severely subnormal and deaf children*. Doctoral Dissertation, Univ. of London.
- Dodd, B. (1976). The phonological systems of deaf children. *JSHD*, 41, 2, 185-197.
- Dodd, B. (1988). Lip-Reading, Phonological Coding and Deafness. In Dodd and Campell, eds., *Hearing By Eye: The Psychology of Lip-Reading*, 177-189.
- Ellegård, A. (1982). *Språket och hjärnan*, Hammarström & Åberg, ISBN 91-7638-031-9.
- Engwall, O., Wik, P., Beskow, J., Granström, G. (2004). Design strategies for a virtual language tutor [pdf], In *Proc of ICSLP*

- 2004, vol. III: 1693-1696. Jeju Island, Korea, 4-8 October. Editors: Soon Hyob Kim and Dae Hee Youn.
- Engwall, O., Bälter, O., Öster A-M., Kjellström, H. (2006). Designing the user interface of the computer-based speech training system ARTUR based on early user tests, to appear in *Journal of Behavioural and Information Technology*.
- Erber, N.P (1974a). Pure-tone thresholds and word-recognition abilities of hearing-impaired children. *Journal of Speech and Hearing Research*, 17, 194-202.
- Erber, N.P (1974b). Visual perception of speech by deaf children: recent developments and continuing needs. *Journal of Speech and Hearing Disorder*, 39:2, 178-185.
- Erber, N.P (1977). Speech perception by profoundly hearing-impaired children, *Research Conference on Speech-Processing Aids for the profoundly hearing-impaired*, Gallaudet Collage, Washington, 23-26 May.
- Eriksson, E., Bälter, O., Engwall, O., Öster, A-M., Kjellström, H. (2005). Design Recommendations for a Computer-Based Speech Training System Based on End-User Interviews [pdf] . In *Proceedings of the Tenth International Conference on Speech and Computers*, pp. 483-486, October 17-19, 2005, Patras, Greece.
- Eskenazi, M. (1996). Detection of foreign speaker's pronunciation errors for second language training-preliminary results, *Proceedings of the international conference on spoken language processing, '96*.
- Eskenazi, M. (1999). Using a Computer in Foreign Language Pronunciation Training: What Advantages?, *Calico Journal*, 16, 3, 447-469.
- Ewertsen, H. W. (1973). Auditive, visual & audio-visual perception of speech (Operation Helen. First preliminary report), *The State Hearing Centre, Bispebjerg Hospital, Copenhagen*.
- Ewing, I. R. (1941). Lipreading for adults, *Teacher of the Deaf*, 39, 3-6.
- Fairbanks, G. (1960). *Voice and articulation drillbook*, 2<sup>nd</sup> edn. New York: Harper & Row. pp124-139.
- Fant, G. (1959). Acoustic analysis and synthesis of speech with applications to Swedish, *Ericsson Technics* No. 1, 1-108.

- Fant, G. (2001). On the speech code, TMH-QPSR, KTH; 61-67, Vol 42.
- Finnerty, J. (1996). *Analyzing the Development of Early Childhood Language*, Educational Software Research Inc, Lexington Mass.
- Fisher, C.G. (1968). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research*, 11, 796-804.
- Fischer, S. (1995). Critical periods for language acquisition: Consequences for deaf education, *18th international Congress on education of the deaf*, Tel-Aviv, Israel, July 16-20.
- Flege, JE. (1989). Using visual information to train foreign language vowel production, *Language Learning* 38, 365-407.
- Flege, JE. (1998). Second-language learning: The role of subject and phonetic variables. *Proc Speech Technology in Language Learning (STiLL 98)*, Marholmen, Sweden.
- Fletcher, S. G. (1986). Visual feedback and lip-positioning skills of children with and without impaired hearing, *JSHR*, 29, 231-239.
- Fletcher, S. G., Dagenais, P. A., Critz-Crosby, P. (1991). Teaching Consonants to Profoundly Hearing-Compromised Speakers Using Palatometry. *Journal of Speech and Hearing Research*, 34, 929-942.
- Fourcin, AJ. & Abberton, ERM. (1971). First application of a new laryngograph. *Med Biol Rev*, 21: 172-182.
- Furth, H. (1964). Research with the deaf: implications for language and cognition. *Psychological Bulletin* 62, 145-164.
- Gold, T. (1980). Speech Production in Hearing-Impaired Children, *J. Comm.Dis.* 13, 397-418.
- Göllesz. (1972). ref. by Mills, A. 1983, see below.
- Granqvist, S. (1998). Spruce signal workstation add-on package, Stockholm, Sweden. See information on web page of Hitech Development AB at <http://www.hitech.se/development/>, april 2006.
- Granström, B. & Öster, A-M. (1994a). Speech synthesis for hearing-impaired persons - in research, training and communication, *Proceedings from 2nd Int. Symposium on Speech and Hearing Sciences*, Sept. 24-25 1994, Osaka, Japan, 49-65.
- Granström, B. & Öster, A-M. (1994b). Speech synthesis for hearing-impaired persons - in research, training and communication, *STL/QPSR 2-3/94*, 93-111.

- Grewel, F. (1963). Remarks upon the acquisition of language in deaf children, *Language and Speech*, 1963, 6, part 1, 37-45.
- Grunwell, P. (1985). *Phonological assessment of child speech (PACS)*. NFER-Nelson, Winsdor, UK/College-Hill Press, San Diego, CA.
- Grunwell, P. (1987). *Clinical Phonology*, Williams and Wilkins, Baltimore.
- Gustafsson, A. (1984). *Svenskspråkiga färdigheter hos specialskolelever, en litteraturoversikt och empirisk studie*, Högskolan i Örebro, Institutionen för psykologi och pedagogik.
- Gårding, E. & Bannert, R. (1979). *Optimering av svensk uttalsundervisning*, Praktisk Lingvistik 1, 1-9. Lund: Department of Linguistics, Lund University.
- Hatzis, A. (1999). Optical Logo-Therapy (OLT), Computer-based audio-visual feedback using interactive visual displays for speech training. Dept. of Computer Science, University of Sheffield, UK.
- Hatzis, A., Green, P. D., Howard, D. (1999). Optical logo-therapy (OLT): Visual displays in practical auditory phonetics teaching, *Phonetics teaching and learning conference (PTLC '99)*, April 1999.
- Hatzis, A. & Green, P. D. (2001). A two dimensional kinematic mapping between speech and acoustics and vocal tract configurations. *Workshop on Innovation in Speech Processing, (WISP'01)*, Stratford-upon-Avon, UK.
- Hincks, R. (2005). *Computer Support for Learners of Spoken English*, Dissertation, KTH, Stockholm.
- Hochberg, I., Levitt, H., and Osberger, M.J. (1983). *Speech of the Hearing-Impaired, Research, Training and Personal Preparation*. Baltimore, MD: University Park Press.
- Hodson, B.W. (1980). *The Assessment of Phonological Processes*. Danville, IL.: Interstate Inc.
- Holmber, E., Sahlén, B. (2000). *Nya Nelli, Pedagogisk Design*, Malmö.
- Hudgins, C. and Numbers, F. (1942). An Investigation of the Intelligibility of the Speech of the Deaf. *Genet. Psychol. Monogr.* 25, pp. 289-392.

- Huggins, A.W.F. (1977). Timing and speech intelligibility, in *Attention and Performance, VII* (ed. J. Requin).
- Ingram, D. (1989). *Phonological Disability in Children*. (2nd ed.), London: Sole and Whurr.
- Jakobson, R. (1968). *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton and Co.
- Jamieson, DG. (1995). Techniques for training difficult non-native speech contrasts. *Proc XIIIth Intl Congress Phonetic Sciences*, Stockholm.
- John, J.E.J. and Howarth, J. (1965). The Effect of Time Distortions on the Intelligibility of Deaf Children's Speech. *Lang. and Speech*, 8, 127-134.
- Javkin, H. (1994). A new Speech Training System: Acoustic/Articulatory Data, Video Games and Synthesized Model Parameters. *2nd Int. Symposium on Speech and Hearing Sciences*, 81-92. Osaka, Japan.
- Kent, R. D., Osberger, M. J., Netsell, R., Goldsmith Hustedde, C. (1987). Phonetic Development in Identical Twins Differing in Auditory Function, *JSHD*, Vol.52, 64-75.
- Kewley-Port, D., Watson, C. S., Elbert, M., Maki, D. and Reeds, D. (1991). The Indiana Speech Training Aid (ISTRA) II: Training curriculum and selected case studies, *Clinical Linguistics & Phonetics*, 1991, Vol. 5, No. 1, 13-38.
- Kewley-Port, D. & Watson, C. S. (1995). Computer Assisted Speech Training: Practical Considerations, *Applied Speech Technology*, Chapter 21, 565-582, Boca Raton: CRC Press.
- Kjellin, O. (1997). *Svenskt uttal i verkligheten*, CD, Hallgren och Fallgren Studieförlag AB, ISBN: 91-7382-750-9.
- Kruger, F., Stromberg, H., Levitt, H. (1972). Synthetic speech as a diagnostic tool. *CSL Research Report No. 2*, June.
- LaRocca, S. (1994). Exploiting strengths and avoiding weakness in the use of speech recognition for language learning, *CALICO Journal*, 12, 1, 102-105.
- Larson Education AB, "Lingus", <http://www.larsoneducation.se>.
- Larsson, T. (1997). Föreläsning om "Hörselskador", *Rehabiliteringsteknik AK ht 1997*, CERTEC, Lund.

- Levitt, H. & Geffner, D. (1987). Communication skills of young hearing-impaired children, Development of language and communication skills in hearing-impaired children, *ASHA Monographs*, 26, 123-158.
- Levitt, H. (1987). Interrelationship among the speech and language measures, in Development of language and communication skills in hearing-impaired children. *ASHA Monographs*, 26, 123-158.
- Levitt, H. (1993). The impact of technology on speech rehabilitation. *Proceedings of an ESCA Workshop on Speech and Language Technology for Disabled Persons*. Stockholm, Sweden.
- Lidén, G. (1985). Audiologi, Almqvist & Wiksell.
- Lindblom, B. and Moon, S-J. (1988). Formant undershoot in clear and citation form speech. *Perilus VIII* (Phonetic Experimental Research, Institute of Linguistics, University of Stockholm), 20-32.
- Ling, D. (1976). Speech and the Hearing-Impaired Child: Theory and Practice. *The Alexander Graham Bell Association for the Deaf, Inc.* Washington, D.C. 20007, U.S.A.
- Lotson, A. (2001). Bättre tal med datorstöd – synsinnet kan ersätta skadad hörsel i talträning, *Computer Sweden*, 34-35.
- Loughlin (2005). <http://users.aber.ac.uk/vil1/approach.htm>.
- Maassen, B. and Powel, D.J (1984). The effect of correcting temporal structure on the intelligibility of deaf speech, *Speech Comm.* 3, 123-135.
- Maassen, B. and Powel, D.J (1985). The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *Journal of the Acoustical Society of America*, 78, 877-886.
- Mahshie, J. J., Herbert, E., Hasegawa, A. (1984). Use of air-flow feedback to modify deaf speaker's consonant voicing errors. *Asha*, 26, 10.
- Mahshie, J. J. (1990). Speech training with deaf children, *Seminar held at the Dept. of Speech, Music, and Hearing, KTH, Stockholm*, 6 April, 1990.



- Mahshie, J. J and Yadab, P. (1990). The Gallaudet University Speech Training and Evaluation System (GUSTES) for deaf children, *Asha*, 32, 10, 75.
- Mahshie, J. J. (1995). The Use of Sensory Aids for Teaching Speech to Children who are Deaf, *Profound Deafness and Speech Comm.*, London: Whurr Publ. Ltd., 461-491.
- Mahshie, J. J. (1996). Feedback considerations for speech training systems, *Proceedings of 4th ICSLP-96*, 153-156, Philadelphia.
- Mahshie, J. J. (1998). Balloons, Penguins, and Visual Displays. SpeechViewer III: Solid Tool for Specialists, *Perspectives in Education and Deafness*, 16, 4.
- Maltby, M. (2000). A new speech perception test for profoundly deaf children, *Deafness and Education International*, 2, 2, 86-101.
- Markides, A. (1985). Type of pure tone audiogram configuration and rated speech intelligibility, *Journal of British Association of the Profoundly Hhearing-impaired*, 2, 33-36.
- Markides, A. (1989). Lipreading: Theory and Practice. *Journal of Brit. Assn. Teachers of the Deaf*, 13:2, 29-47.
- Mártony, J. (1966). Method of correcting the voice pitch level of hard of hearing subjects, *STL-QPSR*, 7(2):19-22.
- Martony, J. (1968). On the correction of the pitch level for severely hard of hearing subjects, *American Annals of the Deaf*, 113, 2, 195-202.
- Mártony, J., Risberg, A., Agelfors, E., and Boberg, G. (1970). Om talavläsning med elektronisk avläsningshjälp, *Intern rapport, Inst. för Talöverföring, KTH*.
- Mártony, J. (1971). Om gravt hörselskadades tal, Fil.lic. avhandling, Inst. för Talöverföring, KTH, Stockholm.
- Martony, J., Risberg, A., Spens, K-E., Agelfors, E. (1972). Results of a rhyme-test for speech audiometry, *Proc. of Int. Symp. on Speech Communication Ability and Profound Deafness, Stockholm*, A.G. Bell Association for the Deaf (ed. G. Fant), 75-80.
- Martony, J. (1974). On a rhyme test, *STL-QPSR* 2-3, 57-71.
- Martony, J., Nordström, P-E. (1995). On vowel production in deaf children, *VIII Int. Congress of Phonetic Sciences, Leeds*, Paper 194.

- Massaro, D. W., Light, J. (2004). Using visible speech to train perception and production of speech for individuals with hearing loss, *Journal of Speech, Language and Hearing Research*, 47,304-320.
- Mavilya, M. (1972). Spontaneous vocalization and babbling in hearing impaired infants, In G. Fant (Ed) *International Symposium on Speech Communication Ability and Profound Profoundly Deafness*, Washington, DC. Alexander Graham Bell Association for the Profoundly hearing-impaired, 163-171.
- McAllister, R. (1986). Tekniska hjälpmedel i Uttalsundervisningen: en delrapport. *PU-rapport 1986:1*, Stockholms universitet, Pedagogiska utvecklingsenheten.
- McAllister, R. (1995). Perceptual foreign accent and L2 production, *Proc XIIIth Intl Congress of Phonetic Sciences*, Stockholm.
- McAllister, R., Flege, J.E., Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian, *Journal of Phonetics*, 30, 229-258.
- McConachie, H. (1990) Early language development and severe visual impairment, *Child: care, health and development*, 16 (1), 55-61.
- McDonald, E. (1964). *Articulation Testing and Treatment*, Pittsburgh, Stanwix House, 1964.
- Merklein, R. A. (1981). A short speech perception test, *The Volta Review*, 36-46.
- Miller, G. A. (1951). *Language and Communication*, McGraw-Hill, New York, p. 39.
- Mills, A. (1983). The development of phonology in the blind child. Dodd and Rampbell, eds., *Hearing by Eye: The Psychology of Lip-Reading*, 145-161.
- Mogford, K. (1983). Lip-reading in the prelingually profoundly hearing-impaired, (Dodd and Rampbell, eds.) *Hearing by Eye: The Psychology of Lip-Reading*, 191-211.
- Monsen, R.B. (1976). The production of English stop consonants in the speech of deaf children. *Journal of Phonetics*, 4, 29-41.
- Monsen, R.B. (1978). Toward measuring how well hearing-impaired children speak. *Journal of Speech and Hearing Research*, 21, 197-219.

- Monsen; R.B. (1983). Voice quality and speech intelligibility among deaf children. *American Annals for the Deaf*, 128, 12-19.
- Moon, S-J. (1991). An acoustic and perceptual study of undershoot in clear and citation-form speech. *Perilus XIV*, 153-156 (Phonetic Experimental Research, Institute of Linguistics, University of Stockholm).
- Neri, A., Cucchiarini, C., Strik, H., Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training, *Computer Assisted Language Learning*, 15, 5, 441-467.
- Nettelblatt, U. (1983). *Developmental studies of dysphonology in children*, Doctoral Dissertation. Lund: Lieber Läromedel.
- Nickerson, R., Stevens, K (1973). Teaching speech to the deaf. Can a computer help? *IEEE Trans. Audio Electro-acoustics*, AU-21, 445-455, 1973.
- Nickerson, R. S., Kalikow, D. N. & Stevens, K. N. (1976). Computer-aided speech training for the deaf, *JSHD*, 41, 120-132.
- Oller, D.K., Kelly, C.A. (1974). Phonological substitution processes of a hard-of-hearing child. *JSHD*, 39, 65-74.
- Oller, D.K., Eilers, R.E. (1981). A pragmatic approach to phonological systems of deaf speakers. *Speech and Language, Advances in basic research and practice*, 103-141, Academic press.
- Oller, D. K, Eilers, R. E, Neal, A. R, Schwartz, H. K, J. (1999). Precursors to speech in infancy: the prediction of speech and language disorders, *Commun. Disord.* 32, 223-245.
- Oller, D. K. (2000). *The emergence of the speech capacity*, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.
- Osberger, M.J. and Levitt, H. (1979). The Effect of Timing Errors on the Intelligibility of Deaf Children's Speech. *JASA*, 66:5, 1316-1324.
- Osberger, M J., Moeller, M. P. & Kroese, J. M. (1981). Computer-assisted speech training for the hearing impaired, *J. of the Academy of Rehabilitative Audiologists*, 14, 145-158.
- Osberger, M., J. (1992). Speech intelligibility in the hearing impaired: Research and clinical implications, in *Intelligibility in Speech Disorders, Theory, Measurement and management*, Edited by

- Raymond D: Kent, John Benjamins Publishing Company, Amsterdam, pp.233-265.
- Öster, A-M. (1985). The Use of a Synthesis-by-Rule-System in a Study of Deaf Speech. *STL-QPSR* 1/1985, 95-107.
- Öster, A-M. (1988a). Datorer I talundervisningen - Ett nytt hjälpmedel?, *Nordisk Tidskrift för Dövundervisningen*, 1, 14-19.
- Öster, A-M. (1988b). Computer-based speech training, *Proceedings of Speech '88, 7th FASE Symposium, Edingburgh, Book 2*, 645-651.
- Öster, A-M. (1989a). Studies on phonological rules in the speech of the deaf. *STL/QPSR* 1/89, 159-162.
- Öster, A-M. (1989b). Applications and experiences of computer-based speech training. *STL/QPSR* 4/89, 37-44.
- Öster, A-M. (1989c). Applications and experiences of computer-based speech training, *Proceedings of European Conference on Speech Communication and Technology (Eurospeech), Paris*, 714-717.
- Öster, A-M. (1990). The effects of prosodic and segmental deviations on intelligibility of deaf speech, *STL/QPSR*, 79-88.
- Öster, A-M. (1991). Phonological assessment of eleven prelingually deaf children's consonant production. *STL-QPSR* 2-3/91, 11-18.
- Öster, A-M. (1992a). *The speech of deaf children - Phonological assessment as a basis for speech training*, Thesis work for the Licentiate Philosophy degree in Phonetics, University of Stockholm, Institute of Linguistics.
- Öster, A-M. (1992b). Phonological assessment of deaf children's existing articulation skills as a basis for speech training. *Proceedings of ICSLP 92, October 12-16, Banff, Alberta, Canada*, 955-958.
- Öster, A-M. (1995a). Principles for a complete description of the phonological system of deaf children as a basis for speech training, *Profound Deafness and Speech Comm.*, London: Whurr Publ. Ltd., 441-461.
- Öster, A-M. (1995b). Teaching speech skills to deaf children by computer-based speech training, *Proceedings of 18th International Congress on Education of the Deaf, Tel-Aviv, Israel*.
- Öster, A-M. (1995c). Teaching speech skills to deaf children by computer-based speech training, *STL-QPSR* 4/95, 67-75.

- Öster, A-M. (1995d). Resultat från datorbaserad röstträning med ett gravt hörselskadat förskolebarn, *Nordisk Tidskrift för Dövundervisningen*, 125-133.
- Öster, A-M. (1996). Clinical applications of computer-based speech training for children with hearing impairment. *Proceedings of ICSPL 96*, 157-160. Philadelphia, USA.
- Öster, A-M. (1997). Auditory and visual feedback in spoken L2 teaching, Reports from the Dept of Phonetics, Umeå University, *PHONUM 4*.
- Öster, A-M. (1998). Spoken L2 teaching with contrastive visual and auditory feedback, *Proc ICSLP, Sydney*.
- Öster, A-M. (1999a). Strategies and results from spoken L2 teaching with audio-visual feedback, *STL-QPSR 1-2/99*, 1-7.
- Öster, A-M., Vicsi, K., Roach, P., Kacic, Z., & Barczikay, P. (1999b). A multimedia multilingual teaching and training system for speech and hearing-impaired children - SPECO, *Proceedings of Fonetik 99*, 149-152. Gothenburg, Sweden.
- Öster, A-M. (2002a). The relationship between residual hearing and speech intelligibility - Is there a measure that could predict a prelingually profoundly deaf child's possibility to develop intelligible speech? *STL/QPSR Vol. 43*, 51-56.
- Öster, A-M., House, D., Protopapas, A., Hatzis, A. (2002b). Presentation of a new EU project for speech therapy: OLP (Ortho-Logo-Pedia), *Proceedings of Fonetik 2002, QPSR, Vol. 44*, 45-48.
- Öster, A-M., House, D., Hatzis, A., Green, P. (2003). Testing a New Method for Training Fricatives using Visual Maps in the Ortho-Logo-Paedia Project (OLP), Umeå University, Department of Philosophy and Linguistics, *PHONUM 9 (2003)*, 197-X, Available online at <http://www.ling.umu.se/fonetik2003/>.
- Owens, E. and Blazek, B. (1985). Visemes observed by hearing-impaired and normal hearing adult viewers, *Journal of Speech and Hearing Research*, 28, 381-393.
- Pickett, J. M. and Constam, A. (1968). A visual speech trainer with simplified indication of vowel spectrum, *American Annals of the Deaf*, 113, 253-258.

- Pickett, J. M. (1980). Tactual Communication of Speech Sounds to the Deaf: Comparison with Lipreading, IEE Press, *Sensory Aids for the Hearing-Impaired*, edited by Levitt, H., Pickett, J. M., Houde, R. A., 262-277.
- Piske, T., MacKay, I., Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review, *Journal of Phonetics*, 29, 191-215.
- Plant, G. (1960). The Plant-Mandy voice trainer - Some notes by the designer, *Teacher of the Deaf*, 58, 12-15.
- Plant, G., and Hammarberg, B. (1983). Acoustic and perceptual analysis of the speech of the deafend", *STL-QPSR* 2-3/1983 pp. 85-107.
- Povel, D. J., Arends, N. (1991). The Visual Speech Apparatus: Theoretical and Practical Aspects, *Speech Communication*, 10, 59-80.
- Preisler, G. (1991). Early patterns of interaction between blind infants and their sighted mothers, *Child: care, health and development*, 17, 65-90.
- Proctor, A. (1995). Tactile Aid Usage in Young Deaf Children, *Profound Deafness and Speech Comm.*, London: Whurr Publ. Ltd., 111-147, Plant and Spens (Ed.).
- Pronovost, W., Yenkin, L., Anderson, D.C., Learner, R. (1968). The Voice Visualizer, *American Annals of the Deaf*, 113, 230-238.
- Protopapas, A. (2004). User's Manual, OLP document QL1971-ILS-IN-C-097-a3.
- Risberg, A. (1968). Visual aids for speech correction, *American Annals of the Deaf*, 113, 2, 178-194.
- Risberg, A. & Mårtony, J. (1970). A method for the classification of audiograms. In G. Fant (Ed), *Speech Communication Ability and Profound Hearing-Impairedness*. Washington, D.C.: A. G. Bell Association for the Profoundly Hearing-impaired, 135-139.
- Risberg, A. (1976). Diagnostic rhyme test for speech audiometry with severely hard of hearing and profoundly deaf children, *STL-QPSR* 2-3, 40-58.

- Risberg, A., Agelfors, E., Florén, Å. (1977). Avläsetest med spondéord, *Preliminär rapport, 770323, Inst. för talöverföring, KTH, Stockholm.*
- Risberg, A. (1979). Bestämning av hörkapacitet och talperceptionsförmåga vid svåra hörsel skador, *Rapport TRITA-TLF-79-2, Doktorsavhandling, Inst. för Talöverföring, KTH, Stockholm.*
- Risberg, A. (1982). Speech coding in aids for the deaf: An overview of research from 1924-1982, *STL/QPSR 4, 65-98.*
- Rooney, E., Carraro, F., Dempsey, W., Robertson, K., Vaughan, R., & Jack, M. (1994). HARP: an autonomous speech rehabilitation system for hearing-impaired people. *In Proc. 1994 International Conference on Spoken Language Processing (ICSLP94), 2019-2022. Yokohama, Japan.*
- Roug, L., Landberg, I., Lundberg, L-J. (1989). Phonetic development in early infancy: A study of four Swedish children during the first eighteen months of life, *Journal of Child Language, 16: 19-40.*
- Saben, C.B. and Ingham, J.C. (1991). The effects of minimal pairs treatment on the speech-sound production of two children with phonologic disorders. *Journal of Speech and Hearing Research, Vol.34, 1023-1040.*
- Schulte, K. (1972). Fonator System: Speech stimulator and speech feedback by technically amplified one-channel vibration, 351-353, in G. Fant (Ed.), *Int. Symposium on Speech Communication Ability and Profound Deafness, A.G. Bell Association for the deaf, Washington.*
- Shriberg, L.D. and Kwiatkowski, J. (1980). *Natural Process Analysis (NPA)*. New York: John Wiley.
- Sjölander, K., Beskow, J. (2000). WaveSurfer - an Open Source Speech Tool, in *Proceedings of ICSLP 2000, Beijing, China.*
- Smith, C. (1975). Residual hearing and speech production in the deaf. *Journal of Speech and Hearing Research, 19, 795-811.*
- Soleymani, A. J. A., Southwood, M. H., McCutcheon, M. J. (1997). Design of Speech Illumina Mentor (SIM) for teaching speech to the hearing-impaired, *Biomedical Engineering Conference, Proceedings of the 1997 Sixteenth Southern, Biloxi, MS, USA.*

- Spens, K-E. (1984). To hear with the skin. Dissertation, TRITA\_TÖM 2-84, ISSN 0280-9850, KTH.
- Stach, B. A. (1998). *Clinical Audiology: An Introduction*, Clifton Park, NY: Singular.
- Stampe, D. (1979). *A Dissertation on Natural Phonology*. In Hankamer, I. (ed), Garland, New York.
- Stoel-Gammon, C. and Dunn, C. (1985). *Normal and Disordered Phonology in Children*, University Park Press, Baltimore, Md.
- Stoel-Gammon, C., and Otomo, K. (1986). Babbling development of hearing-impaired and normally hearing subjects, *JSHD*, Vol.51, 033-041.
- Summerfield, Q. (1979). Use of Visual Information for Phonetic Perception, *Phonetica* 36, 314-331.
- Tanner Dyson, A. (1988). Phonetic inventories of 2- and 3-year-old children, *JSHD*, Vol.53, 89-93.
- Thomas, I. B., Snell, R. C. (1970). Articulation training through visual speech Patterns, *Volta Review*, 310-318.
- Thorén, B. (1994). *Betoningshandboken, Liten hjälpreda för oss som undervisar i svenska som andraspråk*, BT Bättre Svenska, 2:a upplagan, Sundsvall.
- Thore'n, A (2002). Blind children and sighted parents in development and communication, *Dissertation, 2002, ISBN 91-7265-540-2, Stockholm University*.
- Traunmyller, H. (1980). The Sentiphone, a tactual speech communication aid, *Journal Comm. Dis.*, 13, 183-193.
- Upton, H.W. (1968). Wearable Eyeglass speechreading Aid, *American Annals for the Deaf*, 113, 22-229.
- Vicsi, K., Roach, P., Öster A., Kacic, Z., Barczikay, P., Sinka, I. (1999). SPECO - A Multimedia Multilingual Teaching and Training System for Speech Handicapped Children, *6th European Conference on Speech Communication and Technology, Eurospeech '99, Budapest*, 859-862.
- Vicsi, K., Roach, P., Öster A., Kacic, Z., Barczikay, P., Tantos A., Csatári, F., Bakcsi, Zs., Sfakianaki, A. (2000). A Multimedia Multilingual Teaching and Training System For Speech



- Handicapped Children, International, *Journal of Speech Technology*, Vol. 3, 289-300.
- Wallace, V., Menn, L., Yoshinaga-Itano, C. (2000). Is babble the gateway to speech for all children? A longitudinal study of children who are deaf or hard of hearing, *The Volta Review*, Vol, 100 (5) pp. 121-148.
- Watanabe, A., Ueda, Y., Shigenaga, A. (1985). Color display system for connected speech to be used for the hearing, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol ASSP-33, 1, 164-173.
- Watson, C. S., Reed, D., Kewley-Port, D., and Maki, D. (1989). The Indiana Speech Training Aid (ISTRA) I: Comparisons between human and computer-based evaluation of speech quality, *Journal of Speech and Hearing Research*, 32, 245-251.
- Weiner, F.F. (1979). *Phonological Process Analysis (PPA)*, University Park Press, Baltimore, Md.
- West, J.J. and Weber, J.L. (1973). A phonological analysis of the spontaneous language of a four-year-old, hard-of-hearing child. *JSHD*, 38, 25-35.
- Whitehead, R.L. (1983). Some Respiratory and Aerodynamic Patterns in the Speech of the Hearing Impaired. In Hochberg, Levitt, Osberger, eds., *Speech of the Hearing Impaired; Research, Training and Personal Preparation*, Maryland, MD: University Park Press, 97-116.
- Wills, D.M. (1981). Some notes on the application of diagnostic profile to young blind children, *Psychoanalytic Study of the Child*, 36, 217-240.
- Woodward, M.F. and Barber, C.G. (1960). Phoneme perception in lipreading, *Journal of Speech and Hearing Research*, 3:3, 212-222.
- Wundt (1911): ref. by Mills, A. (1983), see above.
- Yamada, Y. & Murata, N. (1991). Computer Integrated Speech Training Aid, *Intern. Symposium on Speech and Hearing Sciences*, Osaka, Japan.
- Yoshinaga-Itano, C., Sedey, A. (2000). Early speech development in children who are deaf or hard of hearing: Interrelationships with

*Computer-Based Speech Therapy Using Visual Feedback*

- language and speech, *Volta Review*, Vol.100(5), monograph, 181-211.
- Youdelman, K., Levitt, H. (1991). Speech training of deaf students using a palatographic display, *Proc. of First Intern. Symp. on Speech and Hearing Sciences*, 1-11, Osaka, Japan.
- Youdelman, K. (1994). Computer applications in teaching speech to deaf children, *Proc. of Second Intern. Symp. on Speech and Hearing Sciences*, 67-79, Japan.
- Ziegler, W., Vogel, M., Teiwes, J., and Ahrndt, T. (1997). Microcomputer-Based Experimentation, Assessment and Treatment, In Ball, M.J., and Code, C. (Eds), *Instrumental Clinical Phonetics*, Whurr Publishers.
- Zimmerman, G., and Rettaliata, P. (1981). Articulatory patterns of an adventitiously deaf speaker. Implications for the role of auditory information in speech production, *JHSR* 24, pp.169-178.

## 12. Appendices

### 12.1. Appendix 1: Diacritics to assess the speech of profoundly hearing-impaired children

(see section 7.3 for references)

Modification of manner of articulation:

| Symbol | Example |   |
|--------|---------|---|
| ~      | ã:      | Nasalized   |
| ≈      | ã̃:     | Heavily nasalized   |
| ω      | ᶜ       | Labialization   |
| ~      | ɫ       | Velarization  |
| ↑      | tIt↑a   | Inhalation  |
| ↓      | tIt↓a   | Expiration  |
| ʄ      | kɔmaʄ   | Very weak production  |
| ()     | (f)     | Silent articulation; absence of air-stream but articulation present |
| ⌈      | t⌈      | Stop with non-audible release                                       |
| F      | Fp      | Stop with unspecified frication                                     |
| ≈      | ᶜp      | Nasal air emission  |
| ○      | ⊙       | Inadequately realized; doubtful symbol                              |
| (( ))  | ((a:))  | Masked by noise   |
| ←      | ←b      | Lip-protrusion  |
| K      | fK      | Fricative with explosive release                                    |

Modification of place of articulation:

| Symbol | Example |               |
|--------|---------|---------------|
| ↑      | ē:      | Tight variant |
| ↓      | ẹ:     | Loose variant |

Computer-Based Speech Therapy Using Visual Feedback

|    |    |                                    |
|----|----|------------------------------------|
| ←  | k: | Advanced variant                   |
| →  | k: | Retracted variant                  |
| ⇄  | e: | Instability between base and tight |
| ⇄  | i: | Instability between base and loose |
| □  | s  | Interdental articulation           |
| ○  | i: | Markedly spread lips               |
| L1 | s  | Apex part of tongue                |
| L2 | s  | Blade part of tongue               |
| L3 | s  | Mid part of tongue                 |
| L4 | s  | Back part of tongue                |
| L5 | s  | Root part of tongue                |

Modification of vocal fold activity:


| Symbol | Example        |                  |
|--------|----------------|------------------|
| o      | b              | Voiceless        |
| ∨      | s              | Voiced           |
| ∨      | s s∨           | Pre/post-voiced  |
| h      | p <sup>h</sup> | Aspiration       |
| h      | h p            | Preaspiration    |
| =      | p <sup>=</sup> | Unaspiration     |
| ^^^    | a:             | Laryngealization |

Modification of duration:

| Symbol | Example          |                |
|--------|------------------|----------------|
| ∪      | m <sub>∪</sub> b | Coarticulation |

|     |           |  |
|-----|-----------|--|
|     |           | Absence of coarticulation: short silence between |
| —   | mIn—nɛ:sa | segments or words                                |
| =   | mIn=nɛ:sa | long silence between segments or words           |
| ≡   | mIn≡nɛ:sa | extra long silence between segments or words     |
| :   | ɑ:        | Long production                                  |
| ::  | ɑ::       | Very long production                             |
| ∨   | t         | Very clipped                                     |
|     | bɔ   j    | Long break                                       |
|     | bɔ    j   | Very long break                                  |
| :   | ɔ:        | Repetition of a segment                          |
| : : | : ɔm :    | Repetition of a word                             |

**Modification of pitch:**

| Symbol  | Example |                       |
|---|---------|-----------------------|
| —   | ā       | High tone             |
| —   | ā       | Low tone              |
| /   | /ā      | Rising tone           |
| \   | ā\      | Falling tone          |
| /\  | /ā\     | Rising - falling tone |
| \ /   | \ā/     | Falling - rising tone |
| — —   | ābābā   | Sequence tones        |
|  | ababa   | Continuous tones      |

## 12.2. Appendix 2: Swedish SAMPA symbols

Definitions from:

[SAMPA home page, UCL Phonetics and Linguistics, University College London](http://www.phon.ucl.ac.uk/home/sampa/swedish.htm) <http://www.phon.ucl.ac.uk/home/sampa/swedish.htm>

### Consonants

There are six plosives:

| Symbol | Word | Transcription |
|--------|------|---------------|
| p      | pil  | pi:l          |
| b      | bil  | bi:l          |
| t      | tal  | tA:l          |
| d      | dal  | dA:l          |
| k      | kal  | kA:l          |
| g      | gås  | go:s          |

There are six fricatives:

|   |       |      |
|---|-------|------|
| f | fil   | fi:l |
| v | vår   | vo:r |
| s | sil   | si:l |
| S | sjuk  | S}:k |
| h | hal   | hA:l |
| C | tjock | COk  |

There are six sonorant consonants (nasals, liquids and semivowels):

|   |      |      |
|---|------|------|
| m | mil  | mi:l |
| n | nål  | no:l |
| N | ring | rIN  |
| r | ris  | ri:s |
| l | lös  | l2:s |
| j | jag  | jA:g |

### Vowels

There are nine long and nine short vowels.

Long vowels (followed by short consonant):

|    |     |      |
|----|-----|------|
| i: | vit | vi:t |
| e: | vet | ve:t |
| E: | säl | sE:l |
| y: | syl | sy:l |
| }: | hus | h}:s |
| 2: | föl | f2:l |
| u: | sol | su:l |
| o: | hål | ho:l |
| A: | hal | hA:l |

Short vowels (followed by long consonant):

|    |      |      |
|----|------|------|
| I  | vitt | vIt  |
| e  | vett | vet  |
| E  | rätt | rEt  |
| Y  | bytt | bYt  |
| u0 | buss | bu0s |
| 2  | föll | f2l  |
| U  | bott | bUt  |
| O  | håll | hOl  |
| a  | hall | hal  |

There are also two pre-r allophones (long and short) of /E/ and /2/  
The following important allophonic variants occur in Swedish which require separate symbolic representation:

|    |        |        |                                   |    |
|----|--------|--------|-----------------------------------|----|
| {: | här    | h{:r   | pre-r allophone of E:             |    |
| 9: | för    | f9:r   | "                                 | 2: |
| {  | herr   | h{r    | "                                 | E  |
| 9  | förr   | f9r    | "                                 | 2  |
| @  | pojken | pOjk@n | schwa vowel allophone             |    |
| rt | hjort  | jUrt   | retroflex consonant, not initial* |    |
| rd | bord   | bu:rd  | "                                 |    |
| rn | barn   | bA:rn  | "                                 |    |
| rs | fors   | fOrs   | "                                 |    |
| rl | karl   | kA:rl  | "                                 |    |

\* in cases where the dental consonants do not change into retroflexes, they are transcribed using the separator sign (ASCII 45): r-t, r-d.

Swedish has two contrasting tonemes, but only in stressed syllables. Tone 1 is indicated by the ordinary stress mark, Tone 2 by a doubled stress mark, e.g.

|                     |       |                     |
|---------------------|-------|---------------------|
| stress and toneme 1 | anden | "and@n (the duck)   |
| stress and toneme 2 | anden | "and@n (the spirit) |

Note on the use of [S] for orthographic *sj* etc.: although [S] is an unambiguous way of transcribing this unusual sound of Swedish, some commentators find this symbol phonetically imprecise. Those who feel this way are free to use more elaborate symbols instead: [s`] or even [x\]

[SAMPA](#) home page, [UCL Phonetics and Linguistics](#) home page, [University College London](#) home page.

### 12.3. Appendix 3: Swedish questionnaire for evaluation of Box of Tricks

Evaluation is based on the answers of eight speech therapists that worked with the system during three months.

The children who were trained with the program were between 4 and 12 years of age and consisted of three groups: hearing-impaired pre-school children, severely hearing-impaired children from a Deaf school and normal-hearing children with a central language impairment.

**1. Do you think that any previous knowledge about computers is necessary to work with the system?**

| A lot | Some | None |
|-------|------|------|
| 0     | 6    | 2    |

No comments:

**2. Do you think that any previous knowledge about articulatory and acoustic phonetics is necessary to work with the system?**

| A lot | Some | None |
|-------|------|------|
| 6     | 2    | 0    |

Comments:

- I would like to recommend a training-course in phonetics for teachers who are going to work with the program.
- I want to know more about the relation between articulatory and acoustic phonetics.

**3. How did the system meet with your expectations?**

| Very well | Well | Not so well | Badly |
|-----------|------|-------------|-------|
| 5         | 3    | 0           | 0     |

Comments:

- The more I work with the system the more I appreciate it.
- I especially valued the fixed vocabularies.
- Interesting.

**4. Was the system easy to work with? If difficult, in what respect?**



| Very easy | Easy | Not so easy | Difficult |
|-----------|------|-------------|-----------|
| 3         | 3    | 2           | 0         |

Comments:

- In the beginning the program was not easy to grasp.
- The program is very big and sometimes it was difficult to find the way to correct exercise.

**5. Did you consider the training as meaningful?**

| Very meaningful | Meaningful | Not so meaningful | Meaningless |
|-----------------|------------|-------------------|-------------|
| 3               | 4          | 1                 | 0           |

Comments:

- If the child is motivated the program seems to be very effective.

**6. What is your opinion of the graphics?**

| Very attractive | Attractive | Not so attractive | Unattractive |
|-----------------|------------|-------------------|--------------|
| 5               | 3          | 0                 | 0            |

Comments:

- The colors and the pictures are beautiful.
- All illustrations are uniform and very clear.
- The children especially loved the feedback of the duck and the flower.

**7. How was the system from a pedagogical point of view?**

| Very pedagogical | Pedagogical | Not so pedagogical | Unpedagogical |
|------------------|-------------|--------------------|---------------|
| 5                | 3           | 0                  | 0             |

Comments:

- The program seems to be very well thought-out and carefully planned.
- To use symbolic cards illustrating each speech-sound was a good idea!
- Some parts of the program are depending on the maturity and ability of the child.

**8. Was the system reliable?**

| Very reliable | Reliable | Not so reliable | Unreliable |
|---------------|----------|-----------------|------------|
| 8             | 0        | 0               | 0          |

Comments:

- The program never failed. It was technically extremely stable.
- Both the visual and auditory feedback worked without any problem.

## Computer-Based Speech Therapy Using Visual Feedback

### 9. How was the interaction with the child?

| Very well | Well | Not so well | Badly |
|-----------|------|-------------|-------|
| 5         | 3    | 0           | 0     |

Comments:

- Big variations but most children were very interested and co-operated.
- The children had great fun imitating different animals in the exercise "Spectrum".
- Every time they came to school they wanted to work with the program.

### 10. Did the child learn something?

| A lot | Some | Nothing |
|-------|------|---------|
| 2     | 5    | 1       |

Comments:

- Especially successful was a girl who had difficulties with the Swedish /TJ/-sound.
- One child got control of her pitch and got a better sustained phonation.
- Too early to tell!

### 11. Did the child understand the feedback from the visual speech pictures (cochleagrams)?

| Very well | Well | Not so well | Badly |
|-----------|------|-------------|-------|
| 5         | 3    | 0           | 0     |

Comments:

- Even if the children didn't understand the acoustic pattern they got an understandable feedback by comparing their pattern with that of the reference speaker.
- In the beginning some children had problems to focus on the most important things.

### 12. Did the child understand the feedback from the visual speech pictures (distribution lines)?

| Very well | Well | Not so well | Badly |
|-----------|------|-------------|-------|
| 8         | 0    | 0           | 0     |

No comments:

### 13. Did you want more instructions in the user manual when working with the system?

| A lot | Some | None |
|-------|------|------|
| 0     | 1    | 7    |

Comments:

- The manual was plain and readable; especially the screen dumps were instructive to study.

**14. Did you miss the possibility to train something? If yes, which one?**

| A lot | Some | None |
|-------|------|------|
| 0     | 0    | 8    |

No comments:

**15. Was the child motivated to train with the system?**

| A lot | Some | None |
|-------|------|------|
| 6     | 2    | 0    |

Comments:

- All preparing exercises for intensity, rhythm, spectrum and voicing attracted the children.
- They wanted to see all the illustrations before starting the training. They were very curious.
- The auditory feedback of their own pronouncing stimulated them a lot.

**16. Additional ideas:**

- It supplies a long-felt want. There are so many computerbased language-training programs but so few speech-training programs on the market!

**17. Try to estimate how much time you have spent for each session:**

Severely hearing-impaired children: 5-10 minutes

Hearing-impaired children: 10-15 minutes

Central language impairment: 5-20 minutes

Thank you for your cooperation!

---