KTH Computer Science
and Communication

The /k/s, the /t/s, and the in-betweens

Novel approaches to examining the perceptual consequences of misarticulated speech

SOFIA STRÖMBERGSSON

Doctoral Thesis

Stockholm, Sweden 2014

TRITA-CSC-A 2014:03
ISSN-1653-5723
ISRN–KTH/CSC/A--13/03-SE
ISBN: 978-91-7595-050-1
Tryckt av Eprint AB 2014

KTH School of Computer Science and Communication
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framläggs till offentlig granskning för avläggande av filosofie doktorsexamen i tal- och musikkommunikation med inriktning på talkommunikation måndagen den 7 april 2014 klockan 10.00 i F3, Kungl Tekniska högskolan, Lindstedtsvägen 26, Stockholm.

Abstract

This thesis comprises investigations of the perceptual consequences of children's misarticulated speech – as perceived by clinicians, by everyday listeners, and by the children themselves. By inviting methods from other areas to the study of speech disorders, this work demonstrates some successful cases of cross-fertilization. The population in focus is children with a phonological disorder (PD), who misarticulate /t/ and /k/. A theoretical assumption underlying this work is that errors in speech production are often paralleled in perception, e.g. that children base their decision on whether a speech sound is a /t/ or a /k/ on other acoustic-phonetic criteria than those employed by proficient language users. This assumption, together with an aim at stimulating self-monitoring in these children, motivated two of the included studies. Through these studies, new insights into children's perception of their own speech were achieved – insights entailing both clinical and psycholinguistic implications. For example, the finding that children with PD generally recognize themselves as the speaker in recordings of their own utterances lends support to the use of recordings in therapy, to attract children's attention to their own speech production. Furthermore, through the introduction of a novel method for automatic correction of children's speech errors, these findings were extended with the observation that children with PD tend to evaluate misarticulated utterances as correct when just having produced them, and to perceive inaccuracies better when time has passed. Another theme in this thesis is the gradual nature of speech perception related to phonological categories, and a concern that perceptual sensitivity is obscured in descriptions based solely on discrete categorical labels. This concern is

substantiated by the finding that listeners rate “substitutions” of [t] for /k/ as less /t/-like than correct productions of [t] for intended /t/. Finally, a novel method of registering listener reactions during the continuous playback of misarticulated speech is introduced, demonstrating a viable approach to exploring how different speech errors influence intelligibility and/or acceptability. By integrating such information in the prioritizing of therapeutic targets, intervention may be better directed at those patterns that cause the most problems for the child in his or her everyday life.

Acknowledgements

First of all, I'd like to thank David House, my main supervisor, who has been my most important guide in this work. Neither the scientific quality, nor my writing would have been half as good without your help! Most of all, I appreciate your allowing me to (letting me think that I was) making my own discoveries. Thank you for showing that confidence! Many thanks also to Åsa Wengelin, my second supervisor, for your sharp-minded and valuable assistance. I've really appreciated your straight-forwardness in challenging my ideas and my writing! I'm also very grateful to Ulrika Nettelbladt, for being a mentor and a source of inspiration. I've truly enjoyed our vivid and inspiring discussions! Many thanks also to Rolf Carlson and Britta Hammarberg, for generously offering your time to reading and commenting on this thesis – I've really valued your advice!

I'm very grateful for having had the opportunity to work at the Department of Speech, Music and Hearing; thanks to all of you – colleagues and friends – for contributing to a nice and creative working environment! The collaboration with Jens Edlund, David House and Anna Hjalmarsson on prosody in questions has been very stimulating, and I truly value the opportunity of having been introduced to this area of research. Thanks also to Giampiero Salvi, for a nice and fruitful collaboration – not only with Study III, but also with assuring the supply of decent coffee! I'd like to thank Jens for stimulating discussions and scientific inspiration – but also for the joy of your company outside work. I also wish to thank my past and current roommates for the nice company – Laura, Catha, and particularly Anna, not only for your professional advice and collaboration, but – most importantly – for being a dear friend and a life coach!

I really value the scientific collaboration with Christina Tånnander at the Swedish Agency for Accessible Media – I already knew you as a great colleague and a dear friend, and I'm happy that we were able to combine work and pleasure in such a fruitful way! Thanks also to Södermalms Talteknologiservice (STTS) for contributions related to this work.

I'd like to thank the Graduate School of Language Technology (GSLT) for providing financial support, PhD courses and great retreats. I'm also thankful to the research training program Hearing, Otorhinolaryngology, Language, Speech (HÖST) at Karolinska Institute, for allowing me to take part in their activities. Thanks also to Promobilia for financial support. My sincere gratitude to all SLPs and teachers who have helped me with recruiting children to my studies, and – of course! – to the children themselves, and to their parents. Thanks also to colleagues, friends and students who have participated in my listening experiments.

I'm really happy for the fruitful theoretical and professional exchanges I've had with Nelli Kalnak, Ulrika Marklund, Cecilia von Mentzer, Martina Hedenius, Olof Sandgren and Ketty Holmström – and to those of you other Swedish-SLPs/PhD-students-who-study-child-language-colleagues, who more occasionally participated in our inspiring meetings. I'm really looking forward to future collaborations!

I've had the fortune to meet and discuss my work with some great giants within the field of speech acquisition – John Locke, Marilyn Vihman, Lise Menn, and Björn Lindblom – I'm truly grateful for this opportunity, and for your generosity in sharing your time and thoughts with me! Furthermore, I'd like to thank Francisco Lacerda and colleagues at the Linguistics Department at Stockholm University for generously inviting me to taking part in their activities. A special thanks to Mattias Heldner, for your commendable way of combining scientific rigor with a healthy playfulness – thanks for that professional inspiration, as well as for all less-professional inspiration!

Finally, a heartfelt thanks to my friends and family, for putting up with me devoting time and energy to this work, but at the same time – and more importantly! – keeping me from complete absorption. A most special thanks to Kristian, for your never-ceasing encouragement, and for your patience with my self-centered focus on work – I'll always be grateful for that! And ultimately, striped-dotted-squared-and-hard embraces to Ester and Vera, for constantly reminding me of more important things than this stupid book.

Contents

Publications and contributors

The studies included in this thesis have been made in collaboration with others. The details of these collaborations are specified below.

Study I Strömbergsson, S. (2013). Children's recognition of their own recorded voice: Influence of age and phonological impairment. *Clinical Linguistics & Phonetics, 27(1)*, 33-45.

David House contributed with the planning and the design of the study, and, together with Åsa Wengelin, with reading and commenting on the manuscript.

Study II Strömbergsson, S., Wengelin, Å., & House, D. (in press). Children's perception of their synthetically corrected speech production. *Clinical Linguistics & Phonetics*.

David House contributed with the planning and the design of the study, and, together with Åsa Wengelin, with reading and commenting on the manuscript.

Study III Strömbergsson, S., Salvi, G., & House, D. (submitted). Acoustic and perceptual evaluation of category goodness of /t/ and /k/ in typical and misarticulated child speech.

Giampiero Salvi performed the acoustic analysis and machine learning. David House contributed with reading and commenting on the manuscript.

Study IV Strömbergsson, S., & Tånnander, C. (2013). Correlates to intelligibility in deviant child speech – comparing clinical evaluations to audience response system-based evaluations by untrained listeners. In *Proceedings of Interspeech 2013* (pp. 3717-3721). Lyon, France.

Christina Tånnander assisted in the planning of the study, and set up the web-based listening test. Jens Edlund contributed with the KDE analysis.

Non-included papers

The author has also written and co-authored other published papers, some of which are related to the contents of the present thesis. These are presented in the list below. Next follows a list of papers related to the research projects on prosody in questions, within which the author has taken part during her time as a PhD student.

Related to the present thesis

Strömbergsson, S. (2012). Identifying children's dialect from isolated words. In *Proceedings of Fonetik 2012*. Gothenburg, Sweden.

Strömbergsson, S. (2012). Is that me? Self-voice identification in children with deviant speech. In *Proceedings of ICPLA 2012*. Cork, Ireland.

Strömbergsson, S. (2012). Synthetic correction of deviant speech – children's perception of phonologically modified recordings of their own speech. In *Proceedings of Interspeech*. Portland, Oregon, US.

House, D., & Strömbergsson, S. (2011). Self-voice identification in children with phonological impairment. In *Proceedings of the ICPhS XVII* (pp. 886-889). Hong Kong.

Strömbergsson, S. (2011). Children's perception of their modified speech – preliminary findings. In *Proceedings of Fonetik 2011*, TMH-QPSR, 51(1), 117-120.

Strömbergsson, S. (2011). Children's recognition of their own voice: influence of phonological impairment. In *Proceedings of Interspeech 2011* (pp. 2205-2208). Florence, Italy.

Strömbergsson, S. (2011). Corrective re-synthesis of deviant speech using unit selection. In Sandford Pedersen, B., Nespore, G., & Skadina, I.

(Eds.),*NODALIDA 2011 Conference Proceedings* (pp. 214-217). Riga, Latvia.

Strömbergsson, S. (2011). Segmental re-synthesis of child speech using unit selection. In *Proceedings of the ICPhS XVII* (pp. 1910-1913). Hong Kong.

Strömbergsson, S. (2009). Development of self-voice recognition in children. In *Proceedings of Fonetik 2009* (pp. 136-139). Stockholm.

Strömbergsson, S. (2009). Modified re-synthesis of initial voiceless plosives by concatenation of speech from different speakers. In *Proceedings of Fonetik 2009* (pp. 198-2001). Stockholm.

Strömbergsson, S. (2009). Self-voice recognition in 4 to 5-year-old children. In *Proceedings of Interspeech 2009* (pp. 820-823). Brighton, UK.

Strömbergsson, S. (2007). Interactional patterns in computer-assisted phonological intervention in children. In *Proceedings of Fonetik 2007*, TMH-QPSR, 50(1), 69-72.

Related to prosody in questions

Strömbergsson, S., Hjalmarsson, A., Edlund, J., & House, D. (2013). Timing responses to questions in dialogue. In *Proceedings of Interspeech 2013* (pp. 2584-2588). Lyon, France.

Edlund, J., House, D., & Strömbergsson, S. (2012). Question types and some prosodic correlates in 600 questions in the Spontal database of Swedish dialogues. In Ma, Q., Ding, H., & Hirst, D. (Eds.), *Proceedings of Speech Prosody 2012* (pp. 737-740). Shanghai, China.

Edlund, J., Strömbergsson, S., & House, D. (2012). Telling questions from statements in spoken dialogue systems. In *Proceedings of SLTC 2012*. Lund, Sweden.

Strömbergsson, S., Edlund, J., & House, D. (2012). A study of Swedish questions and their prosodic characteristics. In *Proceedings of Workshop on Innovation and Applications in Speech Technology (IAST)*. Dublin, Ireland.

Strömbergsson, S., Edlund, J., & House, D. (2012). Prosodic measurements and question types in the Spontal corpus of Swedish dialogues. In *Proceedings of Interspeech 2012* (pp. 839-842). Portland, Oregon, US.

Strömbergsson, S., Edlund, J., & House, D. (2012). Question types and some prosodic correlates in the Spontal corpus of Swedish dialogues. In *Proceedings of Fonetik 2012*. Gothenburg, Sweden.

Strömbergsson, S., Edlund, J., & House, D. (2012). Questions and reported speech in Swedish dialogues. In *Proceedings of Nordic Prosody XI* (pp. 373-382). Tartu, Estonia.

Edlund, J., Beskow, J., Elenius, K., Hellmer, K., Strömbergsson, S., & House, D. (2010). Spontal: a Swedish spontaneous dialogue corpus of audio, video and motion capture. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., Rosner, M., & Tapias, D. (Eds.), In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* (pp. 2992 - 2995). Valetta, Malta.

Beskow, J., Edlund, J., Elenius, K., Hellmer, K., House, D., & Strömbergsson, S. (2009). Project presentation: Spontal – multimodal database of spontaneous dialog. In *Fonetik 2009* (pp. 190-193). Stockholm.

List of abbreviations

ARS	Audience Response Systems
DPD	Developmental Phonological Disorder
ICD	International Classification of Diseases
ICF	International Classification of Functioning, Disability and Health
ICF-CY	The ICF applied to Children and Youth
PCC	Percentage Consonants Correct
PD	Phonological Disorder (Swedish: “fonologisk språkstörning”)
PI	Phonological Impairment, synonymous to PD.
SDCS	Speech Disorders Classification System
SLP	Speech-Language-Pathologist (Swedish: “logoped”)
SSD	Speech Sound Disorder
TD	Typical speech and language development
UG	Undifferentiated lingual gesture
VAS	Visual analog scale

Part I. Background

1. Introduction

Acquiring speech is a formidable task. Even so, most of us achieve this task and become proficient speakers of our surrounding language already before starting school. And once we have reached this level, there appears to be no turning back; once we know a language, we hardly ever hear a spoken utterance delivered in this language without trying to decode what linguistic message it may convey. However, when we as adults set out to learn a new language, we might be reminded of the challenges involved in acquiring our first language. Not only does this task involve the identification of meaningful units from a virtually indivisible sequence of sounds, it also involves the abstraction of linguistic regularities in the language, in terms of what building blocks the language contains (e.g. phonemes, morphemes, or words) and how these are combined (e.g. as described by phonotax, morphology, or syntax). Moreover, acquiring verbal skills also involves the appreciation of the various forms each of these building blocks may take when appearing in spoken messages. As listeners, we have to cope with the acoustic-phonetic variation between different realizations of the same speech sound, i.e. we have to map multiple exemplars to the same phonemic category. In doing so, we have to learn what acoustic-phonetic variation is phonologically significant – e.g. the acoustic-phonetic features distinguishing /r/ from /l/, and what variation does not carry a phonological function – e.g. the acoustic-phonetic variation among different variants of /r/. Depending on our language background, this learning may involve the appreciation of a new phonemic contrast (e.g. the /r/-/l/ contrast in English to a speaker of Japanese), or the adaptation of already established phonemic categories to the acoustic-phonetic norms of the new language (e.g. the acoustic-phonetic realizations of English /r/ to a speaker of Swedish). As speakers, on the other hand, we have to adapt our speech production to this new language, i.e. tune our mappings between the abstract phonemic categories and their articulatory implementation, to approach the acoustic-phonetic norms of the language we are learning. These aspects of phonological knowledge, and how they develop during first language acquisition in childhood, are central topics of the present thesis.

The interplay between speech production and speech perception is intriguing in typical speech acquisition, and central in the study of children with speech impairments. Early in typical speech acquisition, infants and toddlers clearly understand more than they are able to produce themselves. As it happens, this lag between perception and production of speech is often observed also in older children – in typical development as well as in impaired or delayed speech acquisition. For example, consider the following episode, observed in the interaction between an adult and a typically developing child, reported by Berko and Brown (1960):

> One of us, for instance, spoke to a child who called his inflated plastic fish a *fis*. In imitation of the child's pronunciation, the observer said: "This is your *fis*?" "No," said the child, "my *fis*". He continued to reject the adult's pronunciation until he was told, "This is your *fish*." "Yes," he said, "my *fis*."
>
> (Berko & Brown, 1960: p. 531)

This observation – often referred to as the [fɪs] phenomenon – illustrates the apparent gap between children's perception and production of speech, where perception appears more advanced than production. Furthermore, it indicates a difference between children's perception of their own speech production and their perception of speech produced by others, such that they react to errors in others' speech production, while appearing insensitive to errors in their own speech production. These aspects of children's speech perception, and how they relate to impaired speech production, are also central to the present thesis. In exploring how children react to hearing their own speech production – recorded and synthetically corrected – the work presented here provides some new insights into children's perception of their own speech production.

The study of speech requires consideration of different levels of representation. On the one hand, speech may be represented as sequences of phonemes, i.e. categorical entities defined by their linguistic function. In this representational domain, an utterance like “fis” above may be represented as /fɪs/. Descriptions like these represent an abstract level of speech, where the linguistic function of the different units is represented, whereas any acoustic-phonetic variation that may exist between different realizations of the same phoneme is obscured. On the other hand, speech may be represented at a more detailed phonetic level. For example, a narrow phonetic transcription, e.g. [fɪ̠s̻][1], would reveal phonetic details that are not registered at the more abstract phonological level. However, not even this detailed level of transcription is sensitive to all acoustic-phonetic detail in the physical speech signal, details that may only be revealed by physical acoustic and/or articulatory measures. Regarding the [fɪs] phenomenon described above, the interpretation of this observation is dependent on the level at which it is described. Here, the original description solicits an interpretation that the child does not recognize that his own production “fis” is the same as the adult’s production “fis”. However, a more detailed level of description may instead have revealed that the adult’s reproduction of the child’s original utterance is in fact *not* the same as the child’s production – that the adult fails to copy acoustic-phonetic details that may actually be significant. In terms of theoretical implications, this solicits an alternative interpretation: that the adult is insensitive to acoustic-phonetic details in the child’s speech. Hence, our theoretical understanding of speech production and perception is dependent on the level of representation at which they are described. This is another recurring theme in the present thesis.

We use speech to communicate our thoughts and needs to people in our environment. If the speech we produce differs extensively from the acoustic-phonetic norms of our surrounding language, intelligibility will suffer, and we may fail to deliver the messages we intend. This would be the case if you, for example, produce [tu:] both for /sku:/ “sko” (English: *shoe*) and for /ku:/ “ko” (English: *cow*). However, even with a smaller difference between the intended message and the expected phonetic form – possibly even so small that intelligibility is not affected – listeners may still register this idiosyncratic phonetic variation in speech production. So, even if your production [θku:] for /sku/ “sko” (English: *shoe*) is not confused for any other Swedish word, listeners are likely to perceive that your

[1] [s̻] indicates a laminal articulation of [s], i.e. that the air passage constriction is formed by the tongue blade, instead of by (only) the tongue tip.

production of /s/ differs from what is expected from a proficient speaker of Swedish. In doing so, listeners may associate this idiosyncratic phonetic feature with personal characteristics (e.g. age, dialectal background, or even, level of intelligence) that you may or may not feel comfortable with having attributed to you. However, much remains to be learned about what specific speech errors or distortions are most deteriorating – either in terms of speech intelligibility or in terms of speech acceptability. In suggesting a novel approach to the examination of such functional effects of producing misarticulated speech, this topic is also addressed in the present thesis.

1.1 Motivation

The work described in this thesis sprung out of a clinical and theoretical interest in the challenges observed in children with speech impairments to perceive and/or react to the errors in their own speech production (cf. the [fɪs] phenomenon). This fascination, coupled with an interest in developing new technology for modified re-synthesis of children's speech spurred the curiosity regarding the possible benefits of introducing corrective re-synthesis of the child's misarticulations in clinical intervention. Together, these factors instigated the exploration of children's perception of their own speech production – in its online form, in its recorded form, and in synthetically modified forms. During the course of these investigations, however, new questions arose regarding how misarticulations are perceived by other listeners – on the one hand how specific speech sounds are evaluated when presented in isolation, and on the other hand how specific speech errors are perceived when occurring in their natural habitat, i.e. in connected speech. A growing concern regarding traditional categorical descriptions of misarticulated speech – that these instruments are insensitive to information that may be crucial to our understanding of misarticulated speech – motivated the exploration of finer-grained perceptual instruments. Dissatisfaction with the current state of knowledge regarding what misarticulation patterns are most salient or disturbing to listeners instigated the introduction of another new instrument into the field of speech and language pathology, allowing closer exploration of the association between events in the speech signal (i.e. misarticulations) and the distribution of listener reactions. Hence, the common thread running through the present thesis is perception of children's misarticulated speech, and how new technological innovations may contribute to new insights into this domain.

1.2 Research questions

The work described in the present thesis explores the following general research questions:

1. Do children recognize their recorded speech production as their own?

2. What are the children's reactions when presented with synthetically a) modified and b) corrected versions of utterances they have produced?

3. Do listeners perceive misarticulated speech sounds as substitutions or as gradient errors?

4. Can listeners' reactions during the continuous playback of misarticulated speech be linked to triggering events in the speech signal?

5. Does the introduction of new technological innovations into the study of children's speech sound disorders bring benefits that could not otherwise have been achieved?

The answers to these research questions, gained through the studies included in this thesis, will be presented in Chapter 5.

1.3 Thesis outline

This thesis consists of two parts. The first part introduces the reader to the field in which the presented work has been carried out, by describing the population in focus: children with speech sound disorders (SSDs). This description is inspired by the *International Classification of Functioning, Disability, and Health* (ICF) framework (introduced in Chapter 2), in aiming at a holistic view of the child with speech disorder: on the one hand describing the nature and degree of the disorder (Chapter 3), and on the other hand describing the functional ramifications of having the disorder (Chapter 4). Chapter 5 summarizes the findings in the studies included in this thesis, with regards to the research questions presented above. Chapter 6 rounds off the first part of the thesis with general conclusions regarding the findings revealed through this work, in terms of what methodological insights were gained, as well as implications for the understanding of psycholinguistic factors underlying the perception of misarticulated speech. Moreover, clinical implications for the intervention of SSDs are described, and directions for future work are outlined. Finally, Chapter 6 concludes the first part by summarizing the major contributions of this thesis.

The second part of this thesis contains four scientific papers that are central to the work presented. These will recurrently be referred to in the first part by their roman numerals I-IV. For the reader's convenience, the studies are briefly summarized below.

1.3.1 Included papers

Study I describes an investigation of children's perception of their own recorded speech production, in children with typical speech and language development (TD), as well as in children with a phonological disorder (PD). Exploration relates to whether the children recognize themselves as the speaker of recorded utterances, and to whether this ability is affected by the time elapsed since producing/recording the utterance. Furthermore, the study explores whether children's ability to recognize their own recorded voice develops with age in children with TD, and whether or not it is affected by having a PD.

Children's perception of their own speech production is further examined in **Study II**. Here, children with PD are again compared to children with TD, with regards to their perceptual responses to hearing their own speech production – in its online form, in its recorded form, and in synthetically modified forms. A novel method for performing automatic modification of the children's productions is introduced, to explore whether the children notice the modification, and whether their evaluation of accuracy corresponds to the intended modification target. Again, the children's evaluation performance is explored with regards to the time passed since their production of an utterance.

In contrast to the first two studies, where perceptual performance is registered with reference to categorical responses – both in the participating children and in evaluations performed by professionals – **Study III** provides a more fine-grained analysis of the speech material. Here, perceptual evaluations are assessed by means of a visual analog scale (VAS), thus providing a gradient measure of the perceptual correlates to the acoustic-phonetic representation of the children's correct and incorrect speech production. Furthermore, two methods for doing acoustic analysis of the speech material are compared (spectral moments vs. discrete cosine transformation analysis), in order to explore which one corresponds most to the perceptual evaluations performed by human listeners.

In **Study IV**, focus is again shifted, now from the misproductions of specific speech sounds to the communicative context in which they appear. Here, a panel of listeners is exposed to recordings of children with PD participating in free conversation with an adult interlocutor. The speech samples are evaluated with regards to intelligibility/acceptability, by means of an Audience Response System (ARS). Through this procedure, listeners' reactions during playback are continuously registered. The collected data is examined with regards to standard measures of the severity of the speech disorder, as well as to what specific events in the speech signal evoke reactions in the listeners.

1.4 A note on terminology

The terminology used to describe conditions and symptoms of speech production deficits has not always been consistent. In this thesis, the term **speech impairment** will be used to denote any impairment of speech production, regardless of what causes it. Thus, the term will encompass speech production problems that are secondary to cognitive, structural or motor deficits, as well as those problems whose origin is unknown. Furthermore, the term speech impairment will cover deficits in speech production where not only segmental properties of speech are affected, but also aspects of, for example, fluency and voice. The more specific term **speech sound disorder (SSD)** will be used when referring to disorders affecting the production of specific speech sounds, and/or of speech sound patterns. Neither does this term, however, specify whether the child's speech problems are considered his or her primary difficulty, or whether they are secondary to some other condition, e.g. a hearing impairment or a neurological disorder. Furthermore, the term is unspecified with regards to whether the speech problems are primarily caused by restrictions in motor-control, or whether the problems are more related to cognitive functioning. This use of the term is consistent with the definition suggested by the American Speech-Language-Hearing Association (2013). As a side note, the term **developmental phonological disorder (DPD)** has been suggested as an alternative to the term speech sound disorder (Rvachew & Brosseau-Lapré, 2012). This term is used to describe problems with the production of speech sounds (hence: "phonological") in children during speech acquisition (hence: "developmental"). Just as the term SSD, the term DPD does not involve a separation of articulation from other aspects of phonological knowledge, and neither does it exclude speech production problems based on their origin. Instead, both terms may be used to describe speech problems in children as a primary difficulty, or as a problem secondary to some other condition.[2] In these respects, the terms SSD and DPD can be regarded as synonyms. However, whereas the term DPD is explicitly only applicable to children, the term SSD is – at least theoretically – also applicable to older children and adults. Considering the theme of this thesis, where preschool-aged children are the population in focus, the term DPD would be an adequate description of its clinical focus. However, as the term SSD is more widely used and recognized, this is the term that will be used henceforth in this thesis.

All readers may not be comfortable with the choice of describing children's speech production deficiencies as **misarticulations**. However, the choice to use this term reflects a view that significant departures from expected articulatory patterns are best described using terminology from the domain in which they occur, thereby mis*articulation*. When used in this thesis, this term does not entail any implications on the underlying cause of the errors, e.g. whether they are due to speech-motor restrictions or to cognitive-linguistic limitations.

It has long been established that many children with SSDs also exhibit difficulties with regards to other aspects of language, e.g. with morpho-syntax or language comprehension (Pennington & Bishop, 2009; Shriberg, Tomblin, & McSweeny, 1999; Tyler, 2002). Hence, many children within the group of SSD have a concomitant language impairment. However, these aspects of related language skills fall outside the scope of the present thesis, where the focus is on children's speech production, and how impairments in speech production may be paralleled in perception.

[2] Obviously, the information of whether the SSD/DPD is a primary difficulty or secondary to some other condition is crucial both in clinical practice and in clinical research, and should always be specified. In the context of the present thesis, the speech problems exhibited by the children participating in the included studies have been identified as being their primary difficulty.

2. The speech impairment and its context

Producing speech that differs from phonetic and linguistic norms of the surrounding community might have several functional consequences. One potential consequence is that of not being understood, that the linguistic message you are trying to convey is not perceived by the listener. Another potential consequence is the unintentional signaling of personal, social or attitudinal characteristics – features that you may not feel comfortable being associated with. For example, by producing speech on a phonological level of a 4-year-old while in fact being 8 years old, you may signal to the listener that you are younger than you actually are. And by producing unusually slow or slurry speech, you stand the risk of being perceived as tired, as intoxicated, or even, as less intelligent (Walshe & Miller, 2011).

The relation between a child's speech production and its repercussions on how the child functions on a day-to-day basis can be described within the *International Classification of Functioning, Disability and Health* (ICF) framework (WHO, 2001). Within this framework, health-status is described not only through the assessment of the child's body structure and function, but also involves the consideration of what activity limitations and participation restrictions that impairments in body structure and function may entail. In section 2.1, the ICF framework will be described from a general point of view, and section 2.2 will describe the application of this framework to the domain of communicative impairments, and more specifically, to speech impairments in children (section 2.2.1).

2.1 The ICF framework

The ICF has been developed as a framework for describing and measuring health and disability. It is intended to provide clinical guidance for practitioners, as well as constituting an instrument for the evaluation of health care, and to assist decision-makers as a planning and policy tool (WHO, 2001).

The ICF is an attempt at integrating the medical model of functioning and disability with a social model of functioning and disability, such that disability is regarded not only as a feature of the individual, as a feature that can be treated or “corrected” by medical or other treatment, but also as a socially-created problem, which may be resolved through environmental adaptation. WHO refers to this integrated model as a biopsychosocial model of disability (WHO, 2002).

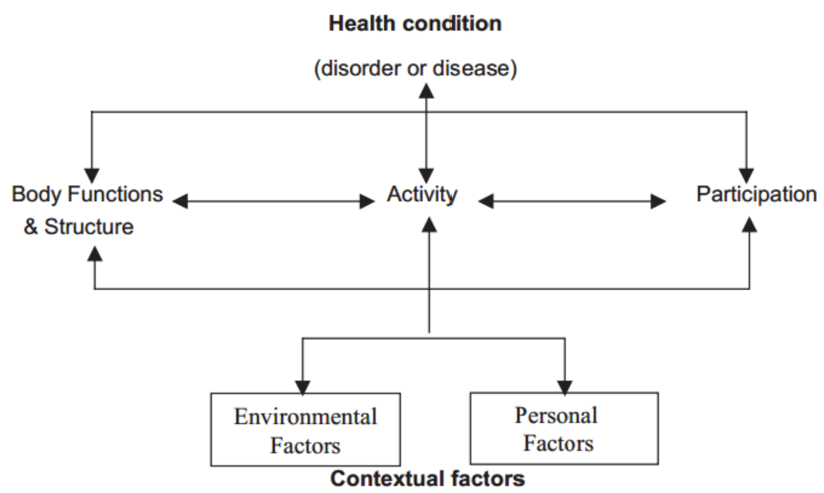

Figure 1. An illustration of the biopsychosocial model of disability underlying the ICF. (From WHO, 2002.) Reprinted with permission from WHO.

Figure 1 illustrates the interactions between an individual’s health condition (e.g. a disorder or disease), and contextual factors. The model implies reciprocal interactions between Health condition, Body functions and Body structures, Activities and Participation. Furthermore, Contextual factors are divided into Environmental factors (e.g. social attitudes or legal structures) and Personal factors (e.g. gender, age, overall behavior pattern). For example, a bone fracture in one of your legs (i.e. affected Body Structure) will limit your ability to walk (i.e. an Activity), which in turn might hinder you from hunting mushrooms with your friends (i.e. restrict your Participation). On the other hand, if you have access to a cane or a wheelchair (i.e. facilitating Environmental factors), and if you are determined and otherwise fit (i.e. positive Personal factors), your possibilities of participating are less restricted.

Two concepts relevant to the assessment of the components Activity and Participation are **capacity** and **performance**, where the former refers to what an individual can do in a standard environment, and the latter describes what the individual does in his or her usual environment. An individual's capacity is assessed by standardized methods in a standardized setting, typically by means of a standardized test, thus allowing comparisons across individuals and populations. An individual's performance, on the other hand, is influenced by Environmental factors, which are unique to each individual and, therefore, cannot be compared across individuals. A gap between capacity and performance, where an individual's performance does not reach the level of his or her capacity, indicates room for environmental adaptations to improve performance. An example of such an adaption would be to lower the noise level in the environment of an individual with a communicative impairment. (The concepts capacity and performance will be revisited in section 4.1.1.1, in the context of intelligibility assessments.)

In response to the need of adapting the ICF to fit the description of health and functioning in infants, children and adolescents, WHO published the *International Classification of Functioning, Disability and Health: Children and Youth Version* (ICF-CY; WHO, 2007). The ICF-CY retains the parts and components of the ICF, but with the addition of categories and domains related to developmental aspects of health and functioning. For example, the ICF-CY acknowledges the importance of the family (or other caregivers) as a context for the growing child, that the role of the family is probably more important while growing up, than in any later point of the individual's lifespan (WHO, 2007). Moreover, variations in the age at which Body functions emerge or skills are acquired are also considered in the ICF-CY, such that developmental lags are acknowledged as delays, and not as permanent features of body structure or function.

2.2 Applying the ICF to communicative impairments

Because the ICF (and the ICF-CY) was designed as a general framework for the description of health and functioning, for a wide range of purposes, researchers and clinicians from different medical fields have described how to apply the ICF to their specific domains. For example, the application of the ICF to clinical practice in speech-language pathology has been described in Threats (2008) and in Threats & Worrall (2004). Although these authors identify some challenges in applying the ICF to communicative impairments, they also identify great benefits. For example, the ICF provides clinicians with a serviceable scheme for documenting effects of targeting factors in a client's environment (e.g. family counseling) on his or her communicative performance; without the use of the ICF, such intervention – although it often constitutes a large part of an SLP's work load – may have gone un-noticed (Threats & Worrall, 2004).

Over the last decade, the ICF and the ICF-CY have become more and more integrated into SLP research and clinical practice related to communicative impairments, with applications to, for example, individuals with voice disorders (Ma, Yiu, & Abbott, 2007) and motor speech disorders (Hartelius & Miller, 2010), and to children with language impairments (W. N. Campbell & Skarakis-Doyle, 2007), and Childhood Apraxia of Speech (Teverovsky, Bickel, & Feldman, 2009). The following section will provide a more detailed description of the application of the ICF (or, specifically, the ICF-CY) to the population in focus in this thesis, namely children with speech impairments.

2.2.1 Applying the ICF-CY to speech impairments in children

In a series of papers, Sharynne McLeod and colleagues have described the application of the ICF-CY to the domain of speech impairments in children (McCormack, McLeod, Harrison, & McAllister, 2010; McLeod & Bleile, 2004; McLeod & McCormack, 2007; McLeod, 2004, 2006). Table 1 illustrates the interrelations between the ICF components and their application to a pediatric speech and language assessment and therapeutic goal setting, as described in McLeod & Bleile (2004). The following paragraphs will explore these components and intervention strategies in more detail.

Table 1. Application of ICF components within the "Functioning and disability" part to speech and language assessment of children with speech impairments. An adapted version of the table presented in McLeod & Bleile (2004).

ICF components	ICF descriptors relating to communication	Possible components of an SLP assessment and analysis	Impairment-based goals	Socially-based goals
Body structures	• Anatomy of nose, mouth, pharynx, larynx • Genetic material	• Oromuscular examination • Audiological assessment	• Surgical repair	• Promote societal acceptance of diversity of the appearance of people
Body functions	• Articulation of phonemes • Expression of spoken language	• Phonological assessment • Analysis of speech and language	• Establish stable and accurate productions	• Decrease use of phonological processes contributing to unintelligibility • Reduce speech patterns that are unusual • Correctly produce phonemes that have a high frequency of occurrence
Activity and participation	• Verbal communication • Engaging in conversation • Interpersonal interactions and relationships • Community, social and civic life	• Intelligibility assessment • Gather qualitative descriptions of successes and difficulties in participation in social situations	• Correctly pronounce the child's name, and as well as the names of significant others • Pronounce words the child would like to say and or/has been teased about	• Collaborate with the child's teacher to promote communicative success, including strategies to repair communication failure • Work with the child's peers on awareness of communicative breakdown

For most children with SSDs, the origin of the disorder is unknown (T. F. Campbell et al., 2003). With regards to the ICF components, this indicates an intact Body structure, and instead an impairment in Body function. By assessing the child's speech production, e.g. by means of tests like the *Goldman Fristoe Test of Articulation* (GFTA; Goldman & Fristoe, 2000), or by estimating the percentage of consonants correct (PCC; Shriberg & Kwiatkowski, 1982), an SLP can describe the nature and severity of the speech impairment. These descriptive data may serve as references when determining impairment-based goals, e.g. increasing the PCC, or establishing stable and accurate productions of specific speech sounds. Therapeutic goals that are selected on the basis of their social relevance are, for example, decreasing the use of phonological (simplification) processes that are known to contribute to unintelligibility (e.g. Hodson & Paden, 1981), or to target correct production of those phonemes that have high frequency of occurrence. Section 4.1.2 will return to this issue, by presenting a deeper exploration of the influence of specific speech errors on intelligibility.

An intelligibility assessment, e.g. by means of the *Intelligibility in Context Scale* (ICS; McLeod, Harrison, & McCormack, 2012), will be informative of the child's ability to make himself/herself understood in different social contexts. As intelligibility is not only dependent on speech production factors, but also by contextual factors, e.g. the listener's familiarity with the child (Flipsen, 1995; for a more elaborated discussion, see section 4.1.1.), an ICS assessment is also regarded as being directed at Activity/Participation-related aspects of having a speech impairment. And as such, socially-based goals aimed at increasing intelligibility might involve collaboration with the child's family and teacher to develop strategies that will facilitate their understanding of the child, e.g. maintaining a diary of significant events that the child may want to communicate, or to encourage the use of visual cues.

Regarding the relation between children's communicative capacity and their communicative performance in their usual environment, there are several aspects to which this distinction is relevant for children with speech impairments. A child's communicative capacity is typically assessed in a clinical setting by the means of a standardized test, e.g. the GFTA (Goldman & Fristoe, 2000), whereas the child's communicative performance is assessed in his/her current environment, in settings representing communicative situations that the person is typically taking part in, e.g. by means of the ICS (McLeod et al., 2012). Hence, assessment of communicative performance necessarily involves people in the child's environment, e.g. parents or teachers, whereas assessment of communicative capacity is usually conducted by an SLP. Not only is the context of the assessment relevant when assessing levels of communicative competence and performance, but also the material upon which the assessment is based. For example, children with speech impairments often display differential results when producing single words versus connected speech (Morrison & Shriberg, 1992) or when producing imitated speech vs. spontaneous speech (Leonard, Schwartz, Folger, & Wilcox, 1978). Careful selection of material, combined with appropriate analysis of the results might provide the SLP with important information regarding characteristics of the speech impairment *per se*, as well as of how it is realized in a more ecologically valid context. On the one hand, if a child is exhibiting more problems when producing isolated words or when imitating speech compared to when (s)he is producing connected, spontaneous speech, it is indicative of a mismatch between communicative capacity and performance in the child. Moreover, it indicates that performance is better than the capacity would predict. This would be the case if the child were using communicative strategies to limit the functional effects of his/her speech impairment, e.g. by using gestures to support verbal communication, or by avoiding words that contain those speech sounds that (s)he misarticulates. If, on the other hand, the child exhibits more problems in spontaneous, connected speech than when producing isolated words or imitated speech, it would indicate a mismatch in the opposite direction, one where communicative capacity surpasses communicative performance. This would be the case if the child produces fewer errors in the more restricted context, and more in the free setting. This could be regarded as an effect of being more focused on the task of producing speech in the restricted setting, whereas when producing spontaneous, connected speech, the child is not only producing speech, but also has to allocate cognitive resources to linguistic planning. From these sketched scenarios, it should be clear that holistic assessments of speech impairments in children are incomplete unless they involve aspects of both performance and capacity. (Aspects of assessment will be discussed further in sections 3.1.3 and 4.1.1.)

2.3 Concluding remarks

This chapter has presented an overview over the ICF, with specific application to children with speech impairments. Hence, the reader should now be acquainted with the idea of describing conditions and functioning from different perspectives: on the one hand focusing on the condition itself, and on the other hand focusing on the functional consequences of having the condition. In the following chapters, speech impairments will be described from these different perspectives, with a narrowed focus on the specific type of speech impairment in focus of the present thesis, namely childhood speech sound disorders (SSDs). Notably, although the ICF promotes descriptions focusing on health and functioning, the following chapters will describe children with SSDs in terms of the difficulties they exhibit, i.e. less-well functioning aspects. So, in Chapter 3, the speech sound disorder *per se* will be described in terms of characteristics related to Body function: how the disorder is identified and assessed. Chapter 4, on the other hand, describes the disorder from the perspective of contextual factors: the functional consequences of having a SSD, and how it may affect the child's ability to participate and function in his or her daily life.

The studies included in Part II in this thesis can be categorized with regards to the different perspectives embraced by the ICF framework: whereas Studies I-III focus on the specific speech sounds that the children are misarticulating – how they are perceived by the child himself/herself (Study I and II), and how they are perceptually evaluated by other listeners (Study III) – Study IV explores the functional consequences of having the SSD, in terms of how misarticulations influence listeners' evaluations of intelligibility. Figure 2 illustrates how the included studies relate to the contents of the following two chapters. With this theoretical background, the reader will be well-prepared for Chapter 5, in which Studies I-IV are summarized and related to the research questions stated in the Introduction.

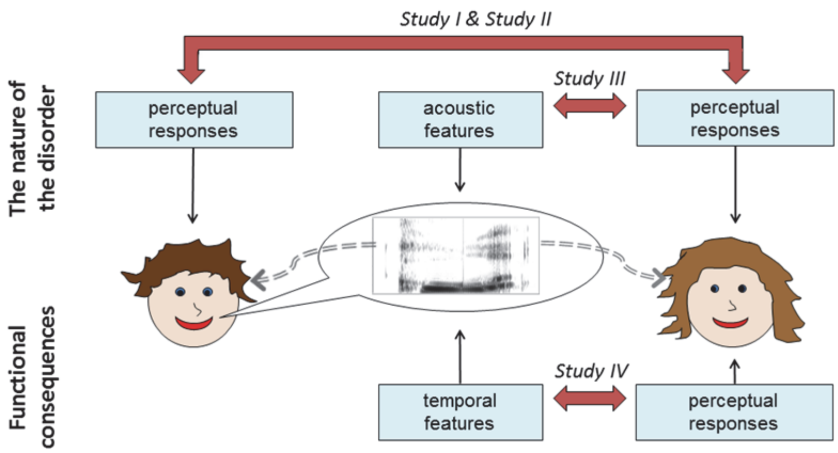

Figure 2. An overview of how Studies I-IV are related to the child's (to the left) speech production, in terms of how it is perceived both by the child himself or herself, as well as by others (represented by the adult female person to the right). Perceptual and acoustic analyses are described both at the level of the disorder itself (thus relating to Chapter 3), and to the functional consequences of having the disorder (thus relating to Chapter 4).

3. The nature of the disorder

The topic of the following chapter is the speech sound disorder *per se*. In ICF-terms, focus will be on Body structure and/or Body function. The group of children with speech sound disorders (SSDs) will be described, and different approaches to the identification of SSD subtypes will be presented. Then, a short note will address the issue of how the term SSD, and the clinical population to which it applies, relates to clinical practice in Sweden (section 3.2). This overview will be followed by a section describing how new insights into the nature of SSDs can be attained if analyses of speech production and perception involve more detailed representations than phonemic prototypes (section 3.3). This will equip the reader with an understanding of what characterizes the population of children participating in the studies included in this thesis.

3.1 SSD classification

The group of children presenting with SSDs is both large and heterogeneous. Reported prevalence estimates vary, but suggest that between 2% and 20% of preschool-aged children are affected (T. F. Campbell et al., 2003; Law, Boyle, Harris, Harkness, & Nye, 2000; McLeod & Harrison, 2009; Waring & Knight, 2013). Moreover, it is the most common diagnostic category among referrals to SLPs (Broomfield & Dodd, 2004; Mullen & Schooling, 2010). The term SSD covers a wide variety of speech problems in children, differing in aspects like severity, speech characteristics, co-occurrence with problems in other linguistic domains, and underlying cause. As the variation within the group of children with SSDs stretches along several dimensions, SSD subtypes can be defined by different features. This might explain why – even though this group of children has been studied and treated for several decades – there is still no universal and agreed-upon classification system for childhood SSDs (Rvachew & Brosseau-Lapré, 2012; Waring & Knight, 2013). Three major approaches to the classification of SSD subtypes can be identified, namely classification by a) etiological features, b) descriptive-linguistic features, or by c) speech processing features. This grouping has influenced the structure of the following sections, in which each approach will be dealt with in more detail.

3.1.1 Etiological features

Etiological (or medical) classifications of SSDs are based on known or assumed origins of the impairment. Broad-based classification systems like ICD-10 (WHO, 2010) do contain diagnostic labels relevant to childhood SSDs. For example, SSDs that are secondary to other medical conditions – like a hearing impairment, or a cleft palate – will be categorized with regards to this primary deficit. However, the criteria defining the diagnostic labels are not specific enough to allow identification of subgroups within the large and heterogeneous group of children whose SSD is their primary difficulty. These children have received a long-held interest, although the labels describing the type of problems they exhibit have varied. The speech disorder has often been referred to as a speech disorder "of unknown origin" (Shriberg et al., 2010), a "functional" speech disorder (Bradford & Dodd, 1996; Munson, Edwards, & Beckman, 2005a; Rvachew & Jamieson, 1989), or as a Developmental Phonological Disorder (Rvachew & Brosseau-Lapré, 2012; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997)[3].

An approach to the subcategorization of SSDs by etiological factors is the *Speech Disorders Classification System* (SDCS; Shriberg et al., 2010). The idea underlying the SDCS is that specific types of speech behavior can be linked to genetic and/or environmental conditions. Table 2 presents an overview over the subtypes identified in the SDCS. Here, SSDs of unknown origin are categorized with regards to three main types: Speech Delay, Motor Speech Disorders, and Speech Errors, each with their own

[3] As referred to in section 1.4, the term Developmental Phonological Disorder (DPD) as specified by Rvachew & Brosseau-Lapré (2012) is not in itself specific with regards to whether it is a primary difficulty, or whether it is secondary to some other condition. However, as these authors specify (*ibid.*, p. 317), their descriptions are focused on cases where the DPD is the child's primary difficulty.

subcategories. Diagnostic markers, i.e. specific symptom composites that differentiate between the eight SDCS subgroups, have been proposed; for example, the combination "Predominantly omission errors with few distortion errors" + "Reduced language test scores" + "Reduced performance on non-word repetition" would classify a speech disorder as a Speech Delay of Genetic type (Shriberg, 2010). However, as of today, the suggested markers are still neither specific nor sensitive enough to serve as accurate diagnostic indicators (Shriberg, 2010).

Although classifications based on etiologic features are necessary when seeking causal factors of SSDs, their clinical utility has been questioned. One aspect that has been criticized is the fact that the majority of childhood SSDs of unknown origin still falls into one category (the "Speech Delay – Genetic" type). (This also applies to the children participating in the studies included in this thesis.) This would not be a problem if this group had been homogenous – e.g. in terms of the speech processing capabilities underlying their speech problems – but evidence suggests that it is not (e.g. Rvachew, 2007). Another aspect that has been criticized is the division between problems with a known biological cause and problems whose cause is yet to be discovered (Rvachew & Brosseau-Lapré, 2012). Furthermore, the same authors caution against the discrimination between environmental and biological causes; the mutual interaction between these two factors, the authors argue, makes it impossible to separate one from the other.

Table 2. Etiological subtypes of childhood speech sound disorders of currently unknown origin and their genetic and/or environmental risk factors. Adapted table from Shriberg et al. (2010).

Type	Subtype	Risk Factors
Speech Delay (SD)	SD – Genetic	Polygenic/Environmental
	SD – Otitis Media with Effusion	Polygenic/Environmental
	SD – Developmental Psychosocial Involvement	Polygenic/Environmental
Motor Speech Disorder (MSD)	MSD – Apraxia of Speech	Monogenic? Oligogenic?
	MSD – Dysarthria	Monogenic? Oligogenic?
	MSD – Not Otherwise Specified	Monogenic? Oligogenic? Polygenic? Environmental?
Speech Errors (SE)	SE - Sibilants	Environmental
	SE - Rhotics	Environmental

3.1.2 Descriptive-Linguistic features

An alternative to classifying SSDs on the basis of underlying causes is classification determined by features observed in the children's speech. In these approaches, surface error patterns in speech production are often described in terms of how they differ from what is expected at a given age, thus accounting for developmental aspects of speech acquisition. One example of a descriptive-linguistic classification system has been suggested by Dodd (2005). An overview of the five subtypes as defined by Dodd is presented in Table 3, together with their characterizing features.

Table 3. Subclassification of childhood SSDs based on error surface patterns, as suggested by (Dodd, 2005). Adapted table from Waring and Knight (2013: p. 32).

Type	Subtype	Features
Phonetic	Articulation disorder	Substitutions or distortions of the same sound across contexts.
Phonemic	Phonological delay	Presence of usual phonological error patterns that are typical to younger children.
	Consistent phonological disorder	Consistent use of some unusual, non-developmental error patterns, e.g. backing or initial consonant deletion.
	Inconsistent phonological disorder	Variability/inconsistency in speech production, as indicated by multiple error forms for the same lexical item while having no oro-motor difficulties.
Motor planning, programming and execution	Childhood Apraxia of Speech (CAS)	Multiple deficits involving phonological planning, phonetic programming and motor programming implementation.

Dodd's classification builds upon the distinction between **phonetic** and **phonological** (or "phonemic") types of speech errors. Phonetic speech errors affect the production of specific speech sounds, typically across different contexts (e.g. in isolation, in isolated words, in connected speech). The most frequently reported phonetic errors regard the production of /s/ or /r/ (Dodd, 2005). (Notably, these problems would be characterized in the SDCS as Speech Errors, see Table 2.) Phonological speech production problems, on the other hand, affect phonological patterns, i.e. classes of speech sounds and phonotactic structure. These types of departures from adult phonology are often expressed as **phonological (simplification)**

processes (Nettelbladt, 1983). Examples of phonological processes that are often seen in typical speech acquisition are **velar fronting** (i.e. a restriction of place of articulation through the substitution of dental/alveolar sounds [t, d, n] for velar sounds /k, g, ŋ/) or stopping of fricatives (i.e. a restriction of manner of articulation through the substitution of stops for fricatives). If such phonological patterns are observed in children when developmental norms suggest that they should have been outgrown, the problems would fall into the subcategory Phonological **delay** in Dodd's differential classification system. Children whose speech exhibits phonological patterns that are unusual (non-developmental), e.g. initial consonant deletion, or backing (i.e. the substitution of the velar [k, g, ŋ] for the alveolar /t, d, n/), will instead be regarded as having a Phonological **disorder** – either of a consistent or an inconsistent type, depending on how the error patterns are distributed.

When applying Dodd's differential diagnosis classification to the population of children with SSD, the majority of cases fall into the category Phonological delay (Broomfield & Dodd, 2004). Evidence has been presented that links the different subtypes to distinct underlying deficits (Crosbie, Holm, & Dodd, 2009; Dodd & McIntosh, 2010), and that indicates increased intervention effects when intervention is matched to SSD subtype (Dodd & Bradford, 2000). Together, these sources of evidence support the validity of the classification system, and indicate that it serves to differentiate clinically meaningful subtypes of SSDs in children.

In terms of descriptive-linguistic features, the speech error types exhibited by the children participating in the studies included in this thesis (velar fronting) would be characterized as Phonemic (see Table 3). However, the participating children varied in whether they exhibited other speech error types in addition to their velar fronting pattern, and in whether the error patterns were consistent or not; therefore, their speech problems would be classified into different subtypes within the Phonemic category.

3.1.3 Speech processing features

Apart from classifications based on etiology or surface error patterns, childhood SSDs may also be categorized with regards to where in the speech production process an assumed breakdown occurs. When applied to typical speech and language acquisition, psycholinguistic models have been suggested that analyze the child's speech behavior at the levels of speech input and output, and of underlying representation(s) (e.g. Hewlett, 1990;

Menn & Matthei, 1992; for an overview over psycholinguistic models, see Baker, Croot, McLeod, & Paul, 2001). In the context of childhood SSDs, the most often cited psycholinguistic framework is the one described by Stackhouse & Wells (1997). Unlike the etiological/medical and descriptive-linguistic classification systems reviewed in the previous sections, Stackhouse and Wells' psycholinguistic framework was not designed as a classification system, but as an explanatory model for speech processing, in children with and without speech and/or reading impairments. Figure 3 illustrates Stackhouse and Wells' psycholinguistic model of different processing steps involved in the perception and production of speech.

By applying this model in the assessment of a child with SSD, an underlying cause can be sought with regards to individual components in the speech processing chain, from peripheral hearing, via auditory discrimination and lexical storage, to articulatory planning and, ultimately, the execution of speech. Although the model is a simplification of the complex interrelations between different skills involved in speech processing, it may prove useful when identifying weak links in an individual child, which may subsequently be targeted in intervention. For example, assessment of speech production might reveal a pattern of velar fronting, i.e. that the child's productions of /k, g, ŋ/ are perceived as [t, d, n]. A test of auditory discrimination – e.g. by asking the child to point to which one of two pictures representing the words *key* and *tea* corresponds to the auditorily presented [tiː] – could then provide information of whether the child perceives the phonological contrast between velar and alveolar plosives. A failure to perceive these contrasts may explain why the child does not react to the errors in his or her speech production, and will indicate this as a target in intervention. (Perceptual abilities in children with SSDs will be discussed in more detail below, in section 3.3.2.) Moreover, more detailed assessment of the production of velar plosives, e.g. whether the child is able to produce the target sounds in isolation, may reveal problems tied to the production of speech. Hence, by gathering information of the different components involved in the perception and production of speech, a psycholinguistic profile can be outlined for the individual child. And, after having identified weak links in the speech processing chain, these can be specifically targeted in intervention.

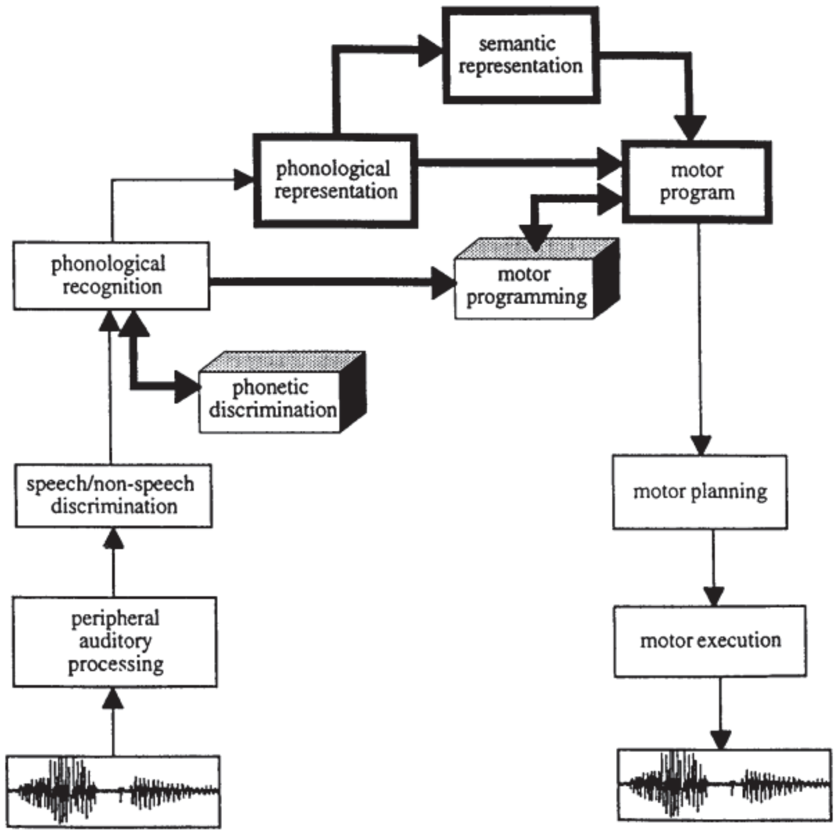

Figure 3. Speech processing model by Stackhouse & Wells (1997). Reprinted with permission from Wiley & Sons.

A distinction that is often made in the context of childhood SSDs is that between **phonological** and **articulatory** speech problems, where the former indicates an assumed deficit on a cognitive-linguistic level, in the child's knowledge of the language sound system, whereas "articulatory" indicates a more peripheral, motor-related, deficit (Dodd, 2005; Fey, 1992; Nettelbladt, 1983; Shriberg & Kwiatkowski, 1988). Thus, in this context, the term "phonological" is used as an explanatory (or "speaker-oriented") label, in contrast to its usage as a descriptive (or "data-oriented") label when categorizing surface error patterns. Although this alternation between a descriptive and an explanatory use of the term "phonological" has been questioned (Hewlett, 1985; Locke, 1983), it is still remnant today, e.g. in Swedish clinical terminology (Nettelbladt, Samuelsson, Sahlén, & Hansson, 2007). Moreover, links are often assumed between phonetic error patterns and articulatory restrictions on the one hand, and between phonological error patterns and a cognitive-linguistic limitation on the other hand (see Section 3.2). However, these suggested links have been questioned (Hewlett, 1990; Rvachew & Brosseau-Lapré, 2012). For example, Hewlett (1990) challenges the view that the type of error, e.g. whether /s/ is realized as [t] or [ɬ] would implicate differential causation, that the former is due to a cognitive-linguistic problem, whereas the latter is due to a problem of speech execution. The reasoning underlying this would be that a substitution of [t] for /s/ (i.e. a phonological speech error; stopping) reflects a failure to appreciate the phonological distinction between the target [s] and the error [t], whereas realization of [ɬ] for /s/ (i.e. a phonetic distortion) instead implies that the child does not have the articulatory skills required to produce an acceptable [s]. The growing body of literature highlighting the gradual nature of speech errors (Edwards, Gibbon, & Fourakis, 1997; Munson, Edwards, Schellinger, Beckman, & Meyer, 2010; Richtsmeier, 2010), which challenges the whole notion of phonological substitutions, can be used as fuel to support this criticism. In the section 3.3 below, the reader will be introduced to methods of analyzing speech from a perspective that speech sounds are continuous rather than categorical creatures. First, however, attention will be brought to how the term SSD, and the clinical population to which it applies, relates to clinical practice in Sweden. Moreover, some phonetic and phonological features of Swedish will be described, to inform the non-native reader of how the misarticulations observed in the children participating in the included studies relate to the research literature, which is primarily based on English data.

3.2 Language-specific considerations

Although SSD is included as a clinical diagnosis in the beta draft of the 11th version of the *International Classification of Diseases*, ICD-11 (WHO, 2014), it is not yet implemented in Swedish clinical practice, which is based on the Swedish version of ICD-10 (WHO, 2010). Hence, the terms used in

Swedish clinical practice today do not directly correspond to the terms used in international clinical research. There is a tradition in Sweden to distinguish between Phonological Disorders and Articulation Disorders, which reflects the view that phonological ability is part of language ability, and distinct from the more peripheral functioning of speech motor control (Hansson, 1998; Magnusson, 1983; Nettelbladt, 1983). Figures of prevalence and incidence of these disorders have been rare, but an extensive survey based on a sample (n = 1645) retrieved from the nation-wide developmental screening of all Swedish 4-year-old children revealed that 12% of these children exhibited phonological problems in speech production (Westerlund, 1994). (This figure did not include isolated problems with the production of /r/ and /s/.) Hence, speech sound disorders occur in a large population of Swedish children, which reflects similar epidemiological figures reported for other countries/languages.

Languages differ not only with regards to what speech sounds (phonemes) are used and how these are combined (phonotactics), but also with regards to what acoustic-phonetic shape these sounds may take. Therefore, it is important to consider how speech error patterns observed in one language relate to descriptions of speech errors in other languages. For example, whereas backing (i.e. the substitution of the velar [k, g, ŋ] for the alveolar /t, d, n/) is characterized as an atypical error pattern in English (Dodd, 2005; Hodson & Paden, 1983), it is observed more frequently in French-speaking children (Rvachew, Leroux, & Brosseau-Lapré, 2014). However, for Swedish just as for English, velar fronting is not only considerably more frequent than backing, it is also one of the most commonly observed speech error patterns in general, both in typical and delayed phonological development (Hansson & Nettelbladt, 2002).

Just like in English, Swedish has three unvoiced plosives, /p, t, k/, and for each of these, there is a voiced correspondent: /b, d, g/.[4] Figure 4 illustrates the differences in place of articulation in standard Swedish production of these plosives. Similarly to English, the set of Swedish consonants produced with a velar place of articulation contains /k, g, ŋ/, and their dental/alveolar correspondents are /t, d, n/. Evidence of slight articulatory/acoustic-phonetic differences have been reported between English and Swedish realizations of these speech sounds (Stoel-Gammon,

[4] The retroflex plosives [ʈ] and [ɖ] also occur in many Swedish dialects; there are, however, different views regarding whether they should be attributed phonemic status or not.

Williams, & Buder, 1994; Wood, 1975); however, these are not considered vital for the interpretation of the work in this thesis and its relation to previous literature.

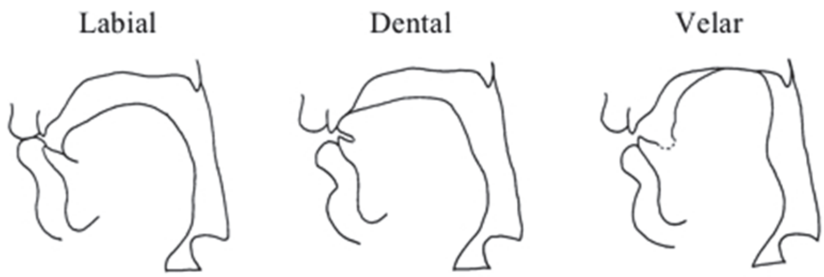

Figure 4. Articulatory configurations for the production of Swedish labial (left), dental (middle), and velar (right) plosives. From Engstrand (2007), based on X-ray images from Björn Lindblom. Reprinted with permission from Studentlitteratur.

3.3 Beyond phonemic prototypes

Critical to the description of speech perception and production in general, as well as in the description of SSDs, is appreciation of the fact that a description must involve (at least) two representational domains (Pierrehumbert, 1990). On the one hand, **phonological representation** involves categorical qualitative distinctions between speech sounds. This involves speakers' implicit knowledge of the finite set of speech sounds available in their language, and the ways in which they are combined. In this domain, speech sounds (phonemes) are identified by their linguistic function. On the other hand, **phonetic representation** involves gradient, physically measurable, differences between events in the physical world – be it an acoustic signal, articulatory movements, or auditory events. In this domain, speech sounds are characterized by their physical features, e.g. spectral characteristics, or the degree of tongue-palate contact. In recognizing the difference between these different types of representations, a description of speech production that only involves phonological and/or phonetic transcription risks being too simplistic. The transcription of speech inevitably involves a reduction of the abundance of information contained in the speech signal – information that, in the context of SSDs, can actually be quite revealing of the nature of the disorder. The following sections will describe some limitations involved in relying on transcription data alone in describing speech, and explore alternative ways of analyzing speech from the perspective that there is more to a speech sound than an IPA character can convey. These issues will first be explored with regards to speech production, then applied to the domain of speech perception.

3.3.1 Analysis of speech production

Phonetic transcription is an important tool in the description of speech. It is difficult to imagine what we would have known about speech at all if we had not had access to phonetic transcription (although a thoughtful suggestion is presented in Munson, Johnson, & Edwards, 2010). However, the use of phonetic transcription to describe speech has also shaped – and restricted – the way we think about speech acquisition and speech disorders. By solely relying on phonetic transcription, we might fail to register subtle signs of progression towards adult-like production in children's speech, as described e.g. by Macken & Barton (1980). In their longitudinal study of the acquisition of initial voiceless plosives in four English-speaking children, Macken and Barton observed a stage in the children's development where the initially absent distinction between voiced and voiceless plosives was acoustically measurable, but still not perceived by adult listeners. This delay between production of contrast and others' perception of contrast corresponds to stage 2 in Figure 5. Such physically measurable contrasts, which are typically not large enough to be registered by listeners, are often referred to as **covert contrasts**, and have been documented both in typical speech acquisition (Li, Edwards, & Beckman, 2009; Macken & Barton, 1980; Scobbie, Gibbon, Hardcastle, & Fletcher, 2000; Smith, 1979), as well as in disordered speech (Forrest, Weismer, Hodge, Dinnsen, & Elbert, 1990; Gibbon, 1999; Hewlett, 1988; Tyler, Figurski, & Langsdale, 1993; Weismer, Dinnsen, & Elbert, 1981). The lack of detail in phonetic transcription might solicit the interpretation that normalization of speech errors is an either/or-issue – that the child goes from not signaling a phonemic distinction at all, to suddenly doing so. However, explorations that include the analysis of acoustic (or articulatory) detail in children's speech production show that children's articulatory knowledge might actually be more advanced than what phonetic transcription would suggest.

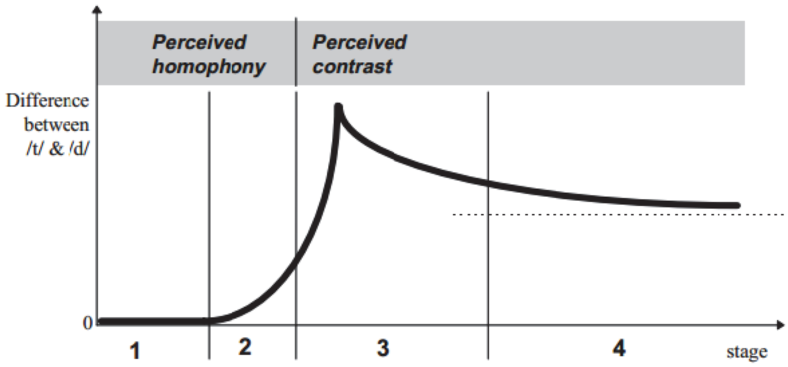

Figure 5. Children's gradual approach towards producing adult-like contrasts between speech sounds, as illustrated in Scobbie et al. (2000), from Macken & Barton's (1980) original suggestion. Reprinted with permission from Cambridge University Press.

By evaluating speech with reference to discrete categories, listeners are forced to disregard acoustic detail in the speech signal. This is true in phonetic transcription, and also in perceptual experiments where listeners evaluate speech samples with regards to whether they are "correct" or "incorrect", whether they are either /t/ or /k/, or with reference to any other categorical decision. Hence, listeners' sensitivity to acoustic detail in the speech signal is potentially masked by the choice of perceptual task. When listeners are instead asked to provide gradual responses, e.g. in rating the category goodness of speech sound stimuli with regards to a visual-analog scale (VAS) as illustrated in Figure 6, they have been found to make use of the entire scale rather than only the endpoints, indicating that they are indeed sensitive to fine phonetic detail (Kuhl, 1991; Munson, Johnson, et al., 2012; Munson, Schellinger, & Urberg Carlson, 2012). Moreover, at least for the /s/-/ʃ/-contrast, listeners have been found to distinguish "clear substitutions" of [s] for /ʃ/ from correct productions of [s] for /s/ (Munson, Johnson, et al., 2012). Hence, covert contrasts are not necessarily as covert as studies involving categorical listener decisions would suggest.[5]

[5] Note, however, that for the /t/-/k/ contrast, findings regarding listeners' ability to perceptually distinguish between [t] as a correct production of /t/ from cases where it is an unsuccessful effort at producing /k/ are inconsistent. Whereas Munson, Johnson, and Edwards (2012) report that listeners could not reliably distinguish [t] as a correct production of /t/ from cases where [t] were unsuccessful efforts at producing /k/, the results reported

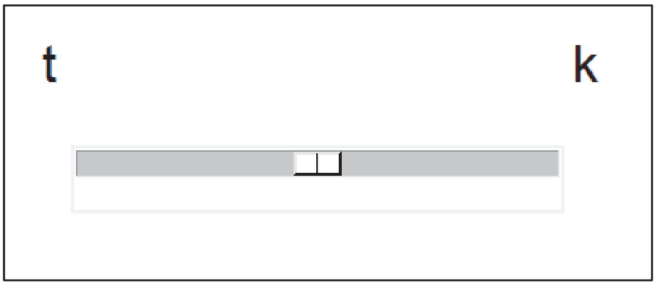

Figure 6. Example visual-analog scale (VAS) for the gradient evaluation of stimuli, where the leftmost endpoint represents a prototypical /t/ and the rightmost endpoint represents a prototypical /k/. From Study III.

One of the findings presented in Study III was that, for children with SSD, even their correct productions of /k/ are less prototypical – i.e. less /k/-like – than those of their TD peers. This finding was based on listeners' evaluations of the children's productions with reference to a visual-analog scale between /t/ and /k/ endpoints (as in Figure 6). And even if the correct /k/s produced by the children with SSD were generally very close to the /k/ endpoint of the scale, they were still significantly further away than the correct /k/s produced by the children with typical speech. Considering that the children with SSD had specific problems producing /k/ (and other velar sounds), these perceptual traces of their fronting pattern may not come as a surprise. More unexpected, however, is that the same pattern was observed for the children's productions of /t/, even though this was not a documented problem for the children with SSD. That is, for the children with SSD, their correct productions of /t/ were perceived as being less /t/-like than the /t/s produced by their TD peers. These findings exemplify patterns that are observable in perceptual evaluations with fine-grained measures, but would pass undetected if using categorical, or otherwise coarse-grained, measures.

in Study III indicate that this distinction is discernable, at least to adult Swedish listeners evaluating productions of Swedish child speakers. For the opposite pattern, both Munson et al. (2012) and Study III report that listeners distinguish [k] as a correct productions of /k/ from cases where [k] is an incorrect production of /t/; however, contradictory results have also been reported (Gibbon et al., 1993).

Even if the perceptual evaluation of children's speech production does not involve phonetic transcription or other categorical decisions, other factors might also mask physically measurable detail in the speech signal. One such factor is the potential **perceptual bias** in the listener, whether it is driven by expectations on the speaker, by lexical predictions, or by the listener's own linguistic experience. For example, listeners tend to interpret speech sounds differently depending on whether they believe that the speaker is a younger child or an older child (Munson et al., 2010); specifically, they tend to judge /s/-sounds as correct if they think that they were produced by an older child. Moreover, the context in which the speech sound is presented also affects listeners' perception of the sound; if ambiguous speech sounds are presented in word and nonword contexts, listeners will prefer phonological interpretations that make words. For example, in synthesized continua (varying in voice onset time) between the word *dash* and the nonword *tash*, listeners show a bias towards categorizing the initial consonant as a /d/, and an opposite bias for continua ranging from the nonword *dask* to the word *task* (Ganong, 1980). The listeners' own level of experience perceiving disordered speech may also affect their perceptual judgments. For example, whereas inexperienced listeners are biased towards interpreting /t/-/k/ productions as the more frequently occurring /t/, experienced listeners do not exhibit the same preference (Munson, Johnson, et al., 2012). Taken together, these findings highlight the need of careful selection of speech stimuli and participating listeners when designing experiments involving the perceptual evaluations of isolated speech sounds, in order to limit the potential influence of a perceptual bias.

In the Introduction (p. 4), reference was made to the commonly cited [fɪs] phenomenon (Berko & Brown, 1960). In this example, an adult listener reproduces a child's misarticulated production of the word *fish* as "fis", whereby the child reacts to this reproduction as being wrong, but still persists in producing "fis" in his efforts to clarify what he intended to say. With phonetic transcription as the instrument of documentation, this example has served as evidence that children – at some stage in speech acquisition – judge speech input, i.e. the speech they perceive, by different standards than they judge speech output, i.e. the speech they produce (see Baker et al., 2001). In light of the research presented in the previous paragraphs, however, an alternative interpretation emerges. A more detailed description may have revealed a fine-phonetic distinction between the child's production "fis" for the target /fɪʃ/, and "fis" as his reproduction of the adult's incorrect "fis", where the target is /fɪs/. If this covert contrast is not recognized – neither by the adult conversation partner, nor in the phonetic documentation of the episode – the puzzle to be solved is indeed why the child perceives errors in others' speech production but not in his or her own. However, if a covert contrast had been documented, this question would no longer be valid. Instead, other questions would arise, for example:

1. By what acoustic cues does the child differentiate between his or her production of /s/ and /ʃ/?

2. Are these acoustic cues also reflected in his or her perception of the /s/-/ʃ/ contrast?

Not only has the level of detail in the description of children's speech production influenced our understanding of SSDs, but the overlooking of covert contrasts can also be assumed to influence clinical practice. For example, Gardner (1997) reports how clinicians' insensitivity to covert contrasts in children's speech might not only cause communicative misunderstandings, but cautions that it may even hold back therapeutic progress.

Considering the limitations involved in describing disordered speech exclusively by means of transcription data and/or perceptual evaluation, it is important to gather supplemental information from other domains (Edwards, Munson, & Beckman, 2011; Kent, 1996). One source of fine-grained, continuous phonetic information can be provided from articulatory measurements, another from acoustic analysis; these approaches will each be described in the following sections.

3.3.1.1 Articulatory analysis

In terms of speech production, the sounds /t/ and /k/ differ by place of articulation. Whereas /t/ is typically produced with the tongue tip/blade at the alveolar ridge, /k/ is typically produced with the tongue body forming a constriction in the posterior or velar region of the palate (Hardcastle, Gibbon, & Jones, 1991). In adults, these articulation configurations are quite distinct, and involve independent control over the different parts of the tongue; the tongue tip/blade and the tongue body are moved independently from one another. Hence, in the typical production of [t], there is no tongue-palate contact except from the constriction between the tongue tip/blade and the alveolar ridge, and in the typical production of [k], the only tongue-palate contact is that between the tongue body and the velum (Hardcastle et al., 1991). In children with SSDs, however, other articulatory patterns have been observed.

Articulatory measures have provided evidence of **undifferentiated lingual gestures** (UGs) in children with SSDs in their production of lingual consonants (for a review, see Gibbon, 1999). These UGs involve an extended tongue-palate contact, so that in their production of alveolar plosives /t/ and /d/, many children with SSD form a constriction that extends from the alveolar ridge back to the palatal and even velar regions of the palate. However, the perceptual consequences of this atypical articulation pattern vary; intended /t/s produced with UGs are often perceived as correct productions of [t], but at times as [k] substitutions (Gibbon, 1999). A critical factor in the perception of velar/alveolar stops produced with UGs is the last tongue-palate contact at the release of the plosive; a release at the alveolar ridge tends to yield an alveolar percept, whereas a more velar release will yield a velar percept (Gibbon, 1999). Based on these observations, Richtsmeier (2010) suggests an articulatory/physiological account for why velar fronting is a common error pattern in children's speech, and particularly so in word-initial position. In producing a word beginning with a velar plosive, a child positions his articulators to produce a /k/. For some children, this will involve an extended constriction along the hard palate, i.e. an undifferentiated gesture. At the release of the plosive, airflow originating from the trachea will break the constriction *from the direction of the force*. Hence, the last point of contact will be at the front end of the constriction, possibly as far as at the alveolar ridge. And consequently, the intended /k/ will be perceived as fronted. It should be noted, however, that regarding the distinction between /t/ and /k/, UGs have been reported only for misarticulation of /t/ and /d/ (backing), which is an atypical speech error, often indicating a more severe impairment (Dodd, 2005; Hodson & Paden, 1983; Nijland, 2009). Descriptions of the articulatory patterns involved in velar fronting have – somewhat surprisingly – been rare. Therefore, articulatory data is still needed to explore whether the reported findings on UGs extend to the pattern of velar fronting in pre-school-aged children – such as the children participating in Studies I and II.

Distributed articulation of alveolar and velar plosives has also been observed in adult speakers, in a fast-rate tongue twister task (Marin, Pouplier, & Harrington, 2010). In their study of the correlation between articulatory and acoustic features of misarticulated /t/ and /k/, Marin and colleagues found that these sounds were often produced with a simultaneous tongue tip and tongue dorsum constriction. Interestingly, this articulatory pattern was found to affect the two plosives differently; whereas for intended /k/s, the articulation resulted in a /k/-like spectrum, intended /t/s produced with the same configuration resulted in more ambiguous spectra, with features typical of both /t/ and /k/. Hence, this articulatory-acoustic asymmetry is biased towards /k/. It is not yet clear, however, whether this bias extends to the articulatory-acoustic relationship in children's speech.

3.3.1.2 Acoustic analysis

In terms of acoustic phonetics, alveolar plosives differ from velar plosives by features of the burst of the plosive, and of the transition into the following vowel (Lindblom, 1963; Löfqvist, 1999; Marin et al., 2010; Stevens & Blumstein, 1978; Sussman, McCaffrey, & Matthews, 1991). Especially for the voiced /d/-/g/ contrast, **acoustic cues** to the distinction between the two have often been sought in the transition from the release of the plosive into the following vowel. For example, Figure 7 displays distinct schematic formant patterns (F1 and F2) for labial, alveolar and velar voiced stops, respectively. Over 60 years ago, these patterns were suggested as being the minimal acoustic information needed to distinguish between the different plosives (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952). However, although these schematic templates make sense for early rule-based synthesizers, where few acoustic parameters are controlled to create different percepts, they are insufficient in the description of natural speech (Blumstein & Stevens, 1979; Fant, 1973). And even if more acoustic dimensions are taken into account, identification of stops by means of mapping spectral features in children's speech to spectral templates based on adults' speech has met limited success (Chapin, Tseng, & Lieberman, 1982; Hewlett, 1988).

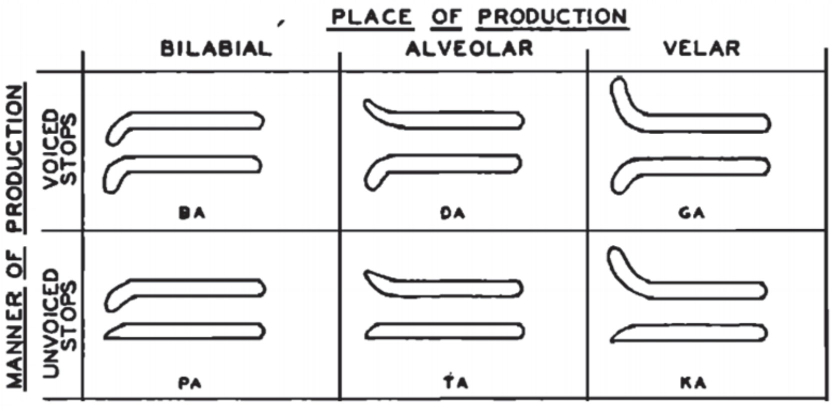

Figure 7. Schematic formant transitions (F1 and F2) into the vowel following bilabial, alveolar and velar plosives. From Cooper et al. (1952). Reprinted with permission from The American Institute of Physics for The Acoustical Society of America.

The acoustic realization of plosives varies with the quality of the following vowel (Delattre, Cooper, Liberman, & Gerstman, 1954; Kewley-Port, 1982). Notably, these co-articulation effects are stronger for velar consonants than for alveolar consonants, i.e. the influence of the following vowel is stronger on velar plosives than on alveolar plosives (Hewlett, 1988). A measure developed to account for the acoustic variation across different vowel contexts is **locus equations** (Lindblom, 1963). This measure captures the relation between F2 frequency at the onset of voicing after the release burst, and target F2 frequency in the following vowel. However, although there are examples where locus equations have been used to, quite successfully, distinguish place of articulation in voiced plosives in adult speech (Sussman et al., 1991), the evidence of successful application to children's speech has thus far been limited (Tyler et al., 1993).

For the voiceless /t/-/k/ distinction, where the absence of voicing limits the possibilities of tracking formants in the transition into the vowel, differentiating acoustic cues are often sought in the release burst of the plosive (Blumstein & Stevens, 1979; Forrest, Weismer, Milenkovic, & Dougall, 1988; Marin et al., 2010). For example, Forrest and colleagues (1988) applied a **spectral moments** analysis to capture dynamic aspects in the interval from the burst release to the onset of the vowel. In their analysis, three spectral parameters – mean frequency, skewness, and kurtosis – were used to successfully (with an accuracy of 92%) classify initial voiceless (correctly produced) plosives by their place of articulation. In the previously mentioned study by Marin et al. (2010), where adult misproductions of /t/ and /k/ were studied both in terms of their articulatory features and of their acoustic features, the acoustic measures were based on time-varying spectral information in the burst, as expressed by the first three **discrete cosine transformation (DCT) coefficients** corresponding to the spectrum's mean, slope and curvature as they change across time in the burst. Furthermore, Marin and colleagues' approach also involved a transformation of the spectral information to the Bark scale, in order to better reflect the human auditory system. Although the spectral moments analysis method and the DCT coefficient analysis method are both measures of temporal changes in the burst spectrum, and, as such, quite similar, there is a value in knowing whether one should be preferred over the other in studies of correctly and misarticulated productions of /t/ and /k/ in children. Study III constitutes (to the author's knowledge) the first comparison between these two approaches, in terms of their capacity to reflect human perceptual responses. For the Swedish data in Study III, the DCT approach proved slightly more accurate than the spectral moments approach when classifying clear correct cases of /t/ and /k/. However, with an accuracy of around 82%, the results show that even the best acoustic analysis method is still not sensitive to all acoustic information that human listeners base their phonological decisions upon. Hence, the methods we use today need to be further improved before they can be used as objective measures of phonetic/phonological accuracy.

Studies of the acoustic characteristics of misarticulated /k/s, i.e. productions that are often perceived as [t], have revealed acoustic evidence of covert contrasts in some children (Forrest et al., 1988; Hewlett, 1988; Tyler et al., 1993). However, for other children – although they did display the same perceived error pattern – the authors did not find any acoustic differentiation between /t/ and /k/. This could indicate either that a) not all children perceived as substituting alveolar sounds for velar sounds differentiate between /t/ and /k/ in their speech production, or b) the acoustic analysis methods applied to date are not sensitive to the acoustic cues these children may use to signal the /t/-/k/ contrast. As cautioned by Munson and colleagues (Munson et al., 2010), evidence of covert contrasts may be difficult to find by means of acoustic analyses, as this type of analysis involves pre-selection of what acoustic parameters to explore. If the hypotheses underlying this pre-selection are inadequate, the chances of finding acoustic evidence of covert contrasts are small. Moreover, articulatory-phonetic strategies may differ from child to child, and analyzing acoustics across children might hamper detection of individual patterns.

Considering that acoustically measurable contrasts – even though they may pass undetected by the listener – are positive prognostic factors (Tyler et al., 1993), there is a need for a more detailed acoustic analysis than is currently available in speech and language clinics. If sensitive acoustic measures were easily available, such information could be used to monitor therapeutic progress, thus influencing clinical decisions on start and withdrawal of treatment. As shown in Study III, however, state-of-the-art acoustic analysis methods still need to be improved before measuring up to the acoustic-phonetic sensitivity of human evaluators.

3.3.2 Analysis of speech perception

Assessment of children with SSDs is not complete unless it contains information about the children's perceptive abilities, particularly with regards to the contrasts that they (seemingly) fail to signal in speech production (Locke, 1980b; Magnusson, 1983; McGregor & Schwartz, 1992; Rvachew & Brosseau-Lapré, 2012). A finding that the child's perception of the misarticulated speech sound(s) is inadequate motivates the inclusion of a perceptual component in clinical intervention (Rvachew, 1994). However, in the assessment of speech perception, it is important to consider two levels of phonological knowledge: **knowledge of higher-level phonological categories** and **perceptual knowledge** (Munson, Edwards, & Beckman, 2005b; Rvachew & Brosseau-Lapré, 2012). The higher-level knowledge involves appreciation of the linguistic function of speech sounds, i.e. how they are used to code meaning. For example, the perception of the difference between the words *tea* and *key* involves recognition of a phonological contrast between /t/ and /k/ – that exchanging one for the other will result in a change of meaning. Perceptual knowledge, on the other hand, involves the knowledge of how the different phonemic categories are represented in the acoustic-phonetic domain. These acoustic-phonetic representations can be thought of as cluster regions in a multi-dimensional acoustic-phonetic space, shaped by the listener's accumulated exposure to speech, and by making associations between the acoustic-phonetic form of speech sounds and their linguistic function (Munson et al., 2005b; Pierrehumbert, 2003). For example, the acoustic-phonetic representation of /k/ will encompass the acoustic-phonetic variety of speech sounds that the listener will perceive as /k/.

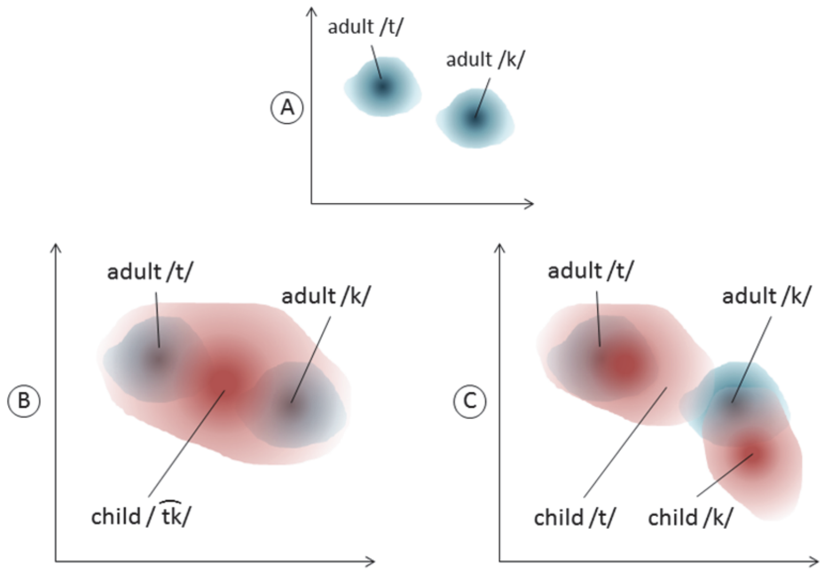

Figure 8. Two alternative hypothetical scenarios accounting for a child's inadequate perceptual knowledge of the /t/-/k/ contrast, where the child's acoustic-phonetic representations (in red) do not match those of adults (in blue). Panel A: Adult acoustic-phonetic representations of the phonemic categories /t/ and /k/. Panel B: The child's phonological system does not have separate phonemic categories for /t/ and /k/; instead, there is one category /t͡k/. Panel C: The child appreciates the /t/-/k/ contrast on a phonological level, but distinguishes perceptually between the two by different acoustic-phonetic criteria than adults do. Hence, the child's acoustic-perceptual representation for /t/ encompasses not only sounds that adults would perceive as /t/, but also some speech sounds that adult listeners would perceive as /k/. (For illustrative purposes, the acoustic-phonetic space is represented as a two-dimensional space, with arbitrary acoustic values on the x- and y-axes. In real life, obviously, natural speech sounds vary along more than two acoustic dimensions.)

In Figure 8, the multi-dimensional acoustic-phonetic space is visualized as a space in two arbitrary dimensions, with cluster regions representing listeners' categorization of the speech sounds /t/ and /k/. Panel A represents how these phonemic categories are represented as attractors in the acoustic-phonetic space, with the centers (i.e. the darker regions) representing prototypical realizations of that specific phonemic category, with boundaries that are gradient rather than discrete.

A failure to perceive a contrast between two speech sounds, e.g. the /t/-/k/ contrast, may explain why a child does not react to his or her own misproductions of the target sound. However, unless further examined, this failure does not reveal the level at which the breakdown occurs. Is it because the child merges /t/ and /k/ into a single phonemic category (as illustrated in panel B in Figure 8)? Or is it because the child's perceptual knowledge of the speech sounds is inadequate, such that some acoustic-phonetic realizations that adult listeners would categorize as /k/ will be perceived by the child as realizations of /t/ (as illustrated in panel C in Figure 8)? As the figure indicates, these children will align with adults in their identification of prototypical productions of /t/ and /k/, as illustrated by the child's acoustic-phonetic representations for /t/ and /k/ embracing the darker regions of the adult acoustic-phonetic representations – i.e. the "prototypical regions" – of the corresponding adult categories.

Among the children who are (perceived as) substituting one speech sound for another, some display problems on the higher-level knowledge of phonemic categories. These children will typically fail on a task like the *Speech Production-Perception Task*, SP-PT (Locke, 1980b), which assesses the child's ability to appreciate the phonological contrast between clear productions of the target sound, the substitution sound and a control sound. In Studies I and II in the present thesis, this task was tailored to the pattern of velar fronting and adapted to Swedish (see Figure 9). Here, a picture illustrating the word "kam" (English: *comb*) was presented to the children, while a randomly generated recording of the phrase "Är det här en X?" (English: *Is this a X?*) was played. Here, X was either "kam" (the target word), "tam" (the target word produced with the substitution sound) or "pam" (the target word produced with the control sound). However, although a few children in these studies did fail on the SP-PT, which has also been reported in other studies with similar assessments (Locke, 1980b; McGregor & Schwartz, 1992; Monnin & Huntington, 1974), most children passed this test without difficulties. This is consistent with the observation that children with SSDs often do not have problems with phonological identification of prototypical productions of the sounds they produce in error (Hoffman, Stager, & Daniloff, 1983; Locke, 1980b; Rvachew & Jamieson, 1989). Hence, for many children with SSDs, their knowledge of higher-level phonological categories is intact. Assessment involving the categorization of less prototypical productions, however, often reveals that many children instead have problems at the level of perceptual knowledge of the sounds they produce in error (Edwards, Fox, & Rogers, 2002; Hoffman, Daniloff, Bengoa, & Schuckers, 1985; Rvachew & Jamieson, 1989). So, even if a child accurately discriminates between prototypical exemplars of a target sound and a substitution sound, the acoustic-phonetic criteria underlying these phonological decisions might not be accurate. Instead, the decision of whether a speech sound should be classified as belonging to one phonemic category or another might be based on acoustic-phonetic cues along dimensions that adults do not normally attend to in their phonological decisions (Locke, 1980a; Rvachew & Brosseau-Lapré, 2012). Assessing this level of the child's phonological knowledge is, however, much less trivial.

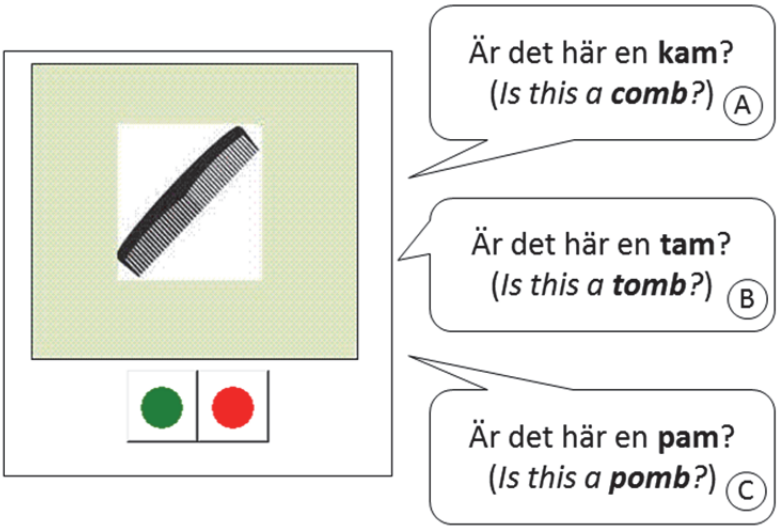

Figure 9. The Speech Production-Perception Task, SP-PT (Locke, 1980b) adapted to the pattern of velar fronting in Swedish. Here, the picture of a comb (Swedish: kam, target: /kam/) is presented while pre-recorded questions are presented in random order, asking whether the object is a) a /kam/ (correct production of the target word), b) a /tam/ (the target word produced with the substitution sound), or c) a /pam/ (the target word produced with a control sound).

The exploration of what acoustic-phonetic criteria underlies a listener's phonemic classification of speech sounds typically involves a) selection of acoustic parameters assumed to be phonologically critical to proficient language users, and b) controlled manipulation of these parameters while registering perceptual effects in the listeners' responses. Figure 10 illustrates a classic example of **categorical perception** (Liberman, Harris, Hoffman, & Griffiths, 1957), where stepwise manipulation of features of the F2 transition from a voiced plosive into the following vowel is reflected as abrupt shifts from /b/ to /d/ to /g/ in the listeners' identification responses.

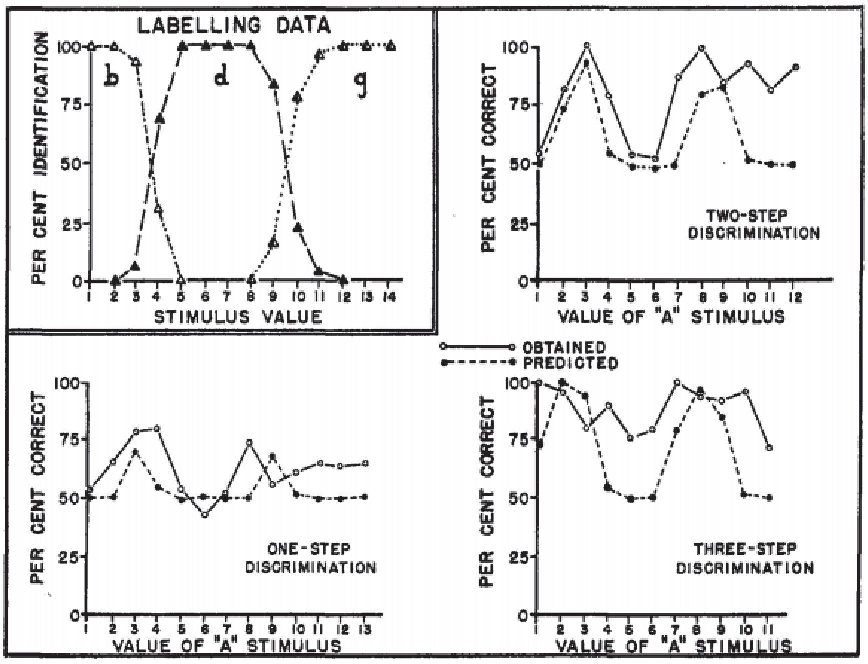

Figure 10. Illustration of categorical perception, from Liberman et al. (1957). As illustrated in the top left panel, the modification of the acoustic parameters (extent and direction of F2 transition) in 14 equal steps (represented on the x-axis) results in abrupt shifts in the subject's identification of the stimuli as /b/, /d/ or /g/(represented on the y-axis). The other three panels illustrate the subject's discrimination responses, indicating that discriminating is more successful between across-category stimuli than between within-category stimuli. Reprinted under fair use copyright provisions.

Studies applying this experimental design have shown that although adult listeners with the same language background tend to agree on where in the acoustic continuum these phonemic boundaries lie (Liberman et al., 1957; Miyawaki et al., 1975), different (groups of) listeners rate the same acoustic changes differently (Hazan & Barrett, 2000; Idemaru & Holt, 2013; Nittrouer, 2005; Rvachew & Jamieson, 1989). For example, Hazan & Barrett (2000) showed that in typical speech acquisition, perceptual strategies develop through childhood, and have still not reached adult norms by the age of 12. Particularly, the boundaries of children's acoustic-phonetic representations are less sharp than those of adults (cf. Figure 8), and they rely on more acoustic redundancy in phonemic identification (Edwards et al., 2002; Hazan & Barrett, 2000). Diminished perceptual skills are even more pronounced in children with SSDs, relative to their typically developing peers (Edwards et al., 2002; Ohde & Sharf, 1988; Rvachew & Jamieson, 1989). Considering the known variation among immature listeners in what acoustic-phonetic changes influence their phonemic categorizations, selecting and modifying acoustic cues on the basis of features that are phonologically critical in proficient language users may limit the chances of identifying the acoustic features that are involved in the phonological decisions of an individual child. A finding that none of the selected parameters influence the phonological decisions of an individual child does not necessarily mean that the child is insensitive to this phonological contrast. The child may instead base his or her phonological decisions on (a combination of) acoustic cues that the examiner has not yet identified.

The finding that children do not attend to the acoustic-phonetic cues that others engage in their phonological decisions does not necessarily imply that they consistently rely on other cues. As some have suggested (e.g. Shiller, Rvachew, & Brosseau-Lapré, 2010), the children's acoustic-phonetic representations might instead be less stable, i.e. more unreliable. With regards to the illustrations in Figure 8, this would be represented as less sharp, or even undefined, boundaries of the red clouds. Evidence that children with SSDs have more difficulties compared to their peers to extract linguistic information from speech produced by a speaker of an unfamiliar accent (Nathan & Wells, 2001) indicates that these children may have a difficulty dealing with acoustic-phonetic variation. Although there is still a paucity of research on how children with SSDs perceive socio-indexical features of speech (Edwards & Munson, 2008), it is conceivable that these children exhibit a more general insensitivity to how acoustic-phonetic variation is used to convey information – not only when associating acoustic-phonetic features to phonemic categories, but also in appreciating acoustic-phonetic cues to speaker identity. In Study I, children with SSD were found not to use the acoustic-phonetic traits of their own misproductions as cues to speaker identity. For example, these children were no better at distinguishing their own misproduction [tat] of the word "katt" (Eng. *cat*, target /kat/) from peer-produced correct productions [kat], than they were of distinguishing their correct productions from peer-produced correct productions. Although salient to proficient language users, these across-category misproductions did not serve as cues for the children to identifying the speaker as themselves. This finding indicates a diminished ability in these children to register idiosyncratic acoustic-phonetic cues to speaker identification. On the other hand, the findings in Shuster (1998) indicate that children with SSDs do attend to idiosyncratic articulation as a cue when determining whether a recorded production had been produced by themselves or not. When their misproduced /r/s had been edited to sound correct, the children in Shuster's study no longer identified the speaker as themselves. However, these two studies differ with respect to the groups of participating children, both in terms of age and surface error patterns. In light of the protracted developmental course of speech perception skills in typically developing children (Hazan & Barrett, 2000; Idemaru & Holt, 2013; Nittrouer, 2005; Rvachew & Jamieson, 1989), the 4-7-year-old children in Study I are indeed expected to use less mature perceptual strategies than the 7-13-year-old children in Shuster (1998). This is further strengthened by the finding that among the children in Shuster (1998), the older children were more accurate in their speaker identification than the younger. However, it should be clear that more studies exploring how children with SSDs perceive cues to speaker identity and social group membership are needed to arrive at firmer conclusions regarding how these children deal with (within-category) acoustic-phonetic variation.

In pursuing the exploration of the interplay between perception and production of speech, alternative routes to the ones described above are also available – routes where the point of departure is the child's perception of his or her own misarticulations.

3.3.2.1 Children's perception of their own misarticulation

The perceptual problems that many children with SSDs have can be assumed to interfere with their acquisition of adequate speech production. Specifically, if the children's perceptual evaluation of their own speech production does not align with the acoustic-phonetic norms in their surrounding language environment, their approach to normalized speech production will be protracted. That is, if the children do not perceive their own speech production as being off-target, they will be less motivated to alter their articulatory strategies (Hewlett, 1990). Moreover, if the children can learn to appreciate whether the speech they produce is phonetically adequate or not, they will be better able to monitor their own speech production. And thereby – assuming that they also have the articulatory skills to correct the errors – they will be better equipped to generalize correct speech production beyond the clinical setting (Koegel, Koegel, & Ingham, 1986). Hence, there are multiple reasons why information regarding the children's evaluation of their own speech production is clinically valuable.

When misarticulating children are asked to evaluate their own speech production in immediate conjunction with the act of speaking, they often appear insensitive to their own misarticulations. This has been reported by Locke & Kutz (1975), and it is also one of the findings in Study II. Many children with SSDs also judge recorded versions of their own misproductions by other standards than adults or peers do. For example, evidence presented by Kornfeld and Goehl (1974) suggests that children can indeed recover their intended productions of words that others perceive as homonyms. And when asked to evaluate whether their own recorded speech production is correct or not, misarticulating children often evaluate what others perceive as misarticulations as correct, as shown by Magnusson (1983) and Shuster (1998), and further corroborated by the findings in Study II. Hence, these children seem to base their evaluations on acoustic information that is disregarded by others. Shuster (1998) suggests that this might be explained by the children's acoustic-perceptual representations being too broad, encompassing not only what others perceive as adequate productions of a sound, but also the child's own misproductions of the sound (cf. panel C in Figure 8). As argued by McAllister Byun (2012) as well as in Study II in the present thesis, this assumed perceptual parallel to the children's production may be a consequence of their extensive exposure to their own misproductions as possible realizations of the target sound. This also fits well with the predictions assumed in the Linked-Attractor model of phonological development (Menn, Schmidt, & Nicholas, 2013), that extensive exposure to the mapping between a phonological intent (i.e. what the child intends to say) and an articulatory/auditory consequence of this intent strengthens this association. It remains to be explored, however, whether the children base their evaluations of correctness of their own speech production on the same physical measures by which they judge others' speech production. As speculated in Study II, based on the evidence that children tune their acoustic-phonetic representations to individual speakers (McQueen, Tyler, & Cutler, 2012), children's evaluation of their own speech production is most likely dependent on an interaction between the acoustic-phonetic criteria underlying their phonological identification and the perceived speaker identity. This hypothesis is supported by the findings in Shuster (1998), where children accurately evaluated corrected versions of their own misarticulated /r/s as being correct, while not accurately identifying the speaker as being him or herself. In other words, what is perceived as a phonetically adequate realization of a speech sound if produced by one speaker (e.g. oneself) might be perceived as incorrect if it had been produced by someone else, and vice versa.

Despite the evidence that some children with SSDs seem more tolerant to their own errors, it is worth noting that not all children who misarticulate can recover their initial intentions (Locke & Kutz, 1975; Panagos & King, 1975), and not all children are oblivious to their own errors (Woolf & Pilberg, 1971). For example, Locke & Kutz (1975) report a study where children, who in their speech production neutralized the /r/-/w/ contrast, could not distinguish whether their recorded productions [wɪŋ] were an intended *ring* or *wing*. Hence, if these children at all signaled the phonemic /r/-/w/ contrast in speech production, the acoustic consequences of this differentiation were so subtle that they were not even perceived by the child himself/herself. (As the behavioral responses in Locke & Kutz' study were not related to acoustic or articulatory measures of the children's productions, there is no way of knowing whether the children did signal the phonemic contrast.) Other studies of children's perception of their own inaccurate speech production indicate that the children who are most consistently misproducing are the ones who are least successful in evaluating the accuracy of their own speech (Aungst & Frick, 1964; Lapko & Bankson, 1975; Shelton, Johnson, & Arndt, 1977; Shuster, 1998). This pattern was also observed in Study II. However, in the absence of correlating physical measures, these behavioral findings also remain inconclusive. Unfortunately, there is still a void of studies relating children's perception of their own productions to physical measures of articulation and/or acoustics. While awaiting firmer evidence on children's perception of their own misarticulated speech, and how it is related to the actual physical representations of the misproduced speech, conclusions cannot be drawn beyond the following conjecture, suggested by Magnusson (1983: 188): "children's ability to make discriminations in their own speech is dependent on the extent to which they differentiate the forms in their production". And, to this one might add, the chances of finding physical evidence of children's differentiation in production are dependent on experimenters' success in identifying the acoustic or articulatory cues involved.

From the above, it should be clear that the acoustic-phonetic representations underlying children's perception of their own misarticulated speech production are often inaccurate. Moreover, just as speech perception in adults is dynamic – in the sense that the same acoustic-phonetic signal may elicit different perceptual responses depending on factors like the listener's lexical expectations (McQueen et al., 2012), or on the ordering of perceptual stimuli (Tuller, Case, Ding, & Kelso, 1994) – there are reasons to assume that children's perception is also dynamic. Although little is yet known about the dynamics of speech perception in children with SSDs, the findings in Study II lend support to this assumption. Here, children with PD were found to evaluate their own speech production differently depending on the time elapsed since their production of an utterance. Specifically, the children often evaluated their misarticulated utterances as correct in immediate conjunction with the act of speaking, and appeared to perceive inaccuracies only when time had passed since their production of the utterances. Recent literature on the dynamic nature of speech perception (Kaschak & Gernsbacher, 2014) invites the interpretation that the assumed bias in the children's acoustic-phonetic representations for a particular speech sound (i.e. one that encompasses their own misarticulated production of the target sound) is enhanced every time they produce this speech sound. With every new misproduction of the target sound, their already established associations between the phonological intent (e.g. /k/) and the acoustic-phonetic product of this intent (e.g. [t]) are renewed and strengthened. As time passes from their production of the utterance, this **perceptual tuning** of the children's acoustic-phonetic representation of the target sound attenuates, and this representation returns to its default state. This default state is shaped not only by the child's previous exposure to his or her own misproductions of the target sound, but also by a myriad of phonetically adequate exemplars of the target sound, as produced by others. Although not yet tested, this interpretation would explain why the children's evaluations of their own speech production changes with the time elapsed since their production of an utterance, such that their immediate evaluations are more permitting to errors, whereas their delayed evaluations are more congruent with those of typically developing peers.

An extension to the exploration of how children perceive their own misarticulated speech is suggested in the previously mentioned study by Shuster (1998), and also in Study II. Here, recordings of children's unsuccessful efforts at producing words containing a specific target sound are modified to sound correct, and played back to the children. Through the use of corrective re-synthesis, Shuster (1998) demonstrated that children's perception of corrected versions of their misarticulated /r/s – i.e. artificially constructed exemplars of what it would have sounded like if the child had been able to produce the utterance correctly – aligned with adult evaluations of the same utterances. In other words, the children perceived the edited-correct utterances as being correct. Interestingly, the children also perceived their original misarticulated productions as being correct. Based on this finding, Shuster suggested that (1998: 947): "these subjects' underlying representation for /r/ is too broad, encompassing at least some of their own incorrect utterances in addition to both their own correct utterances and the correct productions of the other speaker". Notably, this suggestion fits well with an exemplar-based model of speech perception like the one illustrated in Figure 8, but is more difficult to reconcile with a two-lexicon model view on speech perception, which assumes different internal referents for perception and production of speech (e.g. Hewlett, 1990; Kiparsky & Menn, 1977). Hence, the novel experimental design has yielded some important insights into children's perception of their own speech production that could not otherwise have been achieved. The findings in Study II are also consistent with an exemplar-based view on speech perception, although they – as mentioned above – indicate that temporal tuning effects need to be integrated into the model. Furthermore, several reasons can be identified as to why using corrective re-synthesis in clinical intervention of SSDs may prove beneficial. These will be discussed in more detail in section 6.4.

3.4 Concluding remarks

This chapter has presented an overview of the classification of speech sound disorders (SSDs), focusing on the population of children participating in the studies included in this thesis. A speech error pattern shared by all participating children is velar fronting, i.e. the (perceived) substitution of alveolar [t, d, n] for the velar targets [k, g, ŋ], although most children also exhibit additional error patterns. The reader will now be familiar with some of the limitations involved in describing production and perception of misarticulated speech solely on the basis of phonetic transcription, and of the caution that is due when describing "substitutions" of one sound for another. Furthermore, the reader will be equipped with an appreciation of the wide heterogeneity within the population of children with SSDs, with regards to factors like assumed underlying causes, exhibited speech error patterns and perceptual abilities. With this as a backdrop, the following chapter will shift perspective from focusing on the disorder *per se*, to looking at the functional consequences of having the disorder.

4. Functional consequences of having the disorder

So far, focus has been on defining and describing the SSD itself, while paying little attention to the context in which it appears. In this chapter, however, attention will be directed towards the communicative context, embracing aspects related to the ICF components Participation and Activity, as well as to Environmental factors. Specifically, this chapter will explore the repercussions the disorder may have on the child's communicative functioning. While acknowledging that impaired communicative functioning in turn may pose detrimental influence on the child's social relations, school achievement and learning, and possibly also impact his or her psychosocial well-being (McCormack, McAllister, McLeod, & Harrison, 2012; McCormack, McLeod, McAllister, & Harrison, 2009), this description will be limited to the communicative aspects of having a SSD. Verbal communication, however, involves not only the speaker, but also one or many listeners. Hence, communicative success will depend not only on the phonetic (and linguistic) accuracy of the spoken message, but also on factors related to the context in which it is delivered, and to features tied to the listener. One aspect of communicative functioning is **intelligibility**, i.e. the degree to which the intended linguistic message is recovered by the listener. This concept – and factors associated with it – will be explored deeper in section 4.1. However, apart from conveying linguistic information, speech also contains information signaling speaker-characteristic features, e.g. his or her age and emotional state. When speech production is impaired, this may distract the listener and result in reduced **acceptability** of the speech. In section 4.2, this potential functional consequence of having a SSD will be explored deeper.

4.1 Intelligibility

Intelligibility refers to how much of a speaker's intended linguistic message is recovered by the listener. A fundamental goal in verbal communication is making oneself understood, and as this ability is often compromised in children with SSDs, intelligibility is a central concept in the context of childhood SSDs (Flipsen, 2006). Consequently, intelligible speech is often identified as the long-term goal in the intervention of children with SSDs (Dodd & Bradford, 2000; Flipsen, 1995; Gordon-Brannan, 1994; Hodson & Paden, 1983). The value of intelligibility as an index of communicative competence for someone with a speech impairment has been stressed by many researchers (Klein & Flint, 2006; Miller, 2013; Schiavetti, 1992). However, intelligibility is not only dependent on the speaker and the specific errors that the speech signal may contain, but is also influenced by a number of other factors, e.g. the social context, the message content, the physical setting, and by characteristics of the listener (Kent, Miolo, & Bloedel, 1994; Shriberg & Kwiatkowski, 1982). Moreover, in the assessment of intelligibility, the outcome measure will vary depending on what type of speech material is analyzed, and on how the evaluation task is designed. Therefore, an intelligibility score will be difficult to decipher without information regarding these factors (Lagerberg, 2013). The following section will look closer at the assessment of intelligibility in children with SSDs, particularly with regards to the mentioned factors. After that, the reader will be introduced to how specific types of speech problems may interfere with intelligibility.

4.1.1 Assessing intelligibility

In research on speech impairments, assessments of intelligibility are typically based on orthographic transcriptions of audio-recordings of spontaneous speech, which are compared to a reference transcription of the intended words. An intelligibility measure is then estimated as the proportion of words in the evaluators' transcriptions that match the intended words in the reference transcription, i.e. the percentage of words understood (Gordon-Brannan & Hodson, 2000; Gordon-Brannan, 1994). In clinical settings, however, time constrains often preclude orthographic transcriptions, and therefore, estimates of intelligibility most often come in the form of verbal statements describing the overall intelligibility, often with regards to a scale ranging from, say, "not at all intelligible" to "completely intelligible" (Gordon-Brannan, 1994; Lagerberg, 2013). However, such estimates are known to be less reliable than assessments based on the ratio of syllables or words understood (Lagerberg, 2013; Whitehill, 2002). Hence, assessment of intelligibility involves a trade-off between reliability and validity constraints on the one hand, and practicable constraints on the other hand. Furthermore, whenever referring to a measure of intelligibility, it is important to include a description of a) the speech material upon which the measure is based, b) what kind of measure is being used, and c) characteristics of the listener(s) performing the evaluation. These factors are summarized in Table 4, and will each be described in the following sections.

Table 4. Factors influencing assessments of intelligibility.

Speaker	Speech material	Listener responses	Listener(s)
Articulation/Phonology	Conversational	Transcription	Familiarity w/ speaker
Prosody	Read	Forced-choice	Experience w/ speech disorder
Syntax	Re-telling	Rating scale	Familiarity w/ material
Semantics	Isolated sentences	Impressionistic statement	
Pragmatics	Isolated words		
	Imitated vs. read		

4.1.1.1 Speech material

Collecting appropriate speech samples for the assessment of intelligibility requires careful consideration. In order to reflect the child's everyday communicative context – i.e. if aiming at high ecological validity – selecting a sample of spontaneous conversational speech is recommended (Flipsen, 2006; Kent et al., 1994; Kwiatkowski & Shriberg, 1992). However, for highly unintelligible speech, identifying intended words in conversational speech can be difficult, and in such cases, selecting a different type of speech material is warranted (Gordon-Brannan & Hodson, 2000; Kent et al., 1994).[6]

If constraints on ecological validity are relaxed, using predetermined speech materials is a viable alternative to conversational samples, resulting in increased control over the data. Apart from providing the important information of what words are produced, predetermined materials also allow control over what phonological (and syntactic) structures are

[6] Note, however, that methods have been suggested that circumvent the problem of not knowing the intended words in the speech sample, by counting syllables in the unintelligible portions of the speech sample (Flipsen, 2006; Lagerberg, 2013).

collected. For the documentation of short-term changes, preselected word materials are suggested as being more sensitive than connected speech samples (Gordon-Brannan, 1994). Furthermore, preselected materials eliminate the opportunity for the speaker to avoid words or sounds (s)he finds difficult to produce (Lagerberg, 2013).

Different types of speech materials require different degrees of effort from the speaker, which may be reflected in reduced or increased speech production accuracy. For example, producing sentences is more challenging than producing single words, both in terms of linguistic complexity and speech-motor demands. This may explain why articulation is often less accurate in the production of sentences than in the production of isolated words (Lagerberg, 2013; Morrison & Shriberg, 1992; Namasivayam et al., 2013). The demands on the speaker are similarly reduced when imitating a spoken model, compared to when producing speech spontaneously; hence, differential phonological/articulatory patterns are often observed in imitated speech compared to spontaneous speech (Ingram, 1989; Leonard et al., 1978). In ICF terms, the selection of speech material underlying an assessment of intelligibility will reflect either features representative of the child's speech capacity (e.g. as in samples of imitated, single-word material), or of his or her speech performance (e.g. as in samples of spontaneous, conversational speech). (See description in section 2.1.)

By virtue of the contextual cues that may be available for the listener in conversational speech samples – e.g. as provided by an adult conversational partner, or by an understanding of the conversational setting – intelligibility is often rated higher in conversational speech samples compared to imitated word or sentence materials (Gordon-Brannan & Hodson, 2000). However, having access to more speech is only beneficial if the additional speech is reasonably intelligible – hence, the addition of more speech will not aid the listener if a) the speech was quite intelligible even without adding it, or if b) the added speech is unintelligible. Hence, the potential gains in intelligibility that additional verbal information may add are only relevant for mid-range intelligible speech (Miller, 2013).

From the above, it should be clear that different speech materials will have different effects on the evaluation of intelligibility – not only in terms of what can be expected from the speaker, but also of what can be expected from the listener. The choice of material will depend on the purpose of the assessment.

4.1.1.2 Listener responses

As mentioned above, intelligibility measures can be expressed in different ways; whereas some come in the form of a proportion of recognized words (or syllables, or phonemes), others are stated with reference to a rating scale. Sometimes, intelligibility is described by impressionistic estimates without reference to a scale, although this approach is considered both less valid and less reliable (Gordon-Brannan & Hodson, 2000; Kent et al., 1994). Ratio-level estimates, however, are the types of measures that are generally recommended, because of their superiority in terms of validity and reliability (Lagerberg, 2013; Schiavetti, 1992; Whitehill, 2002).

When assessment involves listeners' interpretation of what they hear, these interpretations can be reported in different forms. In some cases, the listeners are provided with a restricted set of alternatives to choose from when providing their responses (a forced-choice task); this is typically done in the context of assessing intelligibility of single-word utterances (Lagerberg, 2013; McHenry, 2011). More often, however, the listener reports what he or she hears in the form of an orthographic transcription (Gordon-Brannan & Hodson, 2000; Lagerberg, 2013). An even more detailed level of transcription – e.g. a phonological transcription – is warranted when exploring the impact of specific speech error patterns on intelligibility (Flipsen, 2010; Kent, Weismer, Kent, & Rosenbek, 1989; Preston, Ramsdell, Oller, Edwards, & Tobin, 2011). According to Flipsen (2010), transcription (or a "write-down" approach) represents the harshest test of intelligibility.

A quite different way of assessing intelligibility is described in Study IV. During the continuous playback of a speech sample, a panel of listeners are given the task of striking a keyboard key – or to "click" – whenever they hear something they do not understand. Hence, the larger the number of clicks, the less intelligible the speech sample. Although this method has still only been used in a small scale, in the evaluation of conversational speech samples produced by children with PD, Study IV reports on a strong correlation between the number of listener clicks for a given speech sample, and the corresponding estimates of the severity of involvement as measured in the Percentage Consonants Correct, PCC (Shriberg & Kwiatkowski, 1982). This is an indication that the number of clicks may serve as a valid index of intelligibility. The potential benefits of using this method for deeper exploration of specific speech events that evoke listener reactions will be described in more detail below (in section 6.2).

4.1.1.3 *Listeners*

The evaluation of intelligibility is influenced also by the listener's familiarity with the speaker. For example, caregivers are better at understanding their children compared to people outside of their family (Baudonck, Buekers, Gillebert, & Van Lierde, 2009; Coplan & Gleason, 1988; Flipsen, 1995). Moreover, listeners' general experience of listening to disordered speech may also contribute to their ability to understand non-typical speech. For example, some reports indicate that SLPs evaluate disordered speech as being more intelligible than inexperienced listeners do (Lundeborg & McAllister, 2007; McGarr, 1983). Listeners who are familiar with the spoken material – e.g. SLPs who are familiar with a preselected read passage – also tend to rate intelligibility higher than those who are not familiar with the spoken material (Beukelman & Yorkston, 1980). Hence, when using predetermined speech materials, it is important to acknowledge the expected effects of the listeners' familiarity with the speech material, the speaker and/or the characteristic traits of the speech disorder.

There are several reasons why familiarity with the speaker aids the listener in understanding the spoken message. Knowing the person who speaks may give the listener an advantage – not only because they may share the same contextual knowledge (Flipsen, 1995), but also because experience with a speaker's way of speaking may have trained the listener in "translating" the speaker's atypical speech patterns (Flipsen, 1995; Tjaden & Liss, 1995).

Apart from the listener characteristics described above, which may contribute to a bias in their evaluation of intelligibility, large variability across everyday listeners is sometimes observed which cannot be attributed to the above-mentioned factors. This has been reported by, for example, McHenry (2011), and it is also observed in Study IV. In these studies, variation could neither be explained by demographic factors like years of education or age, nor by the listeners' varying experience with speech analysis or with interaction with pre-school-aged children. Although McHenry (2011) suggests some other factors that may contribute to listener variation in the evaluation of intelligibility – e.g. musical experience or exposure to other languages or dialects – these factors are not easily quantified, and hence, much of the observed variation across listeners will presumably remain unexplained. Including more listeners – especially when including everyday listeners – is a way of compensating for this variation, to strengthen the reliability of the evaluation.

From the above, it should be clear that the assessment of intelligibility can be approached from different perspectives, and that the decision on which approach to choose depends on the purpose of the assessment. For an individual child, a combination of assessments – based on different materials, evaluated by different listeners, etc. – may provide the most enlightening description of intelligibility (Kent et al., 1994; Miller, 2013). It should also be clear that intelligibility is influenced by multiple factors other than the speech signal itself. Nevertheless, by limiting the influence of these confounding factors, deeper analysis of the speech signal is possible, allowing for the exploration of what specific speech features contribute the most to decreased intelligibility. This will be the focus of the following section.

4.1.2 Intelligibility and speech sound errors

There is an apparent value in knowing how specific speech error patterns contribute to decreased intelligibility, as intervention targeting those error patterns that are most detrimental to intelligibility will potentially be most rewarding in terms of functional gains. Although there is a considerable overlap between the degree of intelligibility and the severity of the disorder (Gordon-Brannan, 1994), these factors do not necessarily go hand in hand. Hence, a child with a more severe SSD is not necessarily more unintelligible than a child with a less severe SSD (Kent et al., 1994). Therefore, merely knowing what types of errors a child produces, and how frequently they occur, will not necessarily translate into a level of (un)intelligibility. And, conversely, a low intelligibility score will not be revealing of *why* the speech is unintelligible. The correlation between these aspects, however, may serve to identify those specific speech error patterns that are most detrimental for intelligibility, thereby indicating these as potential targets in therapy (Bowen, 2009; Hodson & Paden, 1983).

Some intuitive suggestions have been made of phenomena that make speech less intelligible. For example, the neutralization of phonological contrasts, or homonymy, is a potential source of diminished intelligibility (Ingram, 1989). This is apparent in cases where, for instance, a child's production of the word *key* is indistinguishable from his or her production of the word *tea*. Moreover, if the misarticulated sounds occur frequently in the child's language, the effects of the misarticulation on intelligibility are presumably more pervasive than if the misarticulated sounds are less frequent (Brown, 1988). Furthermore, it has also been hypothesized that inconsistent speech errors are more damaging to intelligibility than consistent speech errors, as consistency allows listeners to adapt to the speaker's behavior (Yavas & Lamprecht, 1988). However, suggestions of phonetic/phonological correlates to (un)intelligibility are often based on intuition, and not necessarily on empirical evidence (Klein & Flint, 2006). In the following, a few exceptions will be described.

The exploration of how specific speech error patterns influence intelligibility may be approached from different angles. For example, the point of departure might be the categorization of children into groups by levels of intelligibility. By exploring what speech error patterns the children display, and particularly, if some error types are exhibited more often by the more unintelligible children, conclusions may be drawn on what error types are damaging intelligibility. Hence, the observation that atypical speech error patterns (i.e. error patterns that normally do not occur in typical speech acquisition) occur more often among unintelligible children suggests that atypical speech errors are more deleterious than developmental speech errors (Hodson & Paden, 1981; Yavas & Lamprecht, 1988). In the same vein, omissions of speech sounds are found more damaging to intelligibility than phonetic distortions (Hodson & Paden, 1983), as are syntagmatic (or sequential) errors in comparison to paradigmatic (or substitution) errors (Yavas & Lamprecht, 1988).

An alternative approach to exploring the relative contribution of specific error types is to manipulate the speech signal and study what effects it may have on intelligibility (Klein & Flint, 2006; Maassen & Povel, 1985). With the aim of exploring the relative impact of specific speech error patterns on intelligibility, Klein and Flint generated disordered speech artificially, by having an adult male read aloud pre-processed phonemic transcripts, and letting listeners gloss what they could understand. The three error types under investigation – final consonant deletion, stopping and velar fronting – were each applied to a set of transcribed sentences, acting like different "phonological filters". The results indicated a rank order in how the different error types affected intelligibility, with final consonant deletion being more damaging than stopping, which in turn was more damaging than velar fronting – at least when frequency of occurrence was held relatively low (at 15% or 30%). Interestingly, however, when the level of occurrence of the different error types was higher (at 50%), all three error types were equally disturbing to intelligibility. Although one might question the ecological validity in having an adult male reading aloud manipulated phonological transcripts, Klein and Flint's study constitutes an interesting – and rare – example of how controlled manipulations of the speech signal may be used to explore effects on intelligibility.

Another way of exploring how different speech errors may damage intelligibility has been suggested in Study IV. Here, the registered listener reactions – "clicks" – during the continuous playback of conversational speech serve as pointers to loci of unintelligibility. Provided that speech errors are annotated in the conversational samples, the listener reactions may then be traced back to the events evoking them. Although Study IV only contains a small-scale qualitative analysis, with some limitations regarding the instructions to the listeners, the results are promising regarding the expedience of the method in allowing larger scale quantitative analysis. This suggested approach may respond to the current lack of empirical evidence of the relative effects on intelligibility of different speech sound error patterns (Klein & Flint, 2006).

4.2 Acceptability

Having a SSD might not only affect a child's chances of being understood, it may also affect how he or she is perceived by the listener. Producing speech that calls adverse attention to the speaker may restrict his or her willingness to engage in conversation. For children who are still developing their communicative competence and social skills – notably by interacting with people in their environment – such a restriction is clearly unfortunate. The attitudes evoked in listeners towards the child's speech are linked to the concept of acceptability (Whitehill, 2002). Although the problem of diminished acceptability may be almost as significant to a child as his or her problems of not being understood, the effects of having a speech disorder on acceptability have been considerably less studied than the effects on intelligibility. Moreover, although intelligibility is not trivially assessed, aspects of acceptability are possibly even more difficult to quantify. Furthermore, the ethical aspects of evaluating peer attitudes towards a child as a person require careful consideration. Nevertheless, considering that reduced acceptability may have serious consequences on the child's communicative confidence, peer relations and academic achievements, acceptability is no doubt an important aspect to study. The following paragraphs will review our current state of knowledge on acceptability in the context of childhood SSDs.

One way of assessing listener attitudes towards a speaker is by means of semantic differential forms. Here, speech samples are played to listeners, who are asked to rank the speaker with regards to bipolar scales where the endpoints are positively and negatively loaded, e.g. attractive-unattractive, intelligent-unintelligent, kind-cruel. Using this design to assess young adults' perception of the speech of another young adult with a lateral lisp, Silverman (1976) reported that even this seemingly mild speech impairment may call adverse attention to the speaker. In a similar study, an adolescent speaker substituting [w] for /r/ was also found to be perceived negatively by peer listeners (Silverman & Paulus, 1989). The semantic differential has been used successfully as an instrument of assessing listener attitudes with listeners as young as 8-9 years old (Crowe Hall, 1991).

For younger children, studies of peer attitudes towards speakers with SSDs have been rare. However, Gertner, Rice, & Hadley (1994) assessed peer attitudes in preschool-aged children by means of positive and negative nominations. Here, children were presented with pictures of their peers and asked to point to who they, for example, like to play with (in case of a positive nomination), or do not like to play with (in case of a negative nomination). Using this design, Gertner and colleagues found that 3-6 year-old children preferred peers with normally developing language skills to peers with speech and/or language impairments, and to peers learning English as a second language. Hence, language competence appears to have an impact on peer preferences in preschool-aged children. Notably, however, in these ages, expressive language skills (e.g. speech production) do not contribute to negative peer nominations to the same extent as diminished receptive language skills (Gertner et al., 1994). This may seem to contradict the findings that relatively mild articulation disorders are perceived negatively in peer listeners, in school-aged children (Crowe Hall, 1991; Silverman, 1976; Silverman & Paulus, 1989). However, the age of the listener plays an important role, as older children tend to be more critical to disordered speech than younger children (Crowe Hall, 1991). This, in turn, may reflect the fact that many children still have not achieved adult-like speech production in their preschool years, whereas by the age of 8 or 9, or older, speech production deficits are more unexpected (Gertner et al., 1994).

From the above, it appears that peer attitudes constitute less of a problem for preschool-aged children with SSDs compared to older children with SSDs. On the other hand, children with SSDs do not only interact with peers. Misarticulations have also been found to impact adults', e.g. teachers', perception of a child as a speaker. For example, adult listeners have been found to associate reduced phonological accuracy with reduced cognitive maturity in preschool-aged children (Burroughs & Tomblin, 1990). Moreover, adult listeners exhibit a negative social bias toward children with speech impairments – especially when these are also coupled with reduced language skills (Rice, Hadley, & Alexander, 1993). For older children, second grade teachers have been found to associate reduced speech intelligibility with diminished academic, social and behavioral competence (Overby, Carrell, & Bernthal, 2007). Considering the risk of these expectations becoming self-fulfilling prophecies, these findings highlight the importance of preventing a potential downward spiral – possibly both at the level of the SSD *per se*, but also in terms of promoting acceptance and awareness of speech impairments to people in the child's environment (McLeod & Bleile, 2004; Overby et al., 2007).

Little is yet known about what particular speech sound errors evoke listeners' negative attitude toward the speaker, and if some deficits are more negatively loaded than others. No doubt, many aspects of speech and language affect listeners' perception of the speaker as a person, e.g. fluency, rate of speech, vocal quality, syntax, and turn length (see overview in (Burroughs & Tomblin, 1990). Furthermore, we do know that there is a tendency in adult listeners to integrate expectations on developmental speech production norms in their evaluations of child speakers, such that their evaluations are not negatively affected by speech errors as long as these are developmentally appropriate (Rice et al., 1993). Other than that, however, little is yet known of what types of speech errors have most effect on acceptability (Crowe Hall, 1991). Notably, the issue of what features in speech production contribute to reduced acceptability may also be a concern to other populations than children with SSDs, e.g. to second language learners (Moyer, 2013). The continuous "clicking" task described in Study IV, may constitute a viable approach to the exploration of how different speech errors, or idiosyncrasies, contribute to reduced acceptability, and to tease apart how different aspects of speech and language influence listeners' perception of a speaker as a person.

4.3 Concluding remarks

This chapter has focused on potential communicative consequences of having a SSD, particularly concerning reductions in intelligibility and acceptability. Existing approaches to the assessment of these aspects have been presented, as well as a description of some limitations and difficulties involved in assessing intelligibility and acceptability of speech. An alternative approach to listener evaluations has been introduced, which involves the registration of listener reactions during the continuous playback of speech material (in Study IV). Both regarding intelligibility and acceptability, the potential value of the suggested approach has been described, and how this may address some of the shortcomings in the currently used assessment approaches.

These lines conclude the overview of the context in which the studies included in this thesis were performed. With this background, the reader should now be well-prepared for the next chapter, which will present a summary of the findings gained through Studies I-IV, and how these findings relate to the research questions stated in section 1.2.

5. Summary of findings

The main focus so far has been on setting the scene on which Studies I-IV have been carried out. Now, however, the spotlight will be directed on these specific studies, and on the findings they have provided. In the following sections, each study will be presented briefly; for more detailed descriptions, the reader is referred to the original papers, attached in Part II of this thesis. Finally, this chapter will revisit the research questions (stated in section 1.2), and address each of these with reference to the findings provided by the included studies. This serves to prepare the reader for the general discussion of the implications of this work, and for the overall conclusions, which will be drawn in Chapter 6.

5.1 Study I

Study I explores children's perception of their own speech, and particularly, their ability to recognize themselves as the speaker in recordings of utterances that they themselves have produced. The aim of Study I was to investigate whether this ability varies depending on a) the age of the child, b) whether the child has a phonological disorder (PD) or not, and c) the time between making the recording and listening to it. Additionally, the study explored whether children with PD are better at distinguishing their own recordings from those of peers on misarticulated items, i.e. if they use their misarticulation as a cue to distinguish their own recorded utterances from phonetically accurate productions recorded from children with typical speech and language development (TD). These questions were explored within three groups of children: two groups of children with TD (4-5 years old vs. 7-8 years old), and one group of children with PD (4-7 years old), who all displayed patterns of velar fronting. A recording script of 24 one-word items was used, where half of the words began with /tV/, and half began with /kV/. Hence, the children with PD were expected to produce half of the words in the script correctly, and the other half incorrectly. Each recorded item was presented to the children in a voice line-up setting, where three recordings of children with TD were used as foils. The task for the children was to identify which one of the four recordings that they themselves had produced. This task was performed on two occasions: the first immediately after having produced (and recorded) the utterance, and the second after a period of 1-2 weeks.

The results revealed overall high performance rates, suggesting that children in this age range are generally able to identify themselves as the speaker in recordings of utterances that they themselves have produced. Furthermore, the results showed that a) there is no major development in this ability between the ages 4-5 and 7-8 years, and b) this ability is not affected by the child having a PD or not. However, this ability *is* dependent on the time passing between producing an utterance and listening to it, and the impact of this factor varies with the age of the child; older children are more affected by the presentation delay than younger children. Moreover, children with PD are not better at distinguishing their own *misarticulated* recordings from correctly produced recordings produced by peers, than they are at distinguishing their own *correct* recordings from those of peers. This suggests that the children with PD do not use their misarticulation as a cue to distinguishing their own recordings from those of peers with typical speech.

5.2 Study II

Children's perception of their own speech production is further examined in Study II. Here, children with PD were again compared to children with TD, with regards to their perceptual responses when evaluating their own speech production – in its online form, in its recorded form, and in synthetically modified forms. A novel method for performing automatic modification of the children's productions was developed and introduced, enabling two types of synthetic modification: one where the phonological identity of the first plosive is changed (e.g. changing /tɔp/ into /kɔp/), and one where the identity is not changed (e.g. changing /tɔp/ into a new version of /tɔp/, with a [t] retrieved from another recording). This technique was used to explore whether the children's perceptual evaluation of accuracy corresponded to the intended modification target, and whether these perceptual evaluations were in alignment with those of other listeners. Furthermore, the children's ability to notice synthetic modification was explored. Again, the children's evaluation accuracy was explored with regards to the time passed since their production of an utterance. Experimental tasks were designed to assess the children's perception of phonological accuracy, in different types of stimuli. The stimuli were either a) recordings (original and modified) of utterances produced by other children, b) recordings (original and modified) of utterances produced by the children themselves – the first time presented right after having produced the utterance, and the second time after a delay, or c) utterances they had just produced (i.e. live non-recorded speech), and, presented after a delay, recorded versions of these utterances. A separate task was used to explore the children's ability to detect synthetic modification.

The results showed that children with PD perform on the same level as TD peers when evaluating stimuli produced by other children, whereas in their evaluation of their own speech production, children with PD are less accurate than their TD peers. The following pattern was uncovered in the children with PD: a tendency to evaluate misarticulated utterances (produced by themselves) as correct when just having produced them, and to perceive inaccuracies in the same utterances better when time has passed since their production of them. Regarding the children's ability to detect the synthetic modification, the results showed that for all children – regardless of whether they have a PD or not – synthetic modification that does not result in a phonological change is difficult to detect. This suggests that within-category modifications are too subtle for the children to detect, and serves as an indication of a satisfactory quality of the generated speech.

5.3 Study III

In contrast to the first two studies, where perceptual responses are registered with reference to categorical measures, Study III introduces an alternative instrument: the visual analog scale (VAS). Here, perceptual evaluations of children's successful and unsuccessful efforts at producing /t/ and /k/ were registered with reference to a gradient scale, ranging from "clear [t]" to "clear [k]". By using this instrument, the study explored experienced listeners' perceptual sensitivity to covert contrast in children's "clear substitutions" of one speech sound for another. Moreover, the study addressed the question of whether listeners are sensitive to acoustic-phonetic differences between sounds produced by children with PD compared to those produced by TD peers, even among productions classified as being target-appropriate (i.e. "correct"). In addition, two approaches to the acoustic analysis of the speech material were compared (spectral moments vs. discrete cosine transformation, DCT, analysis), in order to explore which one corresponded most to the perceptual evaluations performed by human listeners. The investigation was based on a speech material consisting of 3627 recorded one-word utterances – all beginning with either /tV/ or /kV/ – collected from 122 children (in the ages 4-8 years old), of whom 40 had been diagnosed as having a PD. A single annotator evaluated the total set of recordings, and a panel of 10 listeners was recruited to explore inter-annotator agreement. Using the two different acoustic measures, acoustic models were derived based on error-free tokens of /t/ and /k/, thus modelling acoustic prototypicality of these consonants. Based on these models, two proximity indices were calculated for each recorded item, to represent the acoustic distance between this item and prototypical /t/ and /k/: one index expressed with regards to spectral moments analysis, and the other expressed with regards to the DCT analysis.

Regarding the question of listeners' perceptual sensitivity to covert contrasts, the results showed that substitutions of [t] for /k/ were indeed rated as being less prototypical (or /t/-like) than correct productions of [t] for /t/. This pattern was also observed for the distinction between substitutions of [k] for /t/ and correct productions of [k] for /k/. Hence, perceptual differentiation of covert contrast was evidenced for substitutions in both directions. The exploration of target-appropriate (correct) productions of /t/ and /k/ within the two groups of children revealed that the plosives produced by the children with PD were evaluated as being less prototypical in comparison to those produced by TD peers. Although this may not be surprising in the case of /k/, which represents a sound not yet (fully) acquired by the children with PD, these results are more unexpected in the case of /t/, which was not a documented problem for the children with PD. Regarding the question of what acoustic measures correspond most to the perceptual evaluations, the findings reveal a slight advantage for the DCT approach. However, neither of the acoustic measures was very successful in mimicking the responses of the human evaluators, suggesting that human evaluators base their perceptual evaluations of category goodness on acoustic features that are not available to the analytic instruments, or even, have not yet been identified. This motivates continued efforts at optimizing current acoustic measures.

5.4 Study IV

In Study IV, focus is shifted from the misproductions of specific speech sounds to the communicative context in which they appear. Here, a novel measure of intelligibility/acceptability was introduced, involving the continuous registration of listener reactions during playback of recordings of children's conversational speech. In representing the first effort at introducing this Audience Response System-based (ARS) technique to the domain of misarticulated speech, one main purpose of the study was to explore the validity of this new measure, by comparing it to a standard measure of severity: the Percentage of Consonants Correct (PCC, Shriberg & Kwiatkowski, 1982). A second question addressed the issue of whether variation in the listeners' response behavior could be attributed to their experience of speech analysis and/or to their experience of communicating with pre-school-aged children. In addition, the study aimed to investigate whether the listeners' reactions could be linked to triggering events in the speech signal. The exploration of these issues was based on 16 conversational speech samples collected from 7 Swedish children, of whom 5 had been diagnosed as having a PD. Two of the 16 recordings were subject to qualitative analysis, where misarticulations were manually annotated. A panel of 20 adult Swedish listeners was recruited, varying with regards to their experience of speech analysis, and to their experience of interacting with young children. The listeners were instructed to listen to the speech samples, and to strike a keyboard key (to "click") whenever they perceived something unintelligible or deviant during playback of the speech stimuli. The distribution of clicks and the distribution of manually annotated speech events were analyzed by means of Kernel Density Estimations (KDEs).

Regarding the first research question, the results revealed a strong correlation between listener clicks and the clinical measure of severity, such that recordings with a large proportion of segmental speech errors evoke a large number of listener clicks. This serves to validate the new measure against a standard clinical measure. Considering the variation in listeners' clicking behavior, this variation could neither be attributed to the listeners' varying experience with speech analysis, nor to their varying experience with interacting with children. Hence, the vast variation across listeners in their clicking frequency remains unexplained. Finally, the preliminary inspection of the relation between annotated events in the speech signal and the distribution of listener clicks revealed quite structured patterns, indicating that larger scale quantitative analysis is a viable way of exploring what types of features in misarticulated speech are the most salient correlates to unintelligibility and/or decreased acceptability.

5.5 Research questions revisited

In the introductory chapter, five general research questions were stated. These questions will be addressed below, with reference to the findings provided by the work presented in this thesis.

1. **Do children recognize their recorded speech production as their own?**

Yes. The findings in Study I confirm that children in the ages 4-8 years old are good at distinguishing recordings of their own speech from recordings of peers, regardless of whether they have a phonological disorder or not. If interpreted from the perspective that children's perception of their own speech is different from their perception of speech produced by other people (i.e. assuming a distinction between internal and external speech perception, referred to in Study I, p. 33), this serves as evidence that listening to recordings of one's own speech indeed involves internal rather than external speech perception.

2. **What are the children's reactions when presented with synthetically a) modified and b) corrected versions of utterances they have produced?**

First, regarding children's reactions to hearing *modified* versions of their own utterances, the findings in Study II shows that children – regardless of whether they have a phonological disorder (PD) or not – generally accept the intended phonological form of modified versions of their own utterances, seemingly without being disturbed by the modification itself. Hence, modifications that do not involve a phonological change (i.e. within-category modifications) are often too subtle for children to detect. Second, regarding children's reactions to hearing *corrected* versions of their utterances, children with PD were found to evaluate most corrected utterances as being correct. (Corrective re-synthesis is, of course, only relevant to the children who misarticulate, and hence, not to children with TD). However, the children's evaluations are dependent on the time passed since their production of the original utterance. In close conjunction with their production of the original (misarticulated) utterance, the children tend to judge the corrected version of this utterance as being correct. When time has passed since their production of the original utterance, however, the children tend to be more critical, and evaluate more corrected items as being incorrect. Hence, children's evaluation of whether an utterance is correct or not changes with the time passed since their production of the utterance.

3. Can listeners perceptually distinguish between clear correct productions of a speech sound and cases where this speech sound is a substitution for another sound?

Yes. As shown in Study III, there was indeed a perceptually measurable difference between [t]s as correct productions of /t/ and [t]s as misarticulated productions of /k/ (i.e. "substitutions"). These misarticulated [t]s were rated as being less /t/-like than the correct [t]s. The same pattern was observed for substitutions in the opposite direction, i.e. between correct [k]s and [k]s when produced as substitutions for /t/.

4. Can listeners' reactions during the continuous playback of misarticulated speech be linked to triggering events in the speech signal?

Yes. Study IV demonstrates a successful application of an Audience Response Systems-based approach to identifying points in time where listener reactions coincide, and illustrates how these points may be linked to potential loci of interest in the speech signal.

5. Does the introduction of new technological innovations into the study of children's speech sound disorders bring benefits that could not otherwise have been achieved?

Yes. The work presented here has demonstrated some areas where new technology can be used to gain extended insights into the perception of children's misarticulated speech. First, consider the corrective re-synthesis technique described in Study II. Although the literature contains one previous case where manual correction of misarticulated speech has been used in exploring children's perception of their own misarticulated speech (Shuster, 1998), the re-synthesis technique developed and introduced in Study II allows correction (and distortion) of children's recorded utterances to be done automatically. Not only does this possibility save working hours, it also enables the study of children's perception of corrected (and distorted) versions of their own utterances in close conjunction with their production of the utterance. Apart from the theoretical insights this possibility brings (described above, and in the following chapter), a number of potential benefits of introducing this technique into clinical settings can also be identified. For example, by presenting a child with recordings of his or her original (misarticulated) utterance contrasted with a corrected version of the same utterance, the child's attention could be brought to an acoustic-phonetic distinction which is phonologically critical to other listeners, but is not yet fully appreciated by the child.

Another technological innovation introduced in this work is the application of ARS into the field of speech sound disorders (described in Study IV). The possibility of registering listener reactions during continuous playback of spoken material opens up a new methodological paradigm in the exploration of connections between specific speech errors and what reactions they evoke in listeners. Rather than exploring indirect links between speech error patterns and their effects on overall intelligibility and/or acceptability, the suggested approach allows exploration of the direct links between individual misarticulations and listener responses. Although the work described here does not demonstrate the full potential of this new methodological approach, it serves to set the direction for future larger-scale investigations.

6. General conclusions

The work described in this thesis covers various aspects of childhood SSDs. On the one hand, focus has been on speech sounds – how they are perceived by the child producing them as well as by other listeners, and how these perceptual responses correlate to the acoustic information available in the speech signal. On the other hand, attention has also been directed toward the functional consequences of producing misarticulated speech, and how specific speech sound errors contribute to listeners' reactions in terms of intelligibility and acceptability. In relating to both of these different perspectives, the work presented in this thesis fits well within a description based on the ICF framework, although – admittedly – this thesis is consistent with the long-held tradition of paying more attention to features related to Body Structure and Functioning than to aspects of Activity and Participation.

A common feature of the studies contained in this thesis is the application of existing – albeit not necessarily widespread – methods and procedures to new domains. For example, standard methods of unit selection speech synthesis were used in the development of a novel approach to automatic correction of children's misarticulated speech production, involving the "phonetic transplantation" of speech segments recorded from different child speakers (Strömbergsson, 2011). In Study II, this procedure is used to gain new insights into children's perception of their own speech production. And in Study IV, an audience response systems approach (ARS), which is frequently used in the TV and film industry to evaluate new productions, was used in the hitherto unexplored context of registering listeners' reactions while listening to conversational speech samples produced by children with SSDs. Hence, one of the main contributions of this thesis is the innovative usage of existing technology and methods in the study of perception of (misarticulated) speech. More important, however, are the theoretical insights gained when applying these methods; these will be discussed in the following sections, in terms of methodological, psycholinguistic and clinical implications. First, however, attention will be brought to some limitations of this work.

6.1 Limitations

Some limitations should be considered when interpreting the findings reported in this thesis. The first concern relates to the relatively small number of children with PD participating in Study II. This group consisted of 11 children, whose behavioral responses were compared to those of a group of 20 TD peers. Considering that variation among children is often larger than that among adults – and even more so in clinical populations – this concern is warranted. However, this sample size is comparable to other similar studies (e.g. Panagos & King, 1975; Locke & Kutz, 1975; Rvachew & Jamieson, 1989), and reflects the difficulties involved in recruiting children who fit the specified inclusion criteria, who are willing to participate, and who manage to complete all tasks included in the experiment. For the initial study of children’s perception of corrected versions of their misarticulated speech, this may suffice; however, firmer evidence requires a larger sample size.

A second limitation is the lack of physical and physiological measures (i.e. detailed acoustic and/or articulatory measures) in Study I and Study II. In order to gain deeper understanding of children’s perception of the speech they produce, e.g. what acoustic cues are involved in their phonological decisions, such detailed measures should be included. However, this limitation is the flipside of the choice to perform recordings in children’s everyday pre-school environment, to maintain an ecologically valid setting. The advantage, on the other hand, is that the reported results reflect the children’s behavior in a representative clinical or educational setting.

Another potential concern relates to the description of the participating adults and children in all four studies; these descriptions do not contain measures of the participants’ hearing. Considering that adequate hearing is a prerequisite for adequate perception, some caution regarding this lack of information is justified. However, individuals – children as well as adults – with known hearing deficits were excluded already before participation. Moreover, as all Swedish 4-year-olds undergo a hearing screening, and as all children participating in the included studies are 4 years or older, it is assumed that significant hearing problems would have been detected before participation.

Finally, the experimental design in Study I is one way of exploring children's ability to identify their own recordings among recordings of other children. However, it is not the only way. Considering the children's high performance, one might suspect ceiling effects. An alternative way of exploring the same question is by means of a classification task, i.e. to play one recorded utterance at a time, letting the children decide whether they themselves are the speaker, or whether it is someone else. This task would be expected to involve reduced memory requirements, and a reduced number of confounding factors, without precluding the possibility of exploring effects of presentation delay and/or the children's sensitivity to production accuracy. Hence, in future studies of children's perception of their own speech production, this experimental design should be considered.

6.2 Methodological implications

The exploration of speech perception may start off from different points of departure. In Study II, a distinction is suggested between **bottom-up** approaches on the one hand, and **top-down** approaches on the other hand. In bottom-up approaches, acoustic features known to be phonologically critical in adult speakers are manipulated to explore perceptual effects in a new group of listeners. This approach has been very rewarding, e.g. in identifying different cue-weighting strategies in different groups of listeners (see section 3.3.2). However, this particular finding – that different listeners use different perceptual strategies – warrants some caution when exploring speech perception in children with SSDs with reference to acoustic features known to be phonologically critical in proficient language users (see section 3.3.2, p. 45). A top-down approach, on the other hand, like the corrective re-synthesis described in Study II, may constitute an alternative route to the exploration of individual perceptual strategies, and of the interplay between production and perception of speech. By starting off from the speaker's own (misarticulated) speech production, and by generating synthetically corrected versions of these utterances, this approach does not involve any *a priori* assumptions regarding which acoustic cues the speaker may engage in his or her production and perception of phonological contrasts. Hence, the "acoustic distance" between an original misarticulated production and the corrected version of this recording will vary – both in degree and nature – from case to case. (And this, of course, is where this approach differs from bottom-up approaches, in which the acoustic distance is carefully controlled.) Closer analysis of the acoustic distance between the original misarticulated utterance and the corrected version of the same utterance may then reveal the acoustic features that are involved in this specific speaker's speech production and perception – without making references to what acoustic cues other speakers rely on. This step of acoustic analysis, however, is not covered in the present thesis, but will be a subject for future investigation. Together with the findings in Strömbergsson (2011), the results reported in Study II point to a high quality of the automatically modified speech; children as well as adults accurately perceive the intended phonological form of the generated speech, and modifications involving within-category substitutions are generally too subtle for the listeners to detect. However, the most important strength of the suggested re-synthesis technique is the possibility of performing correction/modification in real-time. Considering the finding that children's evaluation of their own speech production is dependent on the time interval between producing an utterance and evaluating its accuracy, real-time modification of utterances may constitute a welcome addition to the speech researcher's existing experimental toolbox.

In Study III, an important observation is that for children exhibiting velar fronting, even their productions classified as correct [k] productions were perceptually evaluated as being less /k/-like than those produced by their TD peers. Even more surprising, however, is the finding that the same pattern was observed for the children's [t] productions. Hence, although production of /t/ was not a documented problem for the children with PD, their correct [t] productions were evaluated as being less prototypical than those produced by their TD peers. Not only does this finding serve to identify previously undocumented weaknesses in these children's speech production, it also illuminates the potential benefits of using fine-grained perceptual measures, as these patterns would have been obscured if only categorical, or otherwise coarse-grained, measures had been used. Hence, when designing perception experiments, supplementing categorical responses with the registering of gradient responses should be considered as a way of bridging the gap between continuous acoustic-phonetic space, and the categorical phonological space.

The methodological benefits of using audience response systems (ARS) to register listeners' reactions while listening to misarticulated speech are highlighted in Study IV. Apart from the fact that the collected number of "clicks" per speech sample may serve as a valid measure of intelligibility and/or acceptability, the listener "clicks" may also serve as pointers to loci of interests, e.g. those specific speech errors that are most detrimental to intelligibility and/or acceptability. Moreover, the number of coinciding "clicks" will also serve to indicate the prominence of the evoking speech event – a larger number of coinciding listener reactions can be assumed to reflect a greater prominence. Hence, the distribution of "clicks" will not only identify loci of interest, it will also indicate which loci are more prominent than others. This opens up for new ways of exploring how different speech patterns affect intelligibility and/or acceptability, which is a relevant area of investigation also within other populations, e.g. in adults with speech disorders, or in learners of a second language. However, the large variation observed among the listeners in their sensitivity to speech errors calls attention to the necessity of including a large number of listeners, in order to compensate for these individual differences. Although the results reported in Study IV are preliminary and do not empirically demonstrate large-scale quantitative analysis of the collected data, these initial steps are promising.

6.3 Psycholinguistic implications

One recurring theme in this thesis is children's perception of their own recorded speech. Taken together, Study I and Study II provide evidence that 4-8-year-old children are generally able to distinguish recordings of their own productions of isolated words from recordings of the same words produced by other children, and that having a PD does not affect this ability. Moreover, for the children with PD, their evaluation of recorded versions of their own speech production are more accurate than their evaluations of utterances at the time of producing them (i.e. their self-monitoring). As suggested in Study II, a three-way distinction can be observed, namely between

1. internal perception (i.e. perception of speech at the time of producing it, self-monitoring),

2. pseudo-internal perception (i.e. the perception of one's own recorded speech), and

3. external perception (i.e. the perception of someone else's speech, recorded or not).

This ordering reflects an increasing level of evaluation accuracy, where internal perception is most challenging for the children with PD, pseudo-internal perception is less challenging, and where external perception often is unproblematic. Hence, in psycholinguistic tasks where children evaluate recordings of their own speech, they can be expected to perform differently – less objectively – compared to when they evaluate recordings of speech produced by others. However, as discussed in Study II, the work here leaves questions regarding the acoustic-phonetic criteria underlying the children's decisions unanswered.

A possible interpretation of the findings in Study I is that typically developing 7-8-year-old children rely on echoic memory of the utterances they have just produced when identifying their own recordings, whereas younger children (4-5-year-olds) do not. Effects of just having produced an utterance are also suggested as playing a role in the interpretations of the findings in Study II, where children with PD are more accurate in their evaluations of their own utterances when time has passed since their production of the utterances; here, recency appears to inhibit objective evaluation. Together, these findings highlight the importance of considering the time between children's production of an utterance and their perceptual responses to hearing this – or other – speech stimuli.

An important theme in this thesis is the concern that traditional descriptions of misarticulated speech production are inadequate as instruments in investigations of processes underlying deficient speech production. For example, phonetic transcription may suggest that a child is substituting one speech sound for another, whereas acoustic or articulatory measures – or more sensitive perceptual measures, as in Study III – may reveal the existence of a covert contrast, i.e. that the child is indeed signaling the phonemic contrast, but by other means than the listener is listening for. Hence, a description based solely on phonetic transcription stands the risk of failing to convey acoustic-phonetic variation that may actually be important in describing the processes underlying the child's misarticulations. No doubt, there is value in knowing that a child is perceived as neutralizing a phonemic contrast (i.e. that he or she is perceived as substituting one speech sound for another); this suggests that the child has problems with homonymy in speech production, which may have detrimental effects on intelligibility (Ingram, 1989). Moreover, as illustrated in Studies I and II, in comparing children's perceptual responses with those of other listeners, categorical descriptions of misarticulated utterances (e.g. whether they are "correct" and "incorrect") may provide important insights to whether the child's perceptual evaluations align with those of other listeners. However, phonetic transcriptions of the child's speech production do not provide sufficient information for deciding whether the child's problems are of an "articulatory" or "phonological" nature (recall the discussion on p. 27). This raises concerns that the sharp division between Phonological Disorders and Articulatory Disorders traditionally adapted in Swedish clinical practice (see description on p. 29) may need to be reconsidered.

6.4 Clinical implications

The finding that children with PD recognize themselves as the speaker in recordings of utterances they have produced, and that their evaluation of accuracy of these recordings does not conform to those of typical listeners in the surrounding language environment, indicate this as a potential therapeutic focus. As suggested both in Study I and in Study II, tasks involving the perceptual evaluation of recorded utterances may serve as an intermediate step towards the more challenging task of evaluating one's own production in the act of speaking. In light of the suggested inhibitory effect of just having produced an utterance on the children's objective perceptual evaluation, the time window between production and perceptual evaluation may be narrowed as therapy progresses. This way, self-monitoring may be approached gradually. Regardless of whether the child's problems are due to inefficient self-monitoring, or to inadequately specified internal referents (or to a combination of the two), using the child's own speech as the object of focus is theoretically appealing, as it can be expected to stimulate both the child's inclination to evaluate his or her own speech production, and/or to do so more objectively. Note, however, that a perceptual focus in clinical therapy is only warranted in children where perception is indeed impaired (Jamieson & Rvachew, 1992; Wolfe, Presley, & Mesaris, 2003). Therefore, intervention should never be designed without a preceding assessment of the child's perceptual skills, preferably including not only clear exemplars of the sounds produced in error, but also less prototypical items.

A possible extension of using recordings of the child's utterances as the object of focus in therapy is to introduce corrective re-synthesis. This would enable the generation of synthetic versions of what it would sound like if the child were able to produce the target sound correctly. Multiple potential benefits of contrasting original recordings of the child's incorrect production of the target sound with such individually tailored speech targets are listed in Study II (p. 5). However, the most important benefit is probably the possibility of attracting the child's attention to an acoustic-phonetic contrast that is phonologically critical to typical listeners, while being perceived by the child as within-category variation (if perceived at all). If the child's attention can be brought to these acoustic-phonetic aspects, his or her perception may be tuned to conform better to the norms of the surrounding language environment. Provided that the child also has the productive skills required to meet adequate articulatory targets, therapeutic gains are also expected in speech production. More work remains, however, in exploring how corrective re-synthesis is best integrated into clinical therapy.

The limitations in traditional descriptions of speech production are illustrated by the findings in Study III, where finer-grained perceptual measures revealed signs of covert contrast that would have been obscured in descriptions solely relying on phonetic transcription. For practicing clinicians, this finding highlights the need of supplementing traditional descriptions with more detailed evaluations, and of remaining open to the possibility that the child is indeed signaling a phonological contrast between speech sounds, although by other means than the expected.

As mentioned in section 2.2.1, intervention addressed to remediate speech impairments typically focuses on aspects related to specific speech error patterns, while paying little (if any) attention to the functional consequences of these deficits. As reported by McLeod & Bleile (2004), targets for intervention are most often selected on the basis of stimulability and/or normative data, rather than with reference to social relevance or (other) functional consequences. Study IV sketches a possible way of bridging the gap between impairment-based goals and socially-based goals. By recruiting assumed everyday listeners, and by recording their reactions while listening to speech produced by children with speech impairments, normative data on the specific speech errors that evoke reactions in listeners can be collected. This information may then be used as socially-based norms regarding which speech errors are most detrimental to intelligibility and/or acceptability – knowledge that may be integrated in the prioritizing of intervention targets.

6.5 Directions for future work

Although the work in this thesis has generated some answers, some questions are still not resolved. Moreover, the generated answers, in turn, give rise to new questions. Hence, some directions for future work can easily be identified.

Regarding children's perception of their own recorded utterances, and whether they recognize themselves as the speaker, Study I leaves unanswered those questions regarding which acoustic features children attend to when deciding if a recorded utterance was produced by themselves or not.[7] Furthermore, the finding that children with PD do not use their misarticulations as cues to identifying themselves as speakers would be easier to interpret if similar data for children with typical speech and language were available. For example, one may consider using the same experimental design, but to let the distinction between subjects and foils be one of dialectal features. This would allow exploration of children's ability to identify speakers by dialectal traits, in children with typical speech and language development, as well as in children with poorer speech and language skills. This would respond to the paucity of research on the acquisition of social-indexical knowledge in children with speech and language impairments (Edwards & Munson, 2008).

As discussed above, the availability of technology for performing real-time modification of children's speech opens up for some new routes to the exploration of the speech production-perception interplay. However, the technology is still limited to a specific speech error type, in a specific context. One direction for future work is further development of the technology, to enable it to also handle other speech sounds, and in other contexts. Another direction is to, as discussed above, combine corrective re-synthesis with detailed acoustic analysis, to explore the acoustic difference between incorrect and synthetically corrected productions within

[7] Unpublished exploration of whether the children's selection patterns could be explained by acoustic similarity in terms of f0, spectral features (MFCCs) and intensity did not reveal any such associations.

individual children. In these suggested explorations, it should be acknowledged that "correct production" is not an absolute reference; however, neither is "incorrect production". Hence, since variation is expected both among the children's original incorrect productions and among the synthetically corrected versions of these productions, experiments should involve large enough sample sizes to compensate for this variation. Given the appropriate design, such studies may constitute a new route to revealing patterns of which acoustic-phonetic strategies are involved in an individual child's production and perception of phonological contrasts.

The observation that children's perception is dynamic and susceptible to the influence of just having produced an utterance invites hypotheses that should be tested experimentally. For example, it is still unknown whether these perceptual tuning effects would also be evidenced in the children's evaluations of utterances produced by other children, or whether they are specific to the child's evaluation of his or her own productions – recorded or not. This question may be explored by exposing the children to phonetically similar productions produced by other children, or generated synthetically.

Given the indications of a satisfactory quality of the synthetically generated speech, the technology may be considered ripe for clinical trials. A small-scale pilot-study has already been carried out, where children's misarticulated productions of isolated words were played back to them, and contrasted to synthetically corrected versions of the same utterances. The task for the children was then to – in cooperation with the therapist – evaluate whether the two sounded "the same" or "different". Furthermore, the children were encouraged to – with or without articulatory instructions from the therapist – try to approach the computer-generated corrected version. The qualitative analysis of video recorded therapy sessions identified some pedagogical challenges involved in the therapists' communication with the participating children about acoustic-phonetic differences on the one hand, and phonological contrasts on the other hand. The fact that "different" may refer either to a subphonemic difference or to a phonological contrast often caused confusion in the children. Hence, the administration and presentation of the therapeutic approach needs refinement before new clinical trials are commenced. Furthermore, an additional suggested improvement is to introduce different levels of difficulty, i.e. the target sound in both less and more complex contexts than what the current version allows.

The results reported in Study III highlight the gap between the perceptual sensitivity in human listeners and the insufficient capacity in state-of-the-art instruments of acoustic analysis to mimic this sensitivity. This indicates that there is acoustic information that human listeners pick up that slips unnoticed through the acoustic analysis. Hence, the analytic instruments need considerable improvement before exhibiting the same perceptual sensitivity as human listeners.

Albeit promising, the work presented in Study IV constitutes but the first steps towards a new approach to the exploration of how different speech error types contribute to listener evaluations of intelligibility/acceptability. Hence, larger-scale studies involving the quantitative analysis of correlations between speech events and listener reactions are needed to confirm the practicability of the method. One such study is currently underway, where the limitations identified in Study IV have been addressed (for example, the instructions to the participating listeners have been refined). If proven successful, this study will be followed by more, where the scope is not necessarily restricted to childhood SSDs, but could also extend to other speech, voice or language conditions, and even beyond clinical populations, e.g. in explorations of foreign accent.

6.6 Concluding remarks

The work presented in this thesis has focused on the examination of how children's misarticulations are perceived by other listeners, and by the children themselves. These explorations have involved the introduction of some novel methodological approaches, as well as the innovative use of existing technology to the domain of childhood SSDs. As such, the work constitutes an example of cross-disciplinary study of children's misarticulated speech. Moreover, the present work serves to extend earlier research, which is predominantly based on (American) English, to a new language: Swedish. Through these investigations, conclusions have been drawn that may have implications of both narrow and broad scopes.

For the individual child with SSD, the possible introduction of a new therapeutic tool – corrective re-synthesis – may constitute an attractive addition to existing clinical strategies, in presenting the child with individually tailored speech targets, i.e. what it would sound like if the child himself/herself were able to produce the target sound correctly. For the practicing clinician, raised awareness of the possible existence of covert contrast may stimulate assessment strategies that involve keeping ears open for more subtle acoustic-phonetic detail than phonetic transcriptions convey. As an extension of these findings, clinicians may want to re-consider the traditional dichotomy between phonological and articulatory disorders. For the research community, studies of speech perception are encouraged that involve deeper exploration of the dynamic nature of speech perception, particularly, of how listeners' evaluation of speech may be influenced by the time since their own production of an utterance. Furthermore, the work presented here may stimulate other researchers to invite methods from one area (e.g. speech technology) and apply it to another (e.g. speech-language pathology), to explore possible gains of cross-fertilization across disciplines.

References

ASHA, American Speech-Language-Hearing Association. (2013). Speech Sound Disorders: Articulation and Phonological Processes. American Speech-Language-Hearing Association. Retrieved October 29, 2013 from http://www.asha.org/public/speech/disorders/SpeechSoundDisorders.htm

Aungst, L. F. & Frick, J. V. (1964). Auditory discrimination ability and consistency of articulation of /r/. *Journal of Speech and Hearing Disorders, 29*(1), 76-85.

Baker, E., Croot, K., McLeod, S., & Paul, R. (2001). Psycholinguistic models of speech development and their application to clinical practice. *Journal of Speech Language and Hearing Research, 44*(3), 685-702.

Baudonck, N. L. H., Buekers, R., Gillebert, S., & Van Lierde, K. M. (2009). Speech intelligibility of Flemish children as judged by their parents. *Folia Phoniatrica et Logopaedica, 61*(5), 288-295.

Berko, J. & Brown, R. (1960). Psycholinguistic research methods. In P. Mussen (Ed.), *Handbook of research methods in child development* (pp. 517-557). New York, NY: John Wiley.

Beukelman, D. R. & Yorkston, K. M. (1980). Influence of passage familiarity on intelligibility estimates of dysarthric speech. *Journal of Communication Disorders, 13*(1), 33-41.

Blumstein, S. E. & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America, 66*(4), 1001-1017.

Bowen, C. (2009). *Children's speech sound disorders*. Oxford: Wiley-Blackwell.

Bradford, A. & Dodd, B. (1996). Do all speech-disordered children have motor deficits? *Clinical Linguistics & Phonetics, 10*(2), 77-101.

Broomfield, J. & Dodd, B. (2004). Children with speech and language disability: Caseload characteristics. *International Journal of Language & Communication Disorders, 39*(3), 303-324.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly, 22*(4), 593-606.

Burroughs, E. I. & Tomblin, J. B. (1990). Speech and language correlates of adults' judgments of children. *Journal of Speech and Hearing Disorders, 55*(3), 485-494.

Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Paradise, J. L., Feldman, H. M., Shriberg, L. D., ... Kurs-Lasky, M. (2003). Risk factors for speech delay of unknown origin in 3-year-old children. *Child Development, 74*(2), 346-357.

Campbell, W. N. & Skarakis-Doyle, E. (2007). School-aged children with SLI: The ICF as a framework for collaborative service delivery. *Journal of Communication Disorders, 40*(6), 513-535.

Chapin, C., Tseng, C., & Lieberman, P. (1982). Short-term release cues to stop-consonant place of articulation in child speech. *Journal of the Acoustical Society of America, 71*(1), 179-186.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America, 24*(6), 597-606.

Coplan, J. & Gleason, J. R. (1988). Unclear speech: Recognition and significance of unintelligible speech in preschool children. *Pediatrics, 82*(3), 447-452.

Crosbie, S., Holm, A., & Dodd, B. (2009). Cognitive flexibility in children with and without speech disorder. *Child Language Teaching and Therapy, 25*(2), 250-270.

Crowe Hall, B. J. (1991). Attitudes of fourth and sixth graders toward peers with mild articulation disorders. *Language Speech and Hearing Services in Schools, 22*(1), 334-340.

Delattre, P., Cooper, F. S., Liberman, A. M., & Gerstman, L. (1954). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America, 26*(1), 137-137.

Dodd, B. (2005). *Differential diagnosis and treatment of children with speech disorder* (2nd ed.). London, UK.: Whurr Publishers Ltd.

Dodd, B. & Bradford, A. (2000). A comparison of three therapy methods for children with different types of developmental phonological disorder. *International Journal of Language & Communication Disorders, 35*(2), 189-209.

Dodd, B. & McIntosh, B. (2010). Two-year-old phonology: impact of input, motor and cognitive abilities on development. Journal of Child Language, 37(5), 1027-1046.

Edwards, J., Fox, R. A., & Rogers, C. L. (2002). Final Consonant Discrimination in Children: Effects of Phonological Disorder, Vocabulary Size, and Articulatory Accuracy. *Journal of Speech Language and Hearing Research, 45*(2), 231-242.

Edwards, J., Gibbon, F., & Fourakis, M. (1997). On discrete changes in the acquisition of the alveolar/velar stop consonant contrast. *Language and Speech, 20*(2), 203-210.

Edwards, J. & Munson, B. (2008). Speech perception and production in child language disorders. In R. G. Schwartz (Ed.), *Handbook of child language disorders* (pp. 216-231). New York, NY: Psychology Press.

Edwards, J., Munson, B., & Beckman, M. E. (2011). Lexicon-phonology relationships and dynamics of early language development. *Journal of Child Language, 38*(1), 35-40.

Engstrand, O. (2007). *Fonetik light - kortfattad ljudlära för språkstudier och uttalsundervisning*. Lund, Sweden: Studentlitteratur.

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: The MIT Press.

Fey, M. E. (1992). Articulation and phonology: An addendum. *Language Speech and Hearing Services in Schools, 23*(3), 277-282.

Flipsen, P. J. (1995). Speaker-listener familiarity: Parents as judges of delayed speech intelligibility. *Journal of Communication Disorders, 28*(1), 3-19.

Flipsen, P. J. (2006). Measuring the intelligibility of conversational speech in children. *Clinical Linguistics & Phonetics, 20*(4), 303-312.

Flipsen, P. J. (2010). Factors associated with the intelligibility of conversational speech produced by children with cochlear implants. In P. Rhea & P. J. Flipsen (Eds.), *Speech sound disorders in children: In honor of Lawrence D. Shriberg* (pp. 225-246). San Diego, CA: Plural Publishing.

Forrest, K., Weismer, G., Hodge, M., Dinnsen, D. A., & Elbert, M. (1990). Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children. *Clinical Linguistics & Phonetics, 4*(4), 327-340.

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*(1), 115-123.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110-125.

Gardner, H. (1997). Are your minimal pairs too neat? The dangers of phonemicisation in phonology therapy. International *Journal of Language & Communication Disorders, 32*(2s), 167-175.

Gertner, B. L., Rice, M. L., & Hadley, P. A. (1994). Influence of communicative competence on peer preferences in a preschool classroom. *Journal of Speech and Hearing Research, 37*(4), 913-923.

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech Language and Hearing Research, 42*(2), 382-397.

Goldman, R. & Fristoe, M. (2000). *Goldman Fristoe Test of Articulation* (2nd ed.). Minneapolis, MN: Pearson Assessments.

Gordon-Brannan, M. (1994). Assessment of speech intelligibility: A prerequisite for oral communicative competence. *Topics in Language Disorders, 14*(2), 18-27.

Gordon-Brannan, M. & Hodson, B. W. (2000). Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology, 9*(1), 141-150.

Hansson, K. (1998). *Specific language impairment in Swedish. Grammar and interaction.* Doctoral Thesis: Lund University, Sweden, Lund.

Hansson, K. & Nettelbladt, U. (2002). Assessment of specific language impairment in Swedish. *Logopedics Phoniatrics Vocology, 27*(4), 146-154.

Hardcastle, W. J., Gibbon, F. E., & Jones, W. (1991). Visual display of tongue-palate contact: Electropalatography in the assessment and remediation of speech disorders. *International Journal of Language & Communication Disorders, 26*(1), 41-74.

Hartelius, L. & Miller, N. (2010). The ICF framework and its relevance to the assessment of people with motor speech disorders. In A. Lowit & R. D. Kent (Eds.), *Assessment of motor speech disorders.* Plural Publishing.

Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics, 28*(4), 377-396.

Hewlett, N. (1985). Phonological versus phonetic disorders: Some suggested modifications to the current use of the distinction. *International Journal of Language & Communication Disorders, 20*(2), 155-164.

Hewlett, N. (1988). Acoustic properties of /k/ and /t/ in normal and phonologically disordered speech. *Clinical Linguistics & Phonetics, 2*(1), 29-45.

Hewlett, N. (1990). Issues in speech development and speech disorders: Processes of development and production. In P. Grunwell (Ed.), *Developmental speech disorders* (pp. 15-38). Edinburgh: Whurr Publishers Ltd.

Hodson, B. W. & Paden, E. (1983). *Targeting intelligible speech: A phonological approach to remediation.* San Diego, CA: College-Hill Press.

Hodson, B. W. & Paden, E. P. (1981). Phonological processes which characterize unintelligible and intelligible speech in early childhood. *Journal of Speech and Hearing Disorders, 46*(4), 369-373.

Hoffman, P. R., Daniloff, R. G., Bengoa, D., & Schuckers, G. H. (1985). Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. *Journal of Speech and Hearing Disorders, 50*(1), 46-53.

Hoffman, P. R., Stager, S., & Daniloff, R. G. (1983). Perception and production of misarticulated /r/. *Journal of Speech and Hearing Disorders, 48*(2), 210-215.

Idemaru, K. & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *Journal of the Acoustical Society of America, 133*(6), 4232-4246.

Ingram, D. (1989). *Phonological disability in children* (2nd ed.). London: Cole and Whurr Ltd.

Jamieson, D. G. & Rvachew, S. (1992). Remediating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology, 16*(3), 201-210.

Kaschak, M. & Gernsbacher, M. A. (2014). Changing language. In S. Bischoff & C. Jany (Eds.), *Functional approaches to language* (pp. 149-175). Berlin: De Gruyter Mouton.

Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5*(3), 7-23.

Kent, R. D., Miolo, G., & Bloedel, S. (1994). The intelligibility of children's speech: A review of evaluation procedures. *American Journal of Speech-Language Pathology, 3*(1), 81-95.

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*(4), 482-499.

Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *Journal of the Acoustical Society of America, 72*(2), 379-389.

Kiparsky, P. & Menn, L. (1977). On the acquisition of phonology. In J. Macnamara (Ed.), *Perspectives in neurolinguistics and psycholinguistics* (pp. 47-78). New York, NY: Academic Press.

Klein, E. S. & Flint, C. B. (2006). Measurement of intelligibility in disordered speech. *Language, Speech, and Hearing Services in Schools, 37*(3), 191-199.

Koegel, L. K., Koegel, R. L., & Ingham, J. C. (1986). Programming rapid generalization of correct articulation through self-monitoring procedures. *Journal of Speech and Hearing Disorders, 51*(1), 24-32.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics, 50*(2), 93-107.

Kwiatkowski, J. & Shriberg, L. D. (1992). Intelligibility assessment in developmental phonological disorders: Accuracy of caregiver gloss. *Journal of Speech and Hearing Research, 35*(5), 1095-1104.

Lagerberg, T. (2013). *Assessment of intelligibility in children*. Doctoral Thesis: University of Gothenburg, Sweden.

Lapko, L. L. & Bankson, N. W. (1975). Relationship between auditory discrimination, articulation stimulability, and consistency of misarticulation. *Perceptual and Motor Skills, 40*(1), 171-177.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language and Communication Disorders, 35*(2), 165-188.

Leonard, L. B., Schwartz, R. G., Folger, M. K., & Wilcox, M. J. (1978). Some aspects of child phonology in imitative and spontaneous speech. *Journal of Child Language, 5*(3), 403-415.

Li, F., Edwards, J., & Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics, 37*(1), 111-124.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffiths, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358-368.

Lindblom, B. (1963). On vowel reduction. In *Report No. 29, Speech Transmission Laboratory*. Stockholm, Sweden: The Royal Institute of Technology.

Locke, J. L. (1980a). The inference of speech perception in the phonologically disordered child. Part I: A rationale, some criteria, the conventional tests. *Journal of Speech and Hearing Disorders, 45*(4), 431-444.

Locke, J. L. (1980b). The inference of speech perception in the phonologically disordered child. Part II: Some clinically novel procedures, their use, some findings. *Journal of Speech and Hearing Disorders, 45*(4), 445-468.

Locke, J. L. (1983). Clinical phonology: The explanation and treatment of speech sound disorders. *Journal of Speech and Hearing Disorders, 48*(4), 339-341.

Locke, J. L. & Kutz, K. J. (1975). Memory for speech and speech for memory. *Journal of Speech and Hearing Research, 18*(1), 176-191.

Lundeborg, I. & McAllister, A. (2007). Treatment with a combination of intra-oral sensory stimulation and electropalatography in a child with severe developmental dyspraxia. *Logopedics Phoniatrics Vocology, 32*(2), 71-79.

Löfqvist, A. (1999). Interarticulator phasing, locus equations, and degree of coarticulation. *Journal of the Acoustical Society of America, 106*(4 Pt 1), 2022-2030.

Ma, E. P., Yiu, E. M., & Abbott, K. V. (2007). Application of the ICF in voice cisorders. *Seminars in Speech and Language, 28*(4), 343-350.

Maassen, B. & Povel, D.-J. (1985). The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *Journal of the Acoustical Society of America, 78*(3), 877-886.

Macken, M. A. & Barton, D. (1980). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7*(1), 41-74.

Magnusson, E. (1983). *The phonology of language disordered children: production, perception, and awareness*. Doctoral Thesis: Lund University, Sweden.

Marin, S., Pouplier, M., & Harrington, J. (2010). Acoustic consequences of articulatory variability during productions of /t/ and /k/ and its implications for speech error research. *Journal of the Acoustical Society of America, 127*(1), 445-461.

McAllister Byun, T. (2012). Bidirectional perception-production relations in phonological development: evidence from positional neutralization. *Clinical Linguistics & Phonetics, 26*(5), 397-413.

McCormack, J., McAllister, L., McLeod, S., & Harrison, L. (2012). Knowing, having, doing: The battles of childhood speech impairment. *Child Language Teaching and Therapy, 28*(2), 141-157.

McCormack, J., McLeod, S., Harrison, L. J., & McAllister, L. (2010). The impact of speech impairment in early childhood: Investigating parents' and speech-language pathologists' perspectives using the ICF-CY. *Journal of Communication Disorders, 43*(5), 378-396.

McCormack, J., McLeod, S., McAllister, L., & Harrison, L. (2009). A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology, 11*(2), 155-170.

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research, 26*(3), 451-458.

McGregor, K. K. & Schwartz, R. G. (1992). Converging evidence for underlying phonological representation in a child who misarticulates. *Journal of Speech and Hearing Research, 35*(3), 596-603.

McHenry, M. (2011). An exploration of listener variability in intelligibility judgments. *American Journal of Speech-Language Pathology, 20*(2), 119-123.

McLeod, S. (2004). Speech pathologists' application of the ICF to children with speech impairment. *Advances in Speech-Language Pathology, 6*(1), 75-81.

McLeod, S. (2006). An holistic view of a child with unintelligible speech: Insights from the ICF and ICF-CY. *International Journal of Speech-Language Pathology, 8*(3), 293-315.

McLeod, S. & Bleile, K. (2004). The ICF: A framework for setting goals for children with speech impairment. *Child Language Teaching and Therapy, 20*(3), 199-219.

McLeod, S. & Harrison, L. J. (2009). Epidemiology of speech and language impairment in a nationally representative sample of 4- to 5-year-old children. *Journal of Speech Language and Hearing Research, 52*(5), 1213-1229.

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The Intelligibility in Context Scale: Validity and reliability of a subjective rating measure. *Journal of Speech Language and Hearing Research, 55*(2), 648-656.

McLeod, S. & McCormack, J. (2007). Application of the ICF and ICF-Children and Youth in Children with Speech Impairment. *Seminars in Speech and Language, 28*(4), 254-264.

McQueen, J. M., Tyler, M. D., & Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development, 8*(4), 317-339.

Menn, L. & Matthei, E. (1992). The "two lexicon" account of child phonology looking back, looking ahead. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 211-247). Timonium, MD: York Press, Inc.

Menn, L., Schmidt, E., & Nicholas, B. (2013). Challenges to theories, charges to a model: The Linked-Attractor model of phonological development. In M. M. Vihman & T. Keren-Portnoy (Eds.), *Child phonology: Whole word approaches, cross-linguistic evidence*. Cambridge, UK: Cambridge University Press.

Miller, N. (2013). Measuring up to speech intelligibility. International *Journal of Language & Communication Disorders, 48*(6), 601-612.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18*(5), 331-340.

Monnin, L. M. & Huntington, D. A. (1974). Relationship of articulatory defects to speech-sound identification. *Journal of Speech and Hearing Research, 17*(3), 352-366.

Morrison, J. A. & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research, 35*(2), 259-273.

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge, UK.: Cambridge University Press.

Mullen, R. & Schooling, T. (2010). The national outcomes measurement system for pediatric speech-language pathology. *Language, Speech, and Hearing Services in Schools, 41*(1), 44-60.

Munson, B., Edwards, J., & Beckman, M. E. (2005a). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech Language and Hearing Research, 48*(1), 61-78.

Munson, B., Edwards, J., & Beckman, M. E. (2005b). Phonological knowledge in typical and atypical speech-sound development. *Topics in Language Disorders, 25*(3), 190-206.

Munson, B., Edwards, J., Schellinger, S. K., Beckman, M. E., & Meyer, M. K. (2010). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. *Clinical Linguistics & Phonetics, 24*(4-5), 245-260.

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology, 21*(2), 124-139.

Munson, B., Schellinger, S. K., & Urberg Carlson, K. (2012). Measuring speech-sound learning using visual analog scaling. *Perspectives on Language Learning and Education, 19*(1), 19-30.

Namasivayam, A. K., Pukonen, M., Goshulak, D., Yu, V. Y., Kadis, D. S., Kroll, R., ... Nil, L. F. D. (2013). Relationship between speech motor control and speech intelligibility in children with speech sound disorders. *Journal of Communication Disorders, 46*(3), 264-280.

Nathan, L. & Wells, B. (2001). Can children with speech difficulties process an unfamiliar accent? *Applied Psycholinguistics, 22*(03), 343-361.

Nettelbladt, U. (1983). *Developmental studies of dysphonology in children*. Doctoral Thesis: Lund University, Sweden.

Nettelbladt, U., Samuelsson, C., Sahlén, B., & Hansson, K. (2007). *Språkstörning hos barn utan andra funktionshinder*. In L. Hartelius, U. Nettelbladt, & B. Hammarberg (Eds.), Logopedi (pp. 139-147). Lund, Sweden: Studentlitteratur.

Nijland, L. (2009). Speech perception in children with speech output disorders. *Clinical Linguistics & Phonetics, 23*(3), 222-239.

Nittrouer, S. (2005). Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *Journal of the Acoustical Society of America, 118*(2), 1072-1088.

Ohde, R. N. & Sharf, D. J. (1988). Perceptual Categorization and Consistency of Synthesized /r-w/ Continua by Adults, Normal Children and /r/-Misarticulating Children. *Journal of Speech and Hearing Research, 31*(4), 556-568.

Overby, M., Carrell, T., & Bernthal, J. (2007). Teachers' perceptions of students with speech sound disorders: A quantitative and qualitative analysis. *Language, Speech, and Hearing Services in Schools, 38*(4), 327-341.

Panagos, J. M. & King, R. R. (1975). Self and mutual speech comprehension by deviant and normal speaking children. *Journal of Speech and Hearing Research, 18*(4), 653-662.

Pennington, B. & Bishop, D. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology, 60*, 283-306.

Pierrehumbert, J. B. (1990). Phonological and phonetic representation. *Journal of Phonetics, 18*(3), 375-394.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*(2-3), 115-154.

Preston, J. L., Ramsdell, H. L., Oller, D. K., Edwards, M. L., & Tobin, S. J. (2011). Developing a weighted measure of speech sound accuracy. *Journal of Speech Language and Hearing Research, 54*(1), 1-18.

Rice, M. L., Hadley, P. A., & Alexander, A. L. (1993). Social biases toward children with speech and language impairments: A correlative causal model of language limitations. *Applied Psycholinguistics, 14*(4), 445-471.

Richtsmeier, P. (2010). Child phoneme errors are not substitutions. In *Toronto Working Papers in Linguistics (Vol. 33)*. Department of Linguistics, University of Toronto.

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research, 37*(2), 347-357.

Rvachew, S. (2007). Phonological processing and reading in children with speech sound disorders. *American Journal of Speech-Language Pathology, 16*(3), 260-270.

Rvachew, S. & Brosseau-Lapré, F. (2012). *Developmental phonological disorders. Foundations of clinical practice*. San Diego, CA: Plural Publishing.

Rvachew, S. & Jamieson, D. G. (1989). Perception of voiceless fricatives by children with a functional articulation disorder. *Journal of Speech and Hearing Disorders, 54*(2), 193-208.

Rvachew, S., Leroux, E., & Brosseau-Lapré, F. (2014). Production of word-initial consonant sequences by francophone preschoolers with a developmental phonological disorder. Canadian *Journal of Speech Language Pathology and Audiology, 37*(4), 252-267.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders* (pp. 11-34). Philadelphia, PA: John Benjamins Publishing.

Scobbie, J. E., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. In M. Broe & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 194-203). Cambridge: Cambridge University Press.

Shelton, R. L., Johnson, A. F., & Arndt, W. B. (1977). Delayed judgment speech-sound discrimination and /r/ or /s/ articulation status and improvement. *Journal of Speech and Hearing Research, 20*(4), 704-717.

Shiller, D. M., Rvachew, S., & Brosseau-Lapré, F. (2010). Importance of the auditory perceptual target to the achievement of speech production accuracy. *Canadian Journal of Speech Language Pathology and Audiology, 34*(3), 181-192.

Shriberg, L. D. (2010). Childhood speech sound disorders: From post-behaviorism to the post-genomic era. In R. Paul & P. J. Flipsen (Eds.), *Speech sound disorders in children* (pp. 1-34). San Diego, CA: Plural Publishing.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The speech disorders classification system (SDCS):

extensions and lifespan reference data. *Journal of Speech Language and Hearing Research, 40*(4), 723-740.

Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., ... Wilson, D. L. (2010). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24*(10), 795-824.

Shriberg, L. D. & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders, 47*(3), 256-270.

Shriberg, L. D. & Kwiatkowski, J. (1988). A follow-up study of children with phonologic disorders of unknown origin. *Journal of Speech and Hearing Disorders, 53*(2), 144-155.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech Language and Hearing Research, 42*(6), 1461-1481.

Shuster, L. I. (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech Language and Hearing Research, 41*(4), 941-950.

Silverman, E.-M. (1976). Listeners' impressions of speakers with lateral lisps. *Journal of Speech and Hearing Disorders, 41*(4), 547-552.

Silverman, F. H. & Paulus, P. G. (1989). Peer reactions to teenagers who substitute /w/ for /r/. *Language Speech and Hearing Services in Schools, 20*(2), 219-221.

Smith, B. L. (1979). A phonetic analysis of consonantal devoicing in children's speech. *Journal of Child Language, 6*(1), 19-28.

Stackhouse, J. & Wells, B. (1997). *Children's speech and literacy difficulties: A psycholinguistic framework*. Whurr Publishers Ltd.

Stevens, K. N. & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64*(5), 1358-1368.

Stoel-Gammon, C., Williams, K., & Buder, E. (1994). Cross-language differences in phonological acquisition: Swedish and American /t/. *Phonetica, 51*(1-3), 146-158.

Strömbergsson, S. (2011). Segmental re-synthesis of child speech using unit selection. In *Proceedings of the ICPhS XVII* (pp. 1910-1913). Hong Kong.

Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America, 90*(3), 1309-1325.

Teverovsky, E. G., Bickel, J. O., & Feldman, H. M. (2009). Functional characteristics of children diagnosed with Childhood Apraxia of Speech. *Disability and Rehabilitation, 31*(2), 94-102.

Threats, T. T. (2008). Use of the ICF for clinical practice in speech-language pathology. *International Journal of Speech-Language Pathology, 10*(1-2), 50-60.

Threats, T. T. & Worrall, L. (2004). Classifying communication disability using the ICF. *International Journal of Speech-Language Pathology, 6*(1), 53-62.

Tjaden, K. K. & Liss, J. M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics & Phonetics, 9*(2), 139-154.

Tuller, B., Case, D., Ding, M., & Kelso, J. A. S. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance, 20*(1), 3-16.

Tyler, A. A. (2002). Language-based intervention for phonological disorders. *Seminars in Speech and Language, 23*(1), 69-81.

Tyler, A. A., Figurski, G. R., & Langsdale, T. (1993). Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *Journal of Speech and Hearing Research, 36*(4), 746-759.

Waring, R. & Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Language and Communication Disorders, 48*(1), 25-40.

Weismer, G., Dinnsen, D., & Elbert, M. (1981). A study of the voicing distinction associated with omitted, word-final stops. *Journal of Speech and Hearing Disorders, 46*(3), 320-328.

Westerlund, M. (1994). *Barn med tal och språkavvikelser: En prospektiv longitudinell epidemiologisk studie av en årskull Uppsalabarn vid 4, 7 och 9 års ålder*. Doctoral Thesis: Uppsala university, Sweden.

Whitehill, T. L. (2002). Assessing intelligibility in speakers with cleft palate: A critical review of the literature. *The Cleft Palate-Craniofacial Journal, 39*(1), 50-58.

WHO, World Health Organization. (2001). International Classification of Functioning, Disability and Health. Retrieved October 11, 2013 from http://www.disabilitaincifre.it/documenti/ICF_18.pdf. Geneva, Switzerland: World Health Organization.

WHO, World Health Organization. (2002). Towards a Common Language for Functioning, Disability and Health. Geneva, Switzerland: World Health Organization. Retrieved October 10, 2013 from http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf

WHO, World Health Organization. (2007). International Classification of Functioning, Disability and Health-Children and Youth Version. Retrieved October 8, 2013 from http://apps.who.int/classifications/icfbrowser/. Geneva, Switzerland: World Health Organization.

WHO, World Health Organization. (2010). International Statistical Classification of Diseases and Related Health Problems (10th ed.). Retrieved January 24, 2014 from http://apps.who.int/classifications/icd10/browse/2010/en . Geneva, Switzerland: World Health Organization.

WHO, World Health Organization. (2014). International Statistical Classification of Diseases and Related Health Problems (Beta draft for 11th ed.). Retrieved January 24, 2014 from http://apps.who.int/classifications/icd11/browse/f/en. Geneva, Switzerland: World Health Organization.

Wolfe, V., Presley, C., & Mesaris, J. (2003). The importance of sound identification training in phonological intervention. *American Journal of Speech-Language Pathology, 12*(3), 282-288.

Wood, S. (1975). What is the difference between English and Swedish dental stops? *Working papers, Department of Linguistics, Lund University, 10*, 174-193.

Woolf, G. & Pilberg, R. (1971). A comparison of three tests of auditory discrimination and their relationship to performance on a deep test of articulation. *Journal of Communication Disorders, 3*(4), 239-249.

Yavas, M. & Lamprecht, R. (1988). Processes and intelligibility in disordered phonology. *Clinical Linguistics & Phonetics, 2*(4), 329-345.

Part II. Attached papers