TUCS

Stina Ojala

Towards an Integrative Information Society

Studies on Individuality in Speech and Sign

Turku Centre for Computer Science

TUCS Dissertations
No 135, May 2011

Towards an Integrative Information Society

Studies on Individuality in Speech and Sign

Stina Ojala

To be presented, with the permission of the Faculty of Mathematics and Natural Sciences of the University of Turku, for public criticism in the Auditorium Alpha in the ICT building May 21, 2011 at 12:00.

University of Turku
Department of Information Technology
Joukahaisenkatu 3-5 C, 20014 University of Turku, Finland

2011

Supervised by

Professor Tapio Salakoski
Department of Information Technology
University of Turku
Turku, Finland

Professor Olli Aaltonen
Department of Speech Sciences
University of Helsinki
Helsinki, Finland

Reviewed by

Professor Onno Crasborn
Department of Linguistics
Radboud University Nijmegen
Nijmegen, the Netherlands

Professor Michael O'Dell
Department of Finnish Language and General Linguistics
University of Tampere
Tampere, Finland

Opponent

Professor Markku Turunen
School of Information Sciences
University of Tampere
Tampere, Finland

ISBN 978-952-12-2583-3
ISSN 1239-1883

Abstract

The flow of information within modern information society has increased rapidly over the last decade. The major part of this information flow relies on the individual's abilities to handle text or speech input. For the majority of us it presents no problems, but there are some individuals who would benefit from other means of conveying information, e.g. signed information flow.

During the last decades the new results from various disciplines have all suggested towards the common background and processing for sign and speech and this was one of the key issues that I wanted to investigate further in this thesis. The basis of this thesis is firmly within speech research and that is why I wanted to design analogous test batteries for widely used speech perception tests for signers – to find out whether the results for signers would be the same as in speakers' perception tests. One of the key findings within biology – and more precisely its effects on speech and communication research – is the mirror neuron system. That finding has enabled us to form new theories about evolution of communication, and it all seems to converge on the hypothesis that all communication has a common core within humans.

In this thesis speech and sign are discussed as equal and analogical counterparts of communication and all research methods used in speech are modified for sign. Both speech and sign are thus investigated using similar test batteries. Furthermore, both production and perception of speech and sign are studied separately. An additional framework for studying production is given by gesture research using cry sounds. Results of cry sound research are then compared to results from children acquiring sign language. These results show that individuality manifests itself from very early on in human development.

Articulation in adults, both in speech and sign, is studied from two perspectives: normal production and re-learning production when the apparatus has been changed. Normal production is studied both in speech and sign and the effects of changed articulation are studied with regards to speech. Both these studies are done by using carrier sentences. Furthermore, sign production is studied giving the informants possibility for spontaneous speech. The production data from the signing informants is also used as the basis for input in the sign synthesis stimuli used in sign perception test battery.

Speech and sign perception were studied using the informants' answers to questions using forced choice in identification and discrimination tasks. These answers were then compared across language modalities. Three different

informant groups participated in the sign perception tests: native signers, sign language interpreters and Finnish adults with no knowledge of any signed language. This gave a chance to investigate which of the characteristics found in the results were due to the language per se and which were due to the changes in modality itself.

As the analogous test batteries yielded similar results over different informant groups, some common threads of results could be observed. Starting from very early on in acquiring speech and sign the results were highly individual. However, the results were the same within one individual when the same test was repeated. This individuality of results represented along same patterns across different language modalities and - in some occasions - across language groups. As both modalities yield similar answers to analogous study questions, this has lead us to providing methods for basic input for sign language applications, i.e. signing avatars. This has also given us answers to questions on precision of the animation and intelligibility for the users – what are the parameters that govern intelligibility of synthesised speech or sign and how precise must the animation or synthetic speech be in order for it to be intelligible. The results also give additional support to the well-known fact that intelligibility in fact is not the same as naturalness. In some cases, as shown within the sign perception test battery design, naturalness decreases intelligibility. This also has to be taken into consideration when designing applications.

All in all, results from each of the test batteries, be they for signers or speakers, yield strikingly similar patterns, which would indicate yet further support for the common core for all human communication. Thus, we can modify and deepen the phonetic framework models for human communication based on the knowledge obtained from the results of the test batteries within this thesis.

Acknowledgements

First of all I want to thank my present supervisors Olli Aaltonen and Tapio Salakoski. Olli Aaltonen for allowing me to follow my instincts on this unmapped territory between different disciplines during my first staggering steps and Tapio Salakoski for a warm welcome to the new faculty and providing me with the infrastructure and colleagueship I needed to finally finish this project of mine. One of the past supervisors I have to give a special acknowledgement for supporting me throughout everything I have pursued for - thank you, Risto-Pekka Happonen. He was also amongst the first to show me the true meaning of what it means to be interdisciplinary in research - and what it means to engage in conversations between different research areas.

I could not have accomplished this without the support from many different work groups - representing different areas of research: from medical faculty the PIPARI group and the colleagues within Oral and Maxillofacial Surgery Department, from humanities colleagues from General Linguistics, Esa Itkonen and Terhi Rissanen for their insights on sign linguistics, the close Phonetics Corridor Group for the long and manyfold discussions from the general theory of everything all the way to what would be the best way to record this or that type of data. So the Phonetics Corridor Group consists of the core group of the late Phonetic Corridor: Juha-Pertti, Janne - and in the later stages also Lotta and Outi. One group that I got to know in the very last moments towards the end of my PhD project, but whose importance cannot be undermined is the Acoustic HUT group: Juha-Pertti and Daniel. Still one working group gathers round the coffee table at ICT: Filip, thanks for the important insights on the technological side of things. From University of Helsinki I would like to acknowledge the Phonetics Department group: Mietta and Minnaleena in particular. From Department of Special Education – Riitta, thanks for the support and the opportunity to “practice” the defence from a whole different perspective. Partial thanks for Russ too for providing me that rare opportunity.

I couldn’t have finished this project of mine without the support of two National Graduate Schools and one Nordic graduate school: Finnish Graduate School of Language Technology (KIT), Finnish Graduate School in Language Studies (LangNet) and the Nordic Graduate School of Language Technology (NGLST). These have provided me with financial support throughout my PhD thesis. I have also received various grants from different foundations in Finland and Norway, which I also would like to acknowledge: Alfred Kordelin Foundation, Turku University Foundation and Finnish Concordia Fund in Finland and NordForsk in Norway.

I would like to thank my reviewers professors Michael O'Dell and Onno Crasborn for their careful and detailed comments on my thesis. Many anonymous reviewers have given their comments to my papers and articles in various occasions. Their efforts are also acknowledged here as well as all the participants of my test series and different tasks within them. This type of theses would not exist without the effort and patience of all those volunteers who provided their time and valuable questions. During this project they gave me insights to the manifold questions related to my research.

Last but by no means the least I would like to express my gratitude to my family who have given me love and support throughout all these years. A great deal of understanding it has taken all of them to realise how time-consuming research is – at least for someone like me, who tends not only work when awake but also dream on various themes especially related to theories and philosophy! I sincerely hope that I will be able to spend more time with all my loved ones at home from now on.

To Kerkko, our new explorer of this Universe

Table of Contents

Foreword

The omnipresence of human communication in all its forms requires multidisciplinary research. Usually the study of language and communication results in balancing between several disciplines. The teamwork bringing together expertise from different areas of research requires certain skills in communicating and co-operating. Communicating skills are required when two experts from different areas of research might refer to the same entity with two very different terms. This goes vice versa too – the same word can be used to refer to two different, yet sometimes somehow related entities. The possibility to co-operate with researchers from several different disciplines and areas of research is very rewarding and fruitful for the proceeding of the research. And, as communicating itself, it requires discovering the golden mean - over and over again.

This research has begun an expedition into the outskirts of phonetics by not on studying speech only but first of all attempts to adapt the framework created for the acoustic medium to fit the visual medium. So, in other words I wanted to know whether it could really be done, studying signs within a phonetic framework. This, I hope, will ultimately deepen our understanding of not only sign language per se but will introduce and test some of the phenomena of speech itself too. Outskirts notwithstanding, this thesis is still deeply rooted in phonetics. And by studying the phonetic phenomena and hierarchical structures within signed as well as spoken languages this thesis probes also the outskirts of linguistics. This allows us to seek further knowledge of the rules governing signed languages. With some additional knowledge about signed languages we can add depth to our understanding of spoken languages too. The knowledge about sign languages adds a feedback loop and a testing ground into the study of spoken languages. This new possibility of feedback looping with data from completely different kind of human languages is ultimately needed in testing theories on universals amongst the languages of the world and theories about languages in general.

Another important connection outwards from phonetic research goes towards clinical research framework. Within this thesis one of the major questions is how does individuality affect the production and perception of speech especially in connection to different medical interventions and disabilities, may they be inherited or acquired. One part of the current data within this thesis comes from signed languages and it of course also is linked to the clinical research, deafness being often regarded as a disability. This thesis however, studies sign languages equal to any spoken language that is not as a form of augmentative communication to people who cannot hear. Another link to the clinical framework comes from the studies investigating the effects of different kinds of medical interventions on articulation – in this case the interventions are surgical procedures affecting different parts of the oral cavity.

These results within different studies in this thesis are also discussed with connection to the evolution of language in humans and thus this is a study in part of the ability of humans to communicate. Since results are similar within different language modalities it can be said that we are exploring the outskirts of natural sciences and evolutionary biology from the linguistic/ phonetic framework and the question of universalities within languages are also probed within different study questions.

The different theoretical studies and baselines have yielded connections into information society development and the input and possible starting point for guidelines for developers aiming towards different ways of conveying information via computers and interfaces in various accessible ways for different kinds of people. This thesis corroborates with the strategy of the University of Turku being a truly multidisciplinary study and thus it emphasises the importance of unifying forces from different areas of research.

For reasons stated above this thesis could have been defended in several different faculties and departments within academia, a serendipity is that this thesis is now defended in the Faculty of The Natural and Mathematical Sciences in University of Turku.

1 Introduction

What would the world be like if you couldn't hear? What if your native language would be very different from the majority of people in your native country? It's quite like being a foreigner every day in your native country. And in this case the majority language would not only be different but it would at the same time be for the most part "non-existent", that is invisible. That resembles the situation that signers, that is sign language users, face on an every day basis. The invisibility of spoken languages is a result of the primary carrier channel of spoken language – through acoustics, that is: spoken languages are perceived with the ears. There are nevertheless visible elements on spoken language too, in which hearing people also take advantage of, especially in noisy situations. They are the movements of mouth and face when speaking. However, the visual part of a spoken language results in a much distorted signal, since the majority of information load is carried through the acoustic channel.

The first problem of not being able to hear properly affects two groups: hard-of-hearing and deaf people. The second problem of your native language being very different from the majority of people in your native country affects the latter of those groups: the Deaf. The majority of Deafs do have efficient command of written and read languages, but might face problems with fast-changing texts, for example in subtitling of TV programmes. The subtitling is done with the written form of a spoken language and not with a written form of their native language, sign language, because there is no written form of sign languages yet, though there are attempts towards it (on SignWriting see e.g. Sutton, 1995; on Stokoe Notation see e.g. Stokoe, 1960; on comparison between these two see e.g. Martin, 2000). Subtitling of TV programmes is a great asset for the hard-of-hearing people, because Finnish is their native language and subtitling is a transformation of spoken language into a visible form. The situation for the Deaf in every day life resembles the situation which foreigners have to face when travelling abroad – they are faced with texts written in a foreign language but not their native one.

The basic methodology within this thesis comes from phonetics, or that is speech sciences. This methodology within phonetics is adapted for signed languages to investigate whether those two ways of communication share common characteristics. This thesis is not only studying speech in its acoustic manifestation but also its optical variant – sign language. Here, within this work, signing is regarded as another way of speaking and thus is treated equal in all respects, starting from the framework studied and reaching into the methodological issues within different study questions. In other words: are the phonetic phenomena found in speech also present in signed language and is phonetic methodology flexible enough to be extended into non-acoustic languages?

Individuality is the concurrent theme within this thesis. Within this thesis individuality is used to refer to individual differences and features in speech and sign within a conversation or in a behavioural test setting. All the individual features and differences ultimately are covered by the word borrowed from more philosophical terminology. Here within this thesis it is, nevertheless, regarded as referring to concrete, measurable phenomena within one individual. It engulfs all the individual studies within this thesis and also is an important feature within the each of the research questions as well. The definition of individuality varies slightly between different schools of research but it all boils down to the question: "How are we able to cope with all the differences presented to us?" That question does not only emphasise the differences between individuals but also engulfs the opposite point of view of being able to map and categorise similar enough stimuli together. That requires lots of adaptability which we humans are capable of doing much more so than the computers. We are able to discard small differences that we have learned to be meaningless and to note equally small differences that we have learned to carry meaning to the utterance accordingly. The sources for individuality are manifold:

- anatomical, unchanging (size of hand, larynx in an adult)
- anatomical, changing (size of hand, larynx in a child, plasticity in brain, age)
- physiological, changing (alertness, concentration, senses, adaptability, learning)
- situational, within situation (levels of noise, lighting, space acoustics, technical aids)

The ability to investigate a language which is conveyed in non-acoustic modality is helping us towards a better understanding of the linguistic structure of sign languages and the general underlying patterns that govern both acoustic and non-acoustic languages in the world. The non-acoustic languages of the world are fairly unknown territory to the scientific world even today. The mainstream of general linguistics draws distinct boundaries between signed and spoken languages of the world e.g. by stating grounds on language universals on the basis of acoustic languages only, although there are pioneers amongst linguists too, and the mere existence of signed languages (and the fact that they are true, natural languages) is becoming more and more acknowledged nowadays. There is an increasing group of sign language linguists, but sadly enough the contacts between spoken language linguists and sign language linguists are still scarce. There is an increasing group of sign language linguists, but sadly enough the contacts between spoken language linguists and sign language linguists are still scarce because of the small number of sign linguists. Thus there is a need of promoting sign language as a "testing ground" for linguistic and phonological theories.

As Finnish is a foreign language for signers, so is by default also the "existent" - written version of it. This situation calls for signed flow of information - from official documents to TV programs. Nowadays signers are able to book interpreters for translating documents (Laki vammaisuuden perusteella järjestettävistä palveluista ja tukitoimista 3.4.1987/380), but that process is always unique and does not totally satisfy the purpose of and needs for a written document in which the information should - and is thought to be - stable and unchanging. A translation by an individual interpreter is always correlated with many individually changing parameters by interpreter: knowledge of the background, terminology, competence and fluency in both languages, especially when considering specific areas of expertise, e.g. legal documents or benefit issues. The possibility of booking an interpreter to translate documents belongs to each signer in Finland, but the delays of getting the interpreter might differ from one place of residence to another. And just think about the flow of information that comes to you in written form every day. This translating of written documents is one service which could - and should - be provided by a stable and unchanging automatic process.

As both speech and sign by default are online and highly variable and not the least because of individual differences, written language is generalised, invariant and offline – you can refer back to written documents and the documents are the same to everyone. These two main aspects might be the very reason writing has been developed (For more discussion about writing, please see Coulmas, 1996). This is why writing is one of the most widely used formats for input within language applications. This again relates to the variability in both sign and speech as inputs for language applications. The variability in speech has been acknowledged within application design and here within this thesis the variability and individuality are also tested within sign language and its users.

1.1 Research questions and suggested reading pathways

This thesis aims to provide basic knowledge about sign language parameters at the phonetic level, i.e. on how individuals produce and perceive sign language. This provides basic input for sign language applications. Spoken language can be captured in writing, but sign language cannot since even today it, doesn't have a written form. Thus we need to think about alternative ways of encoding sign language. In the introduction (Chapter 1.4 in particular) sign language basics were explained. The most striking difference between speech and sign is the fact that sign languages are conveyed in three dimensions in space. This research question relates to basic phonetic knowledge of sign languages.

Another aim of this thesis is to provide tools for fundamental design parameters of applications language-independently, to provide the basic rules which in the later stages of application development can further be specified according to the particular languages used, whether signed or spoken. This thesis will provide tools for understanding the modifications needed for visual language interfaces, how the interfaces must be modified according to the needs of languages conveyed in the visual modality. This research question relates to application development.

In this thesis sign and speech as communication methods are compared and the results of the individual test settings provide possible evidence as to the common origins of human evolution of communication. How do the communication methods differ according to the modality and how are they similar independent and in spite of the modalities they are using and used in. From these similarities we can partly try to deduce the origins of the human language faculty. This research question relates to the ultimate quest of providing information on the origins of human evolution.

In addition to the three aims above, this study also takes part in the discussion of existing state-of-the-art language applications: what types of language applications are available, be they for speech or sign and what are the needs for the new types of applications. This study states existing problems and provides tools for solving these problems for its part. For example three-dimensionality has been introduced into the computing world through animation which is the basis for virtual reality and moving avatars. When we investigate the basis of human perception of phonetic categories within a sign language we can get answers on how precise the animation should be as to be intelligible for users. This question relates to developing information society into being even more integrative and accessible for everybody regardless of their language background.

As this thesis does not focus on one single framework or point of view but represents a truly multidisciplinary study, it can provide interesting viewpoints for readers with various research backgrounds and traditions. The inclusion of different themes and research questions in this thesis also provides a possibility of reading the thesis using different reading pathways. The research interests might result in using different reading sequences. Some suggested reading sequences (Figure 1.1) are as follows:

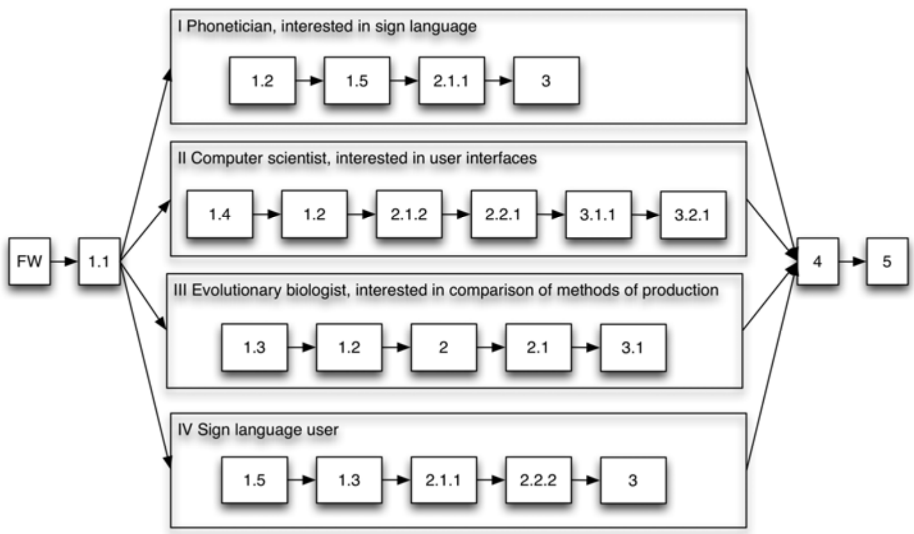


Figure 1.1 Different suggestions for reading pathways according to various interests and backgrounds of the readers.

Following one or other of these suggested reading sequences provide parallel and relevant ways to reach the conclusion. Each reading sequence is though slightly different because of different background knowledge on which the reader might reflect the new information provided by this thesis.

1.2 Phonetics covers speech but also sign

Because speaking and signing appear to be very different by nature when observing only the way the linguistic message is conveyed, it is easy to think that they are representations of two separate processes. However, both communication methods represent the same linguistic ability, which is sometimes addressed as universal language, connecting people (see e.g. Stokoe, 2001; Bellugi, 1980). Speaking is usually understood as producing and perceiving different acoustic signals. It is taken for granted that humans speak in order to be understood and that hearing is the basic prerequisite for understanding messages conveyed by speech. This auditory view is for some the only reasonable solution for speaking (e.g. Stevens, 2000) and undoubtedly it also might be the first things for laymen to think about speech and speaking.

According to the gestural view of speech production (see e.g. Liberman & Mattingly, 1985) the sound is only a carrier signal – a by-product, which turned out to be an excellent carrier of information to the listener from the speaker's thoughts using only the exhaled breaths of air. Speaking enabled to use hands into doing other chores than communicating. Gesturalists presume, that the evolution of speech started from the movements of the vocal tract (lips, tongue, soft palate and larynx), which in humans evolved into gestures of the vocal tract. Here gestures denote those movements of the vocal tract with a linguistic task: they represent simultaneously units of action in speech and the smallest units distinguishing between different meanings (phonemes) (see e.g. Liberman, 1996).

From this gestural view of speech production sound per se is not the source for perception but the dynamic vocal tract movements which produce the sound signal. This is the basis for the automated processes for producing and perceiving speech and speech can easily be perceived in different situations also when speakers are different from each others: others articulate more clearly and others have articulation more sluggish. Nobody speaks exactly the same way on two different occasions, but that is also not important as long as the code – i.e. language – is the same for both parties. The individuality and group differences within language production, even the effects of a situation onto speech or sign are studied by sociolinguistics among others (introductions by Coulmas, 1997 for speech and Lucas 2001 for sign among others). Undoubtedly our vocal tract and hearing apparatus are specialised in speech production and perception respectively. Production and perception of speech must have evolved into such a tightly woven nervous system that it has maintained its position as an inherent communication method for human for tens or even hundreds of thousands of years. Communication via gestures and signs is mainly conveyed by manual movements which then are perceived visually. Similarly signers communicate mainly via hand movements to convey linguistic messages and to be understood. Is it so that hand gestures are analogous to speech gestures according to our brain processes? The latest results in neuroscience research support this view (Willems & Hagoort, 2007).

Communication methods using acoustic or optic channels are learned similarly – by imitation: hearing children learn to imitate the sounds they make, deaf children learn to imitate the gestures they make respectively (Petitto et al, 2000; Goldin-Meadow, 2003). This stage of speech acquisition is called babbling. The easiest sounds to imitate are those with simultaneous visual input. Based on these visual inputs from speech sounds also deaf children can learn to speak words aloud, such as ’mummy’. After tuning in to imitation the expressive language capacity quickly evolves: hearing children learn to combine sounds and deaf children learn to combine hand gestures respectively. The goal for both groups of children is the same: to produce entities of meaning by combining meaningless tokens, i.e. speech sounds or handshapes. This is called the particulate principle, common to all systems, which yield unlimited sets of meaningful information from a fixed, limited set of meaningless units. This is done by means of combining, permutating and repeating the fix set of tokens and the principle holds in chemistry (chemical bases and compounds), physics (atoms and molecules) as well as in formal and natural language (phonemes, words and ultimately language) (von Humboldt, 1836/1972, Abler, 1989, Studdert-Kennedy, 1998; Révész, 1991 on formal languages). Individual speech sounds are then combined to words and handshapes and movements are combined into signs respectively. Thus lexical development progresses according to the same principles in both hearing-speaking and deaf-signing children (Takkinen, 2000).

Although on the surface speaking and signing are based on different motor and sensory systems, the basic structural processes inside the brain involve the same gestural-motor production and perception pathways. (Arbib, 2003; López-Garcia, 2005). Speaking engages the delicate vocal tract organs in coordinated sequences while signers use their hands and body to execute the intertwined coarticulatory speech flow (Wilcox, 1992; Liberman & Mattingly, 1985; Bellugi, 1980). Recently the view of language based on mimetic communication has been widely studied (Corballis, 2002; Arbib, 2002; Korhonen, 1993; Zlatev, Persson & Gärdenfors, 2005). This common underlying mechanism for human and ape communication probably evolved into human speech partly via crucial gene mutation in FOXP2, into which the human language ability has been linked, as shown by various research groups around the world studying families with specific language impairments (SLIs) (Lai et al., 2001, Hurst et al. 1990, Fisher et al., 1998, Lai et al, 2000). Thus the brain processes for both sign and speech would coincide – and seen from inside the only difference would remain to be defined by modality (Nishimura et al., 1999; Petitto et al., 2000).

Ultimately speech perception and production should be understood as one and the same process in the human brain and thus these two ends would not be distinguishable by their motor or sensory dimensions. Similarly, perception and production of signs should be understood as two features of one process inside the brain. Imitation has a very important role within this scenario. A prerequisite for imitation is the integration of sensory and motor processes in both modalities. This integration has been partly explained via the so-called mirror neuron system (for further discussion on mirror neuron system and Broca-Wernicke complex see chapter 1.2 in this thesis).

The similarity between sound and sign gestures lies in the fact, that both require smooth cooperation between the classic speech processing areas, most often referred as Broca's and Wernicke's areas (Brodmann areas 44-45 and 40-42 respectively (Brodmann, 1909), for further information on Brodmann areas see e.g. Amaral, 2000) in the cortex named after clinical researchers Paul Broca and Carl Wernicke who studied language impairments due to lesions in these areas and thus defined the classical symptoms. If there are impairments in the cooperation between these cortical areas or within the structures, this results in different kinds of language impairments, such as aphasias and anomias both in signers and speakers (see e.g. Poizner, Klima & Bellugi, 1987, Damasio et al., 1996, Galaburda, 1994, Emmorey et al. 2002, Hickok et al. 1999 for details on individual studies; on general discussion on speech aphasia see Caplan, 1987; on signers see Corina, 1998 and on general introduction on the neural basis of aphasias see Damasio, 1992). The most interesting results in the scope of this thesis can be obtained when comparing symptoms and lesions over these two language groups: similar symptoms due to similar lesions in both language modalities. That is, that both signers and speakers share symptoms due to a particular type of lesion.

When doing his pioneering work on studying American Sign Language using linguistics framework and methods Stokoe invented the term chereme to correspond to the term phoneme to emphasise the difference between spoken and signed languages. This invention of a new term for the smallest units in sign language was a coine based on the term for phoneme but substituting hand for sound. This was done on the basis of the different surface structures in sign and speech. This also emphasised the temporal and structural structure differences within the two different language modalities. On the other hand, during the evolution of research in sign language linguistics it also has been used to emphasise the fact that sign languages indeed are full-fledged natural languages capable of expressivity on their own completely independent from spoken languages (Stokoe, 1960; Liddell, 1980 (ASL syntax); Brentari, 1995 (SL phonology); Supalla, 1992 (ASL names) international – Finnish: Rissanen, 1985 (FinSL structure); Takkinen, 2002 (FinSL acquisition); Rainó, 2005 (FinSL names); Savolainen, 2000 (FinSL transcription); Jantunen, 2009 (FinSL syllable structure)). Signed languages are composed of optical phonemes which are equivalent to the acoustic phonemes in spoken languages. Some researchers still refer to optical phonemes as cheremes. This is done to emphasize the distinctions between the languages. Nowadays the term chereme is usually replaced by the term phoneme, or more precisely the term optical phoneme (first used by Wilcox, 1992) as to emphasise the similarity of the underlying neural systems for both signed and spoken languages. Also in Finland different researchers use different terms for sign language phonemes depending on their point of view, whether they emphasise the difference between signed and spoken languages (e.g. Rissanen, 1985; 2000) or the similarity of the language processing independent of the modality used (Takkinen, 2002; Savolainen, 2000). The main difference between the signed and spoken languages is the difference in the channel and the medium in which the message is conveyed and received: in spoken languages the message is mediated mainly through an acoustical channel while in signed languages the channel is optical. When thinking about the neural processes in the brain, the phonemes and cheremes are basically one and the same thing, because the recent results in neurosciences suggest that speech and signs are processed in the same cortical area, i.e. the secondary auditory area (Nishimura et al. 2000; Petitto *et al.*, 2000).

Also speaking people sometimes take advantage of the visual cues in speech in some situations. These visual cues are called visemes (Fisher, 1968). They are used and relied on speech perception in situations with background noise to find supportive cues for the information carried within the acoustical signal (Tye-Murray, Sommers & Spehar, 2007). Usually when sounds are acoustically close to each other they differ in their visual cues, i.e. visemes (Carney, 1988; Hazen, 2006 among others), while sounds which are visually close to one another are acoustically more distant. People hard of hearing or non-genitally deaf base their speech recognition usually solely on visemes. This is called speech reading or lip-reading. There are several products and web-based tutorials all over the world for lip reading classes and self study (e.g. http://www.lipread.com.au; http://www.lipreading.net/) along with wide academic interest covering different research frameworks (Hall, Fussell & Summerfield 2005; Santi et al., 2003; Jiang, 2002; Campbell & Wright, 1988, Capek et al., 2007). In Finland visemes are mainly studied in relation to the speech signal (Sams, Möttönen & Sihvonen, 2005; Sams et al., 1991; Möttönen et al., 2002). The research method is based on studies of the McGurk-effect (McGurk & MacDonald, 1976), which is a means to study the importance of visual information in the speech perception process as a whole. The McGurk –effect arises when the visual and auditive stimuli are incongruent, but yield into a combination in perception. When the acoustic information you hear is [ga] and the visual cue gives you [ba], you perceive [da] which is different from either of the original parts of the speech stimulus and furthermore, is actually located in articulatory space in between of [ba] and [ga]. This intersensory integration, that is the ability to combine features and information from more than one sense at one given point in time, begins shortly after birth: infants who are 3 weeks old can perceive the equivalence of auditory and visual inputs (Lewkowicz & Turkewitz, 1980) and 4-months-olds can perceive equivalence of audible and visible syllable inputs (Kuhl & Meltzoff, 1982). The McGurk –effect underlines and focuses on this human ability to gather information of several different senses simultaneously and also gives researchers new vistas on the crossmodality in human action (for further study designs and results with discussion see Spence and Driver, 2004).

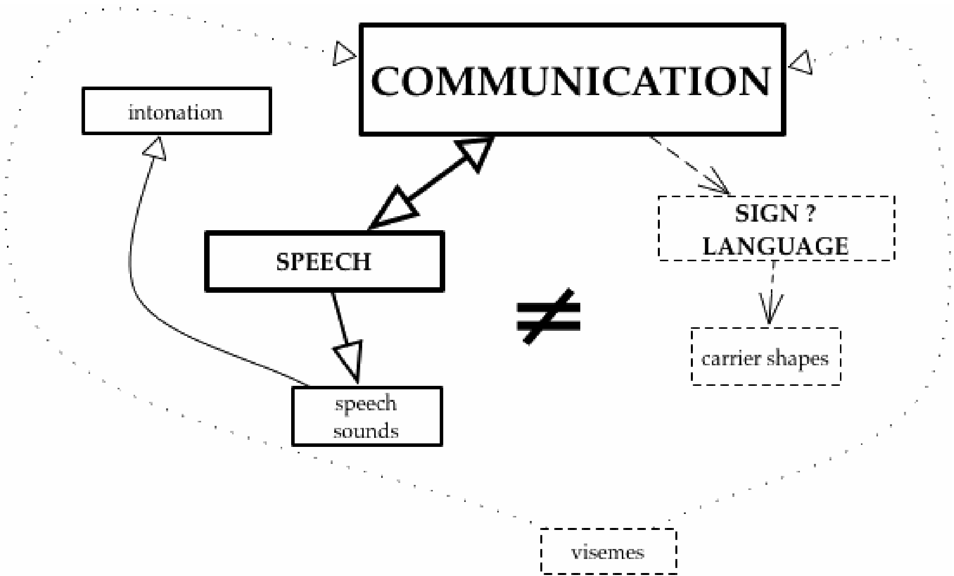


Figure 1.2 The relationships between different ways of communication have not traditionally been very well-defined. In fact, previously it has been thought that signed languages have not been true languages and furthermore, that there wouldn't be any similarities between communication via different modalities.

The next chapter will give an introduction to the basis of the human ability to take advantage of crossmodal action and its evolutionary paths: why are we so keen on copying others with our actions? The human ability to copy others is the glue of human existence, the ability to work as a team. The next chapter also dwells on the evolutionary aspect of human language ability and where it derives from – and what is there to be seen from those bases in human action in general even today.

1.3 Sensory-motor complex, Broca-Wernicke complex and mirror neurons

The notion that communicative actions derived from object-directed actions is not new - Vygotski (1962) already pointed out that the evolution of the pointing gesture was due to attempts of children to grasp far away objects. This pointing gesture is one of the first communicative actions in an infant. That is, the infant wants something that is out of reach and uses the pointing gesture to indicate that a caretaker should hand it over. This indicates that its function in evolution has to be important. Furthermore, pointing gestures also are present when adults speaking different languages with no knowledge of a mutual language are trying to cope and to be understood. This supports the basic need for communication and also emphases the meaning of pointing as a goal-directed movement.

One of the most intriguing findings in neuroscience recently are the so-called mirror neurons, which have been suggested to be the missing link when it comes to human language evolution. They were found to activate in a monkey's brain when it was not only grasping objects but also when it observed objects being grasped. Thus they were discovered to be a class of visuomotor neurons that discharge both when doing a particular action and observing another individual doing a similar action. They are situated in area F5 in the monkey brain. Mirror neurons respond to seeing object-directed actions; that is, they require an interaction between a biological agent (hand/mouth) and an object. In monkeys the same grasping mirror neurons respond to both human and monkey grasping hands and the response is not affected by whether the action is done near or far from the observer. It is in fact of only little importance whether or not the observed action is eventually rewarded; the discharge is of the same amplitude in either case.

The most important group of these observation-activated neurons are those responding to mouth movements and especially those which are activated by e.g. a lip smacking gesture. This for its part supports the role of CV syllables in evolution theory (MacNeilage, 1988). Thus one might argue that these communicative mouth mirror neurons reflect a process of corticalisation of communicative functions but are not yet freed completely from their original digestive basic function.

Mirror neurons represent the neural basis of a mechanism that creates a direct link between sender and receiver of the message, such that actions done by other individuals become messages that are understood by an observer without any cognitive mediation, so much so that Rizzolatti & Arbib postulate that speech evolved mostly from gestural communication. The novelty of this theory consists of the indication of a neurophysiological mechanism creating a common nonarbitrary, semantic link between communicating individuals. This meets the parity requirement. The parity principle states that "what counts for speaker must count for the listener, otherwise there is no communication" (Liberman & Mattingly, 1985). It is likely that a great leap from closed system to a communicative mirror system depended on the evolution of imitation (Arbib, 2002) and related changes in the mirror system: the capacity to respond to pantomimes and to intransitive actions. Further evidence of the importance of pointing gestures comes from primate studies: the precondition for understanding pointing is already present in monkeys, i.e. the mirror-neuron system.

Hand and arm gestures and speech gestures are strictly linked and share, at least in part, a common neural substrate, which is suggested by a number of studies stating that the excitability of the hand motor cortex increases during reading and spontaneous speech and that this effect is limited to the left hemisphere. This effect is not found in motor areas not connected to language, such as the leg motor area. Most interestingly, grasping movements influence syllable pronunciation not only when they are executed but also when they are observed. The syllables tend to be articulated in unison with the grasping movements, whether they are pronounced or observed. Similar examples can be found in numerous experiments which show that hand gestures and mouth gestures are strictly linked in humans and that this link also holds with the oro-laryngeal movements used for speech production.

Well, then, how did the transference from hand to mouth happen during human evolution? Paget (1930) suggests that certain combined sets of motor actions constitute gestures whose meaning is transparent; eating can be denoted as combinations of sounds related to eating, that is, lip-smacks with opening of the oral cavity. The most fundamental step towards speech acquisition per se was when individuals became able to generate the sounds originally accompanied by a specific action without doing the action itself. This, however, would require an echo-neuron system, which Fadiga with his work group have been studying. They recorded muscle evoked potentials (MEPs) from tongue muscles in normal volunteers instructed to listen carefully to acoustically presented verbal and nonverbal stimuli, and the results were intriguing: MEPs from tongue muscles showed an increase when the presented stimuli contained segments needing precise tongue articulatory movements. Similar results have been obtained in connection with different muscles and different kinds of recording techniques (see e.g. Watkins et al. (2003)). One might conclude that when an individual listens to verbal stimuli there is an activation of the speech-related motor centers. Furthermore, one could elaborate these obtained results in concordance with Liberman & Mattingly et al.'s proposal that the echo-neuron system mediates also speech perception processes. The possibility of this being true would fit perfectly with the assumption that speech indeed did evolve from gestures. The claim is further backed up with results from a wide range of studies that show that different layers of speech activate the motor system in the brain (Boulenger et al., 2009 on idiom comprehension; Aziz-Zadeh et al. 2006 on metaphorical sentence reading; Pulvermüller 2001, 2005 on semantic information within language, among others). In some theories the word gesture is used as a technical term for movement atoms and as such it would be the common basic element for sign and speech (Browman & Goldstein, 1992 among others). In this section of the thesis gesture is used in a more traditional way (Kendon, 1972 among others). Within that framework gesture is used when referring to extralinguistic movements. Regardless of the framework, gesture can be considered as to be a common element in speech and sign, be it directly as in technical core element or via evolution of human species as in the more traditional framework.

Humans gesture when speaking: the gestures might denote turn-taking, bring something new to the issues in question, such as directions or manners of movement. Also blind persons gesture when speaking, although they have never observed others doing so. Gestures are usually pictorial in nature and vague so be that they cannot exactly be imitated, because they do not have an observable inner structure, but they are incorporated in the discussion and are thus once in a lifetime –like in that manner. Different cultures have different gesture repertoires, for example in some cultures there is a stronger tendency to use gestures than in others. Culture-bound gestures, which have a certain function within that particular culture are dividable into smaller units, but not in the same way that signs in sign languages are. For example in Finland the thumbs-up gesture has the notion ‘good’ but for divers it is a sign for something being seriously wrong and thus there is a need for urgent surfacing.

The fact that the co-speech gestures are nowadays still in use is a supportive argument for the possibility that gestures would be the basis for human language. This gestural communication has over time developed into full-fledged signed languages among the deaf. Unlike gestures, signs within a particular sign language have a distinctive inner structure and the individual parts of a sign are combinable into meaningful signs. Changing one part of the sign changes the meaning. For example, when changing the handshape of a FinSL sign TIETÄÄ ’to know’ (a flat hand) its meaning changes into

MUISTAA ’to remember’ (a fist hand). This is a feature unique to human languages.

There are also others than deaf who sign[1], for example the Wahlpiri culture in Australia does not allow speaking during mourning but you have to sign at that time period (Sebeok & Umiker-Sebeok, 1987). This has lead to a sign language-like system. This system is, however, bound to spoken Wahlpiri language, so that it is a ”manual code” from a spoken language. Other well-known examples are so-called monastic sign languages (ibid.). Their lexicon has a strong religious content in. These kinds of signed codes have one thing in common: they are used on special occasions, in situations which have a specialised status in one’s life.

The full-fledged sign languages used by deaf people on the other hand are completely separate from spoken languages. For example Finnish Sign Language has very different word order from spoken Finnish. Also the typology, that is the genetic relations of signed languages, is very different from that of spoken languages, for example British and American Sign Languages are not directly related but both of them are related to French Sign Language. Sometimes the grammar and structure of a sign language seems to contradict to what would be intuitively “natural”: according to the structure of American Sign Language (ASL) the intensifying of a meaning is signed with increased velocity, so the sign ”MORE SLOW” is actually signed faster than the sign ”SLOW” (Gibson & Ingold, 1993).

Not only speakers but also signers gesture along with communication. These gestures are much more difficult to notice since gestures and signs are conveyed through the same visuo-spatial channel. Signers also use pantomime, which can in some occasions, e.g. in connection with a sign name, develop into a new sign. Usually name signs are formed based on some specific feature of a signer, usually close to facial features. In the development process the name sign can be two-handed, but fairly quickly it diminishes into a single-handed reduced proper name sign in the central area of sign space, which has no obvious resemblance to the original feature. This is the case especially for so-called descriptive name signs (Supalla, 1992).

Thus, the use of gestures is a common ground for signers and speakers alike. Gestures as such are holistic movement envelopes. Every culture has its own manners of using gestures, so they have a culture-bound lexicon of gestures, which all have their own culture-based meanings. Gestures are also babies’ first

[1] There are also studies where animals have been taught to sign. The most famous signing animals are the gorilla Koko (Patterson & Linden, 1981/1986) and Washoe the chimpanzee (Fouts, 1997/1999) among others.

way of communication: also hearing babies gesture with their hands (Petitto et al., 2000). Hearing babies abandon manual babbling, but for deaf infants it is the first step towards a full-fledged signed language (Goldin-Meadow 2003) independent of the communication method of their parents. Children of signing parents automatically learn their parents' native language, just like the speaking parents' children. The method of communication between speaking parents and deaf children in the early stages are gestures and their combinations, but at the latest by 3½ years of age the combinatorics used by the child in their use of gestures in combinations are more highly developed than that of their mothers. The child's gestures have some degree of inner structure, thus they form a so-called home sign system.

Language has all the properties needed for an adaptation. As such, it can be studied both from a biological and an evolutionary perspective. Unfortunately speech doesn't fossilise but some guiding landmarks can be obtained from cranial fossils which provide information on how brain areas have varied in size. For example even the earliest hominids, such as H. habilis, had larger speech-related brain areas than monkeys and apes. Thus, the increase of speech-related brain areas happened long before the descent of the larynx. This supports the view that not only sounds but also gestures or signs have had an important role during human evolution. Arbib (2002) writes that already the first modern humans used sound-based communication. It was a very primitive system accompanied with a gestural system, but it developed via cultural evolution into the present subtle system of sounds.

According to Rizzolatti Broca's area can be traced back to the ancient brain area related to motor actions. This also supports the gestural theory of speech, which states that speech has developed via hand gestures. Actually it is a combinatory theory, because it suggests that "there are not three separate systems for hand-, sound-, and facial gesture communication but one general system, which operates in different motor-sensory modalities". Mirror neurons provide a possible missing link when discussing the evolution of human language communication. They are not, however, the only explanation for human language ability, but most probably one part of the puzzle.

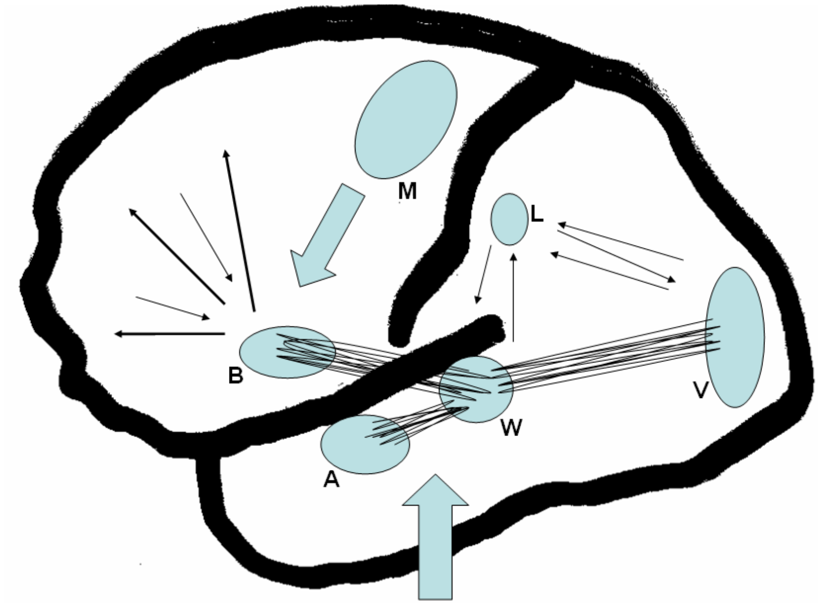


Figure 1.3 A model of the connections between different brain areas governing different functions in a human. Explanations of abbreviations are as follows: M=motor cortex, B=Broca's area, A=auditory cortex, W=Wernicke's area, V=visual cortex, L=lexical processing and reading.

Nevertheless, the Mirror Neuron System (MNS) with sensory-motor integration (Rizzolatti, Arbib et co.) can be seen as the main element of the human language ability or, as the ultimate core of the sensory-motor links between the observed articulatory gestures done by other individuals and the internal articulatory pattern memory traces of gestures we have produced ourselves. MNS presents "the missing link" of the language inside principle (López-García, 2005), which states that both sign and speech are manifestations of the one and the same general inner human language capability. Furthermore, the system was tuned during the the human evolution from a general gesture and trajectory processing system into a system including more specific gestural patterns as well, such as produced when speaking (Corballis, 2002 among others). Recently, new findings in neurophysiological studies have shown that even signs are processed in Wernicke's area (Nishimura et al., 2000, Braun et al., 2001; for a review see Campbell et al., 2007). In speech processing, the so-called Broca–Wernicke complex – the manifold, strong connection pathways between Broca's and Wernicke's areas in the brain – can be seen as the core area with regards to human language ability (Figure 1.3). This Broca-Wernicke –complex is the basic explanation for the possibility of using analogical structures in applications for speech and sign. Therefore, we can use the same basic design in applications for both modalities, the design has only to be adapted for the modality.

This is a collection of studies based on previous findings on speech. The overall structure was designed to be an analogous one to speech studies. All of the studies (the studies discussed in chapters 2.2, 2.3, 3.2 and 3.3 respectively) were done on perceptual continua with only one variable in each one. These continua were all between two perceptually distinct and adjoining categories with no intermediate categories possible for either spoken Finnish or Finnish Sign Language. The situation might differ for other spoken or signed languages. All of the categories had articulatorily defined boundaries as well.

The concordant features in responses within these perception studies over language modalities show the common adaptability of humans to highly variable signals based on the individual differences in speech and sign production, which is discussed further in chapters 2.1 and 3.1 respectively. All in all these small results can be regarded as being concordant with the MNS and the human ability to mirror and adapt to each others' behaviour also language-wise.

1.4 Gestures provide a general background for applications

Gestures (goal-directed movements in space) provide a common ground for both signed and spoken languages, however, the non-linguistic nature of gestures makes them different from both speech and signs in the signed languages. Signs and hand gestures make use of the same visual modality and thus their relationship has not been understood until recently, while the difference between speech and gesturing (so-called co-speech gestures) has been understood longer, because of the different modalities they use (see e.g. McNeill, 1985). - The articulators in sign languages are the same as those involved in non-linguistc reaching and grasping movements. Signs can be perceived also as gestures, especially for those who do not know sign. In the same way signs in foreign sign languages are usually perceived as gestures, just like different spoken languages are perceived as just speech sounds and not anything linguistic (Figure 1.4).

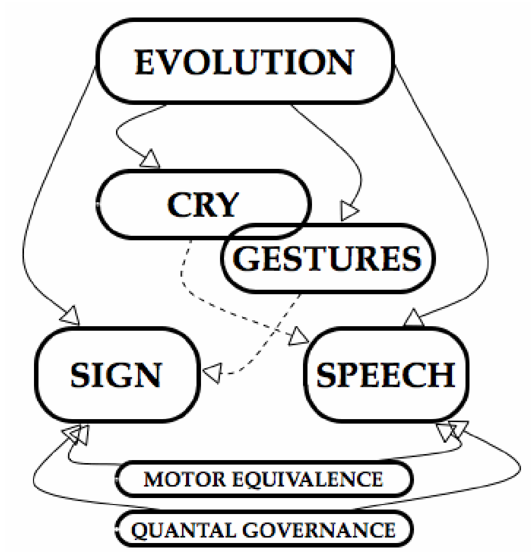


Figure 1.4. Schematised overview of evolutionary aspects on different stages of human language development and the forces governing learning to speak and sign.

In general, both sign and speech can be studied as temporally ordered muscular actions. Also the evolutionary perspective on language also has properties which make it one possible point of view for looking at sign and speech as co-variants of the same inner faculty of language (Hauser, Chomsky & Fitch, 2002). That provides us with a view that the key in language understanding is in organization and control of articulatory movements - whether vocal or visible. There are congruent results from movement control of hands and mouths (compare e.g. Shiller et al, 2005 and Darainy et al., 2007; both indicate voluntary control of stiffness). There are also contradictory results in different experimental settings, which would result in different governing mechanisms for speech and sign (Tremblay et al., 2008 among others), but that is merely an indication of the fact that there is still a long way to go before we can truly understand the human brain and its contributions to all functions within us related to communication. This similarity in articulation control also gives us a common basis for applications in both modalities. Thus, there is a core design with the same structures for applications, which then are modified according to the boundary conditions stated by the differences in modality and perceptual/producing parameters. These boundary conditions are studied in various other fields investigating the physiological and neural basis of human vision, movement and hearing. These boundary conditions also give the limits for application design, they answer questions like: “What would be the optimal speed for movement within segments?”, “How should we modify the natural source for the application to make the application as intelligible and economical as possible?” among others.

While there are individual differences in both sign and speech it's fairly easy for a human to overcome these individual differences and decipher what is said or signed. However, for a machine trying to achieve similar decoding from this phenomenal variation of individual features in both methods of communication is almost beyond capabilities. This provides us with a bundle of research questions in application design concerning individual differences: "What is there within a signal?", "How could we capture parameters that are the bases of individuality?" among others. This, in turn, will enable the application designers to pinpoint the possible sources for problem solving for tackling with individual variation.

The questions stated above hold for both speech and sign applications, however, we are already better equipped with answers to the questions in the speech technology approach, whereas with sign applications the quest for answers is still in its early stages. Just like in speech technology approach, as in the sign applications there are two different approaches to solving the task of providing a language technology application for users: data-driven and rule-based approaches (for an extended review on different approaches on sign translation applications see Morrissey, 2008). When designing a speech-to-sign interface once we have solved the question of which approach we should take on the sign language part of the applications in general we have to proceed to how to encode information for it to be fairly easily translated from one language to another, and in this case it also requires a transition from one modality into another, that is from sign to speech – from the visual to the auditory modality. The choice of approach in the visual modality also is dependent on how we have solved the approach issue on the auditory side of the application.

The studies provided input for a future synthesis development as well. The results from production tests provided us with knowledge on the range of variability in the source signals. The results from perception tests provided further detailed information on the filters for the inputs respectively. These results yield towards designing a rule-based synthesis-recognition system for both sign and speech. The desire for a rule-based synthesis is two-fold. A rule-based synthesis is thriving towards a sieve-type design. Be the input what may, the salient features would be gathered in the sieve without a need for on-going data mining but relying on a finite set of rules.

Technological advancements in computer graphics have provided the core elements for sign language applications by gesture animation in the gaming industry. It has provided the basics for hand and body movement animation, along with facial gesture and mimicry animation. The latter has been further developed by speech technology advancements in talking head –applications, such as the KTH and SynFace joint beta application called EyePhone (SynFace, 2006), which provides a synthesised face for Skype calls and is used as an extension for Skype. It uses the incoming voice from the other party of the Skype call to synthesise the visemes based on the acoustics of the incoming voice. There is a demo software available on the Internet, which can be downloaded free of charge.

The ultimate goal for sign-to-speech interfaces is to provide a possible solution for the communication barrier between signers and speakers. There are difficulties in communication whenever there is more than one language being used within the conversation. This barrier is even harder to overcome when the languages used within the conversation use different modalities, i.e. auditory vs. visual. There is also a lack of public information to the signers. It is most often provided either via teletext applications and subtitling on a tv more generally or by one-to-one basis via sign language interpreters. Teletext and subtitling require high literacy skills and sign language interpreting services are not accessible 24h. Further development for sign applications is needed for providing general information for the signers. Sign applications would also support equality between signers and speakers when it comes to access to information.

1.4.1 On gesture and sign application studies – recognition and synthesis

The research on sign applications has its roots in multiple disciplines and because of that also has experts from those disciplines very often working together. The problem with sign applications is at least two-fold: the quest for sign recognition (perception; equivalent to a speech-to-text –approach) on one hand and on the other hand the quest for sign synthesis (production: equivalent to a text-to-speech –approach). Sign recognition efforts include systematic variation modelling in movement dynamics (Ong, Ranganath & Venkatesh, 2006) and gesture interpretation from video sequences using moving edge detection (Quek, 1994). The movement dynamic modelling approach has mainly been concerned with the systematics in movement patterns governed by the (morpho-)syntactic variations within the sign language. These boundary conditions vary somewhat from one language to another. There are also problems in recognition based on individual variation in signs produced by different individuals. This individuality/lack of invariance problem in production is further investigated and discussed in Chapter 4: Individuality in Production and Perception of Sign in this thesis. The sign recognition approaches have used e.g. HMM modelling of each of the sign language phoneme categories (for further description of fundamentals of sign language, please see Chapter 2 in this thesis) (Braffort, 1996). Most of these recognition systems aim for signer-independent recognition (Ong, Ranganath & Venkatesh, 2006). The background for sign-based applications comes from simpler gesture-based applications used e.g. in emotion related interfaces mainly related to HCI dialogue systems (Gunes et al., 2004; Cassell, 2000) and pointing gesture applications for augmented reality interfaces (Moeslund et al., 2004; Steil et al., 2004) as well as in the gaming industry where gestures are used to play sports games and new approaches are also developed for adaptive games which take into account the emotions you are feeling and adapt games accordingly. Other vistas of gesture recognition include e.g. virtual music in the form of a virtual air guitar developed at the Helsinki University of Technology (Karjalainen et al., 2006).

Although there still are various technical obstacles to tackle when working with applications taking alternative input and giving alternative output from text, such as speech, or more recently gestures and haptics (Turunen et al., 2009, Raisamo et al., 2009 among others), the efforts and advances towards an all-inclusive information society will be achieved through devices and applications which are modifiable through various I/O channels. There are also devices, which are capable of taking info through more than one channel at a time, or in rapid alteration of two channels (Turunen et al., 2006). Some of the devices are especially designed for users with various disabilities, which affect either the individual's ability to input the information needed or to benefit from the output from the device in question (Turunen & Hakulinen, 2007 among others). Sometimes applications are further developed from the viewpoint of discretion (Rantala et al., 2009). Text output can also be modified to be readable with the sense of touch, such as in Braille format using different techniques.

The methods of enabling the gestures and further on, also signed input, are within the focus of this thesis. Currently there are devices able to take multilingual input, but that ability is for the major part restricted to languages that have a written format or are acoustic by nature (text or speech input). The possibility of using gestures as an input method is the first step towards a full-fledged signed input. That would ultimately make signers equal to speakers and enable them to input information to user interfaces using their own native languages in the same manner as speakers do when benefitting from speech interfaces within different devices. But already the possibility of using a gesture-based input would bring user interfaces nearer to the signer's native language input method and would thus be more convenient for native signers to use. Some of the current applications have existing gesture input, and some are even able to cope with blow flow detection as a start signal (Turunen et al., 2010). There are instances when speakers also prefer gestures to speech, such as in instances when using deixis within space. That is, when using expressions like "that thing there", or "I would have to get there", utterances which frequently co-occur with so-called co-speech gestures in normal conversation, even during telephone conversation when the listener cannot see them.

For recognition of signs and movements different research groups use different ways of encoding signs: some of the research groups use phoneme categories, such as handshapes and POAs (Tamura & Kawashima, 1988) while others rely solely on more technological-based approaches (such as HMM and SRN) (Gao et al., 2004). All approaches are aim towards signer-independent dictionaries of isolated signs and thus (reading between the lines) are aiming for closed set databases in the first stages to in the future move on to applications capable of have unlimited input. This requires further basic research on fundamentals of sign languages and their representation in individuals, both in production and perception.

The sign synthesis approaches are mostly statistical requiring fairly large amounts of computing time and space, furthermore they are totally dependent on the size of the database in use. This is the main reason that these statistical approach applications are mainly used in domain-specific applications, such as "supermarket shopping" containing video sequences (Bauer et al., 1999). The video sequences are analysed and the feature vectors are fed into a database for further retrieval and used for recognition (Morrissey, 2008). This approach also has its limits on extensibility. Another approach is to use multiple path architecture, that is to use any way possible to navigate between different options to get to the goal, i.e. whatever the hierarchy of analysis for the synthesis of the particular sign (Huenerfauth, 2006). This architecture has multiple hierarchical layers of information simultaneously available with intertwining paths and crossings to navigate through. The usability of this approach, however, requires user knowledge of multiple notation systems and understanding of the grammar parameters of the particular languages in question. (Morrissey, 2008).

1.4.2 The state-of-the-art of coping with individuality in existing speech technology applications

Jurafsky & Martin (2000; p. 235) put the problem of automatic speech recognition into words very well: "Spoken language understanding is a difficult task, and it is remarkable that humans do as well at it as we do." Speech for us is so automatic that it has become the ultimate user interface amongst us; we usually use speech in communication. For a computer the variation of acoustic speech is almost beyond capabilities, but the latest technical development has brought us somewhat nearer to speech technology application. For a human it is fairly easy to adapt to the needs of a computer application, but for an application adaptation is difficult.

The different areas of research within speech technology relate differently to individual features in speech. They can thus be divided into three different groups according to their relation to individuality: in the first group individuality is something that is present but the applications are not able to cope with it that well. It can, however, be used as a parameter for evaluation of the applications, such as word-error rate or string error rate (Juang et al., 1997). In the second group individuality is acknowledged to be present and the applications are also able to cope with it quite well. The third group of applications is regarding individuality as an asset, something that the applications are able to benefit from. The groups are listed here in diminishing order, from the largest and to the smallest one.

The first group includes for example speech recognition systems (Jurafsky & Martin, 2000). Automatic speech recognition systems do quite well if they are used in speech detection, that is deciding if some sound source is voice or noise (Tuononen, 2008). Speech detection can be used as the first trigger of further stages in applications. After the decision has been made that some sound source consists of speech then it triggers the next stages within that particular application (Rabiner 1989). For a speech detection system it is enough to decide whether something is speech or not and it is not concerned about individual features in speech – it was not designed to do that.

The existing state-of-the art of speech recognition systems are fairly often viable in limited domains. That is they operate within a specific, a priori defined vocabulary, such as the alphabet, command words or more often a built-in vocabulary, such as in dictation softwares. They are also somewhat flexible in that they can be taught to handle a specific user dictating e-mails, letters etc. However, they are language-specific to that detail that most of the speech softwares are nowadays capable of coping with English input only. Speech recognition systems are part of speech-to-text solutions. Using speech recognition systems resembles the situation when talking on the phone – the medium excludes facial expressions and pointing as augmentative features in communication.

The recent development (Cassell, 2009, 2010) has brought human-computer interaction nearer to human-human interaction by the means of embodied conversational agents. They are computer interfaces designed especially for conversation using all modalities, which includes non-verbal means of communication as well, such as capability to recognise and respond to facial expressions or pointing gestures. The agents are also able to deal with conversational functions, such as turn-taking. The development of embodiment within user interfaces, however, brings the confusion element into the conversation where the end user is not sure whether s/he is talking to another human being or a machine (Shneiderman, 1998). The benefits of embodied conversational agents include for example the capability of recognising gestures as a simple pointing gesture can sometimes be used to replace a lengthy verbal explanation about where something is.

The group which benefits from the individual features in speech are speaker recognition systems. The systems are used in situations where the identity of the speaker is important to know. The systems include two different categories of software: speaker identification and verification systems. Speaker identification system tries to answer questions like: “Whose voice is this?” Identification systems can be used for example access control, that is, identifying who will gain access to certain premises, as well as other security applications (Kinnunen, 2005). These systems usually operate within a certain group of individuals. That is they have a certain set of voices and the application decides who within the group the speaker is. Speaker verification on the other hand tries to answer the question: “Is this Bob’s voice?” Speaker verification systems are mostly connected with forensics, so that the automatic methods are combined with human comparison results. Speaker verification also is concerned with various methods of fraud: they have to cope with impostors, efforts to use voice disguise and so on (Kinnunen, 2005, Niemi-Laitinen, 2001).

On the other hand, speech synthesis can be intelligible without any individual features whatsoever. So, if synthetic speech is regarded as something that has to be intelligible, then individual features are not the ones to be concerned about. Another, more specific question is: “Intelligible to whom?” Then again, individual features in speech perception have to be taken into account. There might be regions within the acoustic signal where individual speech sounds might result in confusions (for further information about individuality in speech perception please see chapter 3.1.2. of this thesis, also Savela, 2009), but in a speech signal sounds are not isolate, but embedded into words and phrases, so that individual confusions do not matter that much. This form of synthetic speech does not require lots of computing power, and it is now used in talking computers (screen reader solutions) for the visually-impaired users such as Nuance Talks or Microsoft Narrator. You can manipulate the pitch and speed of the voice. For example Narrator can be used to read aloud the characters you write onto your document. Nuance Talks is a widely-used screen reader software and has specific keyboard shortcut commands to operate the reader. In Finland the most widely used screen reader softwares are Jaws and SuperNova. Jaws is purely a screen reader software while SuperNova is a compilation of both screen reader and magnification softwares.

If we strive towards naturalness in synthetic speech, then individual features in speech are an asset, something to benefit from – we can then adapt the synthetic base signal of the speech towards some individual features in speech, such as prosody, duration, acoustical prototypes of vowel sounds etc. This form of synthetic speech is usually embedded in conversational agents in call-centre applications and information desk applications to enhance naturalness into the conversation. That could also present difficulties when the application is mistakenly presumed a human and frustration of misinterpretation confuses the end-user of the application. This is used also in mobile devices and in navigation system applications, which are operational within a small, definite vocabulary (Saarni, 2010). It might in the future, however, present some ethical questions on voice quality when it approaches naturalness in such a detail that it would be misidentified as a specific individual (Saarni, 2010, p. 145).

1.5 Fundamentals of Sign Languages

Since signed and spoken languages activate the same brain regions, the same circumstances apply in sign language perception as in speech perception. Speech perception goes from the ear through the auditory nerve to the primary auditory cortex and furthermore on to the secondary auditory cortex (Wernicke's area), while sign perception uses an alternative route to the secondary auditory cortex: from the eyes through the optic nerve to the lateral visual cortex and furthermore on to the secondary auditory cortex. Signs, too, are understood even though they are not invariant, just as in speech perception. The shape of a one particular sign differs from one signer to another because of physical differences.

The first alternating parameter is the shape and size of the hand. The second parameter that varies from one individual to another is the size of the signer's signing space. It depends on the size of his or her hands – bigger hands need more space in which to move around. The signing space of an adult is on the average larger than that of a child. Thirdly, the amount of coarticulation in signing varies with the size of the signing space and the rate of signing: the faster the rate the more reduced is the produced and perceived signal. Coarticulation in sign languages is a more complex phenomenon than in spoken languages. Coarticulation in sign languages can be divided into two levels. The coarticulation in the dominant and non-dominant hands forms the lower level distinct units. The coordinated coarticulation of both hands forms the upper level coarticulation entity.

Signing can be understood even if the signing stream is discontinuous or the received signal is disturbed or low in quality. The low quality of the signal can be the result of poor articulation or visual background noise. Visual background noise can consist of a multicolored shirt on the signer or a street with lots of traffic behind the signer. Signs can be perceived at different levels. These levels include e.g. prosodic and segmental levels. One example of prosodic level phenomena is the signing rate, which differs from one sign language to another. Phonetic level phenomena include e.g. the selection of handshape, which can be done by stylistic principles (Peters, 2000). In a situation where several different signed messages could be received we can concentrate on perceiving only one – the same phenomenon in speech sciences is called the *cocktail party effect* (Handel, 1989). On the other hand when walking down the street you automatically direct your gaze towards gesturing hands. After head turning has been completed, you verify the interpretation: only then the gestures are distinguished from linguistic signs.

There is no one, international sign language. The signers must have a shared code in order to understand the message. In the same fashion as a Finn must learn to speak English as a foreign language, so must a signer of Finnish Sign Language learn to sign American Sign Language (ASL) or British Sign Language (BSL). While British and American English are dialects of the same language, ASL and BSL are two completely different languages. Thus an ASL signer does not use his or her native language when signing to a BSL signer, but he or she must learn to sign BSL in order to be understood in the British Isles.
The comparison of different sign languages is difficult, because there is no system of transcription equivalent to the International Phonetic Alphabet (IPA), which is used in spoken language research as a means for comparing different sound systems. There are no writing systems for signed languages available either. The two most widely used systems for transcribing signed languages are the Hamburger Notation System (HamNoSys) and the Stokoe notation, although there are many other systems also in use nowadays, some of which are based on pictograms, like SignWrite. SignWrite is mainly developed for everyday use of native signers and not so much for scientific use, although there are some papers written about SignWrite, but they mainly focus on describing the similarities and differences of the different notation systems (see e.g. Martin, 2000). Many notation systems are based on the handshape as the main root for the notation and other phoneme categories: movement, orientation and location are attached as diacritics into the notation (e.g. in HamNoSys).

Research on sign language began fairly late when compared to research on spoken languages. The first systematic studies using linguistic methods were done on American Sign Language (ASL). The pioneer in the field was an American, William C. Stokoe, whose work *Sign Language Structure* was published in 1960. The publishing of this book has been regarded as the starting point for scientific study on sign languages. In his book Stokoe represented a syntactic analysis of American Sign Language and introduced the first wide-spread notation system for signed languages, the Stokoe notation. Stokoe confirmed that ASL indeed was a natural language of its own, and not based on manual coding of spoken languages. Stokoe's pioneering work gave inspiration to several other researchers of ASL such as Liddell and Klima & Bellugi. Liddell studied the syntax of ASL (Liddell, 1980). The research paradigm Liddell used was based on the Transformational Grammar paradigm (Chomsky, 1957). Liddell also studied the differences between the syntaxes of ASL and Signed English, a manually coded form of spoken English. Klima and Bellugi studied the different constraints that the four dimensions of the visual channel place on communication (Klima & Bellugi, 1979; Poizner, Klima & Bellugi, 1987).

Among the first people who described Finnish Sign Language (FinSL) was David Hirn, who was deaf himself. He published the first dictionary of Finnish Sign Language in 1910. It consisted of three booklets. The signs were presented as pictures which had different kinds of arrows drawn on them to denote movements and other parameters of signs. Hirn's arrow drawing method is still used e.g. in Suvi (Finnish Sign Language online dictionary). This dictionary was, however, not a systematic study but a description and a list of words used in FinSL.

1.5.1 Levels of hierarchy in a signed language

The phonological level is presented as an abstraction in both modalities of language (Brentari, 1995). Acoustic and auditive phonetics as a research framework is replaced with visual and optical phonetics because of the difference in the channel of the information flow. The only distinction between signed and spoken languages is thus the medium of information flow (sound/optical stimulus) and the channel of receiving information (ears/eyes). The most important feature is the similarity of the information flow. In both signed and spoken languages the message is coded into the movements of the articulators. The moving articulators are just in different parts of the body. When speaking, the articulators are usually not visible whereas one cannot help noticing/ seeing the signing hands. Most important is expressing meaning with movement (Wilcox, 2001). As the articulator movements are visible in sign language, the Motor Theory of Speech Perception by Liberman and Mattingly (1985) seems well suited for sign language research as well. According to the motor theory the movement of articulators is the main percept and not the resulting sound alone. According to the Motor Theory evolution has transformed the former gestures and hand movements into the movements of the tongue and the vocal apparatus (Condillac, 1947 [1746]).

Articulatory phonology (Browman & Goldstein, 1992) seems to hold well within signed languages too. Articulatory phonology with a more phonetic framework focuses on the articulatory trajectories as the main source for perception. Other recent, more linguistic-style phonological theories that have been used in signed language phonology research include the Movement-Hold-Movement –theory by Liddell and Johnson (1989), the Hand tier model by Sandler (1993), Prosodic Model by Brentari (1998), Moraic model by Perlmutter (1992) and the Phonological Dependency Model by van der Hulst (1992). While these theories do form an important basis on theoretical discussion on phonology of signed languages, this thesis focuses on the phonetic phenomena and behavioural testing of the individuals, both signers and speakers. Thus, the discussion about different phonological theories and their implementation to the data is, however, out of scope of this thesis. The notion for the relations between articulatory trajectories and their respective representations within the mental mapping suffices here.

The basic units of signed languages are optical phonemes, which can be categorized into the following classes: handshape, location, movement and orientation. Furthermore an important part of a one single sign is mouthing, but its status is still under dispute (Savolainen, 2000). Some researchers treat mouthing as a part of the signs, i.e. belonging to the segmental level, others treat them as suprasegmentals. Mouthing can in some instances cover several segmental signs, and in some cases it differentiates lexical meanings. Signs are categorized mainly according to the handshape. A handshape also has allophones. Some of these allophones are due to physical factors, some are dialect-dependent and some allophones are selected according to the signing style. All of these allophones are, however, understood as instances of one and the same handshape. In the dictionary of Finnish Sign Language the signs are categorized according to their handshape. Signs produced with the same handshape are arranged in sequence and additional categorizing parameters are provided by other phoneme categories: place of articulation (POA), movement and orientation.

Handshapes are also clustered into categories using similarity of appearance to the so-called basic handshapes or primes (Klima & Bellugi, 1979). These basic handshapes are most prominent and easy to distinguish from each other and also those which a child learns first when acquiring a sign language. One might say that the basic handshapes provide a cardinal vowel –like system for sign languages. So there are similarity clusters within the handshapes of sign languages, which are depicted in Figure 1.5 with regards to primary component analysis (PCA). This similarity of appearance clustering is similar to the clustering of vowels in speech – there are similar features in vowel groups, such as the front vowels are all articulated in the front of the oral cavity. Handshape space within sign language has similar physical boundaries as the vowel space in speech: fingers and joints have their unique boundary conditions stated by bending and stretching, abducting and adducting of muscles and joints.

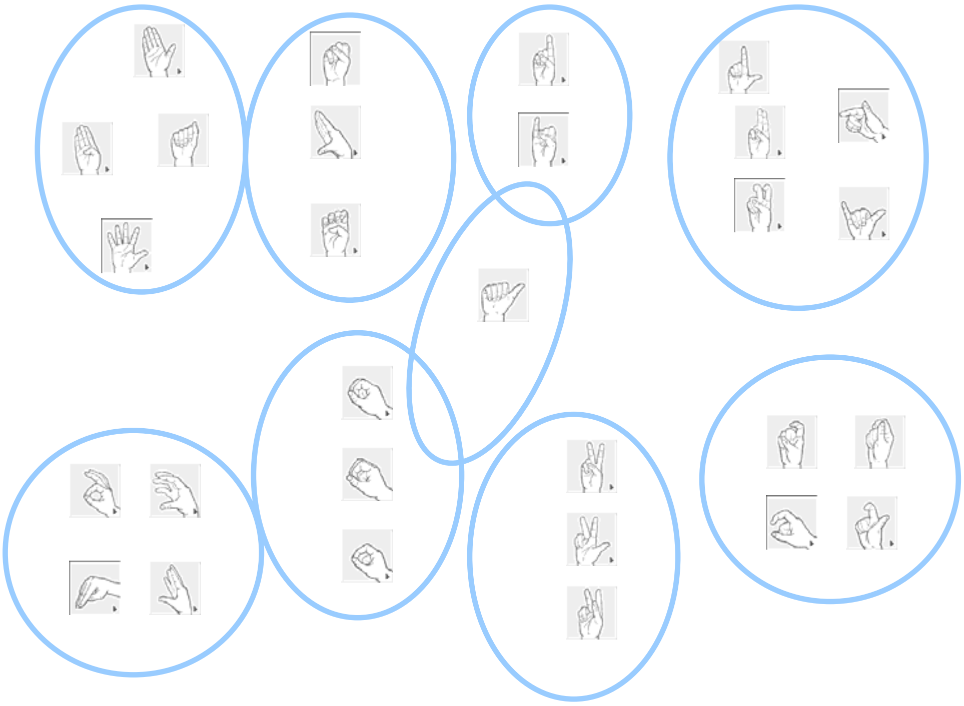


Figure 1.5. The handshape inventory of FinSL with a suggestion for categories. Please note the difficulties within categorization because of multidimensionality of the characteristics.

Some of the handshapes chosen for denoting actions are generally expressed by handling classifier verbs in which the hand configuration depicts how the human hand holds and manipulates the particular instrument. In addition to this, there are instrument classifier verbs, in which the object itself is represented by the articulator and the movement of the sign reflects the stylized movement of the tool or implement. (Gibson & Ingold, 1993; Emmorey et al., 2002 on action and tools; MacSweeney et al., 2002 on spatial processing demands). This is a consequence of the possibilities of sign languages to use iconicity because the articulation is openly visible and thus capable of imitating the actions per se without major requirements on coding visible actions for conveying messages in other media, which has to be done when talking.

1.5.1.1 Places of articulation, movements, orientation, mouthings

Alongside handshape, place of articulation (POA) is the next important phoneme group within signed languages. This is supported by findings that have revealed place of articulation of the signs to be the easiest to remember. Klima and

Bellugi claim that POA might provide key information when associated with serial tasks (Klima & Bellugi, 1974;124). Thus, it might hold a key for sign language acquisition, it being one of the most important features within a sign. POA as a category can further be divided into two subclasses: POAs with and without a body contact. Signs with POAs without body contact are easier to remember as for the memory span than those with POAs with body contact (Emmorey, 2000; pp. 235-236).Their characteristics may differ because of the different spatial stability measures. This thesis focuses on studying the POA class and more specifically, one natural continuum with articulatory boundaries with a constant body contact along the continuum. After division to with/without body contact, POAs are further divided into similarity clusters just as handshapes are. There are clusters of POAs in again perceptually distinct places: face, non-dominant hand and chest area. POAs without body contact are scattered in a larger area while those with body contact are more tightly specified. POAs without body contact are mainly specified by the use of physical planes, which are mutually intersecting and orthogonal. The same system is used in medical sciences and e.g. when discussing brain lesion studies and it is physically based (for references in medical sciences, see Kandel, 2000; for references in sign language studies, see Klima& Bellugi, 1979).

Head	Non-dominant hand	Body contact
forehead	shoulder	upper torso
on top	wrist	waist/legs
eye/ear	palm	no contact
nose	fingers	
mouth	near hand	
neck	side of hand	
in front of		

Table 1.1. FinSL POAs divided into three categories based on primary component analysis (PCA).

POA is most often the anchor point for a sign, it captures the movement within a sign and thus forms a hold unit within a sign according to Movement-Hold phonology (Johnson, 1994). Previous research has suggested that POA might also provide the anchor point for sign language syllables. As with handshapes, POAs also have allophones, so POAs form categories with several different instances. These instances are used interdependently according to coarticulatory requirements determined by boundary conditions stated by surrounding phonemes and their feature sets. POAs in sign also exploit the possibilities given by the 3D space allowed by signing. They are used when depicting locations and start/endpoints in space for different paths of movement.

1.5.1.2 Movements in signs, suggestion for syllable structure for signed languages

Different aspects of movements; e.g. if the movement is slow, fast, unsteady etc. are then again depicted in transitions of POA, and thus lead us to the next category of sign phonetics, namely movement. There are two categories of movement, sign-internal, which is discussed within the phonetic framework here and more general movement patterns, whose use cover all the levels of the linguistic hierarchy, namely from morphology to pragmatics. The phonetic category of movement is, as already stated previously, sign-internal and it has more restricted dimensions in space, i.e. it covers a smaller articulatory space than the more general movements in sign. As movements are multidimensional by nature and multifunctional by their status and place within sign language linguistics they are most difficult to specify and define. Sign-internal movements are, however, crucial to the suggested syllable structure in sign languages, since they are the other category important to the Liddell and Johnson Movement-Hold –theory. The movement category is further divided by the manner, size, direction, contact and hand interaction in two-handed signs. These features are shared for both of the main movement categories. Almost every sign within a signed language has some kind of movement within the sign. Only some of the fingerspelled alphabet and numeral signs can sometimes be articulated without sign-internal movement, but still, there are transitional movements present between subsequent fingerspelled alphabet items. The processing of movement within sign language could be thought as being a special case within a more generalised movement processing system. Movement interpretation seems to be important to online comprehension of sign language, since when the specific movement of the sign is recognised only then can the sign itself be identified. There is thus reason to expect that experience in lexical motion processing also enhances non-linguistic motion processing and indeed, ASL signers do have an enhanced ability to trace nonlinguistic motion, such as writing cao shu (highly artistic way of writing Chinese characters) in those subjects who do not know Chinese writing (Sacks, 1990/1992).

Although handshapes, POAs and movements have established their status within the sign language phonology hierarchy, orientations are still under dispute – some researchers do not give orientation a phonemic status of its own but count it as a parameter of handshape. Here in this thesis, however, we follow mainstream theory and treat orientation as a feature group and thus do not state its status within the phonological hierarchy, since it has not been subject to thorough research yet. The phonological status of orientations seems to be approximately similar to voice onset time, and as such, it might be considered to form a feature set capable of making phonological distinctions, at least in some sign languages. And yet, the status of orientation within the sign language phonological hierarchy is more stable than that of mouthing. At least in FinSL the status of mouthings is still under dispute. Mouthings are further divided into two categories: sign language internal and those which are related to spoken language, for example to spoken Finnish in the case of FinSL. The latter could be contraction of a longer word or a very short word as a whole or just the visemes of the first few sounds.

Thus, the following analogues can be made when considering different elements within sign and speech phonetic hierarchies. The segmental and subsegmental levels within speech do not have but suggestive counterparts in the segmental level within sign phonetic hierarchy.

handshape	vowel
POA	consonant
orientation	e.g.voicing
mouthing	labial
speed, velocity	intensity
facial gesture	intonation
hiding	whispering

Table 1.2. A suggestion for analogic counterparts within sign and speech phonetic hierarchies.

1.5.1.3 3D and durational aspects

The two articulators in sign language (i.e. both hands) are used simultaneously. This feature enables the use of both hands signing different things at the same time, because two-handed signs are understood also when signed one-handed. As an example consider the situation, where one hand signs AUTO ‘car’ and the other signs a pronoun SINUN-OMA ‘your-own’. In addition to these two signs, there is a question mouthing MITÄ ‘what’. This signed utterance is translated in Finnish: *Mitä, onko sinulla auto?* ‘What? Do you own a car?’ [1] The possibility of using all three dimensions in space also enables the signer to paint a scene in which something happens and that in itself incorporates directions and locations into sign language structure more inherently than with speech. In speech again there is a tendency to keep track of the temporal sequences more tightly. This is because speech has more precision in temporal resolution while sign is organized with more complex spatial resolution. Spoken languages do not encode spatial distinctions directly but have developed a range of lexical and pragmatic devices for doing so while in signed languages by contrast space can be used directly for linguistic expression. Signed languages have an ability to manipulate signing space to express both spatial and nonspatial information, as referential use of space employs spatial relations as differentiative for grammatical classes and semantic roles. Space can also be used metaphorically, such as in a situation when higher status is assigned to a higher location in space. There are, however, certain linguistic agreement conditions to parameters: e.g. handshapes for motion and location verbs must agree with the real object features. One example of this is the classificator system. The classifiers are chosen based on the physical dimensions of objects, very much in the same way that the Chinese classificatory system behaves. Similar classifier systems as found in FinSL can be found in various sign languages around the world. Also certain linguistic structures use spatial characteristics of semantic roles and spatial locations topographically. There is a clear distinction in the processing and representing of spatial information as to when space serves a topographical function than when it does not (space for topographic use is different than when used grammatically).

1.5.1.4 Pragmatics and eye gaze

In the dictionary of Finnish Sign Language the signs are divided further according to handedness. Handedness has no phonemic status, as the two-handed signs are understood even if signed with one hand only. There might be situations in which one hand is not available for signing. An example could be that you are holding a baby in your lap or holding a door open to get in, where it is not possible to use the other hand otherwise. In two-handed signs one dominant hand is active and the other non-dominant hand is either passive or active. A passive non-dominant hand is a part of the POA phoneme. An active non-dominant hand moves either as a copy of the dominant hand or as a mirror of it.

Turn taking in sign language is often governed by eye gaze. You tend to break eye contact when you want to keep the turn and not give it away. Eye gaze is very important in sign communication – if eye contact is not present there are no communication possibilities, although there is an exception to this too: tactile sign language, used by deafblind people, but it requires physical contact. In visual sign language the distance between signers defines the size of signing – the further you are the bigger signs you use to make the message as easy to perceive as possible. Here again is a demonstration of human communication following the H&H –theory by Lindblom (Lindblom, 1985).

As to suprasegmental features of speech, intensity (sentence stress) can be directly found also in signed language, although its measurement has been proved to be very difficult because of the variety of simultaneous features expressed with multiple articulators. Intonation patterns are most often expressed by facial gestures: for example in FinSL yes/no -questions are often accompanied with a so-called ”raisin face”: the expression consists of brow frowning and frequently also mouth. Coarticulation functions similarly to speech: segments interact to become as easy as possible to produce and perceive. In this sense both language modalities act according to the H&H (hypo- and hyperarticulation) theory. The opposite forces are thus easiness for the producer (coarticulation) and easiness for the perceiver (intelligibility and explicity) (Lindblom, 1990; Mauk, 2003; Lindblom, Mauk & Moon, 2006).

1.5.2 Neuropsychological studies on sign language and aphasias

Recent findings give additional support to the notion that signed languages are indeed natural full-fledged languages and operate on the same brain processes as speech does. Sign aphasics have similar brain lesions to speaking aphasics and only damage to specific locations in the brain elicit aphasia, so we can conclude, that sign language has the same topographic markers in the brain as the speech loci, because similar brain loci in signers and speakers give rise to similar aphasic patterns. E.g. Corina et al. (1993; also Sacks, 1990, Korhonen, 1993) demonstrated left-hemisphere dominance for ASL but not non-linguistic motion envelopes (waving good-bye etc.). Further evidence can be sought from a case-study involving a patient with dissociative linguistic aphasia, which underlines the difference between linguistic signs and pantomime/gestures. Sign language aphasia is left-dominated just like speech aphasia, while right hemisphere lesions merely affect the discourse abilities and not so much the language structure (Emmorey, 2000). Brentari et al. state that aphasic patients have a breakdown in the phonological level and the Parkinsonian patients have difficulties expressing the phonetic level while at the same time the phonology is unaffected. Furthermore, the sites seem to be very analogous between signers and speakers according to neuropsychological findings and the lesion studies. There is also evidence that neural systems supporting sign language production and pantomimic expression are non-identical. Left hemisphere dominance in language processing is undisputable, but the topographic use of signing space can be impaired also due to lesions in the right hemisphere. However, it might be that the difficulties that RHD patients exhibit in producing and comprehending functions of signing space may stem from a more general problem with encoding external spatial relations into manual representations, particularly when used in bimanual articulation. These are manifested mainly in the so-called "stacking phenomenon" – a more general symptom resulting in agnosia in the contralateral side of the visual field, partly due to neglect (Tompkins, 1997 on more general introduction to the effects of right hemisphere damage to communication; Emmorey, 2000 concerning special issues wrt. signed languages).

2 Individuality in Production and Perception of Speech

Traditionally speech has been understood as being somewhat invariant and not changing except over a very long time. That is, each and every one of us would speak the same way using exactly the same acoustic signals for the same speech sound. This is the mainstream view of the linguistic framework concerning speech. This is also the way of thinking that has allowed us to take up writing – storing something that is by nature online and disappearing.

This chapter of the thesis dwells on the specific characteristics on how individuals learn to speak and how speech can be relearned if the speech organs are affected in one way or another. The view is emphasised that we are individuals – that is there is no standard and invariant way of speaking but the differences between us affect speech in various ways.

It's not only speaking that is individual but also speech perception differs from one individual to another. This in turn, is also affected by the frequent speech sounds we hear in our surroundings. The individuality of speech perception can also be regarded as a natural continuation of the fact that we all speak differently from each other – then it's only natural that we hear differently too.

This chapter is divided into two main chapters by the production and perception studies and then further onto subchapters according to the studies carried out.

2.1 Individuality in Production of Speech

The first part of this chapter takes us through the individual features of learning and the occasionally needed relearning of how to speak. There are many things contributing to that individuality of speech, from the individual anatomical features to variation in feedback – and practice, which becomes important in relearning speech. The first stages of speech acquisition are highly governed by factors we cannot influence. During later stages of speech learning – and especially when re-learning to speak – we can modify the results greatly by practice. The outcome is also affected by individual strategies in articulation.

In this chapter we are dealing with individuality in adapting into different situations – either through acquisition of something new or learning something again. In other words, is individuality something that is present very early on as a set of innate features? This is investigated by acoustic features of pain cry, which is something innate in all of us.

The individual differences have been more widely studied for example within medical sciences (for references see chapters 1.3 for neuropsychological studies, 2.1.1 for maturation and 2.1.2 for innervation). Individual differences are more recognised within speech and language faculties only more recently. This has also emerged studies such as in chapter 2.1.2 where speech sciences and dentistry come together to solve a specific problem with speech in altered circumstances, that is, after a specific oral surgery procedure. The results, in turn, can be used in rehabilitation for the patients if needed.

The further developmental traits for individual learning and adaptation patterns can be observed and measured in the other study, which represents data on individuals with anatomical changes in vocal apparatus, more precisely an intervention in oral cavity altering the physical dimensions for speech production. In this study we notice the individually different learning and coping strategies – which do not necessarily coincide with subjective remarks on success. There are differences in articulation targets, subjective and objective results of articulatory matching and the overall effects to altered circumstances.

2.1.1 Learning to Control Speech Organs – individuality in speech acquisition

There are several factors contributing to the acquisition of speech, some of which are such that we cannot influence them. Others we can modify ourselves through practice and varying movement patterns. In all the first stages of speech acquisition our articulation is highly governed by anatomical and physiological, but individual features. These individual features are present even before speech – they can be found in acoustic studies of cry sounds.

Babies learn to control their movements gradually during maturation. Speaking needs precise control of both the larynx and other speech organs. The larynx is controlled mostly by the central nervous system (CNS). Laryngeal function is the main source of the acoustic properties of cry sound. Thus, the acoustic characteristics of cry can be altered by CNS pathology. Previous studies have found infants with pathological alterations in CNS to have distinctive acoustic features in their cry sounds compared to healthy infants. The altered CNS also affects the latency and duration of the cry with regard to the stimulus presented to the infants along with the fundamental frequency of the cry sound. These disturbances in the CNS can be due to diffuse brain damage, neurological or metabolic conditions, such as prenatal substance exposure.

Along with individual CNS differences also anatomical structures along the vocal tract – and generally throughout the musculoskeletal system – are highly individual (Laaksonen, 2006; Walmsley et al, 1978; Baumel, 1974). Nevertheless, similarities can be observed, especially among relatives with a family resemblance (Hollien, 1990, p. 189). That leads to e.g. similar acoustic characteristics for twins (Loakes, 2008; Van Lierde et al., 2005, Keinänen, 2010) and family members. Some of the results could be explained by the fact that research has been focusing on speech conveyed by the telephone, which has a restricted frequency bandwidth. Fundamental frequency, one of the main characteristics measured in speech, giving the pitch and melody contours of speech, is affected by the length of the vocal tract, and in concordance with that fact, preterm infants have been reported to have a higher fundamental frequency to full-term infants (Michelsson, 1971).

So, the basics for the individual differences in speech are already present in infants and furthermore in the pain cry, which is more controlled by the CNS than other cry types. That by itself provides further evidence on the individual differences in the nervous system control already present in newborns. Just as motor coordination has individual control mechanisms, articulation can also be realised in different ways and still result in similar acoustic patterns in speech.

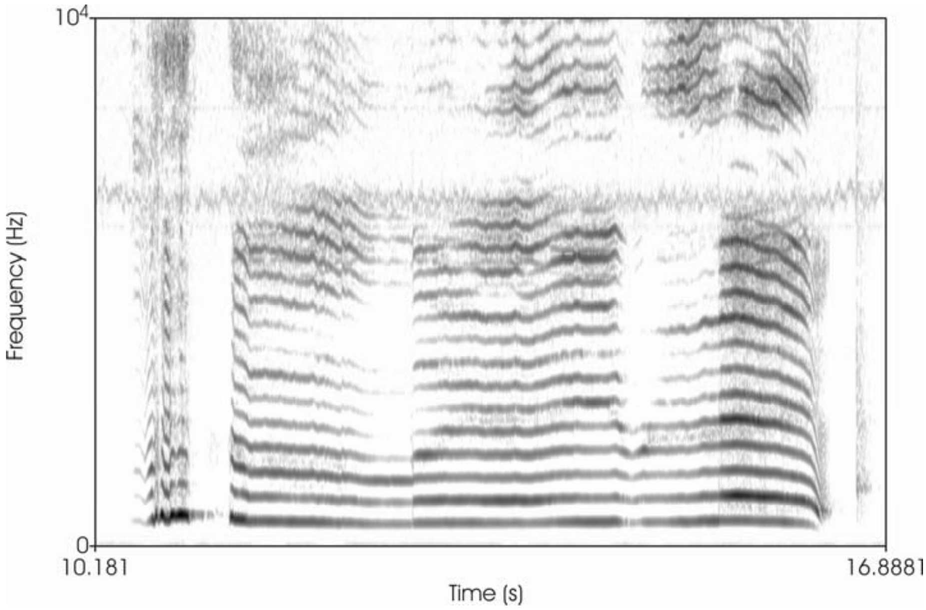


Figure 2.1. A spectrogram of a pathological cry showing a monotonous cry pattern.

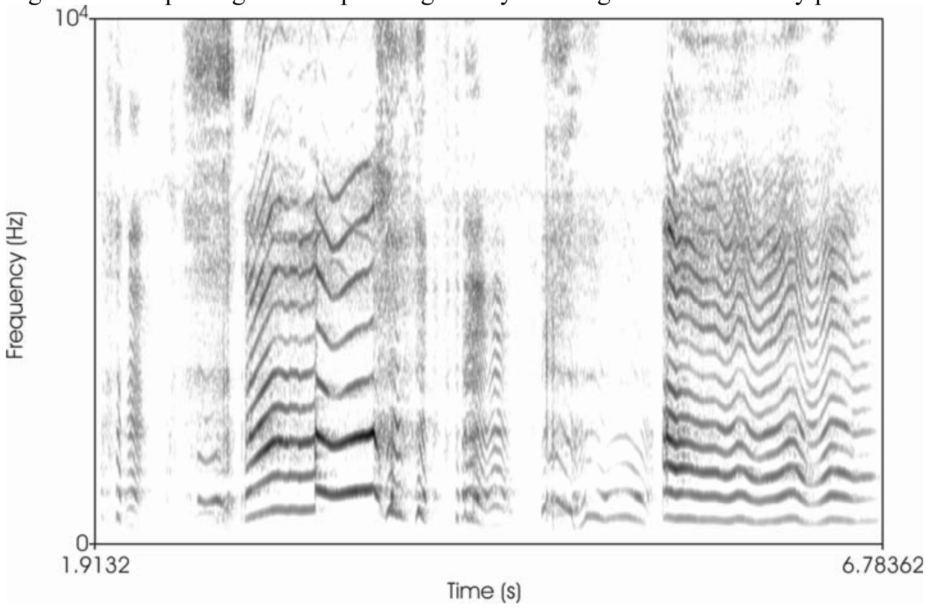


Figure 2.2. A spectrogram of a normal cry showing many different features.

2.1.1.1 Methods

This chapters is divided into 5 subsections. Within the subchapters there are further specifications on: subjects, recording, acoustic analysis, neurological testing and statistical analysis.

2.1.1.2 Subjects

The study group consisted of 21 very-low-birth-weight infants (VLBWI, birth weight ≤1500 g and gestational age <37 weeks at birth), born in Turku University Hospital in 2002 and 2003. The study is a part of a larger multidisciplinary PIPARI study, which was established to investigate the effects of prenatal cerebral blood flow and different NICU (neonatal intensive care unit) procedures on the prognosis of the child, and to measure the VLBI brains with new methods more precisely to further investigate the special characteristics in VLBI brains and the relationships between VLBI and later development in learning, motor skills and schooling. Up till this date the PIPARI study has followed 477 infants from birth to school age.

The study group discussed in connection to this thesis involved 22 VLBWI and 25 healthy full-term infants. 5 VLBWI had normal brain anatomy and 16 did not. These 16 had one or more intracranial findings either in the cranial ultrasound (US) or in the magnetic resonance imaging (MRI). US examinations were performed by a neonatologist in NICU for all study VLBWI at 3 different reference time points during the first month and thereafter monthly during the hospital stay. The detailed US methods used are reported by Maunu et al.(2006). The findings in US were intraventricular hemorrhages (n=7), cysts (n=4) and ventriculomegaly (n=6). Ultrasound imaging was normal for two VLBWI with abnormal findings in MRI. Magnetic Resonance Imaging (MRI) of the brain was performed at term at the same day with the US examination for all study VLBWI. One pediatric neuroradiologist analyzed the MRI findings blinded to both the clinical information and to the result of the US examinations of the infant. For detailed MRI methods see Maunu et al. (2006). The findings seen in MRI were caudothalamic cysts (n =4), various injuries (n =5), hypoplasias (n =2), and ventriculitis (n =1). MRI was normal for 5 VLBWI with abnormal findings in US. That is, 7 VLBWIs had abnormal findings in only one of the two different brain imaging studies while the other showed normal findings. This in part is a demonstration on the fact that both of the different brain imaging studies have to be conducted because they support each other as research methods.

A control group of 25 healthy full-term infants was recruited at well-baby clinics at their appointment for vaccination (with combined vaccine against mumps, measles, and rubella, i.e. MMR) at 1 1/2 years. The MMR vaccine belongs to the Finnish vaccination standard procedure given to 97,5% of Finnish children at that age (THL, 2009). The control group consisted of children who had developed normally, had no major medical conditions, and had their appointments the same day and at the same well-baby clinic with the VLBWI participating in the study. The motor and cognitive development of the controls was evaluated by the physician or the nurse at the well-baby clinic. Exclusion criteria for the controls were the following: (1) use of illicit drugs by the mother during pregnancy, (2) gestational age ≤37 weeks at birth, (3) birth weight more than 2.0 SD below from the mean of Finnish growth charts, (4) any diagnosis of a condition potentially affecting normal development, (5) height or head circumference more than 2.0 SD below from the mean of Finnish growth charts at 1 1/2 years of age, or (6) abnormal motor or cognitive development evaluated by the physician or the nurse at the well-baby clinic. The study protocol was approved by the Ethics Review Committee of the Hospital District of the South-Western Finland. The parents gave written informed consent to participate in the study.

2.1.1.3 Recording and acoustical analysis

We recorded crying after the MMR vaccine was administered at the age of 1 1/2 years by a nurse at a well-baby clinic. The MMR vaccination coverage in Finland is approximately 97% (THL,2009). The duration of the cry response was limited to 30 s. The recording was done using a portable digital minidisk recorder and two condenser microphones with a distance of approximately 10 cm from the infant's mouth for the first 10 samples and 30 cm for the rest of the recordings due to change of the condenser microphone because of function failure. A stopwatch with a beep sound was used to indicate the time stamp for skin perforation by the needle. This provided the reference point for measuring the total duration of crying. The caretaker of the child was asked not to comfort the child during the 30 s of the recording procedure.

One researcher recorded the cries and collected the information about the procedure: the baseline arousal state of the child before vaccination, and site (buttocks or thigh) of vaccination. The baseline arousal state was divided into three mutually exclusive categories (content, not content but not crying, crying). Some of the children received another vaccination after the 30-s recording. As it might have influenced the total duration of crying, the number of inoculations was recorded. The soothing techniques used after the recording by the parent were written down. The background information of medical history, drug abuse during pregnancy, birth weight, gestational age, 5 min Apgar score, current weight, height, head circumference, and an assessment of the current developmental status of the child were obtained from the medical records of the well-baby clinic. The newborn is tested with Apgar procedure, widely used all over the world, two times postnatally: 1 and 5 minutes after birth. The scale of the score is [1 (poor) – 10 (excellent)]. There are five categories in the Apgar test. For further details, please see Apgar (1953).

Cry analysis was performed blinded to the infants' medical history. Acoustic analysis of the cry samples (Table 2.1.) was performed using Praat software version 4.2.05 (Boersma & Weenink, 2003). The sampling rate was 44,100 Hz, and the signal was low pass filtered at 10,000 Hz. The first cry utterance after the vaccination was analyzed. A cry utterance was defined as a cry during expiratory phase with a minimum duration of 0.5 s. Further definitions were based on the continuity of the melody curve and overall spectral properties.

Figure 2.3. Graphical representations of melody patterns within cry samples. The melody patterns found within data are placed on the left hand side and the melody patterns which are possible but not found within data are on the right hand side of the figure.

The cry melody patterns were divided into 3 different groups according to the time points needed to cover the overall shape of the melody pattern. Whenever there was a more complicated time series that was collapsible into a less complicated one according to the intermediate time points needed, the less complicated time series version was used. For the single flat melody type we didn't take into account the different frequency heights, but for all the more complicated melody patterns the beginning frequency was accounted for. In the figure above all the melody patterns found within the given data are depicted with dark lines. We have also added the non-existent melody patterns. They are shown with light lines (Figure 2.3).

A	B	C
F0 min	CoG	shift
F0 max	StD	diplophonia
F0 mean	Skew	shatter
F1	Kurt	glide
F2		vibrato
F3		glottal roll
F4		furcation
		noise
		breaks

A	most frequently used in any acoustic analyses
B	mostly used in acoustic analysis of aperiodic sounds
C	mostly used in acoustic analysis of voice quality

Table 2.1. The acoustic measurements used in this study divided into three groups according to the frequency of use in different types of acoustic analyses.

Furthermore, professional listeners divided the samples into three groups blinded to the medical histories. This was done to find out if there were certain cry characteristics that could be indicators for an abnormal cry and thus would be marked as indicators for prognosis later in life.

2.1.1.4 Neurological testing

The Hammersmith infant neurological examination was given to VLBWI at 12 months corrected age by a physician at the same time in the well-baby clinic. Hammersmith Infant Scale global score ≥73 was regarded as optimal and the scores <73 as suboptimal based on term normative data (Frisone, 2002; Haataja et al., 1999; 2000). The Bayley Scales of Infant Development (Bayley, 1993), was administered to all preterm participants at 12 and 24 months of corrected age by a psychologist to calculate the Mental Developmental Index (MDI) and the Psychomotor Index (PDI). The reaction to an auditory stimulus (turning head towards a sound) was tested for all infants at the well-baby clinics. In addition to this all preterm infants had auditory stimulus reaction tested at 36 weeks of gestational age, at term, and at 1 and 2 months of corrected age. Thirteen of the VLBWI were also tested using brainstem auditory evoked potential (BAEP) measurement.

SAS (version 9.1; SAS Institute, Cary, NC) and SPSS for Windows (version 12.0; SPSS, Chicago, IL) statistical packages were used for analysis. Differences were considered statistically significant if the p-value was below 0.05. The differences between VLBWI and controls were tested using a two-sample t-test for normally distributed continuous variables and a Mann—Whitney U-test for non-normally distributed continuous variables. Analysis of covariance (ANCOVA) was used to further study the group differences in continuous variables adjusted for brain pathology, corrected age, and current weight. ANCOVA analyses were conducted both with and without outliers. A Kruskal—Wallis test was used to compare duration of crying between different arousal states before vaccination. Categorical variables were compared using Pearson's χ^2 test (chi-square test) or Fisher's exact test.

The following patient characteristics were examined as possible predictors of abnormal cry characteristics in exploratory analysis: gender, low 5-min Apgar scores (b5), lengthened duration of ventilator treatment, bronchopulmonary dysplasia (BPD), Bayley scores at 12 and at 24 months of corrected age, Hammersmith Infant Scale rating at 12 months of corrected age as a dichotomised variable, current weight, height, and head circumference. The effects of weight, height, head circumference, gender, and low Apgar scores were studied using the whole sample and separately within VLBWI. Other variables were examined only within VLBWI.

The correlations between continuous cry characteristics and patient characteristics were studied using Spearman's correlation coefficient. Comparisons of two cry characteristics were performed using a two-sample t-test or Mann—Whitney U-test as appropriate. Dichotomous outcomes with at least four events were analyzed using a χ^2 test or Fisher's exact test for the categorical predictors and using logistic regression for the continuous predictors.

2.1.1.5 Results

We looked at the outcomes of the acoustic analyses in three different ways. They all answered different study questions. Firstly, we wanted to know if the VLBWIs were more prone to cry than full-termed controls. The second study question was whether there were certain acoustic characteristics found in the cry sound that would straightforwardly distinguish these two study groups. The third question was whether certain cry characteristics would be more frequent in certain diseases or would co-occur with certain US or MRI findings.

A total of 43 VLBWI were eligible, of whom 22 (51%) were contacted by telephone in time before their appointment for vaccination. All parents of the

VLBWI that were contacted in time agreed to participate in the cry recording. Recording of one VLBWI with normal brain structure was not technically adequate and thus could not be analyzed. The parents of 26 full-term infants were asked about their willingness to participate in the study and only one of them declined. All 25 recordings were performed successfully for the control infants. The corrected age at the time of vaccination was lower in the VLBWI compared to the controls ($p < 0.001$).

The proportion of infants who responded with an adequate cry for acoustic analysis did not differ between groups. The latency could be measured from 13/21 (61.9%) VLBWI and 12/25 (48%) infants in the control group. The baseline arousal state, latency for crying, cry response, and the duration of crying did not differ between groups either. There was no significant difference between the study groups in the number of inoculations ($p = 0.163$) or in the number of soothing techniques after the recording ($p = 0.126$). There was a significant impact of the baseline arousal state of the child on the total duration of crying ($p = 0.032$). Those who cried at baseline cried longer after the vaccination than those who did not cry at baseline.

The minimum fundamental frequency values of the VLBWI were significantly higher than those of the controls ($p = 0.035$). The difference remained significant after adjusting for age, weight, and brain pathology ($p = 0.027$). After adjusting, the F4 values were also significantly higher in the VLBWI than in the controls ($p = 0.030$). No other significant group differences were found in the analysis of the 12 continuous acoustic variables. The adjusted analyses that are presented were done with outliers included. Removal of outliers altered results of F3. When outliers were removed F3 was significantly higher in VLBWI ($p = 0.02$). In the case of kurtosis, assumptions of ANCOVA were not properly met. When comparing the dichotomous variables, there were no significant differences between the groups.

Apgar score <5 at 5 min was associated with shorter duration of cry utterance ($p = 0.003$). The other patient characteristics had no significant effect on the twelve continuous cry variables when examined separately for each background variable in the whole sample. When the effect of patient characteristics on the cry was examined in VLBWI only, low Apgar score was still associated with short cry utterance ($p = 0.007$). In addition, BPD was associated with increasing skewness ($p = 0.025$). High F3 was related to high Bayley Psychomotor Index Scores at 12 months ($p = 0.043$), 5-min Apgar score ≥5 ($p = 0.030$), and currently heavier weight ($p = 0.015$). Heavier weight was also associated with higher F4 ($p = 0.031$) and longer duration of cry utterance ($p = 0.029$). The other patient characteristics did not significantly affect the continuous cry variables. In the dichotomous variables the occurrence of a rise—fall melody pattern was significantly more common in children that were currently heavier ($p = 0.016$)

and the melody pattern was also associated with a larger head circumference (p =0.033). No other significant differences emerged when dichotomous variables were analyzed within the whole study group. Also, when patient characteristics were analyzed within VLBWI only, no significant effects were seen on the dichotomous cry variables. There were 3 infants with cerebral palsy (CP) one of whom also had a hearing deficit. This deficit was currently treated with a hearing aid. The cry patterns of all these three infants were classified as moderately abnormal in a separate classification performed by a researcher blinded to the patient characteristics. Cry samples of VLBWI with or without brain findings were not more often classified abnormal or moderately abnormal than the cries of the controls in this separate classification were.

2.1.1.6 Discussion

This study shows that there are some differences in the acoustic quality of cry between VLBWI and healthy full-term children at the age of 1 1/2 years. The minimum fundamental frequency was higher in VLBWI, and this difference remained significant after adjusting for corrected age, weight, and brain findings. The fourth formant values were also higher in the VLBWI. In addition, cry quality was related to 5-min Apgar scores, BPD, Bayley Psychomotor Index at 12 months, and the current weight and head circumference of the child. We recorded pain cry after MMR vaccination to get as standardised a cry stimulus as possible for all infants. This way we also avoided causing unnecessary discomfort to the child, as MMR is part of the Finnish vaccination schedule. Furthermore, pain cry has been the most commonly studied cry type used to compare group differences in cry acoustics. It has been shown that differences even in the type of vaccine injected can affect cry quality (Ipp et al., 2004). Stress level has been shown to affect cry acoustics through vagal tone (Porter et al., 1988). In our study, the baseline arousal state of the infant could not be standardised and crying as the baseline arousal state prolonged the total duration of crying. As some infants were already crying before the vaccination and all infants did not respond with cry at all, the latency time was measurable for a part of the sample only.

The MMR vaccination in Finland is given at a well-baby clinic visit when the children are close to 1 1/2 years of chronological age. That is why there were differences in the corrected age between VLBWI and control infants in our study. We took this difference into consideration by adjusting the continuous variables as mentioned above. Although acoustic theory suggests that increasing size decreases fundamental frequency (Golub & Corwin, 1985 among others), fundamental frequency has been shown to rise from the age of 0 to 12 months in longitudinal studies (Gilbert & Robb, 1996; Ipp et al., 2004). This rise has been interpreted as an increasing control of cry production as the child grows older. One might assume that the increase in fundamental frequency continues after 12 months of age. In our study, however, minimum fundamental frequency value was higher in premature infants than in controls even though they were younger in their corrected age at the time of the recordings. The effect of prematurity on minimum fundamental frequency remained after adjusting for the corrected age, brain findings, and weight. In contrast to studies of newborn infants (Michelsson, 1971; Michelsson et al., 1982; Michelsson et al., 1983), the maximum or the mean fundamental frequency at 1 1/2 years of age was not related to prematurity or brain pathology.

Patient characteristics had effects on cry quality in our study. Most effects could be seen within the VLBWI only. In a previous study with younger infants, duration of respiratory assistance has been found to be related to cry duration, the occurrence of harmonic doubling and vibrato (Cacace et al., 1995). In our study, BPD had an effect on cry quality but the duration of ventilator treatment did not. We found an association between higher current weight and longer cry utterance duration in VLBWI. However, there was no significant difference between the groups in either the duration of cry utterance or the total cry duration. In a previous study, the duration of phonation in low-birth-weight infants has been found to be longer than in normal weight infants (Michelsson, 1971). Interestingly, the Bayley Psychomotor Index and 5-min Apgar scores were also associated with cry quality. In our study, the only variable that differed significantly between study groups and was also affected significantly by patient characteristics was the fourth formant frequency (F4). Other cry variables that differed significantly between study groups were not affected significantly by the patient characteristics.

Because only one infant had a hearing deficit, the effect of poor hearing on cry was impossible to study further. That would require more testing done for infants with a hearing deficit. This one infant was classified, blinded to patient characteristics, in the moderately abnormal cry group. In earlier studies, cry utterance duration has been found to be longer in hearing impaired infants in comparison to normally hearing infants. In addition, some differences in fundamental frequency and in the number of unvoiced utterances have been found (Clement et al., 1996). The three children with CP were all classified in the moderately abnormal cry group. This study is limited by the relatively small sample size, which restricts the generalisability of findings, reduces the power of statistical methods, and makes chance findings possible. In conclusion, some effect of premature birth on cry quality can still be seen at the age of 1 1/2 years. It seems that the differences in the quality of cry of VLBWI compared to full-term infants relate to prematurity itself, not primarily to brain injury or differences in other patient characteristics. The specific factors in VLBWI causing the differences in cry sound remain unsolved. Potential differences, genetic as well as environmental, in the pathways controlling cry acoustics include several steps beginning from the central nervous system down to the anatomy of the vocal cords.

The cry patterns found within this study varied highly from one individual to another and they were also not so much affected by the pathological status of the baby. Some effects could be found but they were still highly individual. This indicates that as the nervous system is highly individual then also the cry patterns as well as the further development of speech and communication seems to be highly individual too. The individuality thus begins to emerge already in newborn babies reaching its peak during later stages of maturation, both physically and cognitively.

There is a common understanding that when an individual is coherently and repeatedly different from others it is more likely that the explanation would be individual variation and not random noise. On the other hand, a fricative sound /s/ is aperiodic and noise per se, but still there are individual features within that noise (for further discussion on /s/ see chapter 2.1.2).

2.1.2 Re-learning in Speech Production: Effects of transitory lingual nerve impairment on speech: an acoustic study of sibilant sound /s/

Different surgical procedures involving the oral cavity and jaws, e.g. removal of the mandibular third molar, so-called wisdom tooth in the lower jaw, may result in damage to the lingual nerve. The extraction is done because many pathological processes with various symptoms may be associated with the impacted or erupted wisdom teeth including e.g. infections (pericoronitis), odontogenic cysts and tumours, resorption of the adjacent tooth, etc. As a consequence, significant changes may occur in the taste, movement, proprioceptive, tactile and salivatory function of the anterior part of the tongue. The reason why the nerve is occasionally damaged in connection with the extraction is its close anatomical location with the lower wisdom tooth (Behnia, Kheradvar & Shahrokhi, 2000). In these situations, speakers need to modify their normal ways of articulating speech sounds to produce acoustically acceptable speech sounds (Laaksonen, 2008; Niemi et al., 2006, 2004). Previously it has been shown that alterations in acoustic characteristics of speech sounds are observed in connection with a reduction in tactile and proprioceptive feedback mechanisms (Niemi et al., 2002;2004).

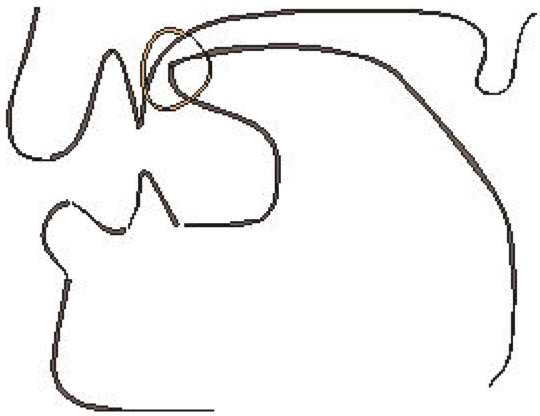


Figure 2.4. A sketch showing the approximate area of constrication in articulation of the Finnish medioalveolar fricative /s/.

The sibilant /s/ requires precise and well-coordinated movements of the tongue in relation to the upper alveolar ridge to produce the turbulences in the oral cavity needed for the sibilant. Compared with other speech sounds, such as vowels and diphthongs, the articulatory manner of /s/ involves fine controlling of the shape and positioning of the tongue blade (Figure 2.4.). Proper articulation of

/s/ requires a precise positioning of the tongue tip near the alveolar ridge, shaping of the anterior tongue blade, and exact placement of the mandible in order to direct the air stream against the incisors through the short midsagittal groove of the tongue. The delicate motor control of the tongue tip is apparently one of the most difficult aspects in speech production. Severe difficulties in articulation of /s/ are often observed in subjects with sensomotor impairments (Hardcastle, 1975). /s/ is also among the last speech sounds learnt by children during speech acquisition, even in Finnish where there is only a single fricative sound in the sound system. In other languages in the world the articulation of /s/ requires even more precision because there are many different fricatives in use (Ojala, 2003).

The white noise in fricative sounds forms no clear peaks that could be extracted. This property makes fricative studies different from those of vowel and other periodic sounds. The different measurement strategies quantifying the overall shape of the spectrum of /s/ include e.g. critical bandwith analysis (Ojala, 2000), spectral shape modeling (Aittokallio, 2002 graphic modeling; Maeda, 1990 vector modeling), defining spectral peaks (Choo & Huckvale, 1997) and so-called spectral moments, which are calculated from a long-time average spectrum (Forrest, 1988; Tjaden & Turner, 1997; Flipsen et al., 1999; Fu et al., 1999). In this study the spectral moments are used to quantify the spectrum. The spectral moments are calculated from the LTA spectrum using a script designed specially for this purpose. The modification used in the script to the one used in the cry analysis was that the signal was high-pass filtered at 1000 Hz to avoid unwanted reverbation from vowel transitions and the electrical hum from the general electric grid at 60Hz. This hum is consistent within Europe, in other continents the grid hum is most often at 50 Hz.

In all likelihood, sibilant production is also more likely to have deviant characteristics compared to the production of many other consonant sounds. It is thus reasonable to expect that speakers who have difficulties with neuromuscular control of their tongue muscles may exhibit deviant control of the delicate movements required for the production of /s/. The effects of reduced orosensory feedback caused by simulated lingual nerve impairment to acoustics and articulation of fricatives were studied by measuring spectral characteristics of the sibilant /s/. Another goal was to examine speakers' capability to compensate for the deviant control of the delicate articulatory movements required for /s/. This knowledge was obtained by asking the subjects to fill out a questionnaire.

2.1.2.1 Methods

This chapter is divided into 6 subchapters. Within the subchapters there are further specifications of subjects, recordings, anaesthesia, acoustic and statistical analyses as well as information on the interviews of the subjects.

2.1.2.2 Subjects

The study group consisted of five healthy Finnish men (aged 24–50 years, mean age 29 years), who reported having no history of speech, language or hearing disorders or difficulties. The study protocol has been approved by the ethical committee of the Turku University Central Hospital, Turku, Finland.

2.1.2.3 Recordings of the material

The recordings were made using a digital audio tape (DAT) recorder and a high-quality condenser microphone in a sound-isolated room. The mouth to microphone distance was 20 cm. The recording level of input was 40 dB VU. The speech samples were recorded from each subject in two sessions with and without local anaesthesia of the right lingual nerve. In both sessions, the subjects repeated the target words in a carrier sentence (‘sano sana _ vielä kerran’ i.e. ‘say the word _ once more’) at their normal speech rate and intonation. The carrier sentence was used to elicit the target word in an unstressed position within the articulatory process. In all, 220 tokens of /s/ were analysed for each subject: before and after anaesthesia 10 repetitions of /s/ in 11 different phonetic contexts embedded in the following eight words (Table 2.2.). The sibilants measured are underlined:

Finnish target word	English translation
ansa	a trap
Aslak	a Lappish name
Israel	Israel
oksa	branch, twig
pyörremyrsky	a hurricane
syyssateessa	in the autumn rain
tupsu	a tassel
veistos	a sculpture

Table 2.2. The Finnish target words and their English translations

2.1.2.4 Local anaesthesia

The local anaesthesia was administered by an oral and maxillofacial surgeon. An injection of 0.8 ml of Ultracain D-Suprarenin (articain hydrochloride 40 mg/ml, adrenaline hydrochloride 5 mg/ml: Aventis Pharma Deutschland GmpH, Frankfurt-am Main, Germany) was administered on the lingual side of the right retromolar region using the standard technique (Haglund & Evers, 1972) (Figure 2.5.).

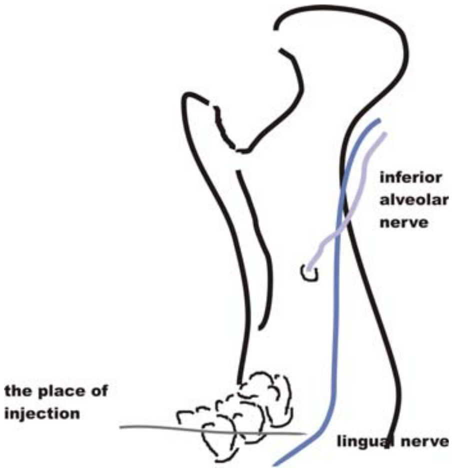


Figure 2.5. The location of anaesthesia administration.

Ten minutes after the subject reported complete loss of sensation, the effectiveness of anaesthesia on the right side of the tongue was controlled by a sharp dental probe. All subjects reported loss of sensation on the right side of the tongue indicating successful anaesthetisation of the lingual nerve alone. The effects of the anaesthesia lasted well beyond the recording session, which took approximately 15 min.

2.1.2.5 Acoustic analysis

Parameters of the Long Time Average (LTA) spectrum (i.e. the centre of gravity, standard deviation, skewness and kurtosis) were analysed using Praat software (version 4.3). LTA was used because of the aperiodic nature of the sibilant sound /s/. The signal was sampled at 22,050 Hz and high-pass filtered at 1000 Hz. High-pass filtering was used to block the possible low-frequency noise generated during recordings of speech samples. A 20-ms Hamming window was used (analysis range 0–10,000 Hz). The spectrum was computed from the waveform using a fast Fourier transform. The LTA spectrum was measured over the entire frication produced (Figure 2.6.). In other words, the initial cursor was placed manually at the onset of the frication and the final cursor at the end of the frication.

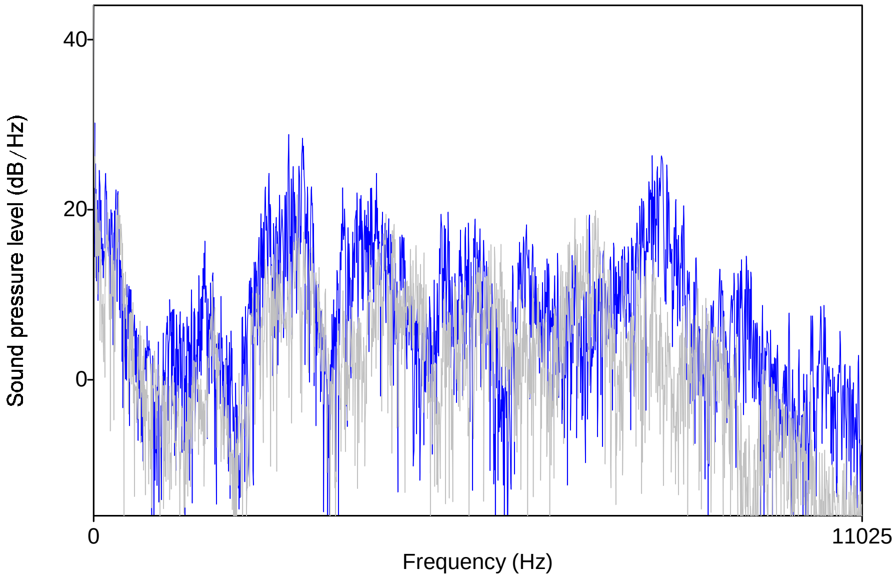


Figure 2.6. LTA analysis of normal and distorted /s/-sounds. Normal /s/ is shown in blue and distorted /s/ in grey line. NB. The distorted /s/ has a lower CoG.

2.1.2.6 Statistical analysis

Statistical tests were carried out in order to compare the four acoustic parameters studied before and after the anaesthesia. The values for each subject were averaged to calculate subject mean values for centre of gravity, standard deviation, skewness and kurtosis. A repeated measures analysis of variance (ANOVA) with time (before versus after anaesthesia) and context type (11 levels) as within participant factors was used to investigate the effects of reduced somatosensory function of the tongue on the acoustic parameters. In addition, non-parametric t-tests (the Wilcoxon matched-pairs test) were used to compare the values of normal and ‘anaesthetised’ conditions in individual participants.

2.1.2.7 Interview of subjects

After the second recording session, the subjects were interviewed orally and they filled in a written questionnaire to determine if they noticed any changes or anything abnormal in their speech production.

2.1.2.8 Results

Averaged data of the acoustic parameters of /s/ produced in the carrier sentence and in different phonetic contexts by the subjects are presented in Table 2.3. At group level, there were no significant changes in any of the spectral characteristics (the centre of gravity, standard deviation, skewness and kurtosis) between the measurements before and after the anaesthesia ($p > 0.05$). The effect of phonetic context was significant for the centre of gravity [$F(10,40) = 3.450$, $p = 0.002$], for the standard deviation [$F(10, 40) = 4.719$, $p < 0.0001$] and for the skewness [$F(10, 40) = 2.241$, $p = 0.035$], but not for kurtosis. The interaction between the time and the phonetic context was insignificant ($F < 1$).

S1diff	S2diff	S3diff	S4diff	S5diff	
531,36	**–335,86**	**–393,87**	–70,82	**238**	CoG
304,15	**–170,70**	**–288,40**	22,17	**198,98**	StD
–0,582	**0,505**	0,18	0,377	**–0,3704**	Skew
–2,192	**2,249**	0,3526	3,507	**–1,379**	Kurt

Table 2.3. Averaged data of the acoustic parameters over different carrier sentence contexts, differences between results before and after the anaesthetisation of the tongue. The statistically significant results are written using a bold font face.

The individual results of the Wilcoxon matched-pairs t- tests showed statistically non-significant changes across the measurements. For subject 1, centre of gravity decreased significantly ($W = 4175$; $p < 0.0001$; mean difference, 244 Hz). Skewness ($W = -2016$; $p = 0.002$; mean difference, 0.3) and kurtosis ($W = -3424$; $p < 0.0001$; mean difference, 2.1) increased when all the contexts were included. No significant changes were observed for standard deviation. The most affected context was ‘after bilabial stop /p/’ (in the word ‘tupsu’). This subject reported that he had to concentrate more on articulation during the ‘anaesthetized’ session compared to the ‘non-anaesthetized’ session, particularly when producing the /s/ sound. For subject 2, no statistically significant changes were found on any parameters analysed ($p > 0.01$). He did, however, report difficulties with articulatory movements on ‘the anaesthetized side’ of the tongue. He had articulated mainly on the unaffected side. He also reported frequent fatigue of the tongue and difficulties, in particular, with production of the /s/ sound during the second, i.e. the anaesthetised, recording session. For subject the 3, the centre of gravity (W = 1907; p = 0.0045; mean difference, 125 Hz) and the standard deviation (W = 1793; p = 0.0076; mean difference, 73 Hz) decreased. No significant changes were observed in skewness and kurtosis. In the context-specific dissection, no particularly sensitive context of significant changes was observed. Subject 3 reported having no speech difficulties at all during the anaesthetized situation. For subject 4, centre of gravity decreased (W = 1625; p = 0.01; mean difference, 268 Hz), skewness increased (W = 2652; p < 0.0001; mean difference, 0.6) and kurtosis increased (W = -2681; p < 0.0001; mean difference, 2.5). No significant changes were observed in standard deviation. The most affected context was 'word-initial position' (i.e. in the word 'syyssateessa'). In this context, centre of gravity decreased (W = 49; p = 0.0098; mean difference, 921 Hz), skewness increased (W = -49; p = 0.0098; mean difference, 1.3) and kurtosis increased (W = -53; p = 0.0039; mean difference, 4.6). He reported no difficulties in speech production. For subject 5, centre of gravity increased (W = -4542; p < 0.0001; mean difference, 705 Hz), standard deviation increased (W = -2252; p = 0.0008; mean difference, 70 Hz) and skewness decreased (W = 3964; p < 0.0001; mean difference, 0.5). No significant changes were observed in kurtosis. The most affected context was 'before dental stop /t/' (i.e. in the word 'veistos'). In this context, centre of gravity increased (W = -55; p = 0.002; mean difference, 1377 Hz), standard deviation increased (W = -51; p = 0.0059; mean difference, 191 Hz) and skewness decreased (W = 55; p = 0.002; mean difference, 1.0). This subject reported that in the beginning of the second recording session the numbness in the tongue resulted in articulatory difficulties, but with time he adjusted to it. He also experienced fatigue of the tongue and reported difficulties pronouncing words containing the /s/ sound.

2.1.2.9 Discussion

The effects of reduced orosensory feedback on phonetic quality and the production of fricative sounds were studied by measuring the spectral characteristics of the sibilant /s/ under normal and reduced conditions. In the latter situation, the tactile information from the tongue was reduced by blocking the right lingual nerve with local anaesthesia. A variety of target words in a carrier sentence was used to provide information on the effects associated with different phonetic contexts. This was done to investigate whether different phonetic contexts were more difficult to produce than others. The four spectral characteristics studied were selected because these parameters have been confirmed to be the most suitable for qualifying the overall shape of the spectrum of /s/ by the concentration (centre of gravity), dispersion (standard deviation), asymmetry (skewness) and peakedness (kurtosis) of energy distributions (Forrest, 1988).

The acoustic analysis showed that reduced tactile sensation has effects on tongue function, resulting in spectral alterations of sibilant /s/. This is due to the decreased accuracy of the speakers' ability to constrict the vocal tract. The speakers presumably have difficulties in adequately maintaining the fixed position and shape of the tongue blade. Distorted function of the tongue caused by the loss of afferent tactile information produces an abnormal obstruction to the airflow normally required for sibilant production, thus modifying the turbulence. The small cavity in front of the constriction would therefore be altered causing deviant acoustic output. Speakers would have to find a way to modify this acoustic production. They are likely to do this by moving their tongue blade to a more posterior position as indicated by lower values for centre of gravity observed in three out of five subjects. For production of /s/, subject 5 used a tongue placement that is further forward than normal, thus producing a smaller resonating cavity anterior to the point of constriction and higher values for centre of gravity. In general, the changes were individual and variable, ranging from slightly affected to completely unaffected.

The variation in the results among the different speakers indicates individual ability to compensate for the effects caused by sensory dysfunction of the tongue. The new articulatory movements did not completely compensate for the effects of distorted tongue function in some subjects in the anaesthetized condition. In other words, the loss of sensory information has an individual effect on the speaker-specific strategies used during adaptation in the altered situation. The large variability among different speakers may also reflect other speaker-specific factors such as vocal tract morphology, the type of information supplied via the lingual nerve (tactile, proprioceptive or both) and the effects of local anaesthesia itself (Borden et al., 1973; MacKenzie, 1997; Zemlin, 1998). It is thus possible that both individual structural and functional differences are operating concurrently to generate the observed changes associated with lowered accuracy.

Speakers' opinions were obtained in order to compare their subjective experiences related to the production of /s/ during the anaesthetized session and the results found in the acoustic analysis. In general, this comparison gave somewhat contradictory results. For example, subject 2 reported difficulties in producing /s/, but no significant changes were found in the acoustic analysis. This person seems to have a good ability to compensate the altered function of the tongue. This further emphasises the fact, that however difficult one might consider a difficulty within pronunciation and articulation it might not be audible to others. On the other hand, even though subject 4 reported no problems in production of /s/ there were statistically significant changes within the acoustic parameters when compared before and after the local anaesthesia administration. These findings are in concordance with the quantal theory of speech production by Kenneth Stevens, who postulates that there are certain areas within the oral cavity which produce identical or nearly identical acoustic signals despite changes in articulation for one reason or another (Stevens, 1988; 2000). The individuality within the results obtained indicates the intra- and intersubject variability within the acoustic data. Supportive evidence for the individuality of speech production is present from very early in life, as can be seen in the acoustic characteristics of infant cries as discussed earlier (see chapter 2.1.1 above) in this thesis.

2.2 Individuality in Speech Perception

In this chapter the thesis continues to explore the theme of individuality. Now the focus concentrates on speech perception and especially on two different aspects of speech perception: the perception of vowels and consonants. As previously shown in Chapter 2.1 within this thesis the production of speech is highly individual and not invariable across individuals. Now the focus turns to the perceptual side of the speech perceived – how we are able to perceive this variety found within speech into a unified understanding of what is said, how the information structure is structured and transformed into the signals.

Perception process doesn't solely consist of classification and division to categories but also includes the ability to perceive category internal differences. These category-internal differences can be referred as individual features. The focus of this research is to study the individual phonetic differences, not the perception of the phonological classes as a system.

This chapter divides into two subchapters investigating two different phenomena in speech perception and the individual differences within them. Both of them further divide into subsections accordingly.

2.2.1 Individual Vowel Categories and the Role of Different Spectral Attributes

Individuality in speech perception has widely been studied previously, but here the focus turns to the relations between perception and categorisation strategies. These strategies are used when there are difficulties in the categorisation process.

Thus far the results within this thesis have shown that speech production is highly individual (studies on infant cries and fricative production). This individuality is already present from very early on in the human development and is partly due to differences in innervation and articulatory processes. There seems to be also individual differences on the compensation aspects to articulation due to various changes in articulation. This part of the thesis investigates the individuality on the perception of speech. As individual differences are present in speech production the hypothesis is that speech perception is highly individual too. This is tested by vowel and syllable listening experiments. Different languages divide the vowel space into various vowel sound categories. The number of these categories differs from one language into another (e.g. Ladefoged & Maddieson, 1996). The Phonetics Laboratory in University of Turku has developed a software for investigating the differences in number and inner structure of vowel categories, namely the Turku Vowel Test, which consists of 386 synthesised vowel stimuli covering the entire vowel space excluding nasals and diphthongs (for further specifications on the architecture of Turku Vowel Test see Savela, 2009).

In this study we aim to investigate how people categorise vowel space according their own language and furthermore, we enlarge the scope into consonant perception by experimenting with consonant continuum. In previous studies vowels in perception have been discovered to induce results demonstrating continua, i.e. that vowel perception is continuous and not so categorical. In other words there are clear-cut boundaries between different vowel categories, but the inner structure of a vowel category resembles that of a mountain in a topographic map: there are isographs of perceptual similarity within the vowel category (Aaltonen, 1997; Rosner & Pickering, 1994 on perceptual prototypes). This has been referred to as perceptual prototype within vowel sounds. According to Kuhl (1997) among others, the prototypical token, or the group of near-prototypical tokens, acts as a magnet for the other non-prototypical vowel sounds and that forms a “hot-spot” within that vowel sound. Thus, the vowel space is not evenly distributed but has density differences along the different dimensions of vowel space. The more recent findings also have provided evidence, that context can shift the category boundaries, but the prototypes are more constant. This has been revealed in so called roving context tests (Iverson & Kuhl, 2000). The basic question in speech perception studies is to understand the role of the different spectral attributes which contribute to perceptual similarity between two sounds. This is crucial in the interpretation of the patterns observed in human performance based on these similarities along the different physical stimuli continua and how speakers of different languages use their knowledge of their native language to categorise the identical continuum differently by number of categories along that continuum and the specific acoustic parameters defining the inner structure of categories. That is: what are the acoustic parameters defining the types of similarity the listener uses if asked to identify the stimuli or to estimate its goodness with reference to category. After gathering knowledge on the used acoustic parameters in different languages we are able to predict the behaviour when presenting people with stimuli they have not perceived before.

There has been an extensive debate within the vowel identification studies framework on the discrepancy between the whole spectral attributes on one hand and formants on another because they provide different information about the physical similarity of the stimulus. Rosner (1994) states that the discrepancy may be related to the difference between general acoustic distance and phonetic distance between the vowel stimuli (also Carlson, Granström & Klatt, 1979). The general view has been that picking formant peaks would be enough to explain the attentive vowel identification responses. More in detail, Savela et al. (2003) demonstrated that the formant peak picking and the general acoustic distance are in fact related to the two different perceptual processes: automatic and post-perceptual respectively. The automatic process uses intuitive knowledge on the salient information within the stimulus, i.e. the formant peaks. The whole spectrum approach information was shown to be post-perceptual language independent general information about the similarity of two stimuli. This knowledge was used when the salient feature mapping was not providing enough information for categorisation process (ibid.).

The aim of the present study is to use comparison of the identification and goodness rating data to show the role of formant based similarity and the whole spectral based auditory similarity between the synthetic stimuli, using familiarity based indexical (goodness ratings) and analytical symbolic strategies. The idea is to identify the attributes sufficient to explain the identification and goodness ratings adequately. The ultimate goal of the present study is to show that vowel identification and goodness rating are based on different spectral attributes in terms of vowel space. The chosen vowels are Udmurt /ɨ/ (unrounded close central vowel) and /u/ (rounded close back vowel). They both cover the same area of the F1-F2 vowel space. This has been revealed as the most challenging area of the vowel space when using of the formant based models. In this particular area of the vowel space the F1, F2 and F3 are distinct and no trivial fusion of the formants can be observed.

2.2.1.1 Methods

The whole test set consisted of 386 synthetic vowels which covered the entire vowel space except for diphthongs and nasal vowels (Figure 2.7). The stimuli were synthesized with a Klatt parallel synthesizer (Klatt, 1980). The vowel space was created by varying F1 from 250 to 800 Hz in steps of 30 mels and F2 from 600 to 2800 Hz in steps of 40 mels. F3 was 2500 Hz if F2 was ≤2000 Hz and it was increased by 200 mels above that. The duration of the stimuli was 350 ms. The pitch within the stimuli rose by 20 Hz from 100 Hz to 120 Hz during the first 120 ms of the stimulus and fell thereafter by 40 Hz to 80 Hz during the rest of the stimulus. The pitch curve was interpolated automatically between the time reference points. The amplitude of the individual formants was not damped, which can lead to higher amplitudes in higher formants. All the stimuli were played back one at a time and the whole test set was self-paced by the subjects with an option to hear the stimuli more than once.

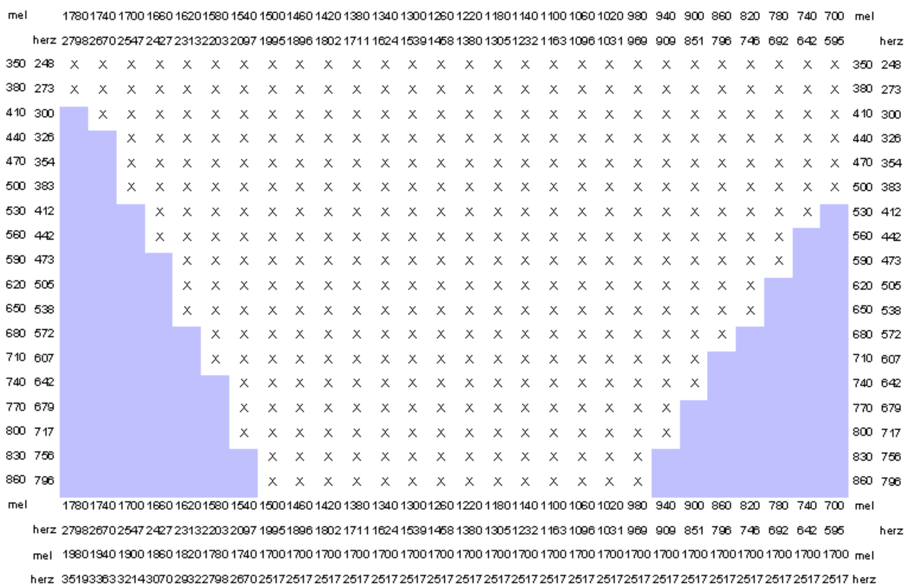


Figure 2.7. General overview of the synthesised vowel space within TVT.

The study group involved in the Udmurt vowel study consisted of six subjects (mean age 28,5). They were asked to identify the vowel (using Udmurt orthography) and thereafter to evaluate its goodness as a member of the identified vowel category (range 1 – 7). The test for the whole database took about 45 minutes. The stimuli were delivered through headphones in a sound insulated room at the University of Turku Language Centre.

2.2.1.2 Results

The results obtained were clear but twofold. In identification data the area of /u/ was divided into two clearly distinct and separate areas divided by /ɨ/ in between whereas the goodness rating data showed only one peak area of high ratings for /u/. This discrepancy between identification and goodness rating data provides evidence against the existence of linear prototype-based categorisation of the vowel stimuli. The prototypes within the identification data might be in some instances also based on the articulatory trajectories. The differences between subjects were systematic and stable. The identification of the category can not be explained by the two lowest formants of prototypical exemplars in some subjects' data and thus an alternative, additional explanation for the finding has to be found.

The four lowest spectral moments in the signals were analyzed using a Praat software script especially designed for this purpose. The basic input analysis for all the stimuli was FFT spectral analysis. The sampling frequency used was 11024 Hz. The spectral moments script used a power spectrum where the magnitudes of the spectral components are squared. This procedure is widely used in fricative studies and has been suggested by e.g. Forrest et al. (1988). The use of spectral moment analysis might give an explanation for the discrepancy of identification and goodness rating results in this case.

The centre of gravity describes the mean frequency of the spectral components in signal, that is, a point in the frequency scale where the energy distribution of the spectrum adds up to the same amount on either side of that point. The standard deviation describes the dispersion of the spectral components around the centre of gravity. It is also defined as the square root of the second central moment of the spectrum. It describes how much the spectral components differ from the centre of the gravity. The (normalized) skewness measures the asymmetry in the shape of the spectrum between the lower and higher areas in the vowel spectrum. The (normalized) kurtosis describes how peaked the spectrum is compared to a Gaussian distribution.

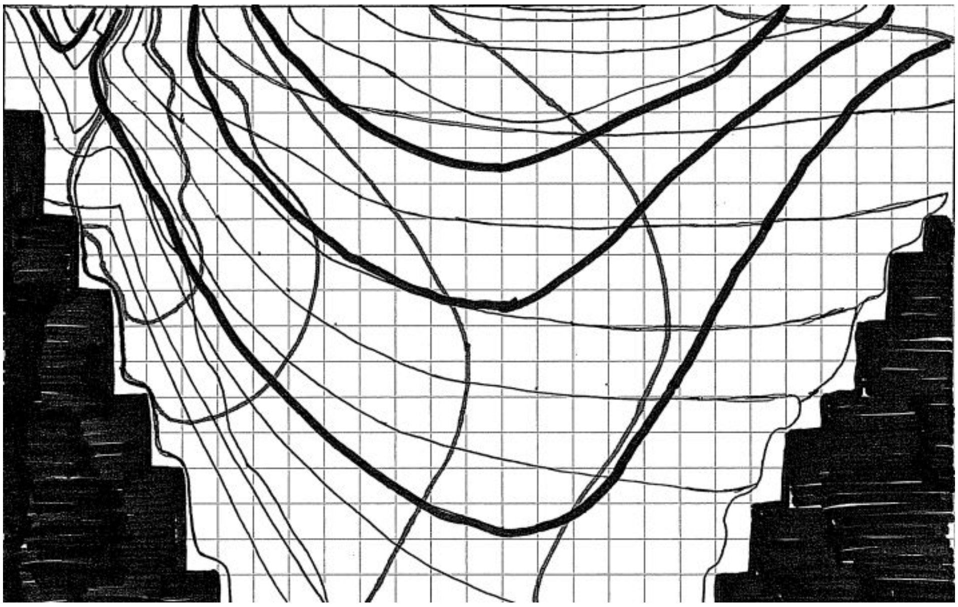


Figure 2.8. The figure shows the relations between different spectral moment function isocurves within the vowel space used in this study. Please note that different widths and colours denote different spectral moment function isocurves. (CoG grey, thin; StD grey, thick; Skew black, thin, Kurt black, thick). The grid denotes the synthetised vowel stimuli within the Turku Vowel Test.

The results show that the identification of vowels was not predicted by formants alone. The subjects were, however, able to assign goodness ratings to stimuli using only formant information within the stimuli. For goodness ratings F1 was more important than in identification. For identification the importance of skewness and kurtosis was more significant. The results showed that the identification and goodness ratings were based on different acoustic criteria. Formants did not always provide enough information for identification in this case, but the subjects gained further information for identification from spectral moments (Figure 2.8). When the identification of the category was fixed, then it was possible to give goodness ratings based on formant information only.

2.2.1.3 Discussion

The present study was designed to show that the goodness ratings and identification of the vowels use different spectral attributes in the vowel space. The results argued against the use of an identical model of predictive acoustic parameters for goodness rating data and identification data. That is, two different models have to be used to explain the discrepancies between the identification and goodness rating data obtained from Udmurt subjects. The formant based similarity does not fit the identification functions of the Udmurt data. It was shown that identification might be based on a post-perceptual process different from goodness rating. The plasticity of the general acoustic features can lead to areas of similarity that are not similar to the proto-typical exemplars of the category areas, but are nevertheless still categorized as belonging to the same category. The general acoustic features can lead to the areas of similarity that are clearly separated from the area adjacent to the prototype of that vowel. The present study showed that many of the questions about the areas of prototypical vowel tokens and their relationship to the rest of the category can be elaborated using an auditory model in which spectral moments are added to the model. Results here are contradictory to Rosner and Pickering's theory in which the vowel categories are always linear in terms of the formant space model (1994). That theory does not explain the identification data within this study. Additionally, this study provides more information about the perception of back vowels in general and furthermore, on the relationship between open and closed back vowels. It might even be, that the perceptual similarity between /u/ and /y/ in some languages may be related to the listeners' habit of using spectral moment information in addition to the formant peak picking, especially if synthetic stimuli covering large areas of the vowel space are used. In conclusion, to obtain a more comprehensive description of vowel categories, the person-independent formant tracking model has to be added and augmented with spectral moment information dependent on person and culture.

Thus, the same theme of individuality seems to govern speech perception too. Perceptual patterns vary from one individual to another but at the same time are near enough to each other to be able to negotiable and flexible enough to adapt to each and every individual's production patterns that we are mutually intelligible when engaged in a conversation. So, since speech production starting from very early in infancy is individual, then speech perception must also be adaptive enough to decipher every production pattern into an intelligible vowel space, there being no empty spaces or "black holes" within it.

2.2.2 Consonant Perception is Different from Vowel Perception

As speech doesn't consist solely of vowels, the perception of consonants has to be taken into account too. Here we show that their perception differs from vowel perception. In the previous chapter we showed that the perception of vowels is highly individual. Categorical perception as such is not specific to speech alone, but is a general phenomenon, which has been studied in various fields of human perception, such as in object (Biederman, 1984) and vision (e.g. Palmer, 1999) perception. Ability to categorise things develops very early on in infants and the basics of the categorisation ability are also present in other species (for a review, see e.g. Bee, 2006). Previously speech researchers have found that vowels and consonants behave differently in perception tests (see e.g. Kuhl, 1994; van Alphen & Smits, 2004 among others). More precisely, vowels tend to be categorised continuously, that is without any clear-cut boundaries, as shown also in the previous section with the Turku Vowel Test battery. Vowels also have an inner structure in the category, which is demonstrated by the individual's ability to rate the goodness of the different vowel stimuli (Aaltonen, 1997, Savela et al., 2004, among others). Consonants, more specifically plosives, are often stated to be more categorical, that is, they have clear-cut boundaries and have no inner structure in the category (Liberman et al., 1967). Consonants are mainly described by their place of articulation, which is the place of the constriction within the oral cavity. One example of places of articulation, namely that of /s/, is shown in this thesis on p. 42. In order to study the perception of different places of articulation in speech, consonant continua have to be used as stimuli. Consonants and especially plosives as such, however, are not suitable for stimuli by themselves, but they have to be incorporated within syllables in order to be heard and only then can they become subject to categorisation. That is already obvious from their name: consonant, which originates from Latin word *con* + *sonāre* with sound, to sound at the same time.

2.2.2.1 Methods

Speech stimuli were constructed with the Klatt synthesizer HLSyn based on the Klatt cascade synthesis method. The cascade synthesis method was used, because it allowed us to model the transitional phase of the vowel more accurately. The continuum consisted of 6 different syllable stimuli along the continuum /pa/-/ta/. Thus the syllables had a CV structure. The stimuli all had a different onset representing different stops while the steady-state vowel was identical in each of them. The onset consisted of a burst followed with a transition. The transition had duration of 30 ms. The transition phase was divided into a burst and onset of 30 ms. The steady state of the vowel began after the transition phase. For the steady state vowel, F1 was 750 Hz, F2 was 1050 Hz, F3 was 2450 Hz and F4 was 3500 Hz. The different stimuli had different formant patterns. The initial F2 at burst of syllable onset varied from 800 Hz to 1300 Hz (six steps of 100 Hz). The initial F3 of syllable onset varied from 2200 to 2700 Hz (six steps of 100 Hz). Spectrograms of the endpoint stimuli in the continuum are depicted in figure 2.9.

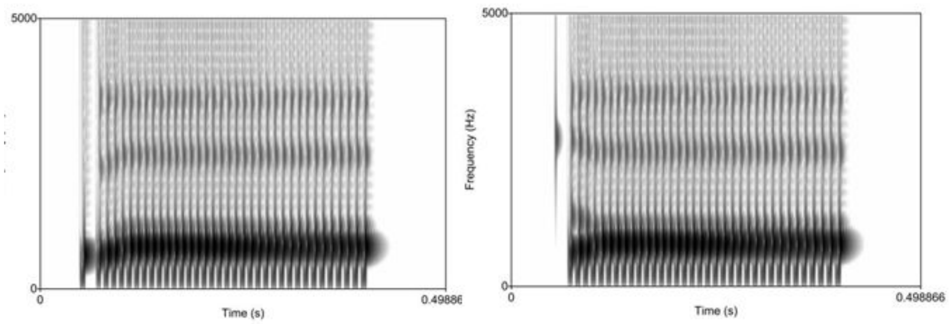


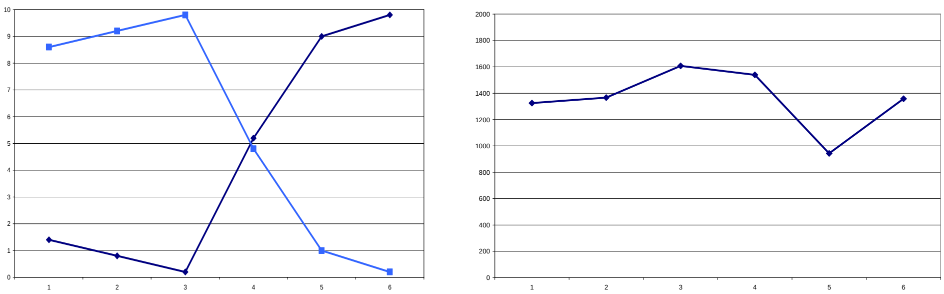
Figure 2.9. The endpoint stimuli of the CV continuum as spectrograms. Please note of the transitions concerning second and third formants when changing from /p/ to /t/.

The study group of listeners consisted of 8 volunteer subjects, all of whom were enrolled for the Finnish Sign Language study as members of the two hearing groups. Therefore, none of the subjects had any reported hearing deficits. Their hearing status was not clinically checked. The subjects were asked to identify the syllable and thereafter to rate its goodness as a member of the labelled category. The identification reaction time was also measured. The stimuli were delivered through passive noise-eliminating headphones in a regular office setting using a laptop computer.

The test took about 15 minutes and consisted of 10 repetitions of each of the 6 different stimuli in randomised order. Answers were given by pressing the keys on a laptop computer. The test was self-paced, but had a limited time window of 25 s for answers to each stimulus, so the maximum time for the completion of the test was 15 minutes including the interstimulus intervals. The test also had a demo version, which was used to familiarise the subject with the procedure if the syllable identification test was the first test in the battery that the individual participated in.

2.2.2.2 Results

All responses demonstrated highly categorical results, where the boundaries between the two given consonant categories were very steep and sharp. The place of the category boundary was in all subjects very similar for all subjects, it only varied between two stimuli, namely stimuli 4 and 5 within the 6-step continuum (Figures 2.10a&b.). The steepness of the category boundary did not vary but was always steeper than those slopes obtained from the vowel studies previously in connection with the Turku Vowel Test project. Both the Turku Vowel Test continua and the continuum in this study were always between two perceptually distinct categories with no intervening ones possible. There was also a steep rise within the response time duration in the category boundary area. All of these results are parallel to previous results on consonant perception studies.



Figures 2.10a&b. showing the averaged results over listeners, a shows categorical boundary between stimuli and b shows the reaction times. In figure 2.10a lighter colour denotes answers [pa] and darker [ta] answers respectively. In figure 2.10b scale is vertical axis is RT in ms. Since these figures are drawn over the individuals as one group, both the category boundary and the reaction time peak are thus smoothed from individual response data.

2.2.2.3 Discussion

The steepness of the slope and the lack of variation within the category boundary area suggest that we do not hear gradual changes of consonant categories, which would coincide with the gradual changes in the acoustic signals but there are quantal jumps in perception from one category to another. This has been the first suggested basis for speech being a special code (Liberman et al., 1967; for a review, see Liberman, 1996). So, even though we are able to discriminate between different consonant stimuli we are able to ignore their differences as long as they belong to the same category. These results also coincide with those obtained from articulation studies stating that there are similar quantal jumps in the acoustics, where there can be quite dramatic changes within articulation but which are not reflected in the acoustic signal (Stevens, 2000). This is especially true for those speech sounds mainly defined by their place of articulation, i.e. the consonants. So, to some extent both the sending and receiving parts of human communication are quantal in nature. The strictly quantal nature of categorisation has been previously found in especially with synthesised plosives (Liberman et al., 1961; Bastian et al., 1959; 1961) and to a lesser degree with synthesised voiced consonants (Eimas, 1963; Studdert-Kennedy et al, 1964).

The nature of consonants is different by nature, since consonants do possess a contact or nearly so between two articulators, for the most part those are the tongue and the upper jaw, i.e. the maxilla, but not always. An example of this would be a bilabial stop /p/, where the lips are the articulators making the contact. /p/ is also included in this study as one of the endpoints in the continuum.

2.3 Individuality of Speech

In this chapter we looked at speech and more precisely at how individuality is represented in both production and perception of speech from various viewpoints, each discussed in separate chapters and subsections respectively.

We started with very early and basic productions of an infant, namely pain cries in connection to vaccination procedure. At first thought, it might be considered as such a basic and innate reaction to pain only governed by the automatic CNS that it wouldn't have individual features in it. However, when studied in more detail, we found the pain cry utterances variable both in duration and inner structure. This suggests that individual differences and features are present very early on in human development. In the next section we looked at adult speech production in normal and altered conditions as a result of an altered nerve function or a simulation of it. This gave us knowledge on the adaptability of the vocal apparatus, the learning strategies and adaptation to changes in articulatory parameters. These results suggest that adaptation and learning processes in an adult are highly individual. Furthermore, the subjective and objective outcomes of the procedure can vary from each other. These results also connect to the speech perception studies, since some of the altered speech sounds were caught by acoustical measurements but not audible for a human listener.

From studies on speech production we moved on to investigating individuality of speech perception. First we looked at vowel perception, that is, how do individuals perceive different synthetic vowels. The vowel space we used covered the synthesised versions of oral vowels in world's languages in F1-F2 plane excluding nasal vowels and diphthongs. We found out, that individuals do not only perceive vowels differently but also use different kinds of strategies in perception which relate to the analytical and holistic view of a vowel spectrum. These results were highly individual. As speech doesn't solely consist of vowels we also looked at consonant perception. Consonant perception was less individual, but nevertheless individual variation was found in consonant categorisation as well.

3 Individuality in Production and Perception of Sign

In the beginning stages of sign language research the only means of storing and saving sign language was an ordinary camera. That provided us with a static picture of sign language. This was the method used in e.g. the first Finnish Sign Language lexicon leaflets (Hirn, 1910/1912). The photographs had arrows and other marks to specify movements and other parameters of signs in them. The development of new kinds of technology, such as VCR, new digital camcorder and various other devices, has enabled us to save and store sign language on-line. This is analogous to the use of spectrograms and spectrum movies in the speech sciences. The sampling rate of an ordinary digital camcorder is limited to 25 Hz. We nevertheless have means which provide us with a possibility to get one still frame from an on-line film. This would be the equivalent of a FFT slice spectrum. This one frame can then be transcribed as such, without any movement coding diacritics. This is a way to store one part of language, but sign language movement that has been analyzed in this way provides us with a kind of language analyzing format similar to what would be obtained if e.g. only the steady states of vowels would be analysed and stored without any transitions of consonants would being analyzed from speech. Nowadays the storage and analysis of a movie is fairly easy. For example the possibility to use a long exposure time in camera results in a picture in which the whole movement envelope from beginning to end of a particular sign is visible in the picture. This can be used as a research method for example in studying the language acquisition of a signing child.

We can analyze the movie one frame at a time with a picture capturing program attached to a video recording. The present technology enables us also to integrate several captured frames into one single picture. The frames are placed on top of each other which results in a movement envelope. In this way we can transcribe and label the envelope by frames. LED cameras and infrared camera provide an even greater increase in sampling rate. This research paradigm is used for instance when studying sign language fluency, coarticulation and signing rhythm (Poizner, 1983). Fluency in sign language does not mean rate and speed as such, but the coordination of fingers and hands with regard to each other (Wilcox, 1992). In sign language fingers and hands are coordinated so that they are always in the right place at the right time in interaction. Fluency is difficult to define in sign language. Difficulties of fluency in sign language can be related to apraxia or the fact that the coordination of fingers is not working for reason or another; for example, they can be crippled or the movement trajectories of fingers can be hindered (Poizner, Klima & Bellugi, 1987). Fluency, however, is generally used by natives to evaluate the language learners' ability to follow a conversation in a particular sign language (McKee & McKee, 1992). The general parameters of fluency in a broad sense, which is commonly used for evaluating sign language learners in a sign language class are discussed in connection to the Finnish Sign Language teaching in Fuchs 2004. Previously it has been found, that to some extent signers alter their language use when people outside the signing community are present in the situation (for an extended review on Deaf culture, the signing world and its contacts with the hearing world, see e.g. Preston, 1994; Malm, 2000). The situation is thus similar to those when non-native language users are present – we tend to alter the use of language into more intelligible one via finding alternative words, sentence structures which are easier to comprehend and also we might talk slower and more accurately (for an extended description of non-native directed speech, see e.g. Gass, 1997).

3.1 Individuality in Production of Sign: Coarticulation and Rhythm in Sign

3.1.1 Rhythm patterns in manual vs. non-manual articulation in sign

Rhythm and oscillatory events are the basis of all human action. In this study we investigated the relationship between manual and non-manual coarticulation in sign and furthermore relate those results to the findings of speech research and the hierarchical structures in oscillatory patterning within speech. We measured coarticulation of visually defined measurement points and discovered two intertwining rhythms within manual and non-manual articulation patterns in sign. The main finding was the existence of several different, interacting rhythm pattern textures.

The importance of rhythm is present in everything we humans do. Without coordinated rhythm no movements would be possible to accomplish. On several occasions, many different rhythms combine within a single rhythmic event. Just think about pieces of music: different groups of instruments, all playing different melodies and these different rhythm patterns every now and then combine to form a stronger beat within the intertwining texture of rhythms (Kokkonen, 1992: 178 on rhythm as a part of musical texture). We rarely stop to think about rhythm and its importance in everything we do unless something disturbs our ability to coordinate the movements we do. It can be a disorder, either genetic or acquired, or temporary difficulty in coordinating movements (Rowland, 2000; concerning signed languages, see Emmorey, 2002).

Usually it is the rhythm texture that identifies a certain composer of music (Brindle 1975, pp. 30-34 on human identification, Lemström, 2000 and Shmulevich et al, 1999 on automatic music retrieval algorithms) and the same is also true when distinguishing speakers from each other: part of the identification is due to different rhythm patterns in speech (for an extended review on rhythm and timing issues concerning speaker recognition, see Nolan, 1985; Tajima, 1997 on durational cues on foreign accents). The difference in rhythm patterns also is one source of foreign accents in second language acquisition (Gass, 1997). The learning process starts very early on since the first rhythmic movement patterns from a linguistic point of view are already present when babies start babbling (Petitto, 2004). It is already then the first attempts to accomplish repetitive coordinated motor patterns start to emerge.

Even before oral babbling stages the rhythmic movement patterns can be observed in baby's hands which usually move in rhythm. These rhythmic patterns in hand waves are the first prerequisites for sign language acquisition. Rhythm is a unifying factor to all human movement, from walking to pointing, from swallowing to talking (Ghez, 2000). Everything we do needs coordination and thus the interplay of different rhythms in unison.

Rhythm in speech can be expressed and observed within two different domains: durations of individual segments and pauses along with the rhythm within segments and how they affect each other via articulatory movements and their coordination. Rhythm changes in speech are highly individual, that is each and every one of us has their own individual speech rhythm, some of us talk slower than others. Some of the rhythm patterns are governed by dialectal variation, usually it is the lengthening of segments in particular positions either within a sentence or within a single word that play a part in giving each dialect its own "figure of speech". Speech rate is sometimes used in giving speech artistic functions, like in poetry recitation (Kiparsky, 1974). The rhythm of a poem is occasionally also written in the text, the words are thought to be spoken out loud and speech can be seen as the ultimate user interface for poetry.

Recent findings within speech rhythm studies have shown that acoustic correlates of speech rhythm can be grouped into rhythm classes and furthermore, that variation between different dialects within a single language can demonstrate differences as great as those between two different languages. These rhythm classes can vary also across individual speakers so that two speakers within one dialect can have very significant differences when it comes to speech rhythm classification (Grabe, 2002).

Sometimes the rhythm patterns within speech are accompanied with gestures, so-called batons or beats (Kendon, 1972). They are used to pace speech, but they are not the only gestures accompanying speech and furthermore, each and every co-speech gesture is uttered in unison with speech at the right times in relation to what is the acoustic signal and the co-speech gesture's relation to it (McNeill, 1992). The other groups of co-speech gestures can vary in their timing in connection to speech, as they are not specifically used in pacing speech, but batons and beats are the speech-pacers, and in co-speech gesture rhythm studies those would be in focus of attention.

Recent sign language research has merely concentrated on rhythm in connection with artistic use of language (Peters, 2000), but there are studies also connected with rhythmic structure in nursery rhymes (Blondel & Miller, 2000) and manual articulation, namely coarticulation (Ojala et al., 2009). The studies dealing with rhythm and artistic use of sign languages have found that some of the sign transitions used in poetry are not found in ordinary signed discourse. Other studies have concentrated on phonological elements of sign language poetry and on how the function of rhythm is produced in sign language songs (Klima & Bellugi, 1976). Klima and Bellugi found, that the patterning of ASL poetry and songs is merely based on individual preferences and not so much conventionalised. There are also several forces affecting the structure of artistic use of sign language, they concentrate on basic devices such as balancing between two articulators and specific ways of using sign space. The latter is often used in the same way that melody is used to superimpose rhythm in songs. As rhythm is present within us since birth in everything we do, the rhythmical elements in sign language are also relevant and should be looked at from a rhythmical point of view. From previous studies (Peters, 2000 among others), that sign poems use rhythmical elements – why should signing in everyday conversations be any different?

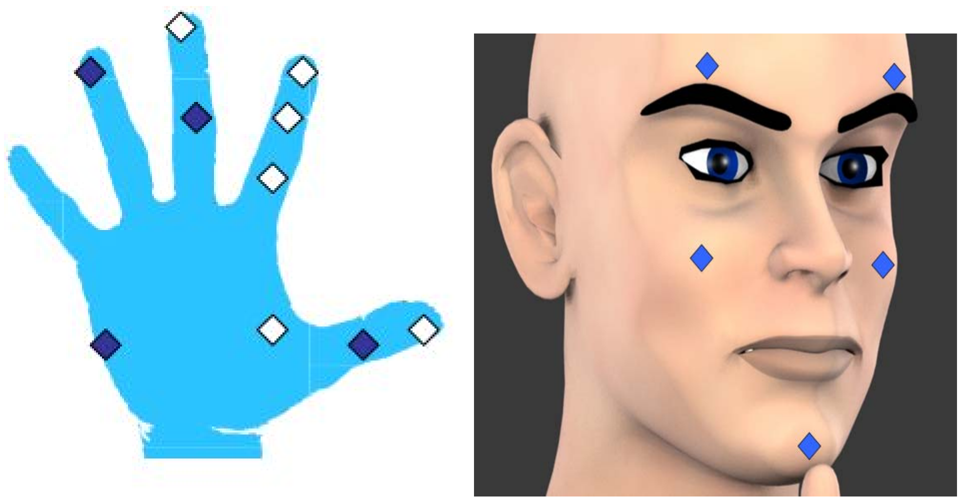
On the other hand, our previous study, which concentrated on coarticulation of manual elements in sign found hierarchical structures in the dominance of hands and fingers (Ojala, 2009). It seems that the index finger has a dominant role in the rate and speed of both coarticulation within one hand and furthermore in the interarticulatory phenomena. The intertwining rhythm structures found in coarticulatory and interarticulatory phenomena can also be found in facial gestures and head nods, although their relationship when discussing hierarchical structures of rhythm phenomena is not yet clear. Headnods are an integral part of sign language non-manual features. Some of the functions of a head nod are by default the same for both signers and speakers. This too is in concordance with the general theme of this thesis, to search for unifying features and common phenomena within languages in different modalities. One head nod is most often read as a positive sign, 'yes', 'I agree', 'I am listening' and so on. There are, however, other grammaticalised functions assigned to head nods within signed languges, which include for example topic – comment structure refining and completion complement functions. (For further discussion on head nods within signed languages, see e.g. Sutton-Spence & Valli, 1998.)

3.1.1.1 Methods

The data of this study consisted of informant interview recordings with a Sony digital video camcorder. The study questions within the interview setting consisted of different tasks and a spontaneous signing passage where the question eliciting the story was: “What did you do during your holiday last summer?” The tasks in question were to describe an imaginary floor plan of an apartment and to give directions based on a street map from one point to another.

This study was one part of a larger research project collecting data on Finnish Sign Language perception and production processes within a phonetic framework. This study concentrated on studying coarticulatory phenomena in sign in relation to manual and non-manual events during spontaneous signing in the production task of the study design. In other words we concentrated on studying the interplay of manual and non-manual rhythmic oscillations within spontaneous signing. The data were gathered by using coarticulation matrices and movement trajectory measurements from digital video recordings.

The measurement points were 10 in each hand (fig. 3.1a) and 5 for non-manual elements (fig. 3.1b) such as head nodding and facial gestures. The measurement points were visually defined and manually marked for each frame in the video sequence with the help of a Matlab-driven GUI programme specially designed for this study. These measurement points were then used as a basis for the 3D coarticulation matrices. These matrices were used to study on one hand the coarticulatory phenomena within the manual coarticulation domain and on the other hand the synchronization of non-manual and manual elements in spontaneous signing in relation to rhythmic patterns within.



Figures 3.1a&b on manual (3.1a) and facial coarticulation points (3.1b) measured from the video recordings.

3.1.1.2 Results

There were two intertwining rhythm patterns, a slower one expressed both non-manually and manually and a faster one expressed primarily with the hands (Figure 3.2). This shows parallel results to the relationship between segmental and prosodic phenomena in speech. It also relates to the intertwining of speech and co-speech gestures in speech.

Facial gestures were mainly linked to accelerations in manual coarticulation. This could relate to accelerations being more prominent phenomena in sign than decelerations, although further investigations are needed to confirm this finding. The facial gestures also often coincide with sentence boundaries, so they seem to have a boundary marking role in sign. They are used to emphasise certain aspects of entities which are referred to, so also in these two cases their role coincides with the role of facial gestures in speech, although in speech the segmental properties can also be altered in connection to emphasising aspects. That role of segmental properties in sign is also present, especially with size and shape specifiers (SASSes), which usually function as adjectives (Baker & Cokely, 1980; Supalla, 1986).

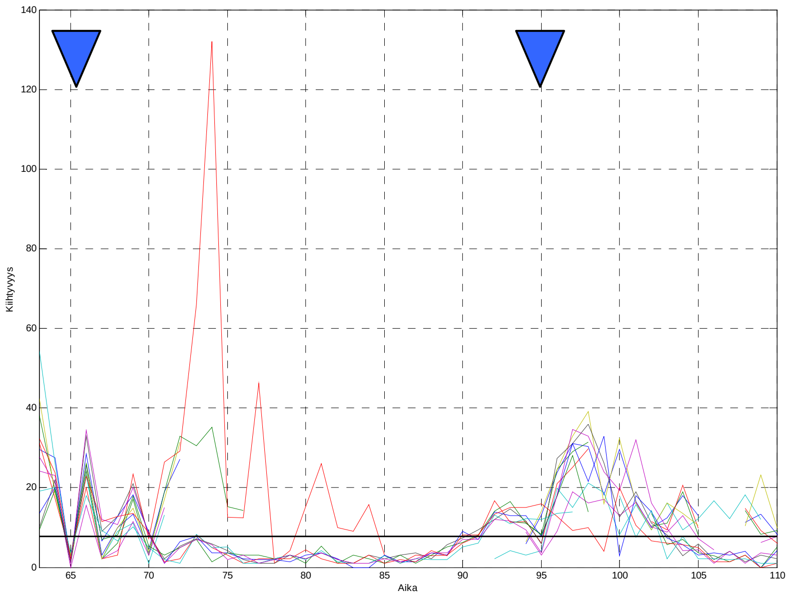


Figure 3.2. The coarticulation matrix of both hands showing the head nods in relation to the accelerations and decelerations of the hands in time.

3.1.1.3 Discussion

In this study the results show that head nods are often linked with junctures, be they utterance-, sentence, turn-taking or pragmatic ones. Head nods also co-occur with the acceleration phases of oscillation in the manual articulation phases. Head nods are most prominent when coinciding with longer pauses within the signed story. Head nods are also used very often as preliminary markers of emphasis – that is, markers of occurrence of specific, important events within the discourse.

In speech the most prominent rhythm elements of all are the so-called beat gestures, functioning as emphatic elements when speaking, furthermore their presence makes the ultimate prominence marker in speech, since they are so different, coming from a channel separate from the auditory. Signers do not have beat gestures, but the Salk Institute research group have suggested that head nods have a similar role in sign discourse (Emmorey, 2000; p. 167). The results presented here show results parallel to the Salk Institute research group's findings when it comes to prominence and head nods in discourse in signed languages. These data were gathered from Finnish Sign Language discourse and Salk Institute group investigated American Sign Language, so this prominence theory of sign language now has supportive evidence from more than one signed language.

All the different phenomena within prosody are referred in the literature using different terminology. The most common are 'stress' and 'accent' in more linguistically oriented literature and 'prominence' in more phonetically orientated literature. The common core to all these words is the fact that something in the flow of information is more salient, attracts more attention than the surrounding signal flow. Lehiste (1972) discusses the features related to stress patterns related to segmental and suprasegmental tiers in speech, while Hayes (1985, 1995) has developed a metrical theory with stress rule patterns for speech. Prosodic properties have been studied in sign languages as well, for example Grosjean (1977) has stated that while signers alter sign rate, speakers alter the pause length if they want to talk or sign faster.

Within this research, however, there are not enough data yet to decipher the nature of these particular properties but it would need further study. Be the limitations what may, nevertheless based on these data the head nods occur in acceleration phases of manual coarticulation, and furthermore that the index finger of the dominant hand has the widest movement envelopes of all fingers. The facial gestures found in this data set coincide with acceleration phases in the manual coarticulation. They also stretch over several manual coarticulated segments, which shows their similarity to suprasegmental features found in speech. They too have their acceleration and deceleration phases along with head nod phases. This gives us further evidence on the rhythmic oscillation events and their presence at different levels of sign language discourse. The facial gestures furthermore provide a broader scope for coarticulation research in sign linguistics. Therefore based on these findings we might suggest a hierarchical structure within non-manual elements within signed languages similar to those found in suprasegmental features (and in broader scope in prosodic events) on one hand and in co-speech gestures on the other.

Head nods and facial gestures if repeated have oscillation with diminishing movement (cf. O'Dell, 2004), which makes them similar to all other repetitive movement with diminishing properties – they also obey the principles of economy found in human actions. That is to say: we all tend to prefer doing less to doing more. In speech research this principle is a part of the so-called H & H theory of speech production. It states that in speech we prefer trajectories which are easier to articulate whenever their use does not disturb the message understanding (Lindblom, 1990). In sign research the main interest has been on the coarticulatory trajectories and their behaviour in normal and fast signing. In sign research economy of trajectories can also be referred to as undershoot of movements (Mauk, 2003). All in all coarticulation, be it speaking or signing, seems to be in accordance with the ease of articulation principle. Coarticulation as a phenomenon is used to make articulation smoother and faster – and it also might reduce the perceptual load of the message receiver since coarticulatory patterns allow us to predict the ultimate articulatory goals the articulators are directed towards. In sign, as in speech, the coarticulatory phenomena within one phonetic parameter very frequently affect other parameters as well – this is simultaneous in sign language (for further discussion, please see Crasborn, 2001) and for the most part consecutive by nature in speech. That is, while in speech these coarticulatory phenomena follow each other in time, in sign they occur at the same time.

A reoccurring finding from several different areas of linguistic research seems to be the interplay of rhythm and basic oscillatory phenomena within human speech. The basic oscillations are found in different areas of research, but they are named using different terminologies: sign research commonly uses movements and holds (see e.g. Kita et al., 1998; Liddell & Johnson, 1989) while speech research talks about consonants and vowels in sound structures (see e.g. MacNeilage, 1998, Lindblom, 1990). More technically-oriented speech research often refers to oscillatory patterning of speech (O'Dell & Nieminen, 2001), but if we put the terminological issues aside and read the literature with more open minds we find that they all refer to the same rhythmic element of human speech and human action (Ghez & Krakauer, 2000). This being the recurrent theme of human research it might hold the key of evolution processes in human communication. This might also suggest a common, united basis for sign and speech via gestures (Wilcox, 2002). This common basis for human speech is also supported within infant language acquisition research (Dolata et al., 2008).

3.1.2 Rhythm in manual movements

In speech the sounds are not discrete phenomena but speech is a continuous string of articulatory movements, where previous and following sounds affect each other. This on-going movement pattern from one sound to another is called coarticulation. Coarticulation is a way to make transitions from one sound to another easier. In this way we are acting according to the ease of articulation principle (e.g. Shariatmadari, 2006). Simultaneously, we also tend to use all the capacity available if needed (Lindblom, 1981). This interplay of coordinating movements makes speech easier and faster. As coarticulation in speech is defined as sequential influences of articulatory gestures on neighbouring segments in sign we have to take into account the simultaneity of signing so within this study coarticulation also takes into account simultaneous movements of different articulators and not only the sequential influences within articulators per se.

When recording speech coarticulatory phenomena are usually controlled by using carrier words and sentences such that different sounds occur in similar positions coarticulatorily. This is a way to control what is said in order to make subjects' productions comparable to each other. Coarticulation provides us with knowledge of how different sounds are represented in different contexts. Coarticulation studies often utilise acoustics in speech because the articulators are not visible. This is different from studies of sign language since in sign the articulatory patterns are visible and thus it is more straightforward to study coarticulation in sign.

As it takes more time to sign because of the trajectories of hands are longer than those of e.g. the tongue, it is only appropriate that sign can be articulated simultaneously just because there are two main articulators (the hands) present. The possibility of using the 3D space in time also adds to the complexity of coarticulation in sign. Coarticulation has the same function in sign as in speech, it functions to make the conveying of the message smoother and more compact. In sign the compactness aspect is even more important since hands are slower as articulators than speech organs. Speech rate usually ranges 90-160 words per minute while within this study signing rate is between 20-30 signs per minute. In signing both hands participate in coarticulation, so there are two levels of coarticulatory patterns – each hand separately and then the interarticulation – both hands together. Coarticulation is also present in facial expressions and gestures but these are left out of this study. There is a distinction between manual and non-manual coarticulation and this study concentrates on manual coarticulation. The controlling of coarticulation in sign language studies is through task design. This is the only way to have control on what is signed because there is no written form of any signed languages. In this study task the design was based on an imaginary floor plan and a map task along with a separate task for spontaneous signing.

Studies on coarticulation in sign first concentrated on fingerspelling research. Fingerspelling is converting text into a visible form by means of a manual alphabet. It has a very limited amount of coarticulation: in most inventories of manual alphabets only one hand participates and the hand has very limited movement patterns. There is however an exception to this: the British Manual Alphabet, which uses both hands. Even so, the effects of coarticulation are great also within the scope of fingerspelling (Wilcox, 1992). Recently scientists have made some initial investigations into the coarticulation of signing: the effects are similar to those in speech – coarticulation makes articulation more fluent and its effects can be seen both in handshapes and places of articulation (Mauk, 2003; Ann, 1996). Thus also signing obeys the ease of articulation principle. On the other hand, Crasborn has studied coarticulation and its individuality in detail (Crasborn, 2001 among others).

Especially the anatomy and physiology of the hand and fingers affect the articulation of handshapes. The economy of articulation is for the most part depended on the dimensions and movement ranges of individual fingers. According to physiological research results the thumb has the widest range of movement patterns and the ring finger is the most restricted in its movement patterns (Ann, 1996).

3.1.2.1 Methods

The data were gathered as a part of a larger study, which concentrates on gathering data from the production and perception of Finnish Sign Language. The data consists of signed answers to a set of questions and tasks. The questions ranged from given tasks to spontaneous signing. The tasks given were: a map task, a ground plan explanation and a spontaneous task answering a question about a previous holiday. This procedure also elicited a sufficient number of handshapes along the desired continua to form the basis of the 3D syntheses for the perception tests discussed later in this thesis. The study design was equivalent and analogous to speech research sound samples, where tasks include different carrier sentences (read speech) and spontaneous speech (informants are asked to tell how they spent their last holidays or they have the option of telling a 2-3 minute story of their own invention). The productions were recorded with a digital video and which were then further processed and analysed with video software.

The data of this study consists of two informants producing the utterance HERE APARTMENT (Here is an apartment), which has a rectangle shape. This utterance was an excerpt from the floor plan task in a series of three tasks. This was the second task. This was chosen because by that time the subjects were adjusted to the thought of their production being recorded onto a video disc for further analysis. They were given the chance to view their productions and sign again if they felt that the first production was not according to their preferences for productions, if e.g. they wanted to avoid excessive repeating of signs or if they felt uneasy with gesturing when planning their story. As stated previously within this thesis signers also use gesturing in the middle of their signing. Usually that is related to planning ahead and to thinking about what to express next. The task consisting of what could be categorised as laboratory speech was chosen for analysis because its structure limits the coarticulation processes across speakers as well as signers. This enabled us to make comparisons between longer stretches of nearly similar productions across the two speakers in question. The signed utterance studied here consists of 3 individual signs and 6 rhythm units. The rhythm units were defined visually from the alternation of accelerations and decelerations of the manual elements of coarticulation. Furthermore a more detailed coarticulation analysis of the manual elements was made frame by frame, that is every 42 milliseconds. In each frame the specified coarticulation points were mapped from the frame and plotted on an x-y-z-matrix and that mapping matrix served as the basis of the study of movements in time for both hands. There were 10 measurement points in each hand in order to gather precise material on how different handshapes are manifested in a continuous sign flow.

In this preliminary study we have measured 6 of those 10 measurement points from each hand. The other 4 measurement points were used to specify the orientation of the hand when it was possible. The orientation of the hand within this study translates as the orientation of the palm of the hand in relation to the body of the signer. The pixel coordinates of the coarticulation measurement points were inserted in a matrix. The matrix was the input for the 3D image of movements of coarticulation points in time. The matrix provided also an input of 2D images of the speed and acceleration/deceleration as changes of each coarticulation point frame by frame. The matrix was furthermore the basis according to which the median of the speed was calculated. All figures and calculations were accomplished by using MatLab scripts.

3.1.2.2 Results

Coarticulation in sign can be studied within two different scopes: a broader one with focus on the interarticulation involving both hands (i.e. the interplay of movement patterns of the two hands) and a more precise one considering coarticulation within the movement patterns of one hand only. Interarticulation seems to effect in such a way that the hands move faster when they are more apart than when they are closer to each other.

Most of the time all fingers move simultaneously and with the same speed, but both index fingers have broader and faster movement envelopes when compared to other fingers. When comparing the index fingers the right index finger has slightly broader and faster movements. A similar handedness effect can be noted in thumb movement patterns but not for other fingers. The movement envelopes become smaller when going from the index finger to the little finger, however, the little finger has a broader movement envelope than the ring finger. This piece of information comes from the wider data set not included in this particular data set. In other words, the ring fingers of both hands have the most compact movement patterns. The thumb has a broader movement envelope than the index finger, but the movements are slower. The overall movement patterns in both hands are quite similar, both in timing and in broadness.

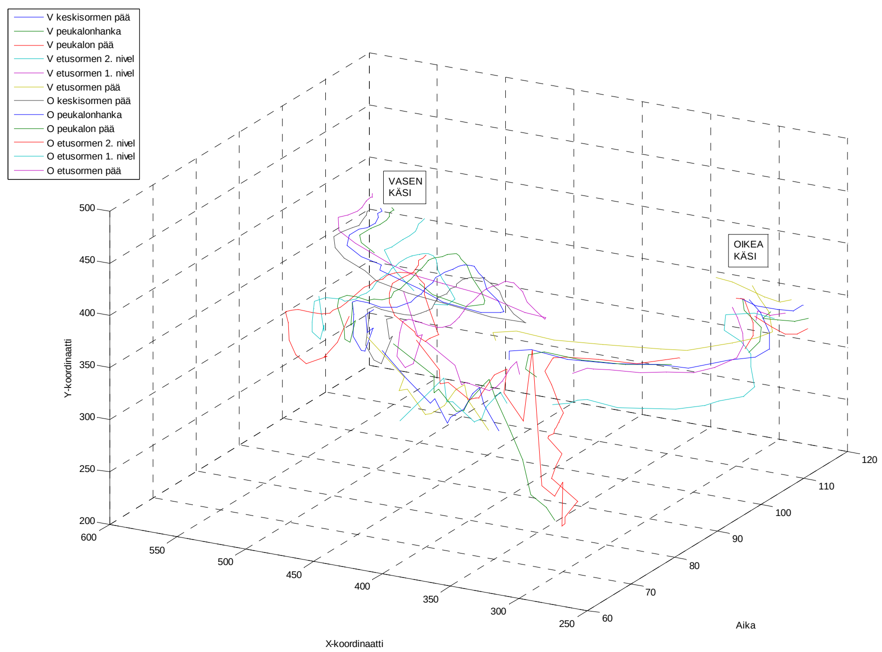


Figure 3.3 on coarticulation matrix (combination of both hands).

Changes in movement were investigated frame by frame. The two dimensions in space (height and width) plus time equals to three dimensions of the graph. The graph in figure 3.3 shows both the speed of individual fingers and the relations between different finger movement patterns. Most of the time within this material the momentary/ instantaneous differences between individual fingers' speeds are so minute that the graph would be sufficient with showing just the movement of one finger, but there are also exceptions to this.

According to results thus far it seems that individual vertical movements are faster than movements along the horizontal plane, but this observation needs further investigation. The graph also shows the tendency to keep movements as slow as possible but the scarce data might distort the results.

3.1.2.3 Conclusion

The alternations of deceleration and acceleration in signing movements are similar pattern to the pattern in speech – also speech has intertwining rhythms at different levels. The bases for these rhythms at least to some extent vary according to the individual language (e.g. O'Dell & Nieminen 2001). In speech the rhythm is achieved with the coordination of articulatory movements and it might be that the alternations between consonant and vowel sounds are one of the very corner stones of human evolution (MacNeilage, 1998). In this preliminary study we have concentrated on the observations of movement patterns on the lower level of coarticulation. Previously the alternations of decelerations and accelerations have been studied by Loomis et al (1983). Rhythm is expressed individually in both signers and speakers, it's incorporated in us, just as is the pace of walking (Ghez, 2000) or clapping of hands (Repp, 1987; Peltola et al., 2007). Grosjean (1977) among others has studied the rates of sign and speech.

The movement patterns in sign seem to have an oscillation pattern and cyclical form as do the articulatory movements in speech (See also Lindblom et al., 2006). The basic rhythm on the higher level of hierarchy in signing materialises in the movements and holds within and between individual signs in on-going signing (Liddell & Johnson 1989; Kita et al. 1998). According to this material the index finger seems to be the determining factor in the amplitude and the rate of signing. Other fingers follow the movement patterns of the index finger, but are more restricted in their patterns. The index finger also has a special task: pointing. Pointing is an important part of both signing and speaking – it is a convenient way to refer to something which is present and visible, be it an object or a person. (For more on pointing, please see Corballis 2002.) The thumb seems to have more independent movement patterns than the other fingers. This demonstrates the tendency of signing to exploit the capabilities in individual fingers' movement patterns as widely as possible whereas the ring finger's more restricted movement patterns demonstrate that signing tends to avoid such patterns that are more difficult to produce. In this way, signing as a form of communication acknowledges the physiological restrictions in the hands and operates accordingly.

3.1.3 Learning to control hand movements – individuality in sign acquisition

Native signers first perceive the movement envelope of a sign (Wilcox, 1992), in a similar fashion as the speaking child perceives first the word as a whole along with the rhythm and intonation typical of speech in their native language (Jusczyk, Cutler & Redanz, 1993). Signing children first produce movement envelopes, i.e. the movements as a whole without small distinctions within a sign (Wilcox, 1992). Thus, both signing and speaking children obey the overall envelopes and gestalts of their own native language. The rhythm of their native sign language can also be distinguished in the child's signing. Children replace difficult handshapes with easier ones (Wilcox, 1992; Takkinen, 2002), for example the handshape K is usually replaced with 1 handshape in a child. This is because handshape 1 is easier to articulate than K is (Figure 3.4). The 1 handshape is also one of the prime handshapes in FinSL (for more on handshape hierarchy, please refer to Chapter 2 of this thesis).

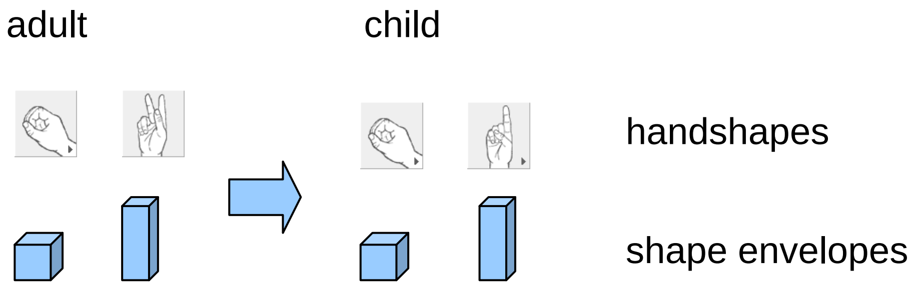


Figure 3.4. Adult production of the OK sign and child's imitation of it, please note the unchanging shape envelopes although the difficult handshape has been replaced with an easier one.

In this way, children are able maintain the movement envelopes despite changes in handshapes, in other words, they are maintaining the overall shape of the word by changing the segments into easier ones, just like speaking children do by changing more difficult speech sounds into easier ones, such as replacing /s/ sounds by e.g. /h/ sounds, which maintains some aspects of the particular speech sound, i.e. hissing, partly while avoiding the more difficult precision of a proper fricative /s/.

3.2 Individuality in Perception of Sign

In this chapter the results found in speech perception within Chapter 3 of this thesis are tested against findings in signed utterances and their perception. Test batteries are used which are analogical to speech perception tests.

3.2.1 Individuality in Perception of Handshapes

Signs are produced individually, as previously stated in section 4.1. and here we wanted to find out whether the individuality also governs the perception of sign phonemes per se. We also wanted to investigate whether the perception process of handshapes is similar or different to that of a) vowels or consonants and b) places of articulation as sign phoneme categories.

The main element of the gestural/motor theory of speech perception is that the main percepts of speech are not the acoustic patterns per se but the articulatory patterns that trigger the acoustic signal Liberman & Mattingly (1985). Recently, sign researchers have also promoted this idea, although in slightly different words, as suggested by Wilcox (1992): "making meaning with movement" - encoding of the meaning in articulatory gestures and trajectories. Both sign and speech are thus encoded in articulatory movements and then the only difference that remains is the channel. Both speech and sign obey coarticulatory effects and follow ease-of-articulation (Shariatmadari, 2006) and H&H-theory principles (Mauk, 2003; Lindblom, 1997) with coarticulatory patterning (Mauk, Lindblom & Moon, 2006). All communication methods in general, be it what might be, follow the so-called parity principle: "what counts for a listener must count for a speaker and vice versa - otherwise there is no communication." (Liberman, 1996). This same phenomenon can be noticed when travelling abroad – if there is no mutual language when there are no interpreting services available communication just doesn't work.

The phoneme inventory of sign includes: handshape, place of articulation (POA), orientation of palm and movement, and some studies regard mouthings as phonemes, as well, although their status in the phoneme inventory of sign is not yet clear. Here handshapes are studied as equivalent to vowels in speech. (For sign language phonetics and analogues to speech in more detail please see section 1.4 within this thesis). Handshapes were chosen as the first phoneme category to be tested within sign language within this project, because in the sign research tradition they are the most thoroughly investigated group of sign phonemes. Handshapes are also used when classifying signs in Sign Language dictionaries that are arranged according to sign properties (e.g. Suvi, FinSL online dictionary).

The purpose of this study was to investigate the identification and discrimination of the handshapes chosen from the handshape continuum generally transcribed in Stokoe notation (Stokoe, 1960) as /G/–/X/. This continuum has physiologically based boundaries: it is impossible to go beyond these two extremes in finger flexion and extension (Figure 3.5). The Stokoe notation is equivalent to IPA symbols in speech research (IPA chart, 1993). As with IPA symbols, the symbols used in Stokoe notation refer to single elements in sign, which do not have a lexical meaning by themselves. Only certain combinations of elements gathered together from the sign language phonemic inventory represent lexically meaningful signs.

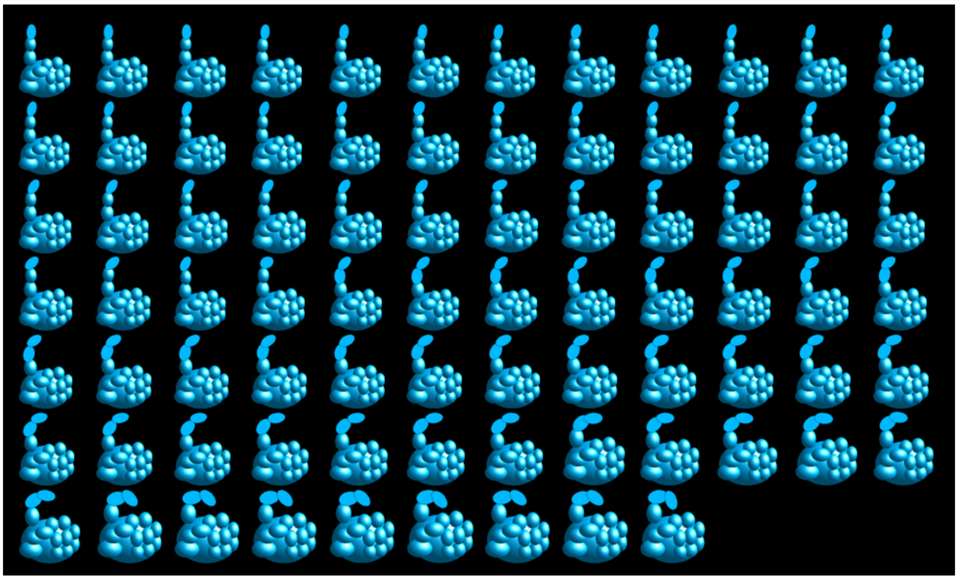


Figure 3.5. Handshapes used in the perception experiment within the continuum.

In visual terms the continuum goes from a fist with protruding, fully extended index finger transcribed as /G/ to a fist with protruding, extremely crooked index finger, transcribed as /X/. In principle this continuum is equivalent to e.g. formant continua in vowels. This implies perceptual phenomena similar to the categorisation of vowels. For vowels, there are large individual differences in the locations of phoneme boundaries and category prototypes (Aaltonen, 1997). If a handshape continuum represents the same for signers as a vowel continuum represents for speakers, then similar effects in the perception of vowels could be expected.

3.2.1.1 Methods

There were 5 subjects from each of the 3 subject groups. The subject groups were: 1. deaf, native FinSL signers, 2. hearing FinSL interpreters, who have learned to sign as adults and 3. naïve speakers who had no command of any signed language. The subject groups were matched by age (range: 26-52, mean age: 33) and sex. All subjects participated in the study as volunteers. The visual acuity of all subjects was tested with a clinical E-chart and it was within the normal range, some of the subjects wore eye glasses for corrected vision (n=7). The normal range is [1.25 – 2.00] which has been stated in DeValois & DeValois (1990).

The first two groups of the subjects participated in the production test set (for details, please see section 4.1. within this thesis), which was done primarily to establish a basis for the synthetic stimuli design for the perception test sets for handshapes and POAs respectively. The production test included three separate tasks: map task, ground plan explanation and free narration. This procedure elicited a sufficient number of handshapes from the desired continuum for the basis of the 3D synthesis for the perception tests. The perception test was further divided into two separate tests: identification and discrimination tests. The AX paradigm was used when designing the discrimination test. The AX paradigm is used in discrimination test where subjects have to attend to stimuli in contrast to the odd-ball paradigm where the subjects are asked to ignore the stimuli and concentrate on a distractive task (for a review on different paradigms in attentive and non-attentive tasks, please see Tuomainen, 1996). By using the AX paradigm we can gather information about subjects' ability to both identify and discriminate the stimuli when attending to the task given. In the identification test the subject was asked to label the stimuli as belonging to either category /G/ or /X/ and furthermore to rate the goodness of the stimulus with the range [1-7] with 7 being the best possible and 1 recognisable. In the identification test response labels, goodness-ratings and reaction times (RT) were measured. In the discrimination test the subjects were asked to state whether the stimuli seen in the stimulus pair were the same or different (AX-paradigm). Subject's discrimination ability (d prime) was calculated along with RT measurements, which were automatically obtained with PXLab presentation software.

The material for the 3D synthesis was gathered based on the native subjects' signed answers in the production tests. The synthetic continuum consisted of handshapes where the only protruding finger was the index finger while the other fingers made a fist. At one end of the continuum the protruding index finger was fully extended (/G/) and at the other end the index finger was as crooked as possible (/X/). The 3D synthesis modelled the original handshapes produced in various natural contexts (three different tasks in the production test set) by blue, oval shape figures. The blue colour was used to avoid any conflicts with naturalness. Semisynthetic stimuli were used in the pilot tests, but they appeared to be a disturbing element for concentration on the tasks in the experiments. The semisynthetic stimuli elicited questions such as: "Who is the individual that has produced this?" "Do I know that person?" and the participants found it difficult to concentrate on the study questions per se. The pilot tests with semisynthetic stimuli also gave results, which made us question whether there were also other, hidden layers of information within the semisynthesic stimuli that were noticeable to the native signers but unknown and unnoticeable to us which might have contributed to the mixed results. Thus, we decided on designing the fully synthetic stimuli devoid of any naturalness to avoid similar effects with presumed naturalness of the stimuli. Also control of fully synthetic stimuli is easier: we can elicit only desired changes within the stimulus continuum and there are no hidden layers of information that might be present in the semisynthetic stimuli originating from the natural basis of semisynthesis (van Hessen & Schouten, 1999). The synthetic stimuli were drawn on the continuum 0°–82° with 1° intervals. The degrees were measured from the tip of the index finger to its base. This measurement principle was possible because the muscles of the index finger are coordinated linearly, as stated in Darling et al. (1990). All the stimuli were included in the identification test, but only every other stimulus was chosen for the discrimination test, i.e. the stimuli were separated by 2° intervals. The intra-pair distance was thus 2° and the pairs covered the whole continuum 0-82°. This was done in order to reduce the concentration load for the participants of the study.

3.2.1.2 Results

Identification and discrimination tests for these three different subject groups showed that the groups differ in discrimination behaviour but not in identification patterns. All subjects had individual differences in the place and steepness of the category boundaries. Also the naïve categorizers, i.e. those who were not previously used to handling the visual input in the identification tests were able to categorise between the different stimuli in the continuum (Figure 3.6).

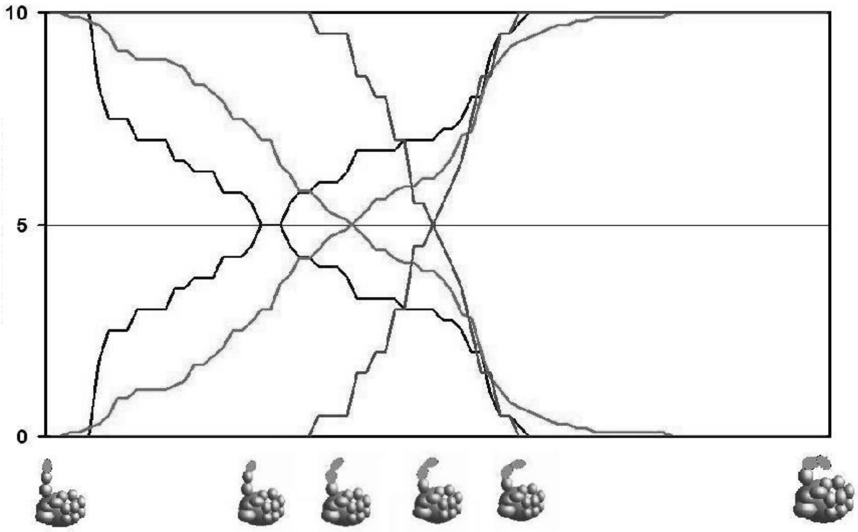


Figure 3.6. The category boundaries for different groups. The leftmost crossing is for the signing group, the middle one is for the naïve group and the rightmost crossing is for the sign interpreters' group. The individual handshapes denote different handshapes along the continuum as follows from left to right: the most upright one, signers' average crossing, naïve group's average boundary, interpreters' average boundary, the stimulus all subjects responded as being crooked and the most crooked one.

Reaction times were faster among the native signers than in other groups. Interpreters had next fastest reaction times. These differences might suggest a difference in familiarity of the stimuli. That is, how used the subject is to categorising visual stimuli. Among the native signers the places varied individually, but when repeated the subjects' category boundaries were very stable. These results are parallel to those obtained in vowel perception experiments. The category boundary along the continuum is nearest to the vertical (mean = 27°) for the signers and furthest away to the crooked end of the continuum for the sign language interpreters (mean = 48°). Naïve subjects' identification function crosses in between of that of signers and interpreters function curves. The identification function curves denote the place within the continuum where the subjects within the particular group on average perceive the change of category from one into another. Attention has to be drawn to the fact that the naïve subjects' study question and instructions differed somewhat from the two other groups. Their study question was based on visual rather than sign language based categories.

3.2.1.3 Discussion

These results for their part provide answers to the questions about the nature of different phonetic categories in sign language; for example whether the handshapes are more categorical by nature than movements or orientation. In speech consonants are said to be very strictly categorical based on the fact that they are very short in duration when compared to vowels. Vowels then again are said to be perceived differently – more in a linear way – because they, being longer in duration – can convey functions other than purely linguistical, e.g. they can act as carriers of information about the health status or emotional factors of the speaker. McCullough with others (2000) have, however, obtained different and opposite results concerning the categorisation of handshapes in ASL: according to their study handshapes are more categorical in nature than e.g. POAs.

According to the present results handshapes seem to be similar to vowels: they are categorised merely as a continuum and they are very strongly are governed by coarticulatory patterns. The POA has a similar function in sign language to that of consonants in speech – that is they are governed very strongly by categories and they are less governed by coarticulation as units. POAs are also the anchor points for sign language – they are the binding force between trajectories, i.e. movements, in a similar fashion to consonants. This might also be ground for the quantal theory of speech for sign languages – Stevens in his quantal theory of speech production claims that in the oral cavity there are such areas which belong to the same quantum – that is, differences in places of articulation within that area result in acoustical differences so minute that they are not perceivable and thus have no or very small significance for perception of speech. In other words, minute changes in POAs within one quantal area do not make any difference at all while a change as minute as the one described previously between quantal areas is very easily perceived as different. This has resemblances with the theory of categorical perception, but here the general idea of changes perceived differently in different areas of continua has been transferred into articulatory terms. Is there a chance we might find quantal areas also within sign languages? POAs seem to have varying sizes in different parts of the signing space. In the same way as in speech there might be such areas of articulation within which differences are easily ignored or they are indifferent.

The patterns found in the results for native signers' data this far have been similar to those obtained from native speakers when categorising their mother tongue. That is, the results vary in steepness and place of the category boundary from one individual to another, but when repeated the test gives stable results for individual categories. The results show improvement of discrimination ability in the vicinity of the category boundary in the discrimination task as well as lengthening of reaction times in the identification task. These results are parallel to those obtained from speakers' data. On the other hand, the absence of discrimination ability improvement in the naïve subjects could be a sign of the categorisation of the continuum evolving from a purely physical phenomenon, i.e. based on the physical differences within the stimuli. The performance of one of the SL interpreters shows a native-like behaviour when identifying and distinguishing between stimulus pairs while the other interpreters' results resemble those of naïve subjects when distinguishing between stimulus pairs. But also in these subjects the identification patterns with goodness-ratings are more similar to those of native signers and one of the interpreters.

This suggests differences in language learning stages of the interpreters. They were all matched with regards to training, hours of interpreting during working weeks and years of employment. This discrepancy between the results of different interpreters raised questions about bilinguals between modalities: how would being a bilingual with Finnish and FinSL affect the identification and discrimination tasks in these two modalities? Would they manifest with native patterns in both modalities or would their ability be biased towards one or another modality? This might further shed more lght on the quest for evolution of language: how did it come about? The prime bases for the evolution of language are further discussed within this thesis in section 1.1. with reference to the relations between sign and speech; section 1.2. with reference to the neural bases; section 1.4. discusses the evolution of signed languages further, but for more general discussion about the origins of human languages, please see Corballis, 2002.

3.2.2 Individuality in Perception of Places of Articulation (POAs)

This study is performed within the phonetic framework and looks at sign language as analogous to spoken language. This similarity between signed and spoken languages as a means of communication has also been suggested by studies from various fields of research (e.g. in brain evoked potential studies Nishimura et al., 1999 and studies on sign language structure Mauk, 2003). The phenomenon of categorical perception has been widely studied in speech sciences for some decades now (e.g. Harnad, 1985). In general, investigations have been made in various areas of human perception, such as object categorization (Biederman, 1985). As a basis for this particular study, studies that have looked at categorical perception of different speech sound continua are to be considered here. There is a discrepancy in results between consonant and vowel continua and their perception as concerns identification and discrimination patterns. While consonants are usually perceived as being strictly categorical, the vowels are perceived as being items in a continuum of changing vowel parameters. Within the project we have used the same design for handshapes, vowels and places of articulation both in speech and sign successfully and the preliminary results gathered have assured us that this paradigm is suitable for studying sign language within the phonetic framework.

The main element of the gestural/motor theory of speech perception as stated by Liberman & Studdert-Kennedy (1985) is that the main percepts of speech are not the acoustic patterns per se but the articulatory patterns which trigger the acoustic signal. This thinking fits perfectly with sign research, because in sign the articulatory movements are completely visible as such and not hidden as in speech. Recently, sign researchers have also promoted this idea, although in slightly different words, as suggested by Wilcox (1992): "making meaning with movement" - encoding of the meaning in articulatory gestures and trajectories. Both sign and speech are thus encoded in articulatory movements and then the only difference remaining is the channel difference. Both speech and sign obey coarticulatory effects and follow both ease of articulation (Shariatmadari, 2006, Lindblom et al., 2003) and H&H-theory principles (Lindblom, 1985) with coarticulatory patterning (Mauk, 2006). Lately, recent findings in neurophysiological studies have shown that signs also are processed in Wernicke's area. Thus one can presume that both sign and speech activate Broca's area. The so-called Broca–Wernicke complex in speech processing can be seen as the core area within the mirror neuron system (MNS) (for further discussion on MNS and BW complex, please see section 1.2. of this thesis). Here MNS represents the main element of the language ability: the core of sensomotor links between observed articulatory gestures and internal articulatory pattern memory traces. MNS presents “the missing link” of the language inside principle (López-Garcia, 2005), that is, that both sign and speech are manifestations of the same human language ability. Sign and speech have traditionally been regarded as two completely separate systems based on the different modalities they use. Sign has also been seen as an augmentative or artificial supporting system to speech for the deaf and not a language on its own. In the present framework within this study sign is, however, treated with the same status and same research methods as speech. This also represents the basis, together with the gestural theory of speech perception, for the experimental phonetic framework used in this study. As the research methods are basically the same, although adapted and modified to fit the different perception patterns in signers, it makes it easier to compare the results from speech and sign research. This study is for its part involved in enlarging the phonetic framework to cover also signed languages, we started our investigations with handshape testing (for results and discussion, please see section 4.2.1. of this thesis), because handshapes are the most prominent phoneme group within sign languages (also section 1.4 of this thesis), and also the primary group by which signs are arranged within a sign language dictionary (Suvi, Online FinSL dictionary). The next phoneme group in the prominence hierarchy is the place of articulation and these two form the basis of any sign. Other phoneme groups within sign languages include movement and orientation, although some research groups also regard mouthings and mouth gestures as phonemes, but their status within phoneme inventory is still under discussion.

Previous results from handshape studies showed that handshapes elicit similar results in identification and discrimination tests as vowels do. This inspired the next study question about whether places of articulation behave in the test setting similarly to consonants, i.e. are more categorical and less continuum-like. There are two distinct result patterns within sign language handshape studies with behavioural test setting obtained, one of which states that handshapes behave similarly as to places of articulation and thus are more categorical (Emmorrey, 1992) and the other states that they are more continuum-like (Ojala et al., 2008). We wanted to test which of these obtained results would coincide with the forthcoming results obtained here. In other words we wanted to test whether handshapes and places of articulation elicit similar or different response patterns in behavioural test settings. The other study question addressed here is whether consonants and places of articulation elicit similar or different responses in the same test battery. In designing sign continua we treated the place of articulation in sign as analogous to consonant loci.

3.2.2.1 Methods

In this study behavioural tests administered to informants are used to investigate patterns of perception in sign and speech. Different characteristics of the senses (i.e. sight vs. hearing) have to be taken into account when designing the individual test batteries. The data were gathered in connection with a larger research project studying production and perception of Finnish Sign Language (FinSL). The preliminary investigations concerning the study paradigm and methods are previously discussedd in an Interspeech conference paper presented in Brisbane in 2008 (Ojala et al., 2008).

There were 2 informants from each of the 3 informant groups. The informant groups were: deaf, native FinSL signers, hearing FinSL interpreters, who have learned to sign in adult age and naïve speakers of Finnish who had no command of any signed language. All of the informants in the two latter, i.e. hearing groups were native speakers of Finnish. The informant groups were matched by age and sex. All informants participated in the study as volunteers. The vision of all informants was tested with clinical E-chart and it was within the normal range. Some of the informants wore eye glasses to correct the eye sight within the normal range. The normal range is [1.25-2.00] which has been stated in DeValois & DeValois (1990). Groups participated in various test settings following the restrictions according to their hearing status. All three groups participated in sign stimuli experiments – this was done in order to investigate whether language background has an effect on the way visual input is dealt with within a behavioural test setting. Including speech stimuli within the study enabled us to study both speech and sign stimuli with controlled, analogous test batteries.

This study paradigm consists of behavioural test settings in both visual and auditive stimuli continua. The informants participated in identification and discrimination tests. In the identification test the informant was asked to label the stimuli as belonging to one of the categories and furthermore to rate the goodness of the stimulus in the range [1-7] with 7 being the best possible and 1 still recognizable. In the identification test the response labels, goodness-ratings and reaction times (RT) were measured. In the discrimination test the informants were asked to state whether the stimuli in the seen stimulus pair were same or different (AX-paradigm). Informant's discrimination ability (d prime) was calculated along with RT measurements, which were automatically obtained with PXLab presentation software.

The 3D synthesis was carried out using Blender software (Online.http://www.blender.org/) based on native signers' productions obtained in given tasks related to the production part of the on-going research project of FinSL, which this particular study is also a part of. The synthetic stimuli chosen for this test procedure covered one natural place of articulation (POA) continuum in FinSL. The continuum also is governed by the boundary conditions in the visual/kinaesthetic, i.e. physiological domain. The continuum is between two phonemes in FinSL: the POAs in the signs 'deaf' (Fig 3.5a) and in the sign 'always' (Fig 3.5b). The particular POAs do not form these signs alone but must be articulated with simultaneous adjacent handshapes and movements. The synthetic continuum in visual terms consisted of a protruded index finger moving from one point to another point along the mandibular bone ridge of an avatar. The endpoints of the particular continuum are depicted in figure 4.5. There were 100 steps along the continuum in the identification test and 50 steps in the discrimination test. This modification to the discrimination test was done in order to ease the concentration load of the informants who participated in the study. In designing sign continua we treated the place of articulation in sign as analogous to consonant loci.

Figures 3.5a (on the left) and 3.5b (on the right). Figures show the endpoint of the sign continuum from the starting point of the signs 'deaf' on the left and 'always' on the right hand side.

In designing the behavioural test sets we used PXLab software to present the stimuli in tests and to analyse the informants' answers to the study questions and reaction times accordingly. The stimuli were displayed in a randomised order in the identification task and also the stimulus pairs were displayed in a randomised order in the discrimination task part of the study. In the tasks the stimuli/stimulus pairs were each displayed 10 times. In the discrimination task setting the number of stimuli was reduced by 50% to ease the concentration load of the informants who participated in the study. This was done by choosing every second stimulus along the continuum to the test set.

3.2.2.2 Results

All informants participated in the visual test setting regardless of the language status, i.e. whether they knew any sign language or not. Identification results for these three groups showed that groups did not differ in identification patterns. The informants exhibited similar category boundary steepness effects, but the results showed that the places of category boundaries were concentrated in two different locations along the continuum. Also the naïve categorizers, i.e. those who were not previously used to handling visual input in identification tests were able to categorise between the different stimuli in the visual continuum. The naïve participants, however, were given the task with visually-based conditions and not requiring any knowledge of sign language. Signers and interpreters received similar instructions based on sign language categories and places of articulation as points of reference.

All three groups had very similar results on category boundary steepness, but for two of the groups category boundaries were bimodally distributed. Signers' boundaries were concentrated around one single point in the continuum while within both hearing groups the category boundaries were concentrated around two different locations. One location, found for one of the SL interpreters, was the same as that found in the signer group. Both of these locations have visually prominent features, but only one of them is found in the results from the signer group. That particular boundary has a linguistic significance in FinSL. The differences between the SL interpreters' category boundary locations suggest that the categorisation of POAs might partly reflect the level of language proficiency in L2 learners also in the visual modality, i.e. signed languages. The individual with categorisation of POAs similar to the deaf informants also had results parallel to the native signers' results in the handshape experiment.

3.2.2.3 Discussion

The identification patterns along the different continua are similar to each other, both in steepness of the category boundary and the tendency of the place of the category boundaries to concentrate near particular points along the continua. In the sign stimuli test setting, however, there were two different places for the category boundary, only one of which was found within the signer group. The ability of the naïve informants to categorise visual stimuli links the results of this study to a larger scope of human capacity to categorise between different kinds of things, such as colours or objects (Palmer, 1999 for colours; Biederman, 1985 for objects). Both places for the category boundary found within the results are visually prominent, and the other one has prominent features anatomically as well. The latter is also a boundary within the structure of FinSL. The place for category boundary found only in the two hearing groups is in the middle of the visual continuum and thus divides the visual continuum with anatomical boundaries in two half. This place is found in one of the interpreters, but not in the other, who divided the continuum similarly to the signer group informants. The nature of this latter place of category boundary cannot, however, be fully understood as being of a purely linguistic nature since it is an anatomically prominent as well. This could partly explain also the fact that it can be found in every one of these groups' results.

The results obtained in part coincide with those of previous research groups, e.g. those obtained at the Salk Institute of Biological Studies. They also found similar effects in categorisation of places of articulation in sign using different continua of places of articulation (Emmorey, 2002). There are, however, differences between the findings of our group and the Salk Institute of Biological Studies findings concerning handshapes. This may be due to bigger differences in the continua used on one hand and on the other hand it may be due to their group using continua consisting of whole signs and ours consisting of purely handshapes detached from whole signs. That is, our stimuli are solely and strictly based on individual phonetic feature categories while the Salk Institute group continua consisted of meaning-carrying whole signs. Emmorey and others state that there are no differences in categorisation of handshapes and places of articulation in sign, but our findings have revealed a discrepancy between categorisation of handshapes and places of articulation in such that handshapes are treated similarly to vowels and this part of the study shows that places of articulation are treated similarly to consonants in the context of a behavioural test setting. There might also be a distinction based on whether the POA continuum used has body contact or not. This issue has, however, not yet been tested but remains for further investigation at later stages.

In the consonant continuum there was just one single category boundary, which could be expexted according to the categorical perception theory for consonants and the fact, that all the informants participating in that setting were native speakers of the same language. Different results might have been obtained from speakers of a language different from the particular one in question. Results from continua with different kinds of stimuli resemble each other both in steepness and compactness in the places of category boundaries. This further strengthens the previously stated similarity of speech and signs as different manifestations of one and the same processing system of human language. Previous findings from different studies comparing sign and speech also demonstrate that the results from these two language modalities coincide closely. This fact might suggest that sign language is less special than widely assumed. The 'specialness' of sign language, it seems, is maybe that it is not so special at all (also Meier, 2002 among others and in various chapters within this thesis).

The obtained results give us a starting point when continuing with further investigations within the phoneme group inventory of signed languages. As so far we have obtained results where handshapes behave like vowels and places of articulation behave like consonants, we want to study whether movements are similar in sign and speech. Also the function of orientations in sign language needs further research. So far, we have managed to establish a solid method for testing both sign and speech with analogous test batteries. These preliminary results obtained so far must also be further strengthened with more informants with different backgrounds, one possible group could be bilingual sign language interpreters. That would enable us to study the elicited results across different languages within one individual.

We carried out this place of articulation perception test set for two main reasons: first of all we wanted to test the infrastructure of the test battery for both consonant and vowel stimuli continua and furthermore, we wanted to operate on two different kinds of stimulus continua in both acoustic and visual modalities using the same procedure and the same software for presentation of the stimuli to participants. This two-fold study design on place of articulation in speech and sign enabled us to state that these study procedures can be used in two language modalities and are analogous in function to each other. We also demonstrated that places of articulation are treated the same way in both modalities used in language, sign and speech. This is also a further statement of the unification of human language as one phenomenon with two different manifestations and not an example of two different and separate language systems according to modality. The perception of handshapes and POAs in sign or vowels and consonants in speech is individual, and thus we have to gather information on the area of unanimity so that we can build applications with ability to ignore minute details in concordance with the quantums of perception (in speech, please see Stevens, 2000).

3.3 Individuality of Sign

In this chapter we looked at sign and more precisely at how individuality is represented in both production and perception of sign from various viewpoints, each discussed in separate chapters and subsections respectively.

We started with a broader scope of investigating sign production, namely looking at rhythmical elements in relation to non-manual and manual articulation of signing. These were found to be interconnected with non-manual articulation being slower and adjacent to both pragmatical and grammatical junctures while manual articulation was faster and thus formed a basis to which non-manuals were anchored in cyclical manner. Furthermore, we found head nods to be connected to the acceleration phases in manual articulation, that is, to such a phase that hands are moving faster than they were. In the second part of investigating sign production we studied the elements of governing coarticulatory patterns in both hands and within a single hand. We found that the hands moved slower when close together and furthermore that the index fingers were the governing element within hands. Sign production was found to be highly individual just as speech production is.

From studies of sign production we moved on to investigating individuality of sign perception. For that purpose we especially designed behavioural test sets that were as closely similar to speech perception studies as possible. Both speech and sign perception tests within this thesis consisted of one continuum in articulation with only one variable parameter in each of them. We treated handshapes equal to vowels and places of articulation equal to consonants. This was because handshapes do not have fixed targets but vary in a continuous fashion from one to another. On the other hand, we chose a place of articulation with a fixed articulatory target, that is, a contact to another part of body. The articulatory contact between two articulators connects us to the way that consonants are uttered in speech production. In the handshape perception study we found that both native signers and naïve informants behaved differently from sign language interpreters. In other words, there might be something within the foreign language learning process that alters the perception of natural continuum. It also revealed a difference in place of the category boundary based on the linguistic background of the informants, whether the continuum in question had a phonetic significance to them or not. And similarly to the consonant perception test within chapter 2, the perception patterns in place of articulation continuum differed from those of the first study.

All in all, both language modalities seem to have common tendencies both in production and perception processes. Furthermore, the design of experiments, use of stimuli, individual results and group findings all seem to converge into something that is within all of us as an all-engulfing trait of behaviour. As we now can perceive, one can not completely separate speech production and perception nor can one separate sign production completely from sign perception. In such a way speech and sign as methods of human communication show similar traits of behaviour as hierarchical systems.

4 General Conclusions on Individuality of Speech and Sign Production and Perception

Having overcome the difficulty presented by design of analogous test batteries for two different modalities and taken account of different properties of the modalities stated for the communication methods the following conclusions can be drawn. However distinct the two modalities may appear at the early stages of research, they still can be studied using analogous test batteries. Of course it goes without saying, that the different properties and boundary conditions imposed by the given modality have to be taken into account when designing the perceptual test sets and the different languages, their fundamental principles of course dictate the task design for the production test set. This provides us means for investigating the phonetic problem further stated in the aims within this thesis.

The re-occurring theme emerging from the various results to different test sets, be they productive or perceptive test batteries, has been the individuality of the results. Similar to the fact, that we do not look identical, we also don't act identically with regards to language but each and every individual has their own ways to produce and perceive speech and sign. Partly this is due to individual differences in physiology, but it also has strong ties to the communities we live in. So, these results support both sides of the well-known "nature versus nurture"debate. Maybe the theme for the debate should perhaps be altered into a "nature and nurture" debate?

Whatever the experimental setting may be, nevertheless, the introduction and procedures of the test batteries must be given in the language of the individual and thus, during this process multiple languages have been used when giving instructions and receiving feedback from the individuals that have participated in the various experiments. It has previously been shown, in relation to the Turku Vowel Test set that the language in which the instruction is given, might affect the results, i.e. shift the results towards the language the instruction is given in. This is the case especially with informants fluent in multiple languages and with deeper inner knowledge of differences between sound systems of those particular languages (Savela et al., 2008). This is a phenomenon similar to the fact that language use is altered when talking with non-native language users.

All experiments in this thesis strive to investigate the nature of human communication. Acoustic recordings of infant cries and their measurements showed that the seeds for the individuality of speech sound are present very early on in the human development. The individuality is present from birth, as can be observed from the patterns in the acoustic study of cry sounds. Crying is already such a multidimensional act of coordinated articulation that it also brings forward the individual differences within the cries. So, the individual differences within acoustical patterns of cries in the first months and stages of human development are brought forward to emphasise that not only speech but also the way we cry demonstrates individual differences – and even though it has not been scientifically studied, the hypothesis can be made that the differences between human cries, and the individual characteristics within cry patterns are further dispersed within the acoustic space when we grow older. Furthermore, the experiments on acoustic quality changes during and after maxillofacial changes due to surgery showed that learning articulation patterns does not stop in childhood but the ability to learn something new is still available to adults. Individuality in production gives reason to believe that perception is also individual and thus is able to normalise across different speakers. Tests with vowel and consonant continua state the case in favour of this individuality also on the perceptual side of communication. This is true not only for the percept as a whole but also there is individuality in perceiving the segments within the whole – individuality especially in vowel continuum perception is a well-stated case, from research starting from the 1960's onwards. Then again, consonants do seem to behave differently, so there is a discrepancy between identification patterns for vowels and consonants. Their production processes differ greatly – while consonants almost always involve some contact, either direct (plosives) or indirect (fricatives, laterals, nasals) vowels are without exception articulated without contact, solely based on proprioception of the individual.

Thus, now, a more precise model of human communication methods and languages can be presented (Figure 4.1).

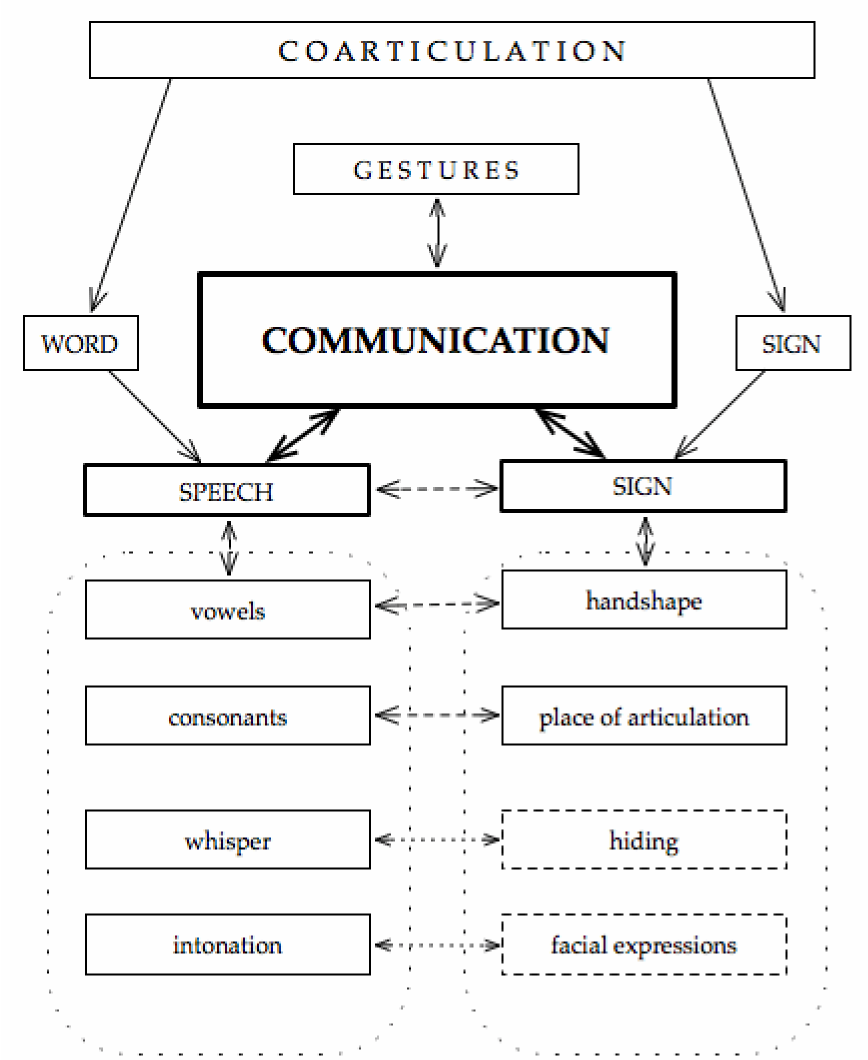


Figure 4.1. A more detailed version of the model for human communication in different modalities and the relations within and across communication methods.

Next we take a detour to the outskirts of phonetics – we explore sign languages, which are conveyed from hands via eyes into the same auditory cortex as speech is. We started by recording the production of informants and investigating the production of sign languages and indeed, we found similar results – individual signs are produced slightly differently according to the context and individual preferences. Also the rhythm patterns of sign are similar to speech: there are several intertwining rhythms present simultaneously. In these premises both speech and sign resemble the melody and rhythmical patterns present in music. Music and vocal singing has also been proposed to be the source of the human language ability (Mithen, 2006). The rhythm textures of music can also be perceived by those who cannot hear with their ears through vibrations in different parts of their body (Palmer & Lahtinen, 2005). This ability of discriminating between different stimuli using touch is also exploited when communicating with deafblind persons (Lahtinen, 2008). Next we experimented on perception of different phoneme continua in sign equivalent to vowel and consonant continua in speech. Individuality was the recurrent theme of the results also in the visual modality when language and communication was concerned.

Individuality in sign – and in gestures even more so – is highlighted further still. Partly it could be because there are no standardised language models for any signed language in the world in the same way as there are for spoken languages. The lack of standardised language models might also be due to lack of a written form associated with the languages. Standardised language models and written languages are tools for finding compromises in conversations and they also act as magnets keeping the individual differences more restricted.

4.1 Implications of Current Research for Design of Applications

There is a trend towards developing new technologies for communication in the field of speech and language technology. As Finland develops into an information society, there is also a need for speech-to-sign applications, this would enable signers to participate within the society more fully, bring information and services closer to them, since when there is a different language in use, there is always a communication barrier which has to be overcome. This is all the more true when the languages do not share the same modality but are conveyed via different systems. Design of speech-to-sign or equivalent applications would promote equality between signers and speakers. One simple example of this is the loudspeaker announcements at train stations – they cannot be perceived by signers but they have to find alternative ways to get that information. One solution to this would of course be captioning of the announcements, but that too wouldn’t be adequate, because Finnish is always a foreign language to signers and the reading skills are affected by the language proficiency of the signers in Finnish.

We strived to design an analogous test set battery for both speech and sign languages based on the recent findings in neuropsychology stating that the brain does not care where the linguistic information comes from – it is still processed in the secondary auditory cortex traditionally thought to process only speech. Succeeding in providing the test batteries with a common core and fundamental design has given us partial answers concerning human language ability. Furthermore, we have been able to design tests such that they have implications for the design of sign applications based on solutions from speech applications. Designing analogous test batteries for languages in different modalities also provides us with clues to similarities beyond the surface structures that different modalities provide us with.

Sign language data also provides a testing ground for speech-based applications, whether they are able to have linguistic data in different modality as the input for their software. Besides not only for speech-based applications, also movement clustering projects could benefit from handling sign language data in testing the validity of the programs.

As this thesis has mainly concentrated on the segmental issues in sign languages, the next step towards an all-inclusive study of sign languages is studying sign language prosody, a more thorough approach to rhythm texture and intensity patterns when signing. The first steps have been taken towards studying sign flow, i.e. continuous signing in the part where we studied the rhythm in non-manual and manual coarticulatory patterns, but the focus was mainly on the segmental level and not on a broader domain. This approach would enable us to model the prosody in sign languages through similar test batteries for prominence and focus in prosody – whether prosody is manifested similarly in sign and speech. Some suggestions are brought by observing signing, but this approach has not yet been scientifically tested.

This thesis aimed at providing basic knowledge about sign language parameters on the phonetic level, i.e. on how individuals produce and perceive sign language. This was accomplished by designing identification and discrimination tests for signers. This provided us answers based on the results of analogous test settings and the conclusion is that the results are parallel to each other, i.e. that both modalities yield similar answers to analogous study questions. This has lead us to providing basic input for sign language applications, i.e. signing avatars and also has given us answers as to the questions on precision of animation and intelligibility for the users.

The phonetic tests using analogous continua in sign and speech and their individual results have provided us tools for fundamental design parameters of applications language-independently, that is, knowledge about what are the most salient handshapes and places of articulation along with the seeds of knowledge relating to the motion envelopes of sign languages compared to the coarticulatory patterns in speech. This also has provided us the basic rules which in the later stages of application development can further be specified according to the particular languages used, whether signed or spoken.

The state-of-the-art of existing sign language applications enables us to pick and choose the best developed solutions to basic questions when designing sign-to-speech applications. This development requires adapting and incorporating parts of the existing solutions – or alternatively adapting and incorporating ideas from those existing solutions – to be used for a FinSL application. The quest for designing what can be basically called a machine translation system is multi-fold and each and every partial solution brings us nearer to the ultimate application. There are several parallel study programs going on around the world and thus the need for international co-operation can not be emphasised enough.

As both results, coming from sign and speech, yield to parallel results, i.e. individuality and stability of the phonetic categories, it points towards the common origins of the human evolution of communication. These results give further support to the idea, that speech and sign do have a common origin, but what it actually is, remains under further investigation. So, for part, this thesis does not only answer questions but also states new ones, maybe providing alternative and new points of view for research on human evolution and the language ability per se.

In conclusion, it is easy to find similarities between speech and signing. The main problems arise from the fact that the methods of research are not stable and standard, but each and every scientist has to develop his/her own methods of research. Neither are there unified standards for terminology which is unlike speech research having a well-established terminology and methodology for phonetic transcription and description as well as methods of study. Despite arisen problems of architecture, design and methodology the comparison of sign and speech is a future vista for phonetics and computer science, which we strive to travel towards solving the problems together with researchers with multiple disciplines and background. This is a truly integrative issue for research.

5 Future vistas

What would the world look like if the obstacles stated above within the introduction of this thesis had been solved altogether? How would people communicate? If individuality could be taken into account when developing user interfaces, what would the programs be like? Would they be capable of learning individual ways of information input and output?

What would the world be like if all the problems stated during this thesis had been solved? How would that world differ from this present one? First of all, it would be more equal between different kinds of languages whatever the modality. Second, people would not be as categorised and divided by the languages they might use as we nowadays are. The lay user of information solutions, such as search engines, info desk applications etc. would perhaps not notice the change at all. Why? Because all the information he might need would be provided in text format anyway. The biggest changes would be noticeable to those individuals who have some other ways for communication or information retrieval than the majority of us. That is if the person is a signer or uses e.g. signed Finnish there would be new ways for information searching, retrieval and providing. That would also provide the solution for the problem stated on page 2 of this thesis about translation of written documents. After those problems stated above were solved the invisible yet existing language barriers would cease to exist at last and for good. That would promote ultimate equality within the information society.

In the future it would be possible to search for information by signing to a computer which would then recognise what has been signed, convert that into e.g. a Boolean search, get results in e.g. text format and then convert those results into a signed avatar input which would then be provided to the person searching for information as a signed utterance. This is not to say that the signers couldn’t search information using e.g. written Finnish, but it is always most natural to use your own native language for information search and retrieval, much more so than any foreign language. And furthermore, who knows, maybe in the future it would be possible to write sign languages. The above-stated scenario would also be possible to implement using spoken document retrieval procedure; only that speech would be replaced with signing as the input method. Part of the problematics for signed document retrieval can be found within the study of spoken language document retrieval - lack of invariance, individuality, coarticulation, signal to noise ratio (SNR), just to name a few of the problems as yet only partially solved.

The problems and themes introduced in this thesis are a collection of seeds for ideas that have to be looked into in greater accuracy and detail when starting the development for sign interfaces. And not only those but some of the issues raised in this thesis can also benefit from doing further studies on different components of either sign or speech. Rhythm is one of such area combining both of the language modalities. The notion of rhythm and coarticulation and their function within the language modalities is not yet fully utilised within speech and sign production and perception testing. There are thus further studies that have to be made before anything more specific can be stated about the overall function of rhythm and coarticulation in communication.

List of Related Original Articles

Laaksonen J.-P., Ojala, S., Aaltonen, O., Niemi, M. & Happonen, R.-P. (2005). Effects of reduced orosensory feedback on spectral characteristics of medioalveolar /s/. *The Journal of the Acoustical Society of America*, 117:4(2), 2571.

Niemi, M., Laaksonen, J.-P., Ojala, S., Aaltonen, O. & Happonen, R.-P. (2006). Effects of transitory lingual nerve impairment on speech: an acoustic study of sibilant sound /s/. *International Journal of Oral and Maxillofacial Surgery* 35, 920-923.

Ojala, Stina (2001). Leukaluukiinnitteisten hammasimplanttien vaikutus suomen kielen medioalveolaariseen frikatiiviin. In Ojala, S. & Tuomainen, J. (eds.) *21. Fonetiikan Päivät 2001.* Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 67, 97-104.

Ojala, S. (2009). Rhythm and Coarticulation in Sign. *Proceedings of the Workshop in Non-manuals in Sign Languages*, Frankfurt-am-Main 4.-5.4. 2009. (in press).

Ojala, S. & Aaltonen, O. (2002). Optinen fonetiikka tutkimuskohteena. Teoksessa Korhonen, P. (toim.) *Fonetiikan päivät 2002.* Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing report 67, 61-66.

Ojala, S. & Aaltonen, O. (2004). Sign Language Phonetics. In Peltola, M.S. & Tuomainen, J.J. (eds.) *Studies in Speech Communication. Dedicated to Ilkka Raimo on his 65th birthday.* Publications of the Department of Finnish and General Linguistics of the University of Turku 72, 61-72.

Ojala, S. & Aaltonen, O. (2004). Similarities and differences of signed and spoken language processing. In Seppänen, T., Suomi, K. & Toivanen, J. (eds.) *Fonetiikan päivät 2004*, 7-9.

Ojala, S. & Aaltonen, O. (2007). Speech and sign – it's all in the motion. *Proceedings of ICPhS XVI 2007 in Saarbrücken*, 2169-2172.

Ojala, S. & Aaltonen, O. (2007). Puheen ja viittomien suhde: foneettinen tutkimus. *Puhe ja kieli* 27:3, 99-107.

Ojala, S., Aaltonen, O. & Salakoski, T. (2008). Investigating perception of places of articulation in sign and speech. *Proceedings of Interspeech'08 in Brisbane, Australia*.

Ojala, S. & Alivuotila, L. (2009). Itkua ja hammastenkiristystä – fonetiikan kliinisiä tutkimuskohteita. In Aaltonen, O., Iivonen, A., Aulanko, R., Klippi, A. Vainio, M. (eds.) *Puhuva ihminen.*Otava.

Ojala, S., Salakoski, T. & Aaltonen, O. (2008). Viittomien koartikulaatiosta. In O'Dell, M. & Nieminen, T. (eds.) *Fonetiikan päivät 2008.* Tampere Studies in Language, Translation and Culture, Series B 3, Tampere: Tampere University Press, 139-146.

Ojala, S., Salakoski, T. & Aaltonen, O. (2009). Coarticulation in sign and speech. *Proceedings of NoDaLiDa 2009, Odense, Denmark*. Electronically available at http://hdl.handle.net/10062/9208

Rautava, L., Lempinen, A., Ojala, S., Parkkola, R., Rikalainen, H., Lapinleimu, H., Haataja, L., Lehtonen, L., & PIPARI Study Group (2007). Acoustic quality of crying in premature infants at the age of 1½ years. *Early Human Development* 83, 5-12.

Savela, J., Ojala, S., Aaltonen, O. & Salakoski, T. (2007). Role of different spectral attributes on vowel perception: the case of Udmurt. *Proceedings of NODALIDA 2007 in Tartu*, 384-388.

Zachau, S. & Ojala, S. (2010). Viittomakieli. In Korpilahti, P., Aaltonen, O. & Laine, M. (eds.) Kieli ja aivot. (in press)

References

Aaltonen, O., Eerola, O., Hellstrom, A., Uusipaikka, E. and Lang, A.H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data, *Journal of the Acoustical Society of America* 101(2), 1090-105.

Ann, J. (1996). On the relation between ease of articulation and frequency of occurrence of handshapes in two sign languages. *Lingua* 98, 19-41.

Abler, W.L. (1989) On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures* 12, 1-13.

Aittokallio, K. (2002). *Etinen, terävä [s] medioalveolaarisen /s/-äänteen varianttina*. Unpublished pro gradu -thesis. University of Turku.

Amaral, (2000). The anatomical organization of the central nervous system. In Kandel, E.R., Schwartz, J.H. & Jessell, T.M. (eds.) *Principles of Neural Science*. McGraw-Hill,

Apgar, V. (1953). A Proposal for a New Method of Evaluation of the Newborn Infant. Current issues in Anesthesia and Analgesia 32, 260-267.

Arbib, M.A. (2002). The mirror system, imitation, and the evolution of language. In Nehaniv, C. & Dautenhahn, K. (eds.). *Imitation in animals and artifacts*. MITPress, 229-280.

Arbib, M.A. (2003). The Evolving Mirror System: A Neural Basis for Language Readiness. In Christiansen, M.H. & Kirby, S. (eds.) *Language Evolution*. Oxford University Press, 182-200.

Armstrong, D.F., Stokoe, W.C. &Wilcox, S.E. (1997). *Gesture and the Nature of Language*. Cambridge University Press. Cambridge, UK.

Aziz-Zadeh, L, Wilson, S.M., Rizzolatti, G. & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology* 16, 1818-1823.

Baker, C & Cokely, D (1980). *American sign language: A teacher's resource text on grammar and culture*. TJ Publishers.

Bastian, J., Delattre, P.C. & Liberman, A.M. (1959). Silent interval as a cue for the distinction between stops and semivowels in medial position. *Journal of the Acoustical Society of America* 31, 1568.

Bastian, J., Eimas, P.D. & Liberman, A.M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval (Abstract). *Journal of the Acoustical Society of America* 33, 842.

Bauer, B., Nießen, S. and Heinz, H. (1999). Towards an Automatic Sign Language Translation System. In *Proceedings of the International Workshop on Physicality and Tangibility in Interaction: Towards New Paradigms for Interaction Beyond the Desktop,* [no page numbers], Siena, Italy.

Baumel, J.J. (1974). Trigeminal-facial nerve communications: Their function in facial muscle innervation and reinnervation. *Archives Otolaryngology* 99, 34-44.

Bayley N. (1993). *Bayley scales of infant development manual*. 2nd edition. San Antonio Psychological Corporation.

Bee, M.A. (2006). Individual Recognition in Animal Species. In Brown, K. (ed.) *Encyclopedia of Language and Linguistics*, 617-626. Elsevier.

Behnia, H., Kheradvar, A. and Shahrokhi, M. (2000). An anatomical study of the lingual nerve in the third molar region. *Journal of Oral and Maxillofacial Surgery* 58(6), 649-653.

Bellugi, U. (1980). Clues from the structural similarity of sign and spoken language. In Bellugi, U. & Studdert-Kennedy, M. (eds.) *Signed and spoken language: Biological constraints on linguistic form.* Weinheim.

Biederman, I. (1985). Human image understanding: Recent research and a theory*, Computer Vision, Graphics, and Image Understanding* 32, 29-73.

Blondel, M. & Miller, C. (2001). Movement and Rhythm in Nursery Rhymes in LSF. *Sign Language Studies* 2(1) 24-61.

Boersma P, Weenink D. (1992-2009). Praat: doing phonetics by computer. Available at: http://www.praat.org/.

Borden G.J., Harris K.S., & Catena L. (1973). Oral feedback II. An electromyografic study of speech under nerve-block anesthesia. *Journal of Phonetics* 1, 297–308.

Boulenger, V., Hauk, O. & Pulvermüller, F. (2009). Grasping Ideas with the Motor System: Semantic Somatotrophy in Idiom Comprehension. *Cerebral Cortex* 19, 1905-1914.

Braffort, A. (1996). ARGo: An architecture for sign language recognition and interpretation. *Proceedings of Gesture Workshop '96*, 17-30.

Braun, A.R, Guillermin, A., Hosey, L & Varga, M. (2001). The neural organization of discourse. An $H_2^{15}O$-PET study of narrative production in English and American sign language. *Brain* 124, 2028-2044.

Brentari, D. (1995). *A Prosodic Model of Sign Language Phonology*. MIT Press.

Broca, P. (1865). Sur le siege de la faculté du langage articulé. *Bulletin de la Société d'Anthropologie* 6, 377-393.

Brodmann, K (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prizipien dargestellt aif Grund des Zeelenbaues*. Barth.

Browman, C.P. & Goldstein, L. (1992). Articulatory Phonology: an overview. *Haskins Laboratories status Report on Speech Research* SR-111/112, 23-42.

Cacace, A.T., Robb, M.P., Saxman, J.H., Risemberg, H., Koltai, P. (1995). Acoustic features of normal-hearing pre-term infant cry. *International Journal of Pediatric Otorhinolaryngology* 33, 213 –24.

Campbell, R. & Wright, H. (1988). Deafness, spelling and rhyme: how spelling supports written word and picture rhyming skills in deaf subjects. *Quarterly journal of experimental psychology. A. Human experimental psychology* 40(4), 771-788.

Campbell, R., MacSweeney, M. & Waters, D. (2007). Sign Language and the Brain: A Review. *Journal of Deaf Studies and Deaf Education* 13(1), 3-20.

Capek, C.M., MacSweeney, M., Woll, B., Waters, D., McGuire P.K., David, A.S., Brammer, M.J. & Campbell, R. (2008). Cortical circuits for silent speechreading in deaf and hearing people. *Neuropsychologia* 46, 1233–1241.

Caplan, D. (1987). *Neurolinguistics and Linguistic Aphasiology: An Introduction*. Cambridge University Press.

Carlson, R., Granström, B. and Klatt, D. (1979). Vowel perception: the relative perceptual salience of selected acoustic manipulations, *TMH-QPSR* 3-4 (Department of Speech, Music and Hearing Quarterly Progress and Status Report, Kunglika Tekniska Högskolan).

Cassell, J. (2000). Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents. In Cassell, J. et al. (eds.), *Embodied Conversational Agents*, 1-27. MIT Press.

Cassell, J. (2006). Conversational Agents, Synthetic. Encyclopedia of Language and Linguistics. Elsevier.
Cassell, J., Bickmore, T., Campbell, L., Vilhjámsson, H. & Yan, H. (2001). More than just a pretty face: Conversational protocols and the affordances of embodiment. Knowledge-Based Systems 14, 55-64.
Clement C.J., Koopmans-van Beinum F.J., Pols L.C.W. (1996). Acoustical characteristics of sound production of deaf and normally hearing infants. *Proceedings of the Fourth International Conference on Spoken Language Processing, Oct 3—6 1996*; Philadelphia, PA, USA.
Chomsky, N. (1957). *Syntactic Structures*. Mouton Publishers.
Condillac, E.B. de (1947 [1746]). Essai sur l'origine des connaissances humaines, ouvrage ou l'on réduit à un seul principe tout ce concerne l'entendement. In *Oeuvres Philosophiques de Condillac*. Georges Leroy.
Carney, A.E. (1988). Vibrotactile perception of segmental features of speech: A Comparison of single-channel and multichannel instruments. *Journal of Speech and Hearing Research* 31, 438-448.
Choo, W. & Huckvale, M. (1997). Spatial relationships in fricative preception. In Smits, R., Rosen, S. & Holland, M. *Speech, Hearing and Language: work in progress 1997*, volume 10.
Corina, D.P., Neville, H.J. & Bavelier, D. (1998). Response from Corina, Neville and Bavelier. *Trends in Cognitive Sciences* 2, 468–470.
Corina, D.P. (1998). Aphasia in users of signed language. In Coppens P, Lebrun Y and Basso A (Eds), *Aphasia in Atypical Populations*. Lawrence Erlbaum.
Corwin M.J., Lester, B.M., Sepkoski, C., McLauglin, S., Kayne, H., Golub, H. (1992). Effects of in utero cocaine exposure on newborn acoustical cry characteristics. *Pediatrics* 89(6), 1199-1203.
Corballis, M.C. (2002). *From hand to mouth: The origins of language*. Princeton University Press.
Coulmas, F. (1996/1999). The Blackwell Encyclopedia of Writing Systems. Blackwell Publishers.
Crasborn, O. (2001). *Phonetic implementation of phonological categories in Sign Language of the Netherlands.* Academic dissertation, Leiden University.
Damasio (1992). Aphasia. *New England Journal of Medicine* 326, 531-539.
Damasio, H., Grabowski, T.J., Tranel, D., Hichwa, R. & Damasio, A.R. (1996). A neural basis for lexical retrieval. *Nature* 380, 499-505
Darainy, M., Towhidkhah, F. & Ostry, D.J. (2007). Control of Hand Impedance Under Static Conditions and During Reaching Movement. *Journal of Neurophysiology* 97, 2676-2685.
Darling, W.G., Cole, K. & Miller, G.E. (1994). Coordination of index finger movements. *Journal of Biomechanics* 27(4), 479-491.
DeValois, K.K. and DeValois, R.L. (1990). *Spatial Vision*. Oxford Science Publications 14. Oxford University Press.
Dolata, J.K., Davis, B.L., and MacNeilage, P.F. (2008). Characteristics of the rhythmic organization of vocal babbling: implications for an amodal linguistic rhythm. *Infant Behavior & Development* 31, 422-431.
Eimas, P.D. (1963). The relation between identification and discrimination along speech and non-speech continua. *Language and Speech* 6, 206-217.

Emmorey, K., Grabowski, T., McCullough, S., Damasio, H., Ponto, L., Hichwa, R. & Bellugi, U. (2004). Motor-iconicity of sign language does not alter the neural systems underlying tool and action naming. *Brain and Language* 89, 27–37.
Emmorey, K. (2002). *Language, Cognition and the Brain: insights from signed language research*. Lawrence Erlbaum Associates.
Emmorey, K., Damasio, H., McCullough, S., Grabowski, T, Ponto, L., Hichwa, R. and Bellugi, U. (2002). Neural Systems Underlying Spatial Language in ASL. *NeuroImage* 17, 812-824.
Fisichelli, V.R. & Karelitz, S. (1963). The cry latencies of normal infants and those with brain damage. *Journal of Pediatrics* 62, 724-734.
Fisher, C.G. (1968). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research* 11, 796-804.
Fisher, S.E., Vargha-Khadem, F., Watkins, K.E., Monaco. A.P. & Pembrey, M.E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics* 18, 168–170
Fitch, W.T., Hauser, M.D. & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition* 97, 179–210
Flipsen, P., Shriberg, L., Weismer, G., Karlsson, H., & McSweeny, J. (1999). Acoustic Characteristics of /s/ in Adolescents. *Journal of Speech, Language, and Hearing Research* 42, 63–677.
Frisone M.F., Mercuri E., Laroche, S., Foglia, C., Maalouf, E.F., Haataja L, Cowan, F. & Dubowitz, L. (2002). Prognostic value of the neurologic optimality score at 9 and 18 months on preterm infants born before 31 weeks gestation. *Journal of Pediatrics* 140(1), 57–60.
Forrest, K., ,Weismer G., Milenkovic, P. & Dougall, R.N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *Journal of the Acoustical Society of America* 84, 115–123.
Fouts, R. & Mills, S.T. (1997/1999). *Simpanssini Washoe – yhteinen matkamme.* Translated into Finnish from original English title: *Next of Kin – What Chimpanzees Have Taught Me About Who We Are*. WSOY.
Fu, H., Rodman, R.D., McAllister, D.F., Bitzer, D.L. & Xu, B. (1999). Classification of voiceless fricatives through spectral moments. *Proceedings of the 5th International Conference on Information Systems Analysis and Synthesis (ISAS'99)*, 307-311.
Fuchs, B. (2004). *Phonetische Aspekte einer Didaktik der Finnischen Gebärdensprache als Fremdsprache*. Academic dissertation, University of Jyväskylä. Jyväskylä Studies in Humanities 21. Jyväskylä University Printing House.
Galaburda, A.M. (1994). Developmental dyslexia and animal studies: at the interface between cognition and neurology. *Cognition* 50, 133-149.
Gao, W., Fang, G., Zhao, D., and Chen Y. (2004). A Chinese sign language recognition system based on SOFM/SRN/HMM. *Pattern Recognition* 37, 2389-2402.
Gass, S.M. (1997). *Input, interaction, and the second language learner*. Lawrence Erlbaum Associates.
Ghez, C. & Krakauer, J. (2000). Organization of movement. In Kandel, E.R., Schwartz, J.H. & Jessell, T.M. (eds.) *Principles of Neural Science*. McGraw-Hill: New York.
Gibson, K., and T. Ingold (eds.) 1993. *Tools, Language and Cognition in Human Evolution*. New York: Cambridge University Press.

Gilbert, H.R. & Robb, M.P.(1996). Vocal fundamental frequency characteristics of infant hunger cries: birth to 12 months. *International Journal of Pediatric Otorhinolaryngology* 34, 237-243.

Goberman, A.M. & Robb, M.P. (1999). Acoustic examination of preterm and full-term infant cries: the long time average spectrum. *Journal of Speech, Language and Hearing Research* 42, 850-861.

Goldin-Meadow, S. (2003). *The Resilience of Language: what gesture creation in deaf children can tell us about how all children learn language*. Essays in Developmental Psychology. Psychology Press.

Golub, H.L. & Corwin, M.J. (1985). A physioacoustic model of the infant cry. In: Lester, B.M., Boukydis, C.F.Z. (eds.). *Infant crying:theoretical and research perspectives*. Plenum Press. pp. 59-82.

Golub, H.L & Corwin, M.J. (1982). Infant cry: a clue to diagnosis. *Pediatrics* 69(2), 197-201.

Grabe, E. (2002). Variation adds to prosodic typology. In B.Bel and I. Marlin (eds.), *Proceedings of the Speech Prosody 2002 Conference, 11-13 April 2002*, Aix-en-Provence: Laboratoire Parole et Langage, 127-132.

Grosjean (1977). The perception of rate in spoken and sign languages. Perception & Psychophysics 22 (4), 408-413.

Gunes, H., Piccardi, M. & Jan, T. (2004) Face and Body Gesture Recognition for a Vision-Based Multimodal Analyzer. In: *Tenth Annual ACM Symposium on Principles of Programming Languages*, 88—98.

Haataja, L., Mercuri, E., Regev, R., Cowan, F., Rutherford, M., Dubowitz, V., et al. (1999). Optimality score of the neurological examination of the infant at 12 and 18 months of age. *Journal of Pediatrics* 135, 153-161.

Haataja, L., Mercuri, E., Guzzetta, A., Rutherford, M., Counsell, S., Frisone, M.F., et al. (2000). Neurological examination in infants with hypoxic-ischaemic encephalopathy at age 9—14 months; use of optimality scores and correlation with MRI findings. *Journal of Pediatrics* 138(3), 332-337.

Hall, D.A., Fussell, D.A. & Summerfield, Q. (2005). Reading Fluent Speech from Talking Faces: Typical Brain Networks and Individual Differences. *Journal of Cognitive Neuroscience* 1**7**:6, 939-953

Halle, M. (1964). On the bases of phonology. In Fodor & Katz, *The structure of language*, pp. 324–333. Prentice-Hall.

Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events*. MIT Press.

Haglund, J. & Evers, H. (1972). *Local anaesthesia in dentistry*. Södertälje Astra Läkemedel.

Hardcastle, W.J. (1975). Some aspects of speech production under controlled conditions of oral anaesthesia and auditory masking. *Journal of Phonetics* 3, 197–214.

Harnad, S. (1987). *Categorical perception.* Cambridge University Press.

Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The language faculty: What is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.

Hayes (1980). A metrical theory of stress rules. PhD thesis. Available as preview and non-printable format at http://hdl.handle.net/1721.1/16066. (Accessed 27.11.2010)

Hazen, T.J. (2006). Visual Model Structures and Synchrony Constraints for Audio-Visual Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 14(3), 1082-1089.

Hickok, G., Wilson, M., Clark, K., Klima, E.S., Kritchevsky, M. & Bellugi, U. (1999). Discourse Deficits Following Right Hemisphere Damage in Deaf Signers. *Brain and Language* 66, 233–248.

Hirn, D. F. (1910- 1916). *De dövstummas åtbördsspråk i Finland - Kuuromykkäin viittomakieli Suomessa.* Vihot I - III. Helsinki.

Hollien, H.F. (1990). *The acoustics of crime: the new science of forensic phonetics. Applied psycholinguistics and communication disorders*. Plenum Press.

Huenerfauth, M. (2006). *Generating American Sign Language Classifier Predicates For English-To-ASL Machine Translation.* Doctoral Dissertation, Computer and Information Science, University of Pennsylvania.

Hulst, H.G. van der (2006). Dependency Phonology. In: Keith Brown (ed.). *The encyclopedia of language and linguistics. 2nd edition. Volume 3*. Oxford: Elsevier, 451-458.

Hurst, J.A., Baraitser, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Developmental Medicine & Child Neurology* 32, 352–355

Ipp, M., Cohen, E., Goldbach, M., Macarthur, C. (2004). Effect of choice of measles—mumps—rubella vaccine on immediate vaccination pain in infants. *Archives of pediatrics & adolescent medicine* 158(4), 323-326.

Ito, M., Tsuchida, J. and Yano, M. (2001). On the effectiveness of whole spectral shape for vowel perception, *Journal of the Acoustical Society of America* 110(2), 1141-1149.

Iverson, P. and Kuhl, P.K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics* 62(4), 874-886.

Jantunen, T. (2009). *Tavu ja lause: tutkimuksia kahden sekventiaalisen perusyksikön olemuksesta suomalaisessa viittomakielessä [Syllable and sentence: studies on the nature of two sequential basic units in Finnish Sign Language].* Academic dissertation, University of Jyväskylä. Jyväskylä Studies in Humanities 117. Jyväskylä: University of Jyväskylä. [with English abstract]

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition* 24, 617–637.

Juang, B.H., Chou, W. & Lee, C.H. (1997). Minimum Classification Error Rate Methods for Speech Recognition. IEEE Transactions on Speech and Audio Processing, 5(3), 257-265.

Jurafsky, D. & Martin, J.H. (2000). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Prentice Hall.

Jusczyk, P.W., Culter, A., & Redanz, N. (1993). Preferences for the predominant stress patterns of English words. *Child Development* 64, 675-687.

Kandel, E.R., Schwartz, J.H. & Jessell, T.M. (eds.) (2000). *Principles of Neural Science*. McGraw-Hill.

Karelitz, S, Fisichelli, V.R. (1962). The cry thresholds of normal infants and those with brain damage. *Journal of Pediatrics* 61, 679 –85.

Karjalainen, M., Mäki-patola, T., Kanerva, A. & Huovilainen, A. (2006). Virtual Air Guitar. *Journal of Audio Engineering Society* 54(10), 964-980.

Keinänen, L. (2010). *Identtisten ja epäidenttisten kaksosten puhe - Monotsygoottisten ja ditsygoottisten kaksosten puheen vertailua kesto-, nopeus-, tauko- ja taajuusmittauksin*. Academic dissertation. Department of Speech Sciences, University of Helsinki.

Kendon, A. (1972). Some relationships between body motion and speech. In A. Siegman and B. Pope (eds.) *Studies in dyadic communication*, 177-210. Pergamon Press.
Kinnunen, T. (2005). Optimizing Spectral Feature Based Text-Independent Speaker Recognition. Academic dissertation, Department of Computer Science, University of Joensuu, Finland
Kiparsky, P. (1975). Stress, Syntax and Meter. *Language* 51(3), 576-616.
Kita, S., Van Gijn, I. & Van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. *Lecture Notes in Computer Science. Proceedings of the Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop, Bielefeld, Germany, September 1997*. Springer Verlag. Volume 1371/1998
Klatt, D.H. (1980). Software for a Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America*, 67(3), 974-995.
Klima, E.S. & Bellugi, U. (1976). Poetry and song in a language without sound. *Cognition* 4, 45-97.
Klima, E.S. & Bellugi, U. (1979). *The Signs of Language*. Harvard University Press.
Kokkonen, J. (1992). *Ihminen ja musiikki. Valittuja kirjoituksia, esitelmiä, puheita ja arvosteluja*. Gaudeamus.
Krakauer, J. & Ghez, C. (2000). Voluntary Movement. In Kandel, E.R., Schwartz, J.H. & Jessell, T.M. (eds.) *Principles of Neural Science*. McGraw-Hill: New York.
Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg U., Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684-686.
Kuhl, P. & Iverson, P. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? Perception and Psychophysics 62(4), 874-886.
Kuhl, P.K. & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. *Science* 218, 1138-1141.
Kuhl, P.K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology* 4, 812–822.
Laaksonen, J-P. (2006). *Control mechanisms of speech production. Evidence from Acoustic Studies of Speech after Neural and Muscular Manipulations of the Tongue.* Academic dissertation. Department of Oral and Maxillofacial Surgery. University of Turku, Finland. Annales Universitatis Turkuensis D730. Digipaino.
Ladefoged, P. & Maddieson, I. (1986). *Some of the Sounds of the World's Languages*, UCLA Working Papers in Phonetics 64.
Lahtinen, R. & Palmer, R. (2005). *Body Story. Creating Musical Images through Touch (CMIT)*. Cityoffset.
Lahtinen, R. (2008). *Haptices and Haptemes. A Case study of developmental process in social-haptic communication of acquired deafblind people*. Academic dissertation. Department of Special Education, University of Helsinki, Finland. Cityoffset.
Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., & Monaco, A.P. (2001). A novel forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519–523
Lai, C.S.L., Fisher, S.E., Hurst, J.A., Levy, E.R., Hodgson, S., Fox, M, Jeremiah, S., Povey, S., Jamison, D.C., Green, E.D., Vargha-Khadem, F. & Monaco, A.P. (2000). The SPCH1 region on human 7q31: genomic characterization of the critical interval and

localization of translocations associated with speech and language disorder. *American Journal of Human Genetics*. 67, 357–368.
Lehiste, I. (1972). Suprasegmentals.
Lemström, K. (2000). *String Matching Techniques for Music Retrieval*. Academic dissertation. Department of Computer Science, University of Helsinki. Series of Publications A, Report A-2000-4.
Lester, B.M. (1987). Developmental outcome from acoustic cry analysis in term and preterm infants. *Pediatrics* 80(4), 529-534.
Lester, B.M., Corwin, M.J., Sepkoski, C., Seifer, R., Peucker, M., McLaughlin, S., et al. (1991). Neurobehavioral syndromes in cocaine-exposed newborn infants. *Child Development* 62(4), 694– 705.
Lewkowicz, D.J. & Turkewitz, G. (1980). Cross-Modal Equivalence in Early Infancy: Auditory-Visual Intensity Matching. *Developmental Psychology* 6, 597-607.
Liberman, A.M. (1996) *Speech: a special code*. MIT Press.
Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review* 74, 431-461.
Liberman, A.M., & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition* 21, 1-36.
Liddell, S.K. (1980). *American Sign Language Syntax.* Mouton publishers.
Liddell, S.K. & Johnson, R.E. (1989). American Sign Language: The phonological base. *Sign Language Studies* 64, 195-277.
Lindblom, B. (1981). Economy of speech gestures. In MacNeilage, Peter (ed.) *The production of speech* pp. 217-246. Springer Verlag.
Lindblom, B. (1990). Explaining phonetic variation: A sketch of H & H –theory. In Hardcastle, W.J., Marchal, A. (eds.). *Speech production and speech modelling*, pp. 403-439. Kluwer.
Lindblom, B. (2000). Developmental origins of adult phonology: the interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica* 57, 297–314.
Lindblom, B., Mauk, C. & Moon, S.-J. (2006). Dynamic specification in the production of speech and sign. In Divenyi, P.L., Greenberg, S. & Meyer, G. (eds.). *Dynamics of Speech Production and Perception*. Volume 374 NATO Science Series: Life and Behavioural Sciences.
Loakes, D. (2008). A forensic phonetic investigation into the speech patterns of identical and non-identical twins. *International Journal of Speech, Language and the Law* 15:1, 97-100.
Loomis, J., Poizner, H., Bellugi, U., Blakemore, A. & Hollerbach, J. (1983). Computer graphic modelling of American Sign Language. *Computer Graphics* 17(3), 105-114.
López-García, A. (2005). *The grammar of genes: how the genetic code resembles the linguistic code*. P. Lang.
Malm, A. (2000). *Viittomakieliset Suomessa*. Finnlectura.
MacKenzie Beck J. (1997). Organic variation of the vocal apparatus. In Hardcastle, W.J. & Laver, J. (eds). *The Handbook of Phonetic Sciences*. Blackwell Publishers. pp. 256–297.
MacNeilage, P. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21, 499-511.
Maeda, S. (1990). Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In Hardcastle, W. and Marchal, A. (eds.). *Speech Production and Speech Modelling*. pp. 187-209. Kluwer.

Mauk, C. (2003). *Undershoot in two modalities: evidence from fast speech and fast signing*. Academic dissertation, Department of Linguistics, University of Texas, Austin.
Maunu, J., Kirjavainen, J., Rikalainen, H., Parkkola, R., Lapinleimu, H., Haataja, L. & the PIPARI Study Group (2006). Relation of Prematurity and Brain Injury to Crying Behavior in Infancy. *Pediatrics*, 118, e57-e65.
Martin, J. (2000). *A Linguistic Comparison: Two Notation Systems for Signed Languages: Stokoe Notation and Sutton SignWriting.*
Marcus, G.F. & Fisher, S.E. (2003). FOXP2 in focus: what can genes tell us about speech and language? *Trends in Cognitive Sciences* 7(6), 257-262.
McKee, R. & McKee, D. (1992). What's so hard about learning ASL? Students' & teachers' perceptions. *Sign Language Studies* 75, 129-157.
McNeill, D. (1985). So You Think Gestures Are Nonverbal? *Psychological Review* 92(3), 350-371.
McNeill, D. (1992). *Hand and Mind: what gestures reveal about thought.* University of Chicago Press.
MacSweeney, M., Woll, B., Campbell, R., Calvert, G.A., McGuire, P.K., David, A.S., Simmons A. & Brammer, M.J. (2002). Neural correlates of British Sign Language comprehension: Spatial processing demands of topographic language. *Journal of Cognitive Neuroscience* 14(7), 1064-1075.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
Meier, P. (2002). Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In Meier, R.P., Cormier, K. & Quinto-Pozos, D. (eds.) *Modality and Structure in Signed and Spoken Languages.* Cambridge University Press.
Michelsson K. (1971). Cry analyses of symptomless low birth weight neonates and of asphyxiated newborn infants. *Acta Paediatrica Scandinavica* Suppl 216, 10– 45.
Michelsson, K., Järvenpää, A.L. & Rinne, A. (1983). Sound spectrographic analysis of pain cry in preterm infants. *Early Human Development* 8, 141-149.
Michelsson K., Raes, J., Thoden, C.J., Wasz-Höckert, O. (1982). Sound spectrographic cry analysis in neonatal diagnostics. An evaluative study. *Journal of Phonetics* 10, 79-88.
Mithen, S. (2006). *Singing Neanderthals. The origins of music, language, mind and body.* Harvard University Press.
Moeslund, T.B., Störring, M and Granum, E. (2004). Pointing and Command Gestures for Augmented Reality. In *Proceedings of Pointing '04. Visual Observation of Diectic Gestures.*
Morrissey,S. (2008). *Data-Driven MachineTranslation for Sign Languages.* Academic dissertation. School of Computing, Dublin City University,
Möttönen, R., Krause, C.M., Tiippana K., & Sams M. (2002). Processing of changes in visual speech in the human auditory cortex, *Cognitive Brain Research* 13, 417–425.
Nartey, J.N.A. (1982). On fricative phones and phonemes: measuring the phonetic differences within and between languages. *UCLA Working Papers in Phonetics* 55.
Niemi, M., Laaksonen, J.-P., Aaltonen, O. & Happonen, R.-P. (2006). Effects of transitory lingual nerve impairment on speech: an acoustic study of diphthong sounds. *Journal of Oral and Maxillofacial Surgery* 62, 44–51.
Niemi, M., Laaksonen, J.-P., Vähätalo, K., Tuomainen, J., Aaltonen, O. & Happonen, R.-P. (2004) Effects of transitory lingual nerve impairment on speech: an acoustic study of vowel sounds. *Journal of Oral and Maxillofacial Surgery* 60, 647–652.

Niemi-Laitinen, T. (2001). Automatic speaker recognition – is it possible? In Ojala, S. & Tuomainen, J.: 21.Fonetiikan päivät, Turku 4.-5.1. 2001.
Nishimura, H., Hashikawa, K., Doi, K., Iwaki, T., Watanabe, Y., Kusuoka, H., Nishimura, T., & Kubo, T. (1999). Sign language 'heard' in the auditory cortex. *Nature* 397, 116.
Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge Studies in Speech Science and Communication. Cambridge University Press.
O'Dell, M. & Nieminen, T. (2001). Speech rhythms as cyclical activity. In Ojala, S. & Tuomainen, J. (eds.) *21. Fonetiikan päivät Turku 4.-5.1.2001* Publications of the Department of Finnish and General Linguistics of the University of Turku 67. pp. 159-168.
O'Dell, M. (2003). *Intrinsic Timing and Quantity in Finnish*. Academic dissertation. University of Tampere, Finland. Acta Universitatis Tamperensis 979. Tampereen yliopistopaino Juvenes Data.
Ojala, S. & Aaltonen, O. (2004). Phonetics of Sign Language. In Peltola, Maija S. & Tuomainen, J. (eds.). *Studies in Speech Communication. Dedicated to Ilkka Raimo on his 65th birthday.* Digipaino Oy.
Ojala, S. & Aaltonen, O. (2007). Speech and sign – it's all in the motion. *Proceedings of ICPhS XVI 2007 in Saarbrücken*, 2169-2172.
Ojala, S., Aaltonen, O. & Salakoski, T. (2008). Investigating perception of places of articulation in sign and speech. *Proceedings of Interspeech'08 in Brisbane, Australia, 2008*.
Ojala, S. & Salakoski, T. (2009). Coarticulation in sign and speech. *Proceedings of Nodalida 2009 in Odense, Denmark.*
Ong, S.C.W., Ranganath, S. and Venkatesh, Y.V. (2006). Understanding gestures with systematic variations in movement dynamics. *Pattern Recognition* 39, 1633-1648.
Paget, R.A.(1930). *Human Speech*. Routledge, reprint 2001.
Palmer, S.E. (1999). *Vision science: Photons to Phenomenology*. MITPress.
Patterson, F. & Linden, E. (1981/1986). *Koko – puhuva gorilla*. Translated into Finnish from original English title: The Education of Koko. Otava.
Peltola, L., Erkut, C., Cook, P.R. & Välimäki, V. (2007). Synthesis of Hand Clapping Sounds. IEEE Transactions on Audio, Speech, and Language Processing 15(3), 1021.
Perlmutter, D. M. 1992. Sonority and syllable structure in American Sign Language. Linguistic Inquiry 23:407–442.
Peters, C.L. (2000). *Deaf American Literature. From Carnival to the Canon.* Gallaudet University Press.
Petitto, L.A., Holowka, S., Sergio, L.E., Levy, B. & Ostry, D.J. (2004). Baby hands that move to the rhythm of language: hearing babies acquiring sign languages babble silently on the hands. *Cognition* 93, 43-73.
Petitto, L.A., Zatorre, R., Gauna, K., Nikelski, E. J., Dostie, D. & Evans, A.C. (2000). Speech-like cerebral activity in profoundly deaf people while processing signed languages: Implications for the neural basis of human language. *Proceedings of the National Academy of Sciences* 97(25), 13961-13966.
Poizner, H. (1983). Perception of movement in American Sign Language. *Perception and Psychophysics* 33(3), 215-231.
Poizner, H., Klima, E. S., & Bellugi, U. (1987). *What the Hands Reveal About the Brain*. MIT Press.
Porter, F.L., Porges, S.W. & Marshall, R.E. (1988). Newborn pain cries and vagal tone: parallel changes in response to circumcision. *Child Development* 59(2), 495–505.

Preston, P. (1994). *Mother, Father, Deaf. Living between Sound and Silence*. Harvard University Press.
Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Science* 5, 517-525.
Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews on Neuroscience* 6, 576-582.
Quek, F. (1994). Toward a Vision-Based Hand Gesture Interface. *Proceedings of .Virtual Reality Software and Technical Conference*, pp. 17-29.
Rabiner, L.R. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257-286.
Rainó, P. (2004) *Henkilöviittomien synty ja kehitys suomalaisessa viittomakieliyhteistössä*. Academic dissertation. University of Helsinki.
Rantala, J., Raisamo, R., Lylykangas, J., Surakka, V., Raisamo, J., Salminen, K., Pakkanen, T. & Hippula, A. (2009). Methods for Presenting Braille Characters on a Mobile Device with a Touchschreen and Tactile Feedback. *IEEE Transactions on Haptics* 2(1), 28-39.
Rapisardi, G., Vohr, B., Peucker, M. & Lester, B. (1989). Assessment of infant cry variability in high-risk infants. *International Journal of Pediatric Otorhinolaryngology* 17, 19– 29.
Ravila, P. (1937). Lasten kielen kehityksestä. *Länsi* 3(4), 139-146.
Repp, B.H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America* 81(4), 1100–1109.
Révész, G. (1991). *Introduction to Formal Languages*. McGraw-Hill: New York.
Rissanen, T. (1985). *Viittomakielen perusrakenne*. Yliopistopaino, Helsinki.
Rissanen, T. (1998). *Categories of nominals and verbals and their morphology in Finnish sign language.* [Unpublished licentiate thesis.] University of Turku.
Rissanen, T. (2000) Suomalaisen viittomakielen lauseoppia. In Malm, A. *Viittomakieliset Suomessa.* FinnLectura: Pieksämäki.
Rizzolatti, G. & Craighero, L. (2004). Mirror-neuron system*.* *Annual Review of Neuroscience* 27, 169-192.
Rizzolatti, G. and Arbib, M. (1988). Language within our grasp *Trends in Neuroscience* 21, 188–194.
Robb, M.P. (2003). Bifurcation and chaos in the cries of full-term and preterm infants. *Folia Phoniatrica Logopaedica* 55, 233–240.
Rosner, S.B. & Pickering, J.B. (1994). *Vowel Perception and Production*. Oxford Science Series 23. Oxford University Press.
Rothgänger H. (2003). Analysis of the sounds of the child in the first year of age and a comparison to the language. *Early Human Development* 75, 55–69.
Rothenberg SJ, Cansino S, Sepkoski C, Torres LM, Medina S, Schnaas L, et al. Prenatal and perinatal lead exposures alter acoustic cry parameters of neonate. *Neurotoxicol Teratol* 1995;17(2):151 –60.
Rosner, B. S. and Pickering, J. B. (1994). *Vowel perception and production*. Oxford, Oxford Univ. Press.
Rowland, L.P. (2000). Diseases of the motor unit. In Kandel, E.R., Schwartz, J.H. & Jessell, T.M. (eds.) *Principles of Neural Science*. McGraw-Hill: New York.
Saarni, T. (2010). *Segmental durations of speech.* Academic dissertation, University of Turku, Department of Information Technology.
Sacks, O. (1990/1992). *Käsien kieli*. Translated into Finnish from 2nd edition. [Original in English Seeing Voices.] WSOY, Juva.

Sandler, W. 1993. A sonority cycle in American Sign Language. *Phonology* 10:243–279.
Santi A, Servos P, Vatikiotis-Bateson E, Kuratate T, Munhall K (2003) Perceiving biological motion: dissociating visible speech from walking. *J Cogn Neurosci* 15:800–809.
Savela, J., Kleimola, T., Mäkelä, L., Tuomainen, J. and Aaltonen, O. (2003). The effects of distinctive features on the perception of vowel categories. *The 15th International Congress of Phonetic Sciences, Barcelona, Spain, Universitat Autònoma de Barcelona.*
Savela, J., Raimo, I., Uusipaikka, E., Aaltonen, O. & Salakoski, T. (2008). The categorisation of synthetic vowels by Swedish speaking listeners in Finland. in O'Dell, M. & Nieminen, T. (eds.) *Fonetiikan päivät 2008.* Tampere Studies in Language, Translation and Culture, Series B 3, Tampere: Tampere University Press, 111-121.
Savela, J. (2009). *Role of Selected Spectral Attributes in the Perception of Synthetic Vowels*. Academic dissertation. Department of Information Technology, University of Turku. TUCS Dissertations 119, June 2009.
Savolainen, L. (2000). Viittoman rakenne. In Malm, A. *Viittomakieliset Suomessa*. Finnlectura.
Shariatmadari, D. (2006). Sounds difficult? Why phonological theory needs 'ease of articulation'. – *SOAS Working papers in linguistics* 14, 207-226.
Shiller, D.M., Houle, G. & Ostry, D.J. (2005). Voluntary Control of Human Jaw Stiffness. *Journal of Neurophysiology* 94, 2207-2217.
Shneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Addison-Wesley.
Steil, J.J., Röthling, F., Haschke, R. and Ritter, H. (2004). Situated robot learning for multi-modal instruction and imitation of grasping. *Robotics and Autonomous Systems* 47, 129–141.
Stokoe, W.C. (1960). *Sign Language Structure*. Reprinted in 1978. Linstock Press.
Stokoe, W.C. (2001). *Language in hand: why sign came before speech*. Gallaudet University Press.
Shmulevich, I., Yli-Harja, O., Coyle, E., Povel, D.-J. and Lemström, K. (1999) Perceptual Issues in Music Pattern Recognition - Complexity of Rhythm and Key Finding. In A. Patrizio, G. A.Wiggins & H. Pain (Eds.) *Proceedings of the AISB'99 Symposium on Musical Creativity*, pp. 64-69. The Society for the Study of Artificial Intelligence and Simulation of Behaviour.
Suvi - online dictionary of Finnish Sign Language. http://suvi.viittomat.net accessed on 12 April, 2009
Sebeok, T.A. & Umiker-Sebeok, D.J. (1987). *Monastic sign languages*. Approaches to semiotics 76. Mouton de Gruyter.
Sams,M., Aulanko, R. Hämäläinen, R., Hari, R., Lounasmaa, O.V., Lu, S.-T. Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex, *Neuroscience Letters* 127, 141– 145
Spence, C. & Driver, J. (2004). *Crossmodal space and crossmodal attention.* Oxford University Press: Oxford.
Studdert-Kennedy, M. (1998). The particulate origins of language generativity: from syllable to gesture. In Hurford, J.R., Studdert-Kennedy, M. & Knight, C. (eds.) *Approaches to the Evolution of Language: Social and Cognitive Bases.* Cambridge University Press.
Studdert-Kennedy, M. (2000). Evolutionary implications of the particulate principle: imitation and the dissociation of phonetic form from semantic function. In Knight,

Studdert-Kennedy, & Hurford (eds.) *The emergence of language: social function and the origins of linguistic form*, pp. 161–176. Cambridge University Press.
Supalla, S.J. (1992). *The Book of Name Signs: naming in American Sign Language.* DawnSignPress.
Supalla, T. (1986). The classifier system in American Sign Language. In C.Craig (ed.) *Noun classification and categorization.* Amsterdam: John Benjamins North America.
Sutton, V. (1995). Lessons in SignWriting. Center for Sutton Movement Writing. Also available online at: http://www.SignWriting.org/lessons/lessons.html Accessed: 28.3.2010
Stevens, K.N. (2000). *Acoustic Phonetics.* MIT Press.
Tajima, K., Port, R. & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25, 1-24.
Takkinen, R. (2000). Viittomakielisten lasten kielenomaksuminen. In Malm, A. (ed.) *Viittomakieliset Suomessa*. FinnLectura: Helsinki.
Takkinen, R. (2002). *Käsimuotojen salat – viittomakielisen lapsen kielenomaksumisesta 2-7 vuoden iässä*. Kuurojen liitto ry.
Tamura, S. & Kawasaki, S. (1988). Recognition of sign language motion images. *Pattern Recognition* 21:4, 343-353.
(THL, 2009) National Health Institute of Finland, webpage: http://www.ktl.fi/portal/3618 (Retrieved 2.7. 2009)
Thyer, N., Hickson, L. and Dodd, B. (2000). The perceptual magnet effect in Australian English vowels. *Perception and Psychophysics* 62(1), 1-20.
Tjaden, K. & Turner, G.S. (1997). Spectral properties of fricatives in amyotrophic lateral sclerosis. *Journal of Speech, Language and Hearing Research* 40(6), 1358-1373.
Tompkins, C.A. (1997). *Right hemisphere communication disorders: Theory and Management*. Clinical Competence Series. Singular Publishing Group.
Tremblay, S., Houle, G. & Ostry, D.J. (2008). Specificity of Speech Motor Learning. *Journal of Neuroscience* 28(10), 2426-2434.
Tuomainen, J. (1996). Psykoakustiikka: akustiikan ja havainnon välinen suhde. In: Revonsuo, A., Lang, A.H. & Aaltonen, O. (1996). *Mieli ja aivot. Kognitiivinen neurotiede.*
Tuononen, M., Gonzalez Hautamäki, R. and Fränti, P. (2008). Automatic Voice Activity Detection in Different Speech Applications. *e-Forensics 2008*, First International Conference on forensic Applications and Techniques in Telecommunications, Information and Multimedia, Adelaide, Australia, Jan 21-Jan 23, 2008.
Turunen, M. & Hakulinen, J. (2007). Spoken and Multimodal Communication Systems in Mobile Settings. In Esposito, A. et al. (eds.): Verbal and Nonverbal Communication Behaviours, LNAI 4775, pp. 227-241.
Turunen, M., Hakulinen, J., Hella, J., Rajaniemi, J-P, Melto, A., Mäkinen, E., Rantala, J., Heimonen, T., Laivo, T., Soronen, H., Hansen, M., Valkama, P., Miettinen, T. & Raisamo, R. (2009). Multimodal Media Center Interface Based on Speech, Gestures and Haptic Feedback. In Gross, T. et al. (eds.). *INTERACT 2009, Part II, LNCS 5727*, pp. 54-57.
Turunen, M., Hakulinen, J., Melto, A., Hella, J., Laivo, T., Rajaniemi, J-P, Mäkinen, E., Soronen, H, Hanesen, M., Pakarinen, S., Heimonen, T., Rantala, J., Valkama, P., Miettinen, T. & Raisamo, R. (2010). Accessible Speech-Based and Multimodal Media Center Interface for Users with Physical Disabilities. In A. Esposito et al. (eds.): *COST 2102 Int. Training School 2009, LNCS 5967*, pp. 66-79. Springer-Verlag.

Turunen, M., Hurtig, T., Hakulinen, J., Virtanen, A. & Koskinen, S. (2006). Mobile Speech-based and Multimodal Public Transport Information Services. *Proceedings of MobileHCI 2006, Workshop on Speech in Mobile and Pervasive Environments (SiMPE2006)*, (no page numbers).

Tye-Murray, N., Sommers, M. & Spehar, B. (2007) Auditory and Visual Lexical Neighborhoods in Audiovisual Speech Perception. *Trends in Amplification* 11, 233.

van Alphen, P.M. & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics* 32, 455–491.

van Hessen, A. J. & Schouten, M. E. (1999). Categorical perception as a function of stimulus quality. *Phonetica* 56(1-2): 56-72.

van Lierde, K.M., Vinck, B., De Ley, S., Clement, G. & van Cauwenberge, P. (2005). Genetics of vocal quality characteristics in monozygotic twins: a multiparameter approach. *Journal of Voice* 19, 511–518.

von Humboldt, W. (1836/1972). Linguistic variability and intellectual development; translated by G.C. Buck and F.A. Raven. University of Pennsylvania Press.

Vygotski, (1962). *Thought and Language*.

Walmsley, B., Burke, R.E., Strick, K., Kanda, K. & Kim, C.C. (1978). Anatomy of medial gastrocnemius and soleus motor nuclei in cat spiinal cord, *Journal of Neurophysiology* 40; 667-680.

Watkins, K.E., Strafella, A.P. & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41, 989–994

Wernicke, C. (1908). The symptom-complex of aphasia. In:A. Chruch (ed.), *Diseases of the Nervous System*, pp. 265-324. Appleton.

Wilcox, S. (1992). *The Phonetics of Fingerspelling: Studies in Speech Pathology and Clinical Linguistics 4*. John Benjamins Publishing Company.

Wilcox, S. (2001). *Conceptual Spaces and Bodily Actions*. Handout in lecture series Conceptual Spaces and Embodied Actions: Evidence from Signed Languages. Turku. University of Turku 12.-13.10. 2001.

Wilcox, S. (2002). The gesture-language interface: evidence from signed languages. In Prillwitz, S, Hanke, T. & Vollhaber, T. (eds.): *International Studies on Sign Language and the Communication of the Deaf.*

Willems, R.M. & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language* 101(3), 278-289.

Zemlin, W.R. (1998). *Speech and Hearing Science: Anatomy and Physiology*. 4th edition. Allyn and Bacon.

Zeskind, P.S., Oberlander, T.F., Grunau, R.E., Fitzgerald, C. & Garber, K.A. (2005). Continuing effect of prenatal SSRI exposure detected by spectral analysis of infant cry sounds at two months of age [abstract]. *Pediatric Academic Societies' Annual Meeting*, May 14—17, Washington, DC.

Zlatev, J., Persson, T. and Gärdenfors, P. (2005). *Bodily mimesis as "the missing link" in human cognitive evolution.* Lund University Cognitive Studies 121.

Software

Avatar basic model online. http://www.katorlegaz.com/3d_models/human_male/0139/index.php accessed on 27 March, 2010.

Boersma P, Weenink D. Praat Software, 2003, Institute of Phonetic Sciences, University of Amsterdam, the Netherlands. Webpage: http://www.fon.hum.uva.nl/praat/ (Retrieved 2.7.2009).Blender software. Online.http://www.blender.org/ accessed on 12 April, 2009

EyePhone. Software. Free trial version available at: http://www.synface.com

HLSyn software, version 2.2. Sensimetrics Corp. Online. http://www.sens.com/hlsyn/index.htm.

Matlab®.Computer software. MathWorks Inc.

PXLab software. Online. http://www.uni-mannheim.de/fakul/psycho/irtel/pxlab/index.html accessed on 12 April, 2009

Turku Centre for Computer Science
TUCS Dissertations

103. **Tapio Pahikkala**, New Kernel Functions and Learning Methods for Text and Data Mining
104. **Denis Shestakov**, Search Interfaces on the Web: Querying and Characterizing
105. **Sampo Pyysalo**, A Dependency Parsing Approach to Biomedical Text Mining
106. **Anna Sell**, Mobile Digital Calendars in Knowledge Work
107. **Dorina Marghescu**, Evaluating Multidimensional Visualization Techniques in Data Mining Tasks
108. **Tero Säntti**, A Co-Processor Approach for Efficient Java Execution in Embedded Systems
109. **Kari Salonen**, Setup Optimization in High-Mix Surface Mount PCB Assembly
110. **Pontus Boström**, Formal Design and Verification of Systems Using Domain-Specific Languages
111. **Camilla J. Hollanti**, Order-Theoretic Methods for Space-Time Coding: Symmetric and Asymmetric Designs
112. **Heidi Himmanen**, On Transmission System Design for Wireless Broadcasting
113. **Sébastien Lafond**, Simulation of Embedded Systems for Energy Consumption Estimation
114. **Evgeni Tsivtsivadze**, Learning Preferences with Kernel-Based Methods
115. **Petri Salmela**, On Communication and Conjugacy of Rational Languages and the Fixed Point Method
116. **Siamak Taati**, Conservation Laws in Cellular Automata
117. **Vladimir Rogojin**, Gene Assembly in Stichotrichous Ciliates: Elementary Operations, Parallelism and Computation
118. **Alexey Dudkov**, Chip and Signature Interleaving in DS CDMA Systems
119. **Janne Savela**, Role of Selected Spectral Attributes in the Perception of Synthetic Vowels
120. **Kristian Nybom**, Low-Density Parity-Check Codes for Wireless Datacast Networks
121. **Johanna Tuominen**, Formal Power Analysis of Systems-on-Chip
122. **Teijo Lehtonen**, On Fault Tolerance Methods for Networks-on-Chip
123. **Eeva Suvitie**, On Inner Products Involving Holomorphic Cusp Forms and Maass Forms
124. **Linda Mannila**, Teaching Mathematics and Programming – New Approaches with Empirical Evaluation
125. **Hanna Suominen**, Machine Learning and Clinical Text: Supporting Health Information Flow
126. **Tuomo Saarni**, Segmental Durations of Speech
127. **Johannes Eriksson**, Tool-Supported Invariant-Based Programming
128. **Tero Jokela**, Design and Analysis of Forward Error Control Coding and Signaling for Guaranteeing QoS in Wireless Broadcast Systems
129. **Ville Lukkarila**, On Undecidable Dynamical Properties of Reversible One-Dimensional Cellular Automata
130. **Qaisar Ahmad Malik**, Combining Model-Based Testing and Stepwise Formal Development
131. **Mikko-Jussi Laakso**, Promoting Programming Learning: Engagement, Automatic Assessment with Immediate Feedback in Visualizations
132. **Riikka Vuokko**, A Practice Perspective on Organizational Implementation of Information Technology
133. **Jeanette Heidenberg**, Towards Increased Productivity and Quality in Software Development Using Agile, Lean and Collaborative Approaches
134. **Yong Liu**, Solving the Puzzle of Mobile Learning Adoption
135. **Stina Ojala**, Towards an Integrative Information Society: Studies on Individuality in Speech and Sign

Turku Centre *for* Computer Science

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www. tucs.fi

University of Turku
Faculty of Mathematics and Natural Sciences
- Department of Information Technology
- Department of Mathematics

Turku School of Economics
- Institute of Information Systems Science

Åbo Akademi University
Division for Natural Sciences and Technology
- Department of Information Technologies

ISBN 978-952-12-2583-3
ISSN 1239-1883

Stina Ojala

Towards an Integrative Information Society

Stina Ojala

Towards an Integrative Information Society

Stina Ojala

Towards an Integrative Information Society: Studies on Individuality in Speech and Sign