Stockholm University

Swedish voiceless fricatives

A multidimensional investigation of adult and child productions

Carla Wikse Barrow

Doctoral Thesis in Linguistics at Stockholm University, Sweden 2024

Swedish voiceless fricatives

A multidimensional investigation of adult and child productions

Carla Wikse Barrow

Academic dissertation for the Degree of Doctor of Philosophy in Linguistics at Stockholm University to be publicly defended on Friday 20 September 2024 at 14.00 in Hörsal 7, hus D, Södra huset, Universitetsvägen 10.

Abstract

Voiceless fricatives are articulatorily and acoustically complex, and relatively late acquired by children. There are a plenitude of descriptions of voiceless fricatives in other languages, which have revealed language-specific patterns in realisation and acquisition. However, studies of Swedish adult's voiceless fricatives are dated and small-scale, and knowledge concerning Swedish children's acquisition of these complex sounds is limited.

This dissertation is based on four papers, three of which investigate acoustic characteristics of adult and child productions of Swedish voiceless fricatives /f, s, ɕ, ɧ/. Static and dynamic, individual and group-level acoustic patterns (primarily spectral features) are described, and between-fricative contrasts are quantified for individual speakers. The fourth paper explores the influence of lexical context and experience on perceptual ratings of children's voiceless sibilant fricatives /s, ɕ/. Specifically, listeners with and without clinical experience of assessing child speech (i.e., speech-language pathologists and laypeople) provided gradient ratings of sibilants presented in lexical and non-lexical contexts (i.e., words or CV-syllables).

This work contributes to a better understanding of the characteristics and variability of /f, s, ɕ, ɧ/, and provides new insights into children's acquisition of voiceless fricatives. A number of spectral parameters and statistical models were utilised in this work. The results of the fourth study also have relevance for perceptual assessments of speech in the clinic.

Keywords: *fricatives, Swedish, spectral analysis, speech acquisition, perceptual assessment.*

Stockholm 2024
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-231895

ISBN 978-91-8014-859-7
ISBN 978-91-8014-860-3

Department of Linguistics

Stockholm University, 106 91 Stockholm

SWEDISH VOICELESS FRICATIVES

Carla Wikse Barrow

Stockholm University

Swedish voiceless fricatives

A multidimensional investigation of adult and child productions

Carla Wikse Barrow

ISBN print 978-91-8014-859-7
ISBN PDF 978-91-8014-860-3

Cover illustration by John Adam Barrow, 2024.
Printed in Sweden by Universitetsservice US-AB, Stockholm 2024

For Robin and Bo.

Contents

List of Papers

The following papers, referred to in the text by their Roman numerals, are included in this thesis.

I **Static and dynamic spectral characteristics of Swedish voiceless fricatives**
Wikse Barrow, C., Włodarczak, M., Thörn, L. & Heldner, M. (2022). *Journal of the Acoustical Society of America*, **152**, 2588–2600.
doi: 10.1121/10.0014947

II **Variability in Swedish voiceless fricative contrasts**
Wikse Barrow, C., Włodarczak, M., Heldner, M. & Strömbergsson, S. (2023). *Proceedings of he 20th International Congress of Phonetic Sciences*, 813-817.

III **Individual variation in the realisation and contrast of Swedish children's voiceless fricatives**
Wikse Barrow, C., Strömbergsson, S., Włodarczak, M., & Heldner, M. (2024). *Journal of Phonetics*. Advance online publication.
doi: 10.1016/j.wocn.2024.101351

IV **Exploring the effect of lexicality and experience on listeners' gradient ratings of Swedish sibilant fricatives**
Wikse Barrow, C., Ottosson, L. & Strömbergsson, S. (2024). *Clinical Linguistics & Phonetics*. Advance online publication.
doi: 10.1080/02699206.2024.2371121

Non-included publications.

The following work is relevant to the subject matter at hand and was conducted during the time that the dissertation was written, but is not included in the present doctoral thesis.

1. **A survey of Swedish speech-language pathologists' practices regarding assessment of speech sound disorders**
   Wikse Barrow C., Körner, K. & Strömbergsson, S. (2023). *Logopedics Phoniatrics Vocology*, **48(1)**, 23-34.
   doi: 10.1080/14015439.2021.1977383

2. **A multidimensional investigation of covert contrast in Swedish acquiring children's speech - a project description**
   Wikse Barrow, C., Strömbergsson, S., & Heldner, M. (2019). *Proceedings of Fonetik 2019*, 79-83.

Author contributions

The PhD project described herein is a result of fruitful collaboration with my supervisors and students. To allow evaluation of my work, I list my contributions to the project here, first related to project administration and planning, and then in relation to writing and the publication process.

Project administration and planning
I formulated the idea for this project together with Mattias Heldner and Sofia Strömbergsson. I wrote the first draft of the ethical permit applications (one original and two amendments), under the supervision of Sofia Strömbergsson and Mattias Heldner. I took the lead regarding experimental design, preparation and participant recruitment for all studies, and conducted all speech recordings of adults and children. I supervised the annotation of the adult speech data (performed by phonetics student Lisa Thörn) and performed annotation and transcription of the child data. The speech perception task in *Paper IV* was designed by myself and Sofia Strömbergsson. I prepared the speech stimuli and implemented the experimental paradigm in Psychopy. The speech perception data were collected by a speech-language pathology student (Lina Ottosson), whom I supervised. I took the initiative to many methodological choices regarding acoustic and statistical analysis, but these choices were always discussed extensively with my supervisors (primarily Marcin Włodarczak re statistics).

Writing and publications
The Kappa of this book was written by me. All papers included herein were written in collaboration with others. Author contribution statements are published in *Papers I*, *III* and *IV* and my specific contributions to each paper are described below. I was the corresponding author for all papers and took the lead regarding revision and editing.

Paper I: I designed the study with Mattias Heldner. I collected and curated the data, and wrote the original draft of the manuscript. I performed the statistical analysis and wrote the description of method and results together with Marcin Włodarczak.

Paper II: I wrote the original draft, summarised and visualised the results. The statistical analysis was performed by myself and Marcin Włodarczak. All authors contributed to conceptualisation, and to the review and editing of the manuscript.

Paper III: I designed the study together with Sofia Strömbergsson and Mattias Heldner. I collected the data, organised the database, performed the annotations and transcriptions, summarised and visualised the results and wrote the first draft of the manuscript. I wrote the scripts and performed the statistical analysis with Marcin Włodarczak. All authors contributed to editing and review of the manuscript, and approved the submitted version.

Paper IV: I designed the study and prepared the experiment in collaboration with Sofia Strömbergsson. I wrote the original draft, performed the statistical analysis (with input from visiting Professor Florian Jaeger) and prepared the data and visualisations for the paper and for the publicly available dataset and code. I reviewed and edited the manuscript with input from Sofia Strömbergsson. All authors read and approved the final version of the manuscript.

Abbreviations

AoA	Age of Acquisition
BF	Bayes Factor
CLP	Cleft lip and/or palate
CP	Categorical Perception
CS	Central Swedish
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
GAMM	General additive mixed model
IPA	International Phonetic Alphabet
LINUS	LINköpingsUnderSökningen
MTS	Multitaper spectrum/a
OoA	Order of Acquisition
PCC	Percentage Consonants Correct
RSE	Residual Speech Error
SLP	Speech-Language Pathologist
SSD	Speech Sound Disorder
SVANTE	Swedish Articulation and Nasality Test
VAS	Visual Analogue Scale

List of Figures

List of Tables

Acknowledgements

I would like to begin by thanking my three supervisors Professor Mattias Heldner, Associate Professor Sofia Strömbergsson and Associate Professor Marcin Włodarczak. I have thoroughly enjoyed working with you and am endlessly grateful for all you have taught me.

Mattias, thank you especially for your calm and steady "it'll work out" attitude, for sharing your profound phonetic knowledge and for your kindness. Sofia, thank you for your analytical eye, our many fruitful and enjoyable discussions, and your positive energy. Marcin, thank you for your patience, your humour and all your help with statistics (GAMM Thursdays for the win!).

I am thankful for the input from visiting professor Florian Jaeger during the statistical workshop at Stockholm University in 2023.

Thank you also to all my past and present colleagues at the Department of Linguistics, in particular my fellow PhD-students. You all contribute to the vibrant, welcoming and stimulating environment at the Department. Special shout-outs to Caroline, Christoffer, Signe, Petter, David, Johan, Amanda and Anna. Thank you for fun and discussion!

I would also like to thank my family for their love and support, particularly my mother, Maya.

Thank you to the love of my life, Robin, who has stood by me through PhD-related frustration, and who continues to encourage me every day, even though I'm pathologically anxious.

Finally, thank you to my dearest boy, Bo. Your smile and your cuddles remind me everyday of what is truly important in life.

1. Introduction

The overall aim of this thesis is to describe characteristics and variability of Swedish adults' and children's voiceless fricatives. At the outset of this work, my primary interest in voiceless fricatives was related to children's acquisition of them. Given that child speech is habitually compared or related to the adult target, I began by looking for accounts of Swedish adults' voiceless fricatives. Alas, descriptions were scarce. Thus, *Paper I: Static and dynamic spectral characteristics of Swedish voiceless fricatives* was born. Ideally, phonetic descriptions should include both group-level mean values of the features of interest, as well as accounts of individual variation. Hence, *Paper II: Variability in Swedish voiceless fricative contrasts* came to be. Finally, children's voiceless fricatives were described, using similar (but unfortunately not identical) methods as for the adults, in *Paper III: Individual variation in the realisation and contrast of Swedish children's voiceless fricatives*. To conclude, adult's perception of children's voiceless fricatives was investigated in *Paper IV: Exploring the effect of lexicality and listener experience on gradient ratings of sibilant fricatives*, although focus lay on the perceptual task rather than the relationship between acoustic characteristics and listeners' perception.

The following chapters provide background and context for the four included papers, beginning with a chapter on Swedish phonology (Chapter 2). The third chapter presents an overview of articulation and acoustics of voiceless fricatives (Chapter 3), and the fourth chapter concerns children's acquisition of voiceless fricatives, in the context of general speech development (Chapter 4). A brief review of the perception of voiceless fricatives is provided in Chapter 5. The four papers are summarized in Chapter 6, and a general discussion with concluding remarks follows in Chapter 7. A Swedish summary is provided at the end.

This thesis presupposes some knowledge regarding basic phonetic concepts (e.g., *waveform*, *phoneme* and *phonetic features*) and analysis (e.g., *phonetic transcription* and *spectrograhic analysis*) but does not require a previous understanding of child speech acquisition or spoken Swedish.

2. Swedish phonology

This chapter serves as brief introduction to Swedish phonology. A detailed description of the acoustics of Swedish voiceless fricatives can be found in Section 3.2.7.

Swedish is spoken by over thirteen million people (as L1 by 9,911,930 and as L2 by 3,157,400) across several continents (Ethnologue, 2024), although primarily in Sweden and surrounding countries. There are seven main regional varieties of Swedish, which roughly correspond to seven geographical regions of Sweden (Bruce, 2010). The dialects primarily vary with respect to prosody (specifically intonation and lexical word accent; Riad 2014), although segmental differences, such as in the realisation of voiceless fricatives (Lindblad, 1980), are also found. The recordings for this thesis were conducted in Stockholm municipality, recruiting speakers who were born and raised in the Svealand region so as to elicit Central Swedish (CS).

Swedish has been described as "typologically fairly mainstream" with a "medium-large and medium complex" phonological inventory (Bruce and Engstrand, 2006). Cross-linguistically notable features include the populous central-front rounded vowel space, the crowd of voiceless fricatives and prosodic features like lexical word tone and the quantity system for stress. The quantity system necessitates that stressed syllables are heavy, which, on the segmental level, is manifested in the vowel or in the vowel and the following consonant. A stressed syllable must contain one, and only one, long segment; either a long vowel or a long consonant (after a short vowel). Hence, segmental quantity is complementary (Riad, 2014).

The Swedish vowel inventory includes nine vowel phonemes, each with short and long variants/allophones (Riad, 2014, see Figure 2.1 for the CS vowel space). Some of the vowel pairs are very distinct (e.g., [ɑː]-[a], [eː]-[ɛ], [ʉː]-[ʊ]) although most are qualitatively similar. The short allophones of /e/ and /ɛ/ have merged as [ɛ̝] resulting in eight short allophones rather than nine. The Swedish vowel space is rather crowded in the high central/front region, incorporating both in-rounded and out-rounded vowels /y, ʉ, ø/ (Riad, 2014).

With respect to consonants, Swedish has 18 phonemes, of which 16 occur in long and short variants (i.e., all except /ɕ, h/ which are never long and never appear in postvocalic position), that are qualitatively alike (Riad, 2014). The consonant inventory of CS is displayed in Table 2.1. The throng of alveolo-

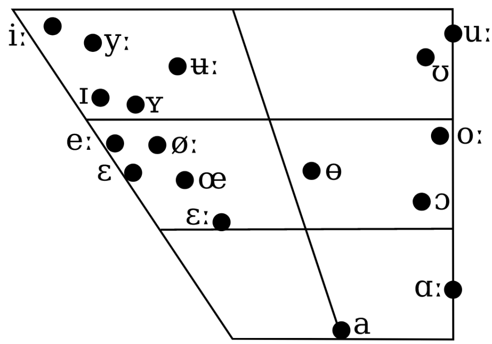


Figure 2.1: Central Swedish vowel space. "Swedish monophthongs chart" from Wikipedia (shared with a CC-SA 3.0 license), based on the description in the Handbook of the International Phonetic Association (Engstrand, 1999).

palatal and velar voiceless fricatives, /s/, /ɕ/ and /ɧ/, is a typologically special feature. The velar /ɧ/ shows large allophonic variation, and in speakers of CS, it is customarily realised as [ɧ] in prevocalic position and as retroflex sibilant [ʂ] in postvocalic position (Lindblad, 1980). The approximant /ɹ/ also exhibits significant variability across dialects and word positions, with realisations such as an alveolo-palatal fricative [ʑ], a uvular trill [ʀ] and a uvular fricative [ʁ] (Riad, 2014). The retroflexion rule [1] of Swedish states that when /ɹ/ precedes /s, t, d, n, l/, retroflex sounds are generated; [ʂ, ʈ, ɖ, ɳ, ɭ] respectively (Riad, 2014). Hence, the voiceless sibilant /s/ before /ɹ/ would be produced as [ʂ], which has been described as perceptually indistinguishable from [ɕ] (see e.g., Shosted, 2008, and references therein).

[1]The rule is applicable to most dialects north of Småland, including Central Swedish (Riad, 2014)

Table 2.1: Central Swedish consonants sorted by place of articulation (columns) and manner of articulation (rows). Unvoiced consonants are placed to the left and voiced consonants to the right in the tables cells. The table was adapted from the description of Swedish in the Handbook of the International Phonetic Association (Engstrand, 1999).

	Bilabial	Labiodental	Dental	Alveolar	Palatal	Velar	Glottal
Plosive	p b		t d			k g	
Nasal	m		n			ŋ	
Fricative		f v	s		ʝ		h
Approximant				ɹ			
Lateral approximant			l				

ɧ (Voiceless dorso-palatal/velar fricative) ɕ (Voiceless alveolo-palatal fricative)

3. The production of voiceless fricatives

Voiceless fricatives are articulatorily complex and acoustically noisy. The present chapter provides a brief overview of the articulation (Section 3.1) and acoustics (Section 3.2) of voiceless fricatives. As this dissertation primarily employs acoustic analysis, articulation is only briefly reviewed. Acoustics of voiceless fricatives are discussed in more detail, including various methodological aspects of acoustic analysis, centered on spectral analysis. The current state of knowledge concerning Swedish voiceless fricatives is detailed in Section 3.2.7.

3.1 Articulation of voiceless fricatives

The primary noise source in fricative sounds is turbulent air-flow (i.e., frication) that is created as the laminar air-stream from the lungs is forced through a constriction in the vocal apparatus. This *manner of articulation* is common for all fricatives, although *place of articulation*, which is expressed in relation to the constriction, varies from labio-dental (as for [f]), through dental (e.g., [s]), palatal (for [ç]) and velar (in [x]) to glottal (e.g., [h]). In articulation of *sibilant fricatives*, an additional noise source is introduced, as the jet of air from the constriction hits the front teeth (Johnson, 2012).

The constriction need be narrow enough and the air particle velocity high enough for turbulence to arise (Johnson, 2012). In lingual fricatives (i.e., fricatives produced through a constriction made by the tongue, such as [s]), the shape and size of the constriction, as well as grooving of the tongue affects (and directs) airflow (see e.g., Narayanan et al., 1995). Bracing of the tongue and contact with the palate posterior to the constriction has also been cited as important in the articulation of anterior sibilants (Howson and Redford, 2022).

The size and shape of the cavity anterior to the primary constriction can vary as a consequence of jaw position, lip protrusion, presence of a sublingual cavity and location of the constriction. Naturally, fricatives with an anterior place of articulation will have a smaller front cavity than fricatives with relatively more posterior place of articulation. Rounding of the lips will elongate the front resonating cavity while spreading them results in a smaller chamber, with higher resonant frequencies. Interdental and labiodental fricatives, such as [f, θ], have no front cavity to speak of, although the lip horn can influence the noise generated (see e.g., Johnson, 2012; Shadle, 2011). As regards lingual fricatives, the tongue can lower to close off the sublingual cavity, as in production of [s], or raise to open the cavity, as for [ɕ] (Lindblad, 1980), thus modifying the size and shape of the anterior resonating chamber. Finally, jaw position will also affect the size of the front cavity, and can change during the course of a fricative articulation.

Fricatives are hence complex and are thought to have more specific articulatory targets than other sounds, given the required precision in the articulators' position and configuration (see Howson and Redford, 2022, and references therein). Although the present description is not nearly comprehensive, it is nevertheless hopefully clear that the articulation of voiceless fricatives is a complicated business.

3.2 Acoustics of voiceless fricatives

There have been many endeavours to differentiate between voiceless fricatives through acoustic features. Generally speaking, duration, intensity and spectral features of the fricative and the fricative-vowel transitions discriminate well between fricative place of articulation and between sibilants and non-sibilants in adult speech (Behrens and Blumstein, 1988b; Forrest et al., 1988; Jongman et al., 2000; Nirgianaki, 2014). For instance, sibilants are generally longer and more intense than non-sibilant fricatives (Jongman et al., 2000).

The present exposition will focus on spectral characteristics of voiceless fricatives (although see descriptions of duration and within-speaker intensity levels for Swedish adults' voiceless fricatives in *Paper I*). First, the relationship between articulation and acoustics is reviewed, after which spectral estimation methods, spectral moments analysis and alternative approaches to spectral analysis are discussed. Then, spectral dynamics, individual differences and acoustics properties of Swedish voiceless fricatives are considered. Acoustic characteristics of children's voiceless fricatives are detailed in the following chapter (Section 4.2.4).

3.2.1 Acoustic correlates of articulation

The articulatory configurations described above determine acoustic characteristics, some of which are general for all fricatives. The fricative *manner of articulation* leads to aperiodic (noisy) spectra, with energy spanning across a large range of frequencies (from under 1000 Hz in [ɧ] to over 10 000 Hz in [s]).

More specific spectral characteristics vary with place of articulation, sibilance, vowel context, effort level and speaker-specific characteristics.

Changes in the size of the front resonating cavity will influence fricative spectra, such that enlarging the cavity (e.g., through lip-rounding, jaw lowering or presence of a sublingual cavity), will lower the resonance frequencies, while decreasing the size of the cavity (through jaw raising or closing of the sub-lingual cavity) will raise them. By and large, the frequency location of prominent spectral peaks are related to resonant frequencies of the front cavities, although resonances can be cancelled or merged with neighbouring peaks. For example, [s] before the rounded vowel [u] is known to have lower resonance frequencies than [s] before [i], likely due to anticipatory coarticulation (Johnson, 2012). The lack of front resonant cavity in labio-dental fricatives results in relatively flat spectra without pronounced peaks and valleys (cf. [f] in row one, Figure 3.1). However, see Shadle (2011) for observations on the subtle effects of the lip horn - which is, effectively, a tiny front cavity.

As the noise source of the voiceless fricative arises somewhere along the vocal tract, the turbulent noise can interact with chambers both anterior and posterior to the source/s (Shadle, 2011). These chambers have resonances and anti-resonances, which can cancel each other or give rise to pronounced peaks and troughs in the spectrum, depending on the linguo-palatal constriction area. When the constriction is small enough, the cavities are decoupled, and back cavity resonances are dampened or cancelled (Shadle, 2011). If the constriction area increases, so do the back-cavity resonances, which are often visible in the beginning and end of fricative segments (Shadle, 2011), presumably as the constriction area changes to accommodate for neighbouring sounds. Nevertheless, back cavity resonances generally have little influence on fricative spectra (Johnson, 2012).

As described above, sibilant fricatives have an additional noise source located at the front teeth, which gives rise to pronounced high frequency energy in the spectra and overall louder fricative noise (e.g., Johnson, 2012). Raising of the jaw will alter the position of the turbulence-producing obstacle (i.e., the front teeth) in relation to the jet of air coming from the lingual constriction, which can affect the noise source activation.

An increase in high frequency energy is often seen in the middle of the fricative (sometime referred to as the "steady-state" of the sound). This increase could be the result of a decreased constriction area, an increased flow velocity, raising of the jaw (in sibilants), an increased effort level, or a combination of the above (Shadle, 2011).

The random nature of the fricative noise source implies that "...for a *fixed* tract configuration and lung pressure, time series (and hence the frequency spectrum) produced by the system will be different over two intervals, despite having identical statistical properties" (Blacklock, 2004, p. 23). That is, fricatives are inherently noisy, a feature which spectral analysis must take into consideration.

3.2.2 Spectral estimation

The random fluctuations of the fricative waveform makes estimation of their spectra more complicated than for periodic sources, such as vowels (e.g., Blacklock, 2004). The present section briefly reviews different ways in which to form a spectral estimation from an acoustic waveform.

Historically, spectra have most commonly been estimated by applying the Fourier Transform (Discrete or Fast; DFT and FFT, respectively) to the acoustic waveform. However, DFT and FFT are known to have large estimate variance (i.e., large mean squared error, or poor precision, see Blacklock, 2004; Reidy, 2015).

To reduce variance, one can average DFTs across multiple productions (ensemble averaging) or individual windows within the same sound (time-averaging) (Shadle, 2023). However, such averaging comes with specific prerequisites; the signal must be stationary for time-averaging to be valid, and need be ergodic to meet the assumptions of ensemble averaging (for a discussion, see Blacklock, 2004; Shadle, 2011). As fricatives are dynamic (see Section 3.2.5) and variable across productions, neither of these requirements are met.

An alternative method to generate reduced-variance spectral estimates is through multitaper analysis. A multitaper spectrum (MTS) is generated from a single window, which is multiplied by a set of orthogonal tapers. DFTs are computed from the output of the tapers, and then averaged to produce a MTS. The MTS does not assume stationarity or ergodicity and is, therefore, more flexible than the spectral averaging procedures detailed above. Additionally, the MTS method has high resolution for time and frequency, and low error, making it is well-suited for the study of fricatives (Blacklock, 2004).

While the importance of spectral estimation has received much attention (e.g., Shadle, 2023), it has been questioned whether choice of spectral estimation method has any linguistically relevant effects. Reidy (2015) investigated whether different spectral estimation methods (DFT and MTS) influenced two spectral features; spectral centre of gravity (M1, see next section) and degree of sibilance ($AmpD_{M-LMin}$, see Koenig et al., 2013), in English sibilants /s, ʃ/. In addition, he tested whether the estimation method affected tests of linguistic contrast between the sibilants. He found that M1 was not sensitive to spectral estimator, but $AmpD_{M-LMin}$ was. Reidy (2015) suggests that "...estimation strategy is un-likely to affect conventional analyses of sibilants..." (p. EL254) given that the same linguistic contrasts (place of articulation and the interaction between gender and place of articulation) were uncovered with both methods.

The analysis of adult fricatives in *Paper I* and *II* of this thesis used FFT spectra, while MTS were created for the child productions in *Paper III*.

3.2.3 Spectral moments analysis

The most widespread manner in which to characterise fricative spectra is by spectral moments analysis (see e.g., Forrest et al., 1988; Jongman et al., 2000; Kochetov, 2017; Nirgianaki, 2014). In spectral moments analysis, the fricative spectrum is considered a probability distribution, from which four moments are derived; spectral mean (or centre of gravity; M1), standard deviation (M2), skewness (M3) and kurtosis (or peakedness; M4). Note that calculation of the moments requires a normalisation such that the area under the spectrum sums to one. For an illustration of the moments see Figure 3.2.

M1 is related to place of articulation, with lower values associated with more posterior place of articulation. As such, M1 is expected to be higher for alveolo-dental /s/ than for postalveolar /ʃ/ (Jongman et al., 2000; Nissen and Fox, 2005). However, M1 varies greatly in non-sibilant fricatives, such as glottal /h/ and interdental /θ/, due to their flat spectra, and cannot, therefore, cast light upon their place of articulation.

The second spectral moment reflects spectral diffuseness and is generally found to differentiate between sibilants and non-sibilants. For example, labio-dental /f/ has high M2 (cf. panel one of Figure 3.2), which can be interpreted as a result of the absence of a front resonating cavity. However, this observation does not hold for all non-sibilants; the spectral envelope of Swedish velar non-sibilant /ɧ/ is markedly different from the English non-sibilants upon which the antecedent descriptions are based (see panel four of Figure 3.1 and Section 3.2.7 for further discussion).

The third and fourth spectral moments are defined as representing spectral skewness and kurtosis, respectively (although see Shadle, 2023 for a more accurate rendition of what distributions the moments can describe) and are less commonly reported, although Jongman et al. (2000) showed that skewness could distinguish between all four places of articulation for English sibilants, and Nirgianaki (2014) found higher M3 for more the posterior fricatives in Greek. High M4 has also been associated with posterior (palatal and velar) place of articulation in Greek.

Despite extensive use in research on voiceless fricatives, spectral moments analysis has received harsh criticism (Holliday et al., 2010; Koenig et al., 2013; Shadle, 2023; Shadle et al., 2023), primarily because it provides a coarse-

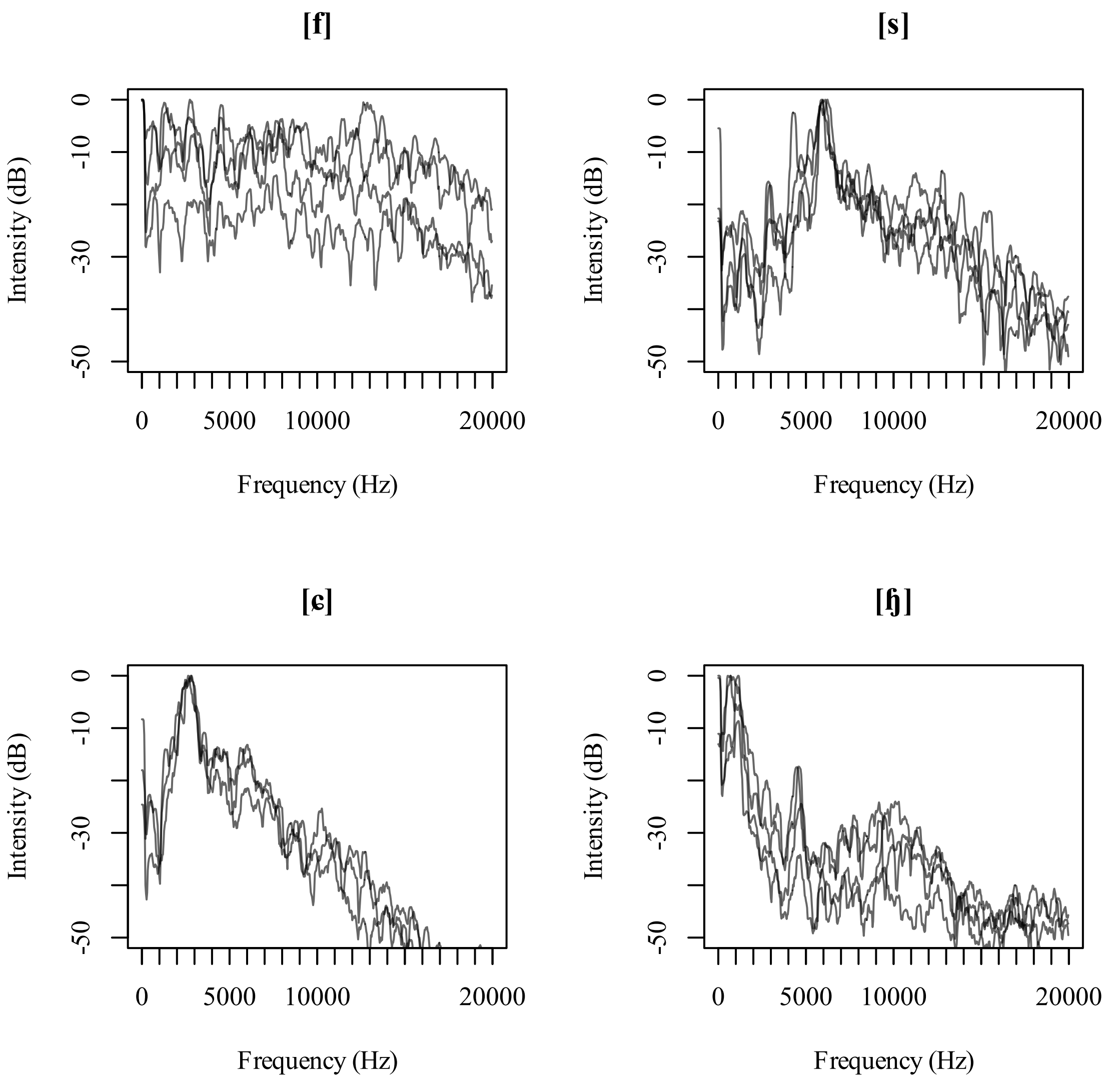


Figure 3.1: Multitaper spectra (k=8, n=4) of [f, s, ɕ, ɧ] (3-4 tokens per sound) preceding [ɑː] from one adult female speaker.

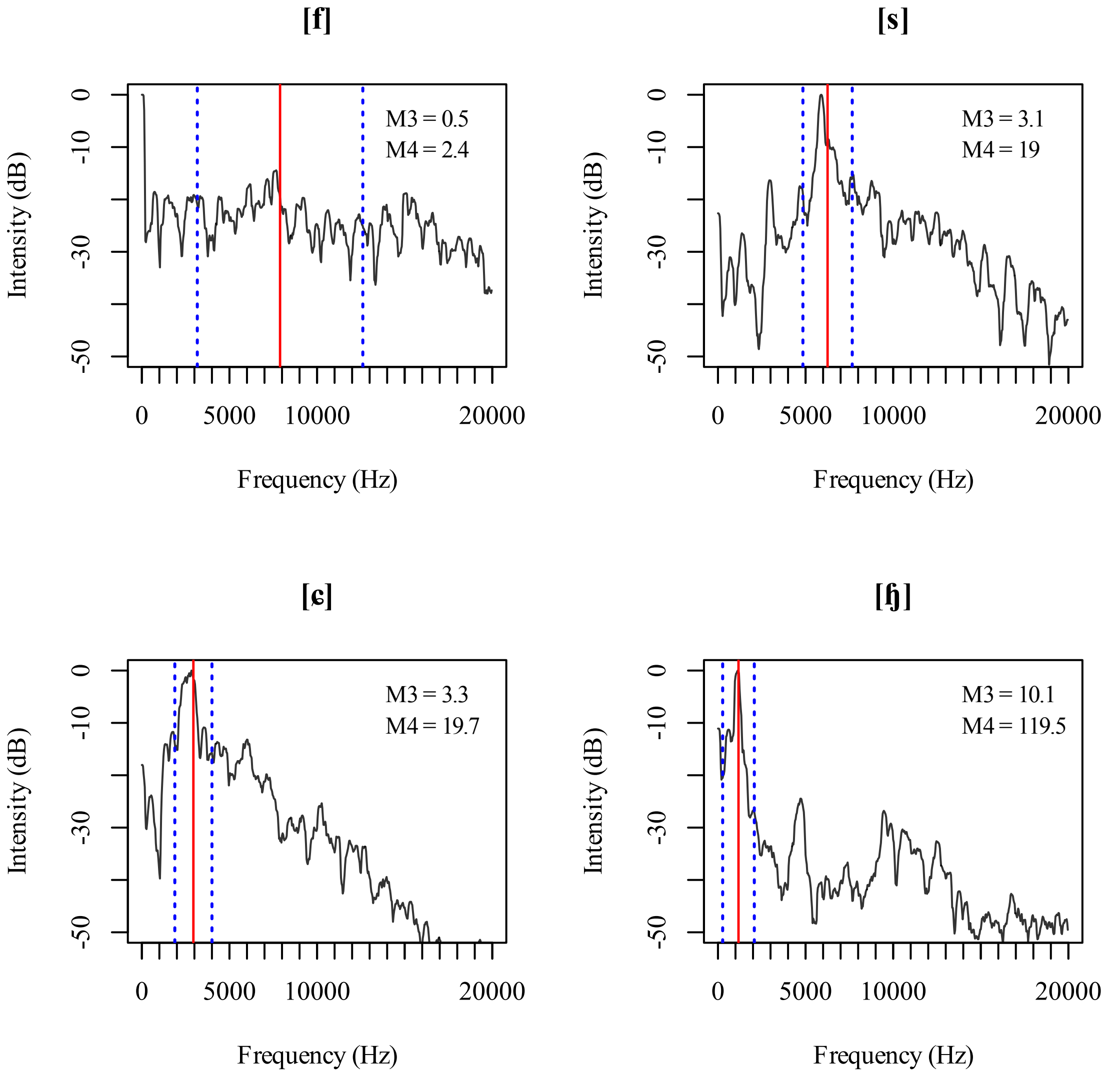


Figure 3.2: Multitaper spectra (k=8, n=4) of one production of [f, s, ɕ, ɧ] preceding [ɑː] from one adult female speaker. M1 is shown in solid red and M2 (centered around M1; M1 $\pm$ M2) in dotted blue. M3 and M4 are displayed in the right upper corner of each panel. A frequency range of 0.3-20 kHz was used.

grained description of fricative spectra, and does not faithfully reflect articulatory behaviour. For instance, although the first spectral moment (M1) does separate voiceless fricatives by place of articulation, its relation to front cavity size is ambiguous. That is, different aerodynamic and articulatory phenomena, such as stronger noise source activation and/or coupling with the back cavity can influence M1, without necessarily affecting front cavity size or constriction location. In addition, the frequency range and spectral estimation method (e.g., FFT or MTS) can affect spectral moment values (Shadle, 2023). Due to the fact that methodology differs substantially across studies, direct comparisons of spectral moments are problematic, which inhibits cross-linguistic meta-analyses. Moreover, it has been argued that the normalisation of spectra necessary for computing spectral moments can mask potentially important differences in peak and/or through amplitudes (Shadle, 2023).

3.2.4 Alternatives to spectral moments analysis

A number of alternative measures for spectral analysis have been proposed by Shadle and colleagues, based on mechanical models and acoustic studies (see e.g., Jesus and Shadle, 2002; Koenig et al., 2013; Shadle, 2023; Shadle et al., 2023). The measures were all designed to capture features of the noise source or filter. The most prominently featured filter parameter is the F_M parameter, which aims to represent the lowest uncancelled spectral resonance of the fricative. The F_M parameter is defined as the frequency of the maximum amplitude peak within a fricative-specific mid-frequency range (e.g., 3-7 kHz for English [s]) and is argued to reflect the size of the anterior resonating cavity more directly than M1 (see more in Shadle, 2023). However, the definition of the F_M parameter does not include a specification of peak prominence or amplitude, leaving some questions concerning implementation unanswered. That is to say, the algorithm proposed by Shadle and colleagues will not always choose the lowest resonance if several peaks of similar amplitude are present within the predetermined range, nor can it guarantee that the peak identified is sufficiently prominent to be considered a resonance of the front cavity in cases where there are no obvious peaks in the range.

Other spectral peaks have been documented in the analysis of fricative spectra (sometimes described as fricative formants), and the most frequently reported peak is the frequency of the maximum amplitude peak across the entire frequency range (which, as mention above, varies substantially across studies). Such *global* measures of spectral peak have been reported to vary with place of articulation (more posterior place was associated with lower peak in Jongman et al. 2000). However, global spectral peak, like the F_M parameter, is not always associated with front cavity size if spectra contain multiple high amplitude peaks.

Additionally, spectral balance measures, that is measures of the relative energy in different frequency bands in the fricative spectra, have also been advanced for analysis of voiceless fricatives (described by Shadle, 2023, as noise source parameters). For instance, Koenig et al. (2013) explored the differences between the density levels of the low (0.55-3 kHz), mid (3-7 kHz) and high (7-11 kHz) frequency ranges in their study of adolescent productions of /s/, and proposed that the measures could capture the evolution of sibilance during the production of the fricative. For a more comprehensive review of alternative spectral measures (e.g., spectral slope and peak-trough amplitude differences), the reader is referred to Shadle (2023), Koenig et al. (2013) and Holliday et al. (2010).

3.2.5 Spectral dynamics

Although fricatives have often been analysed as "steady state" phenomena, they are not static but rather dynamic in nature. Articulatory studies show that the jaw and tongue move throughout the production of fricatives, both as the constriction reaches its maximum and due to coarticulation, to and from the surrounding segments (e.g., Iskarous et al., 2011). As detailed in sections 3.1 and 3.2, movement of the tongue and jaw will affect the dual noise sources in sibilant fricatives (e.g., by changing the area of the primary constriction or the position of the front teeth in relation to the jet of air from the constriction).

Several studies have shown how spectral features vary across fricative segments (often measured in three windows from the beginning, middle and end of the fricative; Kochetov, 2017; Maniwa et al., 2009; Munson, 2004; Nirgianaki, 2014). These studies show an overall increase in M1 towards the middle of the fricative, and a decrease at the end. In a study of the time-varying relationship between articulation and acoustics in English prevocalic and preconsonantal /s/, Iskarous et al. (2011) compared x-ray microbeam data and acoustic analysis of audio recordings from 24 speakers, taken from the Wisconsin XRMB database. Results revealed that M1 (extracted from nine 30 ms evenly spaced windows) rose in the middle of the fricative. The observed increase was primarily related to jaw movement.

Language- and fricative-specific patterns in spectral dynamics were unearthed by Reidy (2016b) who explored word-initial sibilants in English (/s/, /ʃ/) and Japanese (/s/, /ɕ/), by means of a psycho-acoustic measure of spectral peak (peak ERB_N number). The peak ERB_N number trajectories were derived from 15 evenly spaced intervals along the fricative, and modeled using orthogonal polynomial growth-curve models. The findings illustrated differences in the shape of ERB_N number trajectories between sibilants across and within languages. The curvature of the trajectories was more pronounced for the anterior sibilant as compared to the posterior one in both languages, and although the overall level of ERB_N number for [s] was similar in English and Japanese, language-specific differences in slope and curvature were found.

Despite differences in methodology, it is clear from previous work that the spectral characteristics of fricatives change across their duration. In *Paper I* of this thesis, the spectral dynamics (M1 in Mel) of Swedish adults voiceless fricatives are explored, though a Generalised Additive Mixed Model.

3.2.6 Individual differences

Different people will arrive at similar acoustic output through different articulatory configurations due to anatomical differences in the vocal tract, tongue and dentition. As such, fricative *place of articulation* is a coarse articulatory description that may not reflect all individual speakers' productions (Shadle, 2011). Additionally, differences in fricative acoustics have been found related to social status, group membership, gender and sexual orientation (for a recent review of sociophonetics and fricatives, see Chappell et al., 2023). For instance, gender differences have been widely reported for voiceless fricatives, predominantly for /s/, for which female speakers reportedly have higher M1 and spectral peak than male speakers (Fox and Nissen, 2005; Jongman et al., 2000; Maniwa et al., 2009; Nittrouer et al., 1989). Such gender-related patterns in voiceless fricative acoustics are larger than expected given vocal tract differences, and are found to interact with other social variables (e.g., age and SES; Stuart-Smith 2007), suggesting involvement of socio-cultural factors.

3.2.7 Swedish voiceless fricatives

Prior to the work presented in *Paper I* and *II*, the most comprehensive description of the acoustics of Swedish voiceless fricatives was found in Lindblad's doctoral thesis from 1980. Lindblad presented detailed descriptions of perceptual, acoustic and articulatory features of Swedish voiceless fricatives, including a large variety of dialectal variants. Schematic spectra and tongue tracings were provided, obtained via spectrographic analysis (from recordings of five speakers[1]) and cinefluorographic X-ray (from two speakers). Lindblad's reference descriptions of the sounds relevant for this dissertation are presented below, in order from anterior to posterior place of articulation.

The labio-dental [f] is described as low in intensity with evenly distributed energy across the spectrum. Articulation of [f] can vary substantially without

[1]Note that the speakers produced /ɧ/ as [ʂ]

strong perceptual consequences, both with regards to the area of the labiodental constriction and the position of the teeth.

The anterior sibilant [s] is produced with a predorsoalveolar place of articulation. The tongue blade makes a narrow constriction and the tongue tip rests on the front teeth (where a second source of noise is introduced), thus closing the sublingual cavity, which leads to a small resonating cavity. The spectral energy of [s] is reportedly amassed in a broad frequency band with a sharp lower limit, around 4 kHz. The spectra often show wide peaks, and energy is present above 8 kHz (i.e., the upper cut-off in Lindblad's drawings).

The predorsoalveolar sibilant [ɕ] is produced with the tongue body raised to the palate. The tongue tip is lifted to guide the air stream to the front teeth, where the secondary noise source is produced. The sublingual cavity is described as a small pocket, and the front resonating cavity is larger than in the production of [s]. The spectrum of [ɕ] is characterised by a distinct and broad peak around 3-4 kHz preceeded by an energy trough (under 2.7 kHz).

The retroflex sibilant [ʂ] is described as an apico-alveolar fricative with a larger front cavity than that of [s] and [ɕ], which includes a larger sublingual cavity than for [ɕ]. The spectrum of [ʂ] shows a broad plateau of energy (appr. 5-6 kHz), with a peak in the lower and upper ranges (around 2.5-3.5 kHz and 7-8 kHz). All sibilants are described as perceptually intense, and both [s] and [ʂ] can exhibit a whistling quality.

The velar [ɧ][1] is described as dorso-velar, with the primary constriction produced at the front of the soft palate. The spectral representation of [ɧ] shows a clear and distinct peak around 1 kHz and small plateaus of energy at higher frequencies. The sound is described as perceptually dark and weak, but not harsh (as opposed to [x]).

Lindblad observed acoustic variability in connection with vowel context, primarily related to lip position, and noted that fricative spectra varied substantially even if spoken by same speaker in the same context. His work provides the foundation upon which this dissertation stands. However, given that his description is dated and based on a small number of speaker, a revisit with updated acoustic measures and more speakers is warranted. Note, for example, that Lindblad describes the spectra of [s, ɕ] as lacking prominent peaks (p. 66

[1]Note that choice of symbol to denote the Swedish velar voiceless fricative is not uncontroversial. The International Phonetics Association describes it as "Simultaneous ʃ and x" and Lindblad (1980) claims that some variants of /ɧ/ are produced with dual sound sources. However, the presence of double articulation has been questioned (e.g., Shosted, 2008). Moreover, because [ʂ] occurs in all dialects and is stable in postvocalic position, it is claimed to be more appropriate choice for the main allophone (Riad, 2014). In this thesis, I use /ɧ/ as symbol for the main allophone, following the IPA notation (Engstrand, 1999).

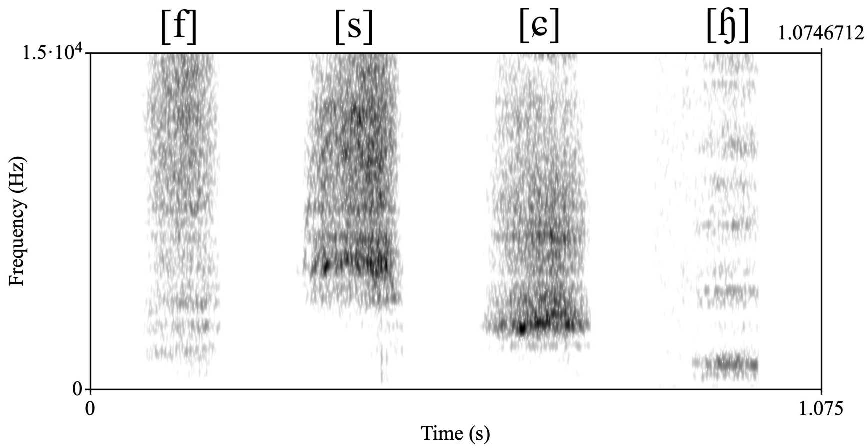


Figure 3.3: A spectrogram of [f], [s], [ɕ] and [ɧ], extracted from words in which the fricative preceded [ʉː], from one female speaker. Note that this production of the word "sjuk" [ɧʉːk] was relatively loud, chosen to illustrate the spectral difference between the voiceless fricatives. The velar was most often produced with lower intensity in relation to the other fricatives.

and 71), while the spectra from the present thesis (shown in Figure 3.1) show clear peaks for these sounds. This discrepancy is likely due to differences in recording procedure and spectral estimation.

In a more recent addition to the literature, Shosted (2008) investigated the acoustics of voiceless fricatives produced by a single Swedish speaker who produced [ɕ], [h] and allophones of the voiceless velar fricative [ʂ, ɧ] (note that the light allophone of /ɧ/ was denoted as [ʃ] in the article). The speaker produced the fricative in carrier phrases, in neutral and emphatic contexts. Results indicate that allophones [ʂ] and [ɧ] were separated by M1 when produced both with and without emphasis. Mean M1 for the anterior allophone was around 2 kHz, while mean M1 values for [ɧ] were less than 1 kHz. M1 for [ʂ] overlapped significantly with [ɕ] in the non-emphatic condition. Although emphasis increased category separability (through increase in M1 for [ɕ]), overlap between the sounds was still visible (cf. Figure 8 in Shosted 2008), which was taken to indicate that place of articulation for the two sibilants was similar.

As an illustration of the spectral differentiation between [f,s, ɕ, ɧ], see Figure 3.3, which displays spectrograms of the four fricatives, extracted from words including [ʉː]. Clear differences in energy distribution and intensity are visible for the fricative targets, which overall match descriptions form Lindblad (1980) and Shosted (2008).

Paper I and *II* provide an updated and elaborated description of the acoustic characteristics of Swedish voiceless fricatives from a larger group of speakers than previous accounts. Spectral moments 1-4 and (global) spectral peak are presented for adult speakers in *Paper I* and *II*, in addition to information concerning fricative intensity and duration in the first paper.

Summary

This chapter has attempted to provide a brief summary of the articulation and acoustics of voiceless fricatives. The noise source in fricatives is turbulent airflow, produced by a narrow constriction located somewhere along the vocal tract. Many lingual fricatives have strict articulatory requirements, making them complex to produce.

Fricative spectra are stochastic in nature, a fact which acoustic analysis needs to handle. There is ongoing debate as to what the best methods of spectral analysis are, although spectral moments analysis is currently the most prevalent method. Moreover, spectral characteristics of voiceless fricatives change across their duration and individual differences in production are profound.

As regards Swedish fricatives, an updated account of fricative characteristics is motivated given that previous descriptions are small or employ out-dated acoustic measures.

4. The acquisition of voiceless fricatives

The present chapter reviews children's acquisition of voiceless fricatives. To provide context for development related to these specific sounds, the chapter begins with a general description of speech acquisition (Section 4.1), including an overview of select milestones in speech development, and the acquisition of speech sounds, speech sound contrasts and coarticulation. The introductory section also includes a short synopsis of speech development when it does not progress as expected (more specifically, speech delay and disorder), and concludes with a brief review of Swedish children's speech acquisition. Finally, a description of children's acquisition of fricatives is presented from a general cross-linguistic perspective (Section 4.2), followed by a review of the acoustics of child productions (Section 4.2.4) and Swedish children's acquisition of voiceless fricatives (Section 4.2.5).

A note on terminology

The attentive reader will have noticed that I use "speech acquisition and - development" rather than "phonological development". In my understanding, the term phonological development involves speech production (articulation and speech planning) as well as speech perception and phonological processing. These abilities are undeniably interlinked in development, as are they linked to more general linguistic and cognitive maturation. For illustrative examples see Hearnshaw et al. (2023) regarding speech perception, speech production and vocabulary in children with and without Speech Sound Disorder (SSD), and Stoel-Gammon (2011) on the relationship between early lexical and phonological development. However, as the studies presented in this dissertation focus solely on speech output and how the output is perceived by adult listeners, the present chapter is restricted to children's speech production.

4.1 General speech acquisition

Speech acquisition is characterised by substantial variation within and across individuals, social groups and languages. Nevertheless, some developmental milestones seem robust to many sources of variation. This section presents three developmental milestones in speech acquisition, as well as a number of potentially influential individual and environmental factors that are cited as influential for speech acquisition. This abbreviated rendition aims to illustrate the complexity of speech acquisition, but is not exhaustive. References for further reading are provided throughout.

4.1.1 Developmental milestones

To provide context for the coming discussion, and stake out general trends in development, three important milestones in speech acquisition are reviewed: canonical babbling, the emergence of first words and speech intelligibility.

- *Canonical babbling*
  Children come into this word making sounds, and in the first months of life progress from reflexive cries and shrieks, through voluntary coos of comfort, vowel- and consonant-like sounds to speech-like sound combinations. Around 6 to 8 months, canonical babbling, or the production of canonical syllables (consonant-vowel combinations with transitions of adult-like speed), such as [ba], [bababa] and [badiga], begin to appear. When children reach 10 months of age, a majority will produce canonical babbling (e.g., Cychosz et al., 2021a; Oller et al., 1999). There are striking cross-linguistic similarities regarding both age of onset and the sound repertoire used in canonical babbling, with nasals /m, n/ and anterior stops /t, d, p, b/ frequently occurring (see Morgan and Wren, 2018, and references therein). The fact that onset of canonical babbling is similar in different languages and cultural contexts (e.g., Cychosz et al., 2021a) and delayed onset of canonical babbling is associated with later vocabulary acquisition and risk of speech- and/or language disorder (e.g., Lohmander et al., 2017a; Morgan and Wren, 2018; Oller et al., 1999), makes babbling an important developmental milestone, and a precursor to spoken language.

- *First words*
  Around a child's first birthday her first words appear, signifying an important step in deciphering the speech code (Kuhl, 2004). The transition between babbling and words is smooth, and early words are often not distinguishable from a child's canonical babbling. The first words are produced with the same sounds and structures as babbling (McCune and Vihman, 2001; Stoel-Gammon, 2011), but are differentiated through symbolism: words have clear and consistent referents whereas babbling sequences do not. Nevertheless, determining what constitutes a word

is not easy. Stoel-Gammon (2011) states that "...the child's phonetic form(s) must be systematically linked with the context(s)" (p. 3) to be considered a word, all the while acknowledging that the child's production may differ significantly from the adult target. The restricted sound repertoire of the year-old child leads to many homonymous productions, further complicating the picture. The onset of first words is nevertheless robustly documented around 12 months across languages, and is a significant speech-language milestone as it represents the transition from pre-linguistic vocalisations to meaningful speech.

- *Speech intelligibility*
  During the second year of life, children begin to combine words and gestures, reaching the *two-word stage* around their second birthday (e.g., Berk and Lillo-Martin, 2012). The ability to combine linguistic units leads to an exponential increase of the child's linguistic repertoire, and constitutes a speech-language milestone in its own right. Once children reach this stage it can be difficult to separate effects of different linguistic abilities such as morpho-syntactic, prosodic or pragmatic skills from speech accuracy, in appraisals of functional communication. However, *intelligibility of speech* is often described as a key measure of functional speech communication (Lagerberg et al., 2021). Intelligibility expresses how much of the speakers message is understood by the listener (Weismer, 2008), and is, thus, an amalgamation of speech characteristics and listener experience, expectations and adaptation. Indeed, adults are very good at adapting to the children that they know/spend time with (see e.g., Yu et al., 2023), which is perhaps why developmental norms for speech intelligibility are often based on conversations with strangers (e.g., McLeod et al., 2012). Traditionally, intelligibility of speech has been reported to be around 50% at two years of age, and 100% at four (Coplan and Gleason, 1988). However, a recent large scale exploration of intelligibility in isolated words and in connected speech (Hustad et al., 2021) revealed larger variation than was previously described. Despite individual differences, speech intelligibility was at least 50% by 48 months (Hustad et al., 2021). Given that the main point of speech is communication, reaching speech intelligibility levels that enable conversations with unfamiliar interlocutors is a noteworthy milestone in speech development.

4.1.2 Speech sounds: age and order of acquisition

Norms for speech acquisition are often based on analysis of the perceived accuracy of individual sounds, most often consonants. Overall, some sounds seem to be acquired early and easily, while others are acquired later and with increased effort. The reasons for differences in order and age of acquisition (OoA and AoA, respectively) are much debated and the formal criteria for "acquiring", "mastering" or "establishing" a speech sound vary across studies. For instance, criteria for AoA differ such that some studies apply a 90% criterion (i.e., the sound needs to be produced correctly in 90% of given opportunities) while other use 50% or 75% for AoA norms (McLeod and Crowe, 2018).

Generally speaking, vowels are established early (around three years of age; for an overview, see Stoel-Gammon and Pollock 2008), and children will have acquired most consonants of their ambient language(s) by age five (McLeod and Crowe, 2018). However, children are not expected to consistently produce all sounds correctly until years later (e.g., Smit et al., 1990).

Language universal tendencies in AoA are widely attested (e.g., Edwards and Beckman, 2008b; International Expert Panel on Multilingual Children's Speech, 2012), many of which can be understood in light of physiological constraints on speech production, and speech motor development. By and large, more articulatorily complex sounds are established later than sounds that require simpler articulatory configurations. For instance, the sounds of babbling (e.g., /b, m/) are simple and require no separation of tongue and jaw, which is natural given early constraints on the coordination of articulators (Green et al., 2000; Namasivayam et al., 2020).

By contrast, consonants articulated with the tongue (lingual consonants) are often acquired later. Indeed, acquiring motor control and coordination of the tongue is infamously challenging (Kent, 1992; Namasivayam et al., 2020) and on the whole, children tend to start out with less specific and more variable tongue gestures as compared to adults (Gibbon, 1999). These undifferentiated lingual gestures involve increased tongue-palate contact and lack of separation between tongue tip, body and dorsum for lingual consonant targets. Hence, the acquisition of sounds that demand more complex tongue configurations, such as rhotics and lingual fricatives, is protracted across typologically varied languages.

Language-specific patterns in OoA and AoA are also well documented and the phonological structure of the ambient language(s) is known to affect the order and age at which children acquire sounds and structures (e.g., Edwards and Beckman, 2008b; Munson, 2001; Stokes and Surendran, 2005). For example, in an investigation of word-initial obstruents in two to three-year old children learning Greek, Cantonese, Japanese or American English, Edwards and Beckman (2008b) found that both phoneme frequency and phoneme sequence frequency (i.e., phonotactic probability) influenced the children's speech accuracy. Such typological tendencies can be found as early as in the babbling stages (see e.g., de Boysson-Bardies and Vihman, 1991), although cross-language differences are amplified with age.

On the whole, language-specific frequency effects interact with universal physiological constraints on speech production (and perception). However, due to large individual variability, normative data on speech sound acquisition should be considered with care as the patterns described (if reported without quantification of variation) may not reflect the development of any individual child. Different language- and speaker-specific developmental trajectories can thus be expected, which is why it is important to quantify the boundaries of variation in typical speech development across languages.

Overt and covert contrasts

It is now widely accepted that speech sound acquisition is gradual, and acquiring phonetic contrasts involves gradient differentiation between contrasting sounds. That is, children gradually converge on the articulatory configurations to produce the acoustic output that is perceived as correct and *overt contrast* between perceptually similar sounds emerges. Children may produce systematic distinctions between contrasting sounds before the divergence is readily perceived; "... the point at which the child starts to articulate a detectable contrast and the point at which the child provides useable cues to the contrast for the speech community or transcriber need not be synchronous." (Scobbie et al., 1996, p.44). Such *covert contrasts* have been found in children's productions of sounds differentiated through place of articulation (Baum and McNutt, 1990; Li et al., 2009; Munson et al., 2010; Strömbergsson et al., 2015) and voicing (Macken and Barton, 1980). Covert contrasts can be captured with acoustic analysis (inasmuch as the appropriate acoustic cues are chosen) and through perceptual analysis, given fine-grained rating tasks, but "...gradient acquisition is difficult to observe if we rely solely on phonetic transcription." (Edwards et al., 2011, p. 38). Nevertheless, the vast majority of normative studies of speech acquisition are based on transcription of child speech.

The development of speech sound contrast is related to age, but is not necessarily linear. Young children tend to produce less differentiated sound categories, which gradually diverge. Once they reach overt contrast they may continue use inappropriate or exaggerated cues (described as "immature contrast" in Scobbie et al. 1996), before they settle on less, but sufficiently, distinctive productions. Hence, although most speech sounds are acquired by the time children enroll in school (i.e., around six years of age, McLeod and Crowe, 2018), contrasting sounds continue to diverge past the point at which they are transcribed as correct - a *covert fine-tuning*, if you will.

As such, children will often differentiate between sounds before contrast is evident in transcription, and continue to refine their production after the sounds are transcribed as correct.

Coarticulation and speech in context

Needless to say, speech is not a neat and linear stream of sounds with interjected pauses or clear boundaries. Speech is flowing and variable, with speech sounds influencing and being influenced by their neighbours. Accordingly, children learn sounds in context. *Coarticulation*, or overlap between speech gestures, can manifest through articulatory modifications in anticipation of the following sound (anticipatory coarticulation) or as adjustments that persist from the previous sound (perseveratory coarticulation). Although descriptions diverge, most studies reveal more coarticulation in young children's speech, as compared to adults (e.g., Nittrouer et al., 1989; Noiray et al., 2018; Zharkova et al., 2011). That is, children's sound productions harmonise more strongly with surrounding sounds than do adults'. Degree of coarticulation seems to be related to age and experience as coarticulation has been found to decrease with age (Zharkova et al., 2011) and children who vocalise more, coarticulate less (Cychosz et al., 2021b).

Segment-specific effects on coarticulation have been found, such that consonants with higher articulatory demands on the tongue exhibit more coarticulation (see e.g., Zharkova, 2018, who investigated coarticulation via ultrasound in children between three to thirteen years olds) and certain sounds are described as particularly resistant to coarticulation (e.g., voiceless fricatives; Howson and Redford, 2022). As such, the phonetic context is important, and contextual constraints on speech development are evident in the later acquisition of consonant clusters as compared to their constituent sounds in isolation, and in word position effects on speech sound accuracy (e.g., Fox-Boyer et al., 2022; Lohmander et al., 2017b; McLeod and Masso, 2019). Nevertheless, much previous research has taken a *phonemic perspective* (see Howson and Redford, 2022) on speech acquisition, thus disregarding positional and coarticulatory effects on speech sound accuracy.

4.1.3 Individual differences and environmental influences

The extant literature has revealed substantial individual differences in speech acquisition (e.g., Kidd and Donnelly, 2020; Vihman et al., 1986) and although much of the observed variation seems idiosyncratic, demographic features such as age, sex and/or gender, socio-economic status and multilingualism are often discussed as factors in speech acquisition.

Age is, in all likelihood, the most well studied demographic component and speech norms are often presented in relation to chronological age (c.f.

AoA in Section 4.1.2). Clear increases in overall speech accuracy and intelligibility are found with increasing age (e.g., Hustad et al., 2021; McLeod and Crowe, 2018), as could be expected given gains in speech motor control, cognitive maturation and experience. However, large variability across same-age peers has been attested (e.g., *Paper III*) and the division of children into groups based on chronological age (years or months) is somewhat arbitrary.

With respect to gender, boys' speech and language development is often described as slower than girls (on group level), but differences are rather small and individual variation sizeable (see, e.g., Rinaldi et al. 2023). Nevertheless, male gender has been identified as a risk factor for speech disorder Wren et al. (2016). A number of recent perceptual experiments indicate that gendered speech emerges as early as 2:6 years (Fung et al., 2021; Munson et al., 2022). For instance, Munson et al. (2022) explored adults perceptions of child speaker gender, and revealed differences in gender-typicality ratings of speech produced by children who were assigned male at birth and those assigned female at birth. The difference was found at 2:6-3:6 years, but was more pronounced at 4:6-5:6 years. Differences in perceived gender typicality have also been linked to the child's gender identity (e.g., Munson et al., 2015), evidencing the involvement of a social component in the development of gendered speech.

In addition to individual factors, a plenitude of environmental components are known to influence speech acquisition. Socio-Economic Status (SES) of the child's caregivers has been extensively studied with respect to language acquisition and child directed speech (e.g., Schwab and Lew-Williams, 2016), and low SES has been identified as a predictor of speech disorder (e.g., Wren et al., 2016). High SES is often reported to be associated with higher scores on speech and language tests, as compared to children with low SES backgrounds (see e.g., Pace et al., 2017). However, cross-study comparisons are hindered by differences in the quantification of SES (e.g., parental education or income) and the over-representation of high SES families in normative speech studies make estimates of variation difficult.

Multilingualism is also often discussed as a factor in speech development (e.g., Fabiano-Smith and Goldstein, 2010; Goldstein and McLeod, 2012; Hambly et al., 2013) and early studies showed protracted speech development for multilingual children as compared to their peers. However, recent reviews of multilingual speech acquisition paint a more complex picture; in comparison to monolingual peers, multilingual children can present with both more advanced and less advanced speech skills (see Goldstein and McLeod, 2012; Hambly et al., 2013; Unsworth, 2013). Exposure (both quantity and quality) to the languages is critical, and speech sound skills tend to be more advanced in the child's dominant language. Although there do not seem to be clear differences in rate of acquisition, qualitative differences between mono- and multilingual children are discernible, and transfer (i.e., cross-linguistic influence) between the languages is often present (Hambly et al., 2013; Unsworth, 2013). Nonetheless, there is a scarcity of knowledge concerning multilingual speech-language development, particularly for minority languages, which can lead to consequences for the care of multilingual children (e.g., delayed referral to speech-language pathology services; Hambly et al., 2013; Nayeb et al., 2015).

Despite discussing these factors in separation above, an individual child is more than the sum of their demographics. In a recent cross-continental study, (Bergelson et al., 2023) explored the influence of a number of factors on early child speech (amount, not content or form). They found no predictive effect of SES, gender or multilingualism on child vocalisations, although adult talk and child age were found to be predictive of amount of child vocalisations.

Moreover, much research on speech and language acquisition is conducted in Western, Educated, Industrialized, Rich and Democratic (W.E.I.R.D; Henrich et al., 2010) settings. As such, our understanding of the influence of environmental factors is largely restricted to a select privileged group. Hence, the source of speaker-specific variation is not always known, and what is true for a group may not hold for the individual child. What we do know, however, is that children's speech varies on many levels. Understanding both individual variability and group-level trends is important for theoretical accounts of speech acquisition as well as for identification of children who are at risk of developing speech- and/or language disorders.

4.1.4 Speech delay and disorder

For most children, acquiring speech is a gradual process that follows a general developmental trajectory akin to the path described in Section 4.1. However, for some the acquisition of speech is protracted. Deciding when a child's speech difficulties are to be regarded as a delay or a disorder is not trivial and is, in the clinic, parallel to the question of who to "watch and wait" and who to refer to speech intervention.

Speech sound disorder (SSD) is an umbrella term for speech impairments, that is widely used in English speaking countries [1]. There has been significant debate regarding terminology and differential diagnosis of SSD (see e.g., Lit-

[1] SSD is not currently widely used in a Swedish context. The Swedish terminology to describe children with speech impairments of unknown origin primarily differentiates between *articulation disorder* (Swe: artikulationsstörning) and *phonological language impairment* (Swe: fonologisk språkstörning). See Wikse Barrow et al. (2023a) for a discussion.

tlejohn and Maas, 2023; Stringer et al., 2023; Waring and Knight, 2013, for discussions), and children included under the SSD umbrella are highly heterogeneous regarding speech output, speech perception abilities, and involvement of other linguistic faculties. Nevertheless, all children with SSD produce speech that does not match expectations given the child's experience and developmental level.

It has been suggested that different types of speech errors[1] can aid in differentiating delay from disorder. All children make errors in speech production during development, if we accept the definition of an error as a deviation from the adult target form. That is, adults are expected to produce the word "sun" as [sʌn], and if the child says [tʌn] - the initial sound is produced in error. During the first few years of life, these errors are common and gradually decrease with age and experience. Early speech errors will often affect syllables or word structure, for instance through reduplication of a CV-syllable or omission of stressed syllables (Nettelbladt and Salameh, 2007; Vihman and Croft, 2007). As children get older, errors involving single segments grow in frequency, for example through deletion of word-final consonants, consonant cluster reductions, substitution errors (e.g., fronting; [t] for /k/, or stopping; [t] for /s/) or distortions (i.e., a production that is audibly inaccurate, but cannot be described using another of the languages phonemes) (e.g., Nettelbladt and Salameh, 2007). Speech errors often pattern in systematic ways (e.g., /s/ is always realised as [t]), but can also be inconsistent within the same child.

In the context of SSD assessment, distinctions are often made between "typical" vs. "atypical" and "consistent" vs. "inconsistent" speech errors. For example, Dodd (2014) proposes that children who exhibit typical speech error patterns (i.e., patterns akin to those found in younger children with typical speech development) could be regarded as having speech delay, while children who display atypical speech error patterns (often in combination with an increased amount of typical speech errors) could be classified as having speech disorder. Inconsistent speech errors are unusual in typical speech development (around 80% of words produced by children over three years of age are consistent across multiple repetitions; Holm et al. 2023 and references in Dodd et al. 2023) but are cited as a key feature of specific SSD subtypes; inconsistent phonological disorder (Dodd, 2014; Dodd et al., 2023) and childhood apraxia of speech (Iuzzini-Seigel et al., 2017) [2].

[1] How to classify and interpret developmental speech errors has been subject of much debate, and terminology varies substantially. In this thesis the term *speech error* is used to describe cases in which an observer perceives a mismatch between the child's production and the adult target.

[2] Note that speech inconsistency can refer to either token-to-token inconsistency, as described above, or to phonemic inconsistency, that is, inconsistent realisation of a

Regardless of differences in speech characteristics, SSD will often lead to functional consequences for the child, and some children will go on to have long-standing academic and/or social difficulties (Daniel and McLeod, 2017; McLeod et al., 2013; Tambyraja et al., 2020; Wren et al., 2023, 2021). Understanding how listeners perceive child speech can therefore aid in our understanding of the everyday significance of the disorder (see Chapter 5 for a brief review of perceptual analysis of voiceless fricatives).

4.1.5 Swedish children's speech development

With a background in general trends, focus now moves to Swedish children's speech acquisition. In view of the current context, this review will focus on acquisition of consonants. A summary of AoA for Swedish consonants is presented in Table 4.1, which shows that the stops /p, b, t, d, k/, nasals /m, n/, approximants /i, j/ as well as the labiodental and glottal fricatives /f, h/ are acquired by three years of age. Note that different criteria for AoA were used for the two samples upon which the description is based. With a 50% AoA criterion, the voiced fricative /v/ and the voiced plosive /ɡ/ are also established at three (Lohmander et al., 2017b), while higher AoA was found by [][four and five years, respectively]Lundeborg-Hammarstrom:2019aa, who used an AoA criterion of 90%. The voiceless fricatives /ɕ/ and /ɧ/ are reportedly acquired by the age of five, while the anterior sibilant /s/ continues to be elusive even at six years of age (if adhering to the higher AoA criterion, see Table 4.1). Similarly, the rhotic /ɹ/ looks to be established at five, given the lower criterion, but else not until six years of age. Keep in mind that the two speech material upon which this description rests (the phonological test LINköpingsUnderSökningen, LINUS; Lundeborg Hammarström 2019, and the Swedish Nasality and Articulation Test, SVANTE; Lohmander et al. 2017b) differ with respect to number of words included and consequently the number of opportunities to produce each sound. Note also that the target consonants were presented in a variety of word positions in both tests, and AoA is therefore not position-specific. Nevertheless, Lohmander et al. (2017b) explicitly states that consonants in medial and final position were established to a notably higher degree than were word-initial consonants. Altogether, average speech accuracy, expressed in percentage of consonants correct (PCC; Shriberg and Kwiatkowski, 1982), was 82 (SD = 14.26) for three-year-olds and 96.3 (SD = 5.44) for five-year olds in Lohmander et al. (2017b), with broad ranges (40-100 and 83-100, respectively) displaying large within-age group variability.

The LINUS manual also includes a description of a number of speech error patterns, which reveals that over 80% of children displayed cluster reduction

phoneme across different phonetic contexts.

Table 4.1: A summary of consonant AoA, from LINUS (Lundeborg Hammarström, 2019) at three, four, five and six years and from SVANTE (Lohmander et al., 2017b) at three, five and seven years. Consonants are presented at the age at which they are established, based on the predetermined AoA criteria (90% in LINUS and 50% in SVANTE).

	Sounds acquired (accumulative)				
	3 yrs	4 yrs	5 yrs	6 yrs	7yrs
LINUS	p, b, t, d, m, n, h, j, l, f, k	v	ɡ, ŋ, ɕ, ɧ	ɹ, s**	-
SVANTE*	p, b, t, d, m, n, v, h, j, l, f, k, ɡ	-	ɕ, ɹ, s	-	all consonants

*ɧ and ŋ not included **88%

at 36-41 months. A steady decline of occurrence was visible with increasing age (16,5% of children at 72 months) and similar findings were presented for omission of singleton consonants (from 79% at three to 22% at six). Omissions of pre-stressed syllables, omission of final consonants and epenthesis were all relatively common in three-year olds (46%, 27% and 56%, respectively), and relatively rare in the oldest age group (over 72 months; 1%, 10% and 6%, respectively). Unfortunately, these broad descriptions provide no further details regarding what target sounds or word positions are affected.

In two recent unpublished master theses from the speech-language pathology program at Karolinska Institutet, speech errors in Swedish three-year-old and five-year-old children were chronicled in more detail. The results are based on a reanalysis of recordings collected for the SVANTE norms (Lohmander et al., 2017b). For three year old children (n=30), omission of sounds was relatively common (23% of observed errors), primarily for /ɹ/ in word-final position and /s/ in word-initial position (Palo, 2022). Weakening/gliding (i.e., substitutions of [j] or [l] for /ɹ/), comprised 17.8 % of the observed errors, and substitutions of /s/ were also common, including both fronting, backing and stopping errors. Cluster reduction made up 8.8% of all errors in the three-year-old sample. In five year old children (n=30), the majority of speech errors observed were omissions or weakening of /ɹ/ in word-final position, and inderdental realisations of /s/ in word-initial position (Göthlin, 2022). Substitutions of [j] for /ɹ/ occurred in eight of 30 children, and eleven children produced [θ] for /s/. Large individual differences were revealed, as six of the 30 children produced 49% of the total speech errors.

In summary, we know the age and order of acquisition for Swedish sounds (although not specific to word position) and have recently gained knowledge concerning details about some developmental speech errors. Less is known about individual variation, development of contrast, and speech in challenging contexts.

4.2 Children's acquisition of voiceless fricatives

Voiceless fricatives are articulatorily demanding, relatively late-acquired and are often the subject of residual speech errors (i.e., speech errors that persevere beyond 8 years of age; Flipsen Jr 2015; Shields and Hopf 2023; Veríssimo et al. 2012). In the present section, a general description of cross-linguistic trends in fricative acquisition is provided (Section 4.2.1), followed by a review of acoustic analysis of children's voiceless fricatives (Section 4.2.4), including brief audits of fricative contrast and gender differences. Finally, Swedish children's acquisition of voiceless fricatives is described, in Section 4.2.5).

4.2.1 Cross-linguistic similarities and differences

As alluded in the general description of speech acquisition above, voiceless sibilant fricatives, such as /s, ʃ/, are late-emerging across many typologically diverse languages. The overall later emergence of sibilant fricatives is unsurprising, in view of their articulatory complexity; their production requires a narrow and precise lingual constriction, which is demanding to control (see Section 3.1). As for non-sibilants, labio-dental [f] and glottal [h] are often acquired earlier, although /θ/ is the last fricative to be established in English-acquiring children (McLeod and Crowe, 2018).

In a cross-linguistic study of consonant acquisition across 27 languages, mean age of acquisition (90%) for /f, ʂ, ɕ, x, h/ was found to be 3;0-3;11, and 4;0-4;11 for /s, ʃ/ (McLeod and Crowe, 2018). As mentioned in Section 4.1.2, the phonological structure of the ambient language(s) affects the order and age of speech sound acquisition, and voiceless fricatives are no exception. Language-specific developmental trajectories of voiceless fricative acquisition are thus to be expected. For example, Polish and Putonghua have three voiceless sibilants each, but Polish order of acquisition is /ɕ/>/s/>/ʂ/ (Żygis et al., 2023), while the sounds are established in the sequence /ɕ/>/ʂ/>/s/ in Putonghua (Li and Munson, 2016). The observed patterns reflect the phoneme frequency of the languages.

Language-specific patterns are also evident in developmental speech errors involving voiceless fricatives. Japanese and English each have two sibilant fricatives (/s, ɕ/ and /s, ʃ/ respectively) but commonly observed substitutions diverge, such that Japanese-acquiring children often produce posterior [ɕ] for target /s/ while English-acquiring children are prone to substitute anterior [s] for /ʃ/ (Li et al., 2009).

4.2.2 Contrast

As is true for other speech sounds, contrast between voiceless fricatives develops gradually, and numerous perceptual studies have uncovered fine phonetic detail (or *covert contrast*) in children's voiceless fricatives using gradient rating scales (e.g., Holliday et al., 2010; Munson et al., 2010, and Chapter 5). Acoustic analysis has revealed covert contrasts between children's productions of English /s/ and /θ/ (Baum and McNutt, 1990), and English and Japanese children's productions of sibilant fricatives (/s, ʃ/ and /s, ɕ/, respectively; Li et al. 2009). Articulatory studies of Scottish English sibilants have similarly showed "... some evidence of articulatory contrast in production preceding contrast in perception." in three-year-olds speech (Zharkova, 2021, p.1126).

As an example of developing acoustic contrast, Figure 4.1, shows MT spectra of two Swedish children's productions of /s/ and /ɕ/ before [uː]. The three-year-old child's productions (left panel) were transcribed as [pθ] or [s/θ], regardless of target fricative, while the four-year-old's productions were transcribed as correct (i.e., [s] for /s/ and [ɕ] for /ɕ/). The spectra of the younger child's productions of the two sibilants show dispersed spectral energy and a lack of acoustic differentiation between the target sounds (i.e., neutralised contrast). The older child's fricatives have more pronounced peaks and troughs. The two sibilant targets have different peak frequency locations, and spectral contrast between the fricatives is visible.

In a innovative study on the development of contrast between English sibilants /s/-/ʃ/, Holliday et al. (2015) explored the acoustic overlap between the word-initial sibilant targets (measured by M1) in two- to five-year old children. Contrast robustness was calculated as the proportion of a given child's productions of a sibilant target (i.e., /s/ or /ɕ/) that was correctly predicted by a mixed-effects logistic regression built on all children's (sibilant) productions. The productions (more specifically, the extracted initial CV-syllables) transcribed as correct were also presented to adult listeners so that the contrast robustness measure could be related to category goodness ratings. The acoustic contrast between sibilants was found to increase with age and was correlated with perceived category goodness.

In a study of older children (9-14 years of age), preceding that of Holliday et al. (2015), Romeo et al. (2013) investigated within-fricative dispersion, between-fricative overlap and overall discriminability for word-initial English sibilants /s/-/ʃ/. The measures were all based on M1. Results showed no age-related decrease in dispersion, although between-category distance did increase with age. Thirteen-year-old children had adult-like overall discriminability, although their sibilant productions had larger between-fricative distance and were more disperse than those of adult speakers. The authors hy-

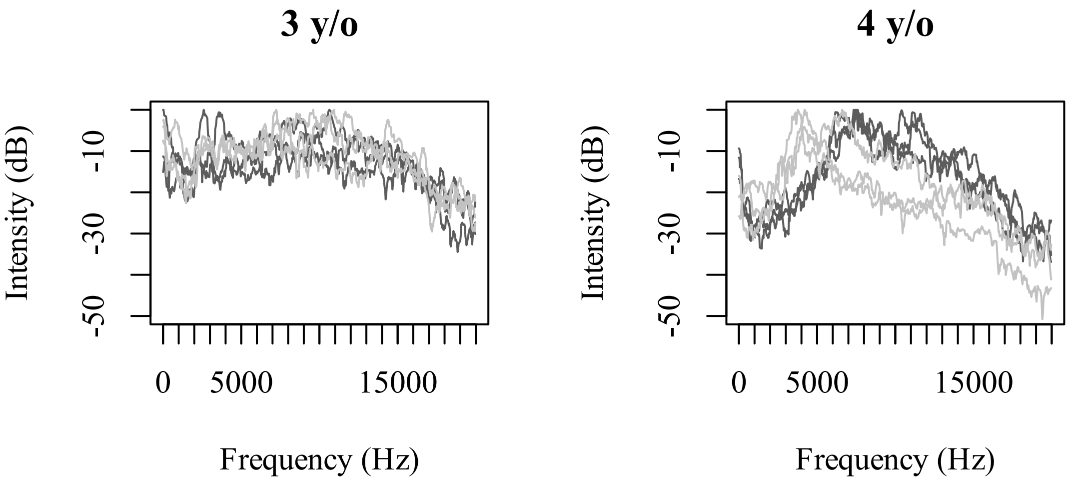


Figure 4.1: Multitaper spectra (k=8, n=4) of Swedish target sibilants extracted from words in which the fricative preceded [uː]. Productions from a three-year-old girl are displayed in the left panel (n=6), and productions from a four-year-old girl in the right panel (n=6). Target /s/ is shown in black and /ɕ/ in grey. The three-year-old's productions were transcribed as [pθ] or [s/θ] for target /s/ and [s/θ] for target /ɕ/. All the four-year-old's productions were transcribed as correct. Both children were recorded in a laboratory setting.

pothesise that category structure continues to be refined after the studied age.

4.2.3 Gender differences

Analogous to adults, there is evidence of gender specific patterns in children's voiceless fricative productions (see e.g., Bang et al., 2017; Fox and Nissen, 2005; Li, 2017; Li et al., 2016; Romeo et al., 2013). Specifically, /s/ produced by girls tend to have higher M1 than those produced by boys, and larger between-fricative contrast has been observed for girls (Holliday et al., 2010; Romeo et al., 2013). Li et al. (2016) showed that sex/gender differences in English /s/ are not directly associated with the children's height (as proxy for vocal tract size), which evidences a socio-cultural element in play.

4.2.4 Acoustic analysis of children's voiceless fricatives

Many of the acoustic analyses described in Section 3.2 above have been used to analyse fricatives produced by children as well as adults. However, the application of measures derived from adult speech to child speech is not without issue, because children are more variable than adults, both within and across individual speakers (Koenig et al., 2008; Lee et al., 1999; Munson, 2004), and may be highly inconsistent with respect to fricative realisation. For instance, see Figure 4.2 which shows productions of sibilants [s, ɕ] preceding [ɑː] from one adult and one child speaker. Although all productions were transcribed as correct, the child is more variable across repetitions of the same word and the adult shows more compact and pronounced spectral peaks and troughs than the child. See also Figure 4.3 for examples of within-fricative variability in children's productions.

Children may use other cues than those employed by adults to convey contrast between fricatives, or produce less distinct contrasts, with more acoustic overlap, as compared to adult speakers. Moreover, children may exhibit different coarticulatory patterns, and display lengthened or disrupted consonant-vowel transitions, thus hindering annotation and analysis of fricative duration. Additionally, as children's voiceless fricatives have higher resonant frequencies than those produced by adults, the choice of frequency range for spectral measures of child speech requires consideration.

Notwithstanding these challenges, acoustic analysis offers an objective description of children's speech which can supplement perceptual analysis, such as transcription. Moreover, as some acoustic parameters have clear articulatory correlates (e.g. the second vowel formant for tongue advancement), they enable an indirect study of speech motor maturation (see e.g. Li and Munson, 2016).

Spectral analysis: methods and challenges

The vast majority of previous work concerning the acoustics of children's voiceless fricatives has employed spectral moments analysis (with or without additional acoustic measures, see e.g., Bang et al. 2017; Holliday et al. 2015; Howson and Redford 2022; Kehoe and Philippart de Foy 2023; Li et al. 2009; Li and Munson 2016; Luo et al. 2023; Miodonska et al. 2022; Nissen and Fox 2005; Zharkova 2021; Żygis et al. 2023 and the studies described above). For example, Nissen and Fox (2005) explored the development of voiceless fricatives in three to six-year-old American English children, and found that differences in spectral variance (M2) was the only measure to differentiate between the four places of articulation in child productions. Li and Munson (2016) showed that children acquiring Putonghua used both M1 and F2 onset to distinguish [s, ɕ, ʂ], and Żygis et al. (2023) similarly found that the same parameters were most influential in distinguishing between Polish-speaking children's word-initial sibilants. Many studies report on several moments, al-

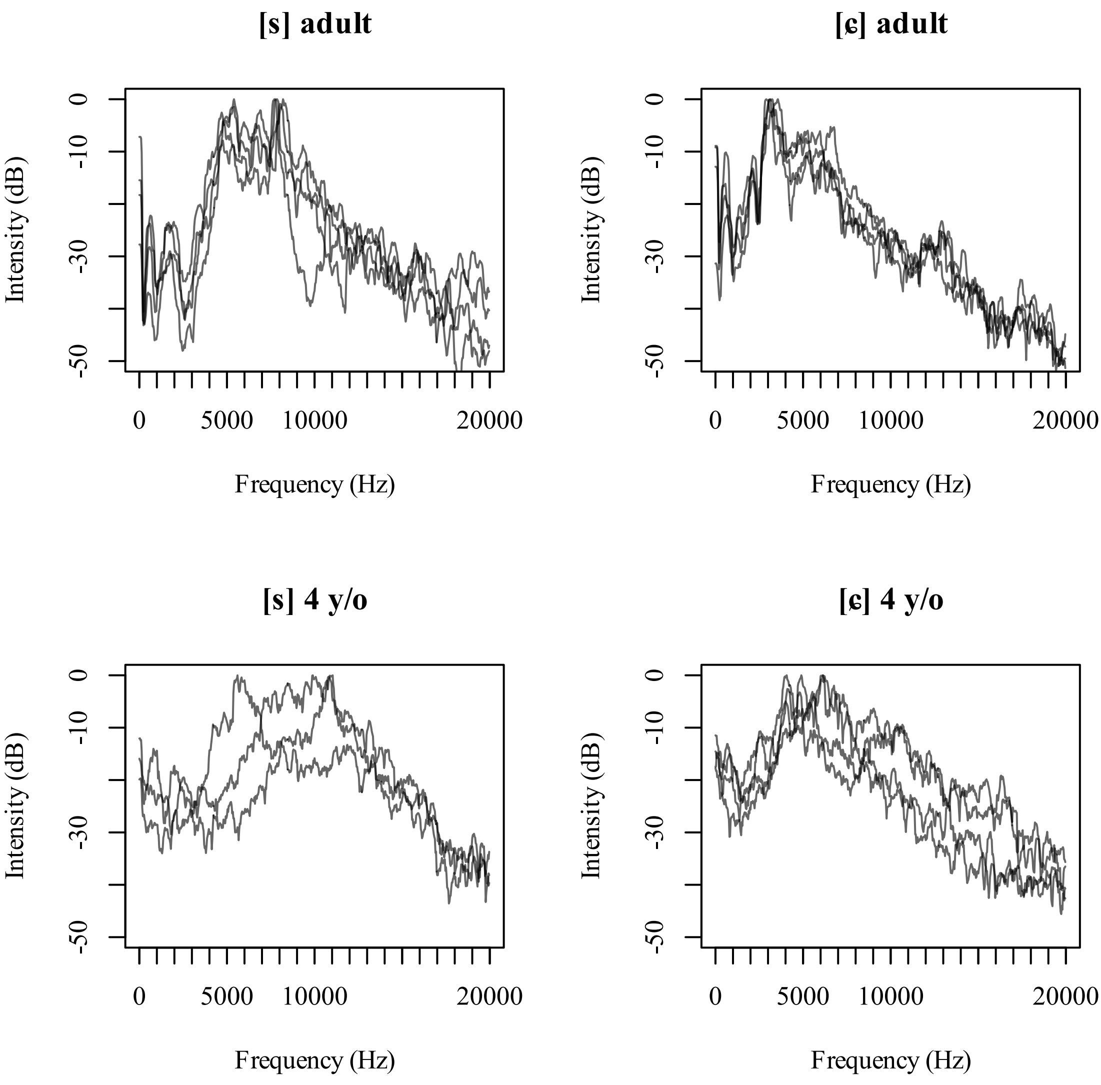
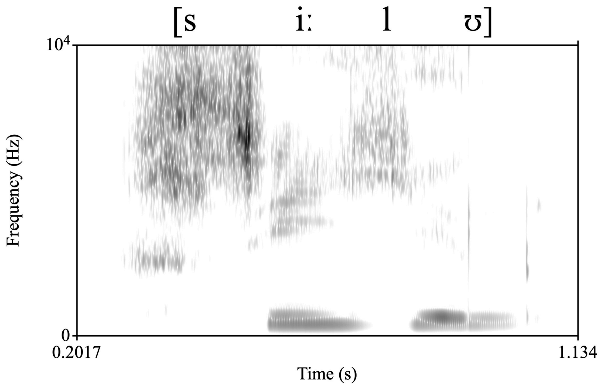
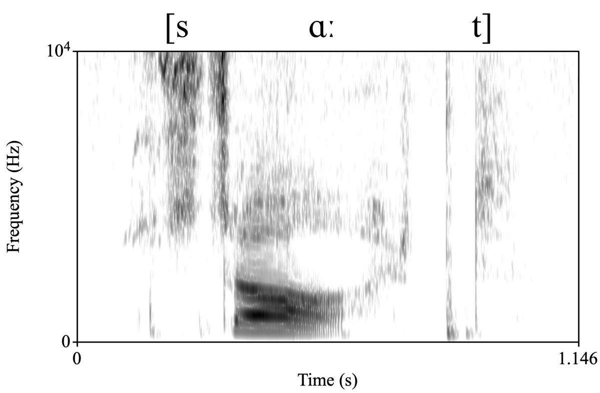


Figure 4.2: Multitaper spectra (k=8, n=4) of Swedish [s, ɕ] preceding [ɑː] from one adult female (top row, total n=8) and one four-year-old girl (bottom row, total n=7). All productions were transcribed as correct.



(a) Spectrogram showing spectral variability in the initial fricative of a three-year-old boy's production of "kilo" [ɕiːlʊ].



(b) Spectrogram showing intensity fluctuations in the initial fricative of a four-year-old boy's production of "tjat" [ɕaːt]. Note a click from background noise before the burst in [t].

Figure 4.3: Example illustrations of variability in child fricative productions.

though the English sibilant contrast [s-ʃ] is often described using M1 alone (see descriptions of Holliday et al., 2015; Romeo et al., 2013, above).

The previously mentioned criticism of spectral moments analysis also holds for research on child speech, and Holliday et al. (2010) argue that while M1 can be informative, "it may not be an appropriate measure for assessing articulatory development in children" (p. 1886). However, although spectral moments are a gross estimation of fricative spectra, that does not mean that they are useless. On the contrary, in analysis of highly variable child fricatives, their coarse character could be seen an advantage. Moreover, alternative approaches to spectral analysis in child speech research are rather exiguous (although see examples in Koenig et al. 2013 and Holliday et al. 2010) and seldom critically evaluated. Given differences in methodology, for example frequency range, spectral estimation and spectral measures, cross-study comparisons of the characteristics of children's voiceless fricatives are awkward.

4.2.5 Swedish children's acquisition of voiceless fricatives

As previously described, voiceless fricatives are among the last sounds to be acquired by Swedish-speaking children (Lohmander et al., 2017b; Lundeborg Hammarström, 2019). Labio-dental /f/ is acquired by age three, while /ɧ/ and /ɕ/ are established at around five, and /s/ at six or later (see summary in Table 4.1). Knowledge concerning common developmental speech errors that affect voiceless fricatives is limited, although Lohmander et al. (2017b) noted /s/ distortions in more than 50% of three-year olds, and around 25% of five-year-olds. Additionally, approximately 10% of Swedish adolescents exhibited "problems" with /s/ (Lohmander et al., 2017b), indicating either continued articulatory refinement or socio-linguistic variation in pronunciation.

In an in-depth analysis of 30 three-year old children's speech errors, Palo (2022) found that /s/ was often realised as [ʃ] or [ɕ], and that substitutions of [θ] for /s/ was the most common distortion in the sample. Moreover, /ɕ/ was sometimes fronted (a total of 25 errors across 30 children), while few substitutions of /ɧ/ were uncovered (n = 9, realised as [f] or [h]).

For five-year-old children, Göthlin (2022) found that 16 of 30 included children occasionally produced [θ] for /s/, and backing of /s/ was also reported - primarily in word-initial position. Although Palo (2022) and Göthlin (2022) revealed new details concerning developmental speech errors in Swedish, for example by including word position in their analyses, the number of opportunities for each voiceless fricative target was unbalanced in the speech material. This uneven distribution is unsurprising, given that /s/ is more frequent (Witte and Köbler, 2019), and motivated, as the tests aims to uphold phonological balance. However, investigating the realisation of the two more posterior lingual fricatives /ɕ, ɧ/ in more detail is motivated.

With respect to the acoustic characteristics of Swedish children's fricatives, Karlsson's 2006 thesis is, to the best of my knowledge, the only acoustic description prior to *Paper III*. Karlsson investigated the acquisition of /s/ in isolation and in /s/ + plosive consonant clusters in 22 Swedish children between 1;6 and 4;6. On balance, children's productions of initial /s/ became more adult-like with increasing age as measured by M1, M3 and F2 onset of the following vowel, but not in M2 and M4. As detailed in Section 4.1.5, /s/ is often acquired after 4;6, suggesting that there may be developmental patterns yet unaccounted for. Moreover, the acoustic characteristics of /f/, /ɕ/ and /ɧ/, and the acquisition of fricative contrast were, as of *Paper III*, uncharted.

Thence, there is a paucity of knowledge concerning the details of Swedish children's acquisition of these complex sounds. In *Paper III* of this thesis, children's correct and incorrect productions of word-initial voiceless fricatives are described, collected via a speech task with a balanced number of targets for /s, ɕ, ɧ/. The third paper also describes acoustic characteristics of Swedish voiceless fricatives and between-sibilant contrast, by means of spectral moments, spectral peak (as in *Paper I* and *II*) as well as spectral balance in low-mid and mid-high frequency ranges. In *Paper IV* experienced (SLPs) and inexperienced adult's gradient ratings of children's /s/ and /ɕ/ are explored.

Summary

Speech development is a gradual process during which children converge on adult-like speech output through experience, maturation and practice. Despite immense individual variability, some universal tendencies in typical speech development are evident, such as canonical babbling before 10 months, first words around 12 months and speech that is intelligible to strangers around 4-5 years of age. Broad trends in AoA of speech sounds are also attested, many of which are related to speech motor maturation and articulatory complexity. Differences in OoA across languages are well documented and are associated with language-specific phoneme frequency and structure. For children who encounter difficulties on their route to acquiring functional speech, clinicians may struggle to identity those at risk of longstanding difficulties, and those who will "catch up" to their peers.

The acquisition of voiceless fricatives is often protracted, and it is not uncommon for speech errors involving these complex sounds to persist until school age. Language-specific patterns in acquisition are evident, and acoustic contrast between sibilants is often late-emerging.

As for Swedish, general trends in speech sound acquisition are well understood, although fine-grained development of contrast, as well as detailed descriptions of typical speech errors are lacking, both with respect to voiceless fricatives and other sounds. Detailed language-specific developmental norms that include diverse groups and report on individual variation are important for our general understanding of speech acquisition and, in extension, the identification of SSD.

5. The perception of voiceless fricatives

Articulatory and acoustic studies provide insights into speech behaviour but if one wants to understand what variation actually matters for speech perception (and in the extension for functional speech communication), perceptual studies are warranted. However, speech sounds "... can be labelled differently depending on a variety of social, pragmatic, and linguistic factors."(Munson et al., 2017b, p. 58). Moreover, there is widespread variation regarding task design, participant instructions and speech material in perceptual research, all of which can affect listener responses and ecological validity.

The following overview will, in the interest of space, briefly describe the relationship between acoustics and perception, and the effect of listener characteristics and the perception task, because these factors have relevance to the investigation in *Paper IV*. As the fourth paper includes perceptual ratings from both inexperienced listeners and speech-language pathologists, a brief review of perceptual analysis in the clinic is also provided. The focus will be on fricatives throughout, with special attention paid to the perception of children's voiceless fricatives.

5.1 Acoustic cues to perception

Many different acoustic features have been studied in relation to the perception of voiceless fricatives (e.g., F2 of the surrounding vowels; Żygis et al. 2023 and intensity; Behrens and Blumstein 1988a; Hedrick and Ohde 1993) although spectral features are consistently reported to carry strong perceptual cues to voiceless fricative identity[1]. Most acoustic-perceptual studies have focused on spectral centre of gravity (M1). For instance, Julien and Munson (2012) found a strong correlation between M1 and the ratings of children's

[1]A number of studies have used psycho-acoustic spectra (i.e., spectral representations that mimic the human ears response to speech), which are naturally pertinent for speech perception research. However, the present thesis primarily uses the traditional Hertz scale (with the exception of Hertz-to-Mel converted spectra in the model of spectral dynamics in *Paper I*) for reasons of convenience.

English sibilant fricatives /s, ʃ/ using Visual Analogue Scale (VAS; a gradient rating scale comprised of a horisontal line with a sound on each side, see Figure 6.2) Perceived goodness of children's English sibilants was similarly found to correlate with a sibilant contrast robustness measure based on M1 in Holliday et al. (2015). Żygis et al. (2023) studied the influence of fourteen acoustic features (including fricative duration, spectral moments 1-4 and spectral peak) on Polish children's classification of word-medial voiceless sibilant fricatives produced by themselves. Overall, spectral mean (M1) and F2 onset of the surrounding vowels were found to be most influential for sibilant categorisation.

The perceptual consequences of spectral variability of voiceless sibilants has also been subject of investigation. Newman et al. (2001) studied the effect of within-speaker acoustic variability in English adult's productions of /s, ʃ/ on listeners' reaction times and accuracy in a classification task. The stimuli consisted of CV-syllables from speakers with varying degrees of acoustic separation between the sibilants, as measured by M1 and M3. An increase in reaction time was found for the speaker with category overlap (in M1), suggesting slower perceptual processing for less distinct categories. Newman and colleagues interpreted their results as indicating that robustness of contrast, and particularly overlap between categories, might influence intelligibility of speech.

Perceptual cue-weighting strategies may also be idiosyncratic, as was shown by Kong and Edwards (2011), who observed "... individual differences in responses to subphonemic detail and that these differences may be systematically related to sensitivity to different acoustic cues." (p. 1126). Nevertheless, individual differences are seldom reported, as most previous work focuses on group-level analysis.

Although many studies only investigate some select feature/s in relation to perception, it is conceivable that multiple acoustic cues contribute to identification of voiceless fricatives, given the complex acoustics of voiceless fricatives, and individual variability in production. In perceptual studies of child speech, where more acoustic variability is expected, nuanced and comprehensive representations of fricative acoustics are particularly motivated.

5.2 The influence of the listener

It is beyond dispute that people's perceptions of the world are influenced by their experiences and expectations. As such, it is to be expected that listeners carry biases with them into speech perception experiments. In fact, listeners have been shown to rate speech differently depending on who they believe the speaker to be. Gender bias has been attested in the perceptual classification of English non-sibilant (Babel and McGuire, 2013) and sibilant voiceless fricatives (Munson et al., 2017a) , in such a way that fricative identification was affected by whether the listener believed that the speaker was a man or a woman. With respect to children, the inferred age of child speaker has been shown to affect judgements such that /s/ productions are rated as more prototypical if believed to be produced by an older child (Munson et al., 2010).

Listeners' language background has also been found to influence perception. Li et al. (2011) compared English (n=19) and Japanese (n=20) listeners classification of children's sibilant fricatives to acoustic features of the fricatives (M1-M4, F2 onset). They revealed that acoustic cues were weighted differently depending on the listeners language background, indicating language-specific patterns in perception of children's voiceless sibilants.

In addition, *familiarity effects* (i.e., facilitative effects of knowing the speaker) are widely attested in speech perception research (e.g., Souza et al., 2013). With respect to children, it is well established that that caregivers understand their child better than unfamiliar interlocutors (Baudonck et al., 2009; Flipsen Jr, 1995), although it is less clear how experience of interacting with children affects processing and perception of child speech. Yu et al. (2023) investigated whether different types of experience with children affected listeners' ability to transcribe children's speech in noise. Single word productions from adults and children (between 30 and 72 months) were presented via an online task to four listener groups: inexperienced listeners (n=48), mothers of young children (n=48), early childhood educators (n=48) and speech-language pathologists (SLPs, n=48). The accuracy of transcriptions was similar across all groups except SLPs, who performed better than the other listeners. Nevertheless, Yu and colleagues 2023 state that "... the ability to transcribe child speech is not modulated by experience [with children]." (p. 446), as the SLP also performed better in transcription of adult speech, indicating a task-related advantage for the clinicians.

Clinical experience has also been considered in perceptual studies of children's voiceless fricatives. Munson et al. (2012a) investigated the influence of experience on VAS ratings of phonetic detail in children's productions of /θ/-/s/, /d/-/g/ and /t/-/k/. Forty-two listeners (21 SLPs; 21 inexperienced listeners) rated CV-syllables on a VAS specific to each contrast (e.g., /θ/-/s/). Overall, SLP ratings had higher reliability and were more strongly associated with acoustic characteristics of the sounds (for fricatives; peak ERB, compactness index and peak loudness). Meyer and Munson (2021) further explored the role of clinical experience in an online experiment designed to elicit gradient ratings of child productions of word-initial /θ/, /s/, /ʃ/, /d/, /g/, /t/ and /k/. Participants with no clinical experience (n=20) , and SLPs with up to 10 years (n=42) or more than 11 years (n=39) of clinical experience, rated CV-sequences on a nine-point interval scale, anchored by one of the above sounds on each end (e.g., /s/-/ʃ/). Experience was not found to have clear effects on accuracy but differences were revealed with regards to categoricity of response, such that SLPs with more than 10 years of experience used the endpoints of the scale to a higher extent than did less experienced listeners. In other words, more clinical experience was related to more categorical responses.

In summary, experience and expectations can affect perceptual ratings, and although it is clear that experience can influence perception, how different types of experience affect perception of child speech remains partially obscured.

5.3 Task-related bias

Bias in perception can be related to the individual listener's experiences and expectations (as discussed above), but can also be induced or exaggerated by the speech perception task itself. Hence, instructions to participants, speech material and task design need to be taken into account when assessing experimental results.

Categorical perception (CP; Liberman et al., 1957) has been highly influential in speech perception research. In broad strokes, CP claims that perception of speech is a process of converting the variable speech signal into discrete categories (i.e., phonemes), and in doing so, listeners disregard sub-phonemic ("irrelevant") information. CP is traditionally measured through classification and discrimination of acoustically manipulated sounds (e.g., a continuum between [sa]- [ɕa] with decreasing M1 in the initial fricative). Despite profound impact, the CP phenomenon has been contested (see e.g., McMurray 2022, Schouten et al. 2003 and Apfelbaum et al. 2022); CP seems to be a consequence of task characteristics rather than a reflection of human speech categorisation. That is, given appropriate (continuous) tasks, listeners provide gradient ratings for gradient speech phenomena, as evidenced by the plethora of studies that have found covert contrasts in child speech (Harel et al., 2017; Julien and Munson, 2012; Munson et al., 2017b; Munson and Urberg Carlson, 2016; Schellinger et al., 2017; Strömbergsson et al., 2015; Urberg-Carlson et al., 2009).

Munson et al. (2017b) investigated whether they could influence perceptual ratings of children's English /s/ and /θ/ in CV-sequences through task manipulation. They elicited categorical and continuous ratings of childrens fricatives through a binary classification task and a VAS. The fricative rating tasks were presented in two conditions, interspersed either with a (categorical) rating of the vowel produced or a (continuous) rating of gender typicality. The two conditions were compared for each task, and results showed that the continuous VAS task was less affected by the biasing conditions, than the categorical binary classification task, indicating that scale type might affect resilience to bias.

There are also biases grounded in linguistic knowledge and expectations. The *Ganong effect* (Ganong, 1980), denotes an effect of stimuli lexicality on the perceptual classification of acoustically ambiguous sounds. More specifically, if one were to present a listener with a sound that is acoustically intermediate between the real word "task" and the non-word "dask", the effect predicts that the listener will classify the stimulus as a real word. Previous VAS studies on the perception of children's speech have primarily included word fragments (extracted initial CV syllables) bereft of lexical meaning (children's fricatives studied in Holliday et al. 2015; Meyer and Munson 2021; Munson et al. 2012a; Munson and Urberg Carlson 2016; Schellinger et al. 2017 and other consonants in Ancel et al. 2023; Strömbergsson et al. 2015). Notable exceptions, with whole word stimuli, are presented in Harel et al. (2017) and Coniglio et al. (2022). The use of CV-sequences permits cross-linguistic studies of perception, but lacks ecological validity, given that children speak in words and sentences, not syllables.

In *Paper IV*, the effect of lexicality on experienced and inexperienced listeners ratings of Swedish children's sibilant fricatives is explored by manipulating the lexical status of the speech stimuli presented. That is; does the rating of the same acoustic stimulus (a child's production of a sibilant fricative) differ depending on whether it is presented in a whole word or in a word fragment? As the study recruited both inexperienced listeners and SLPs as listeners, a possible interaction between lexicality and experience is also investigated.

5.4 A brief note on speech assessment in children with SSD

There is substantial variation with respect to clinical assessment of speech in children with suspected SSD (see Joffe and Pring, 2008; McLeod and Baker, 2014; Skahan et al., 2007; Wikse Barrow et al., 2023a, for examples). Perceptual assessment of speech (primarily by means of transcription) is, nonetheless, a key component of SSD assessment, and phonetic transcription of a representative speech sample is said to "...inform all aspects of clinical management [of children with SSD]." (Child Speech Disorder Research Network, 2017a, p.1). Perceptual assessments in the clinic are conducted for other purposes and in other conditions than assessments in perceptual research. In the clinic, it may be motivated to use varying levels of detail (e.g., phonemic or phonetic transcription), depending on the goal of the analysis, although narrow transcription is necessary to capture phonetic errors (see Child Speech Disorder Research

Network, 2017b, and references therein.). Indeed, surveys of clinical practice indicate that clinicians often use broad rather than narrow transcription, or a combination of the two (Knight et al., 2018; Wikse Barrow et al., 2023a). Knight et al. (2018) found that among clinicians in the UK (n= 759), many felt insecure about their transcription ability, partly due to lack of continued education. A total of 57% of the sample did not feel that they were supported in the maintenance of their transcription skills in their workplace. Given that the reliability of perceptual assessment is contingent on training and practice (Klintö and Lohmander, 2023; Lee et al., 2009), this lack of post-qualification transcription training could be regarded as troubling.

Additionally, as transcription in itself cannot reliably capture all relevant fine phonetic variation in child speech (see Section 4.1.2), it has been suggested that clinical transcription should be supplemented with acoustic analysis and/or perceptual judgements from (naïve) listeners (Edwards and Beckman, 2008a). Clinical implementation of acoustic and/or articulatory analysis would indeed augment speech assessments, but in view of time constraints and technical hurdles, it is unlikely (and probably unreasonable) on a day-to-day basis. Continuous ratings scales such as the VAS, which have the potential to capture gradient progression in the production of speech sound contrast, are a more accessible complement to traditional transcription (see e.g., Munson et al., 2012a). However, the extent of VAS use in clinical assessment of children's speech, internationally and in Sweden, is unbeknownst to me. Moreover, many studies of VAS use CV-sequences as stimuli, and clinicians assess speech in context (i.e., words or sentences, not syllables), questions concerning the effects of lexical bias remain open.

Summary

Perceptual studies have revealed that spectral centre of gravity (M1) is important in identification of voiceless fricatives, primarily for sibilants. Different characteristics of the individual listener such as language background or experience with the speaker can influence her perception, and listeners' ideas about speaker age and gender have been found to impact perceptual judgements of voiceless fricatives. The specifics of the perception task, such as the rating scale and speech stimuli used, can also influence listener responses, and careful task design is, consequently, mandated. In clinical assessment of child speech, SLPs use transcription in ways which differ from research practice. Augmentative assessment tools, such as acoustic or articulatory analysis, and fine-grained perceptual rating scales have been proposed to increase the quality of clinical speech assessment, but some research-practice gaps persist.

6. The studies and their contributions

This chapter describes the scientific contributions of the four papers included in this thesis. The papers all describe the characteristics of Central Swedish (CS) voiceless fricatives, although do so from different perspectives; acoustic characteristics of adult productions are described in *Paper I* and *II*, children's fricatives are detailed in *Paper III*, and adults' perception of child productions are investigated in *Paper IV*. The following renditions are highly condensed. For further details, the reader is referred to the relevant publications.

6.1 Paper I: Static and dynamic spectral characteristics of Swedish voiceless fricatives

Authors: Carla Wikse Barrow, Marcin Włodarczak, Lisa Thörn & Mattias Heldner.
Journal: The Journal of the Acoustical Society of America.
Status: Published.

The acoustics of voiceless fricatives have received ample international attention, with descriptions of English, Japanese, Greek and European Portuguese contributing to the knowledge base concerning universal and language-specific fricative features (Behrens and Blumstein, 1988b; Forrest et al., 1988; Hughes and Halle, 1956; Iskarous et al., 2011; Jesus and Shadle, 2002; Jongman et al., 2000; Koenig et al., 2013; Munson, 2004; Nirgianaki, 2014; Reidy, 2016b; Romeo et al., 2013). Descriptions of Swedish fricatives were, prior to this paper, limited to aged and small scale studies, such as Lindblad's thesis (1980) and Shosted's single-case investigation (2008).

Although most previous studies have focus on "steady state" or static characteristics of voiceless fricatives, fricative spectra are known to vary across their duration (see e.g., Iskarous et al., 2011; Zharkova et al., 2018) in language-specific ways (Reidy, 2016b). In *Paper I*, a description of static and dynamic acoustic characteristics of word initial Swedish voiceless fricatives /f, s, ɕ, ɧ/ is provided, motivated by the lack of up-to-date descriptions.

Methods. Twenty native speakers (10 self-reported female and 10 self-reported male) of CS were recorded. The speech stimuli included 28 fricative initial words with varying vowel contexts. Four spectral moments (M1, M2, M3 and M4 - see Section 3.2 for a description of the method) and the frequency of the highest amplitude peak (global spectral peak) were obtained from a 20 ms Hann window from the middle of the fricative segment (frequency range = 0-11 kHz). Duration of the fricatives was estimated from the annotated onset and offset, and z-scored intensity was calculated within each speaker. For analysis of spectral dynamics, 15 spectra were extracted from Hann windows evenly spaced across the fricative. The spectra were converted to the mel scale and temporal change in M1 for /s, ɕ, ɧ/ was modelled using a Generalized Additive Mixed Model (GAMM), including effects of fricative, lip rounding (rounded-neutral-spread), speaker and duration.

Results. With respect to static acoustic features, M1 differentiated between the two sibilants [s, ɕ], and velar [ɧ], although labiodental [f] was not distinguishable by M1 alone. With the addition of the second spectral moment (M2: spectral standard deviation) discriminability of [f] increased. Overall, the sibilant fricatives [s, ɕ] were longer and had higher intensity than non-sibilants [f, ɧ]. The model of M1's temporal variation similarly showed clear separation of [s, ɕ, ɧ] in the overall level (i.e. intercept) as well as in their shape over time (i.e. smooths). The M1 trajectories were increasingly dynamic for increasingly anterior sounds, such that [s] was most curved, followed by [ɕ] and finally [ɧ], which was rather flat. Due to a coding error the levels of z-scored intensity published in the original article were slightly warped. An errata was submitted shortly after I discovered the error (see Wikse Barrow et al., 2023b), and is included following *Paper I*. The data and code used to generate the results are publicly available on Zenodo (link in pdf).

The major novelty of this study is the comprehensive and up-to-date description of the acoustic characteristics of Swedish voiceless fricatives, including both static and dynamic spectral features. The paper and the associated dataset are a contribution to the cross-linguistic study of voiceless fricatives, and may be useful as reference in speech acquisition research.

6.2 Paper II: Variability in Swedish voiceless fricative contrasts

Authors: Carla Wikse Barrow, Marcin Włodarczak, Mattias Heldner, & Sofia Strömbergsson.
Proceedings: Proceedings of International Congress of Phonetic Sciences 2023.
Status: Published.

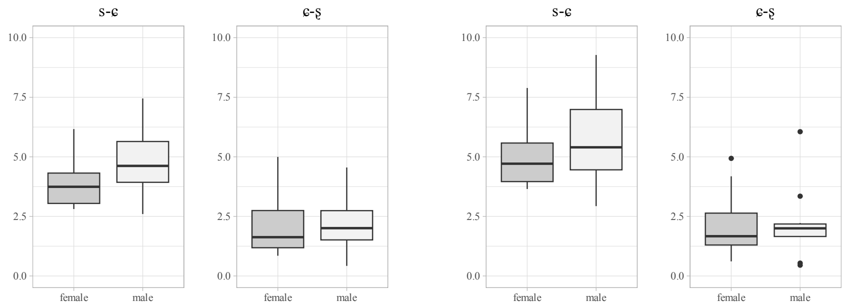
Pre-dorso-alveolar /ɕ/ and velar /ɧ/ are two of the select segments that exhibit dialectal variation in Swedish. Generally speaking, most speakers of CS will use two allophones of /ɧ/; velar [ɧ] in prevocalic/word initial position, and retroflex sibilant [ʂ] in post-/intervocalic position (with the exception of complex onsets; Riad, 2014). The phonetic realisation of [ʂ] is perceptually very similar to [ɕ] (e.g., Shosted, 2008). As of yet, contrast between voiceless fricatives of CS and individual differences in fricative realisation has not yet been investigated.

In *Paper II*, individual variation and fricative contrast in the voiceless fricatives of adult Swedish speakers is explored, by reanalysing the recordings described in *Paper I*.

Methods. Speech material was selected from the recordings of the 20 speakers from *Paper I*. In addition to the four fricative phonemes previously described, the light allophone of /ɧ/; [ʂ] was also included in analysis. Spectral moments 1-4 and global spectral peak were calculated from a 30 ms window in the centre of the fricative. The first two spectral moments were chosen to illustration the acoustic differentiation and dispersion of voiceless fricatives in individual speakers, because the measures discriminated well between [f, s, ɕ, ɧ], on a group-level, in *Paper I*. In order to investigate contrast further, a contrast robustness measure was adopted from Romeo et al. (2013) and used to quanitfy the acoustic contrast between the sibilant pairs [s-ɕ] and [ɕ-ʂ].

Results. Generally, [s, ɕ, ɧ] were distinct from one another in M1, with velar [ɧ] being most clearly differentiated for a majority of speakers. Overlap between the sibilant fricatives was visible for a few speakers but overall, the sounds were clearly separable. Individual differences were found with respect to degree of separation between fricatives in the M1-M2 dimension as well as for within-category variability. M1 for [ʂ] was very similar to [ɕ], but was distinctly lower for some individuals.

A coding error led to the illegitimate exclusion of tokens in the calculation of the published contrast robustness measure, which was addressed during the presentation of the paper at the International Congress of Phonetic Sciences in the summer of 2023. The original and updated contrast robustness scores can be seen in Figure 6.1. The two figures show similar trends, although the



(a) Published contrast figure (adapted) **(b)** Updated contrast figure

Figure 6.1: Contrast robustness for [s-ɕ] and [ɕ-ʃ] for female and male speakers.

dispersion of male speakers' contrast is notably smaller in the updated figure (Figure 6.1b). The figures show that male speakers have higher overall discriminability scores for both sibilant pairs, even though differences are small. This gender difference is the opposite pattern as to what has previously been reported, namely that women produce more distinct sibilant fricatives than men (e.g., Romeo et al., 2013), although direct comparisons with previous studies are hindered by differences in methodology.

The primary contribution of the second paper is the description and quantification of individual variation in the spectral characteristics of Swedish adults' voiceless fricatives.

6.3 Paper III: Individual variation in the realisation and contrast of Swedish children's voiceless fricatives

Authors: Carla Wikse Barrow, Sofia Strömbergsson, Marcin Włodarczak & Mattias Heldner.
Journal: Journal of Phonetics.
Status: Published (Advance online publication).

Descriptions of group-level developmental trends in speech acquisition are of great value. Nevertheless, studying and quantifying heterogeneity in speech acquisition is important for our understanding of the underlying mechanisms of development, as well as for informing clinical decision-making in the care of children with SSD. Prior to this work, the approximate age and order of acquisition of Swedish voiceless fricatives was known, as were acoustic characteristics of /s/ in children up to the age of four (Karlsson, 2006; Lohmander et al., 2017b; Lundeborg Hammarström, 2019). However, acoustic descriptions of /f, ɕ, ɧ/, and the contrast between Swedish children's voiceless fricatives was unexplored.

Paper III presents a small-scale study of Swedish children's voiceless fricatives, including both transcription and acoustic analyses of group-level features and individual differences.

Methods. Thirty-one children between three and eight performed a digital audio-prompted picture naming task. They were recorded in a sound treated room at the Department of Linguistics (n=24), or in their homes (n=7). Recordings of their productions of twenty-two fricative initial mono- or bisyllabic words in a variety of vowel contexts were transcribed and acoustically analysed. Transcriptions were summarized and described in terms of correct productions and inaccurate realisations (i.e., perceptual mismatches with the adult target). The spectral characteristics of fricative and affricate productions were analysed using spectral moments (frequency range = 0.3-15 kHz), spectral peak and spectral balance measures (Forrest et al., 1988; Jongman et al., 2000; Koenig et al., 2013; Shadle, 2023), estimated from multitaper spectra. The relative importance of the spectral measures was evaluated through random forest classification (Breiman, 2001). For each child, contrast between the sibilant pair [s-ɕ] was quantified as the proportion of correctly predicted sibilant tokens by a random forest trained on all other children's correct sibilants.

Results. The transcription results showed that fronting errors and affrication errors were common in off-target productions. Acoustic analyses of correct productions revealed similar mean and range of M1 and spectral peak across age groups. The spectral measures M1, spectral peak and low-mid spectral balance were ranked as most important in the classification of children's on-target productions, for all age groups over four. Although spectral measures for correct productions were similar across age groups, the accuracy of the random forest models increased with age, suggesting more consistent spectral patterns for the older children. A clear increase in the magnitude of acoustic contrast (i.e., prediction accuracy) between sibilants was also visible with increasing age, although individual differences were substantial. All code and anonymous acoustic data (spectral measures) can be found on Zenodo (link in pdf).

Contributions. This study provides a description of Swedish children's realisation of voiceless fricatives. Novelties of *Paper III* are the quantification of individual differences in the acoustics of Swedish children's fricatives, and a continuous description of developing acoustic contrast, which is based on a more spectral measures than previous work. The paper and the associated data set provide a new language source for research on cross-linguistic tendencies in speech acquisition.

6.4 Paper IV: Exploring the effect of lexicality and listener experience on gradient ratings of Swedish sibilant fricatives

Authors: Carla Wikse Barrow, Lina Ottosson & Sofia Strömbergsson.
Journal: Clinical Linguistics & Phonetics.
Status: Published (Advance online publication).

Sibilant fricatives, such as Swedish /s/ and /ɕ/, are cited as late emerging in many languages (McLeod and Crowe, 2018). Previous studies report on gradual differentiation and increased contrast between sibilant pairs with age (e.g., Holliday et al., 2015; Li, 2008, see also *Paper III*). Hence, many children will produce sounds that perceptually fall somewhere "in-between" the target sibilants during acquisition. Studies of the perception of such fine phonetic variation in English children's sibilants (/s, ʃ/) have frequently employed Visual Analogue Scale (VAS) (Munson et al., 2010, 2012a,b; Munson and Urberg Carlson, 2016; Schellinger et al., 2017) to elicit ratings of category goodness. The VA scale allows for fine-grained ratings of speech and is a practicable alternative to transcription of speech in the SLP clinic. However, possible effects of lexical bias in VAS ratings of speech remain uncharted. It is well known that lexical knowledge influences speech perception (e.g., Ganong, 1980) but as most previous VAS studies have, by design, used CV-syllables devoid of lexical meaning, the influence of lexicality on VAS ratings of speech is yet unknown. It has been hypothesised that SLPs might be "...less susceptible to lexical bias than inexperienced listeners." (Munson et al., 2012a, p.137), as their ratings have higher reliability and better correspond to the acoustic characteristics of the sibilants, than ratings from inexperienced listeners. In *Paper IV*, this very question is explored; how are experienced (SLPs) and inexperienced listeners' ratings of Swedish children's voiceless fricatives affected by lexicality?

Methods. Inexperienced (n=27) and experienced (i.e., SLPs; n=18) listeners took part in a speech perception task, using a digital VAS (see Figure 6.2). The listeners were presented with [s] and [ɕ] initial words and word fragments (i.e., the initial CV syllables of each word), and were asked to rate the first sound of each token on a scale from [s] to [ɕ]. The speech material was extracted from a subset of the recordings described in *Paper III*. A total of 250 tokens (125 words, 125 syllables) from two minimal and two near-minimal word pairs were used, chosen to cover the perceptual continua between [s] and [ɕ]. Ratings were visually inspected and effects of experience (SLP vs. non-SLP) and lexicality (whole word vs. word fragment) were explored with a

Bayesian mixed-effects Beta regression. The model analysed mean and precision of transformed VAS ratings as a function of experience and lexicality, also including random by-participant and by-item intercepts as well as slopes for lexicality (for both) and experience (for items).

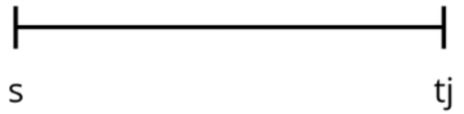


Figure 6.2: The digital VAS used in Paper IV

Results. Inspection of responses show individual differences in scale use (e.g., categoricity of responses), although most listeners showed similar response patterns and used the whole scale. Intra- and inter-rater reliability were higher for the SLP group, but variation was considerable. Large overall effects of lexicality were found; words were rated as more prototypical (i.e., closer to the intended fricative) and with higher precision (i.e., less variance) than syllables. Experience also showed robust effects, such that SLPs provided more precise and target-like ratings than inexperienced listeners. The interaction between lexicality and experience also received support, showing a smaller effect of lexicality for SLPs than for inexperienced listeners. Data and code used to produce the results in *Paper IV* are obtainable via Zenodo (link in pdf).

Contributions. This paper is, to the best of my knowledge, the first direct comparison of VAS ratings in lexical and non-lexical contexts. The results of this paper indicate that lexicality does seem to matter, and suggests that effects of experience may have been exaggerated in previous VAS studies. The results can contribute to a better understanding of how previous studies using different stimuli relate to one another.

7. General discussion and conclusions

This thesis includes three papers that detail acoustic characteristics of Swedish voiceless fricatives in adults and in children, including descriptions of group-level (age, gender) trends and individual variation in fricative realisation and contrast. Prior to the studies presented in this thesis, knowledge concerning Swedish adult's voiceless fricatives was limited to studies based on a small number of speakers from a time when acoustic analysis was more cumbersome and less informative. Descriptions of children's speech were restricted to transcription-based studies covering broad trends in development, and one acoustic study of /s/ in young children. The three studies herein provide an updated and elaborated description of the acoustic properties of Swedish voiceless fricatives, using modern acoustic and statistical methods.

The fourth and final paper investigated the influence of lexicality on VAS ratings of fricatives by experienced (SLPs) and inexperienced adult listeners, thus providing a clinically relevant perspective on fricatives in acquisition. The effect of lexicality was found to be robust, and the findings from *Paper IV* have bearing on the implementation of VAS in the clinical care of children with SSD.

In this final chapter, the main findings from the four included papers are considered in relation to previous research (Section 7.1), followed by a general methodological discussion (Section 7.2). Ethical considerations and future directions for research are outlined in Sections 7.3 and 7.4, and the chapter closes with some concluding remarks (Section 7.5).

7.1 Main contributions

Taken together, the major contributions of this thesis can be summarised as 1) a comprehensive descriptions of adult and child productions of Swedish voiceless fricatives, with resources in the form of acoustic data (descriptions, not audio) made available for cross-linguistic research, and 2) an investigation of lexicality on experienced and inexperienced listeners ratings of voiceless sibilants using VAS. The main findings are discussed below, with reference to relevant work on Swedish and other languages.

7.1.1 Swedish adults' voiceless fricatives

The first paper of this dissertation presents static and dynamic acoustic characteristics of Swedish voiceless fricatives. The results confirm that some of the observations in Lindblad (1980), hold in a larger group of participants. For example, the overall spectral shapes of the voiceless fricatives are similar to Lindblad's descriptions. However, methodological differences are so considerable that comparisons of more detailed aspects are difficult. For instance, many of the prominent peaks observed in adult productions of /s/ (*Paper I*) were outside the bounds of the spectra displayed in Lindblad (1980).

Overall, the first study showed that spectral centre of gravity (M1), measured in the middle of the fricative, separated the three lingual voiceless fricatives /s, ɕ, ɧ/ rather well, even on a group level. The labio-dental /f/ was distinguishable if spectral standard deviation (M2) was included, although the sounds exhibited high spectral variability. Shosted (2008) utilised similar acoustic measures (e.g., M1) in his case-study of Swedish voiceless fricatives, although the frequency range used was larger than in this work. Similar to the findings presented here, he found that spectral centre of gravity (M1) discriminated well between the allophones of the velar voiceless fricative [s̹, ɧ], and that [s̹] and [ɕ] has similar M1 values, in both neutral and emphatic conditions.

In relation to other languages, Swedish adult's productions of [s] was similar to English, Russian and Greek [s] in M1, M2 and spectral peak (Jongman et al., 2000; Kochetov, 2017; Nirgianaki, 2014). The posterior sibilant [ɕ] had slightly lower mean values for M1 than the English [ʃ] (Jongman et al., 2000) and Greek palatal [ç] (Nirgianaki, 2014). The acoustic properties of /f/ were generally similar to reports from other languages (Jongman et al., 2000; Nirgianaki, 2014) and Swedish (Lindblad, 1980); the spectrum of /f/ is dispersed across a large frequency range, with high spectral standard deviation (M2) and low kurtosis (M4). As mentioned previously, characterising spectral features of non-sibilant fricatives is challenging (Shadle et al., 2023). However, the Swedish velar non-sibilant [ɧ] is distinct from English non-sibilants such as [f, θ], in that it has a clearly defined peak and a skewed energy distribution, rather than a flat and featureless spectrum (cf. Figure 3.1). The spectra of [ɧ] have high M3, low M2 and a clear peak around 1 kHz, akin to previous descriptions of the Swedish velar (Lindblad, 1980; Shosted, 2008). Compared to the Greek velar [x], Swedish [ɧ] has lower peak and M1 (Mean M1 for for [x] was 3397 Hz in Nirgianaki 2014, which is approximately 2 kHz higher than M1 for [ɧ]), and higher M3. However, as Nirgianaki (2014) excluded frequencies below 1 kHz in their analysis, the descriptions are not directly comparable. Given that the notation of the Swedish velar is somewhat controversial, comparisons of posterior voiceless fricatives across a number of languages would be interesting.

Spectral dynamics

Spectral dynamics (M1 in Mel) of the lingual voiceless fricatives were modeled using a GAMM. The results showed that [s, ɕ, ɧ] were differentiated through level and curvature of the M1 trajectories, which were all concave. Similar to previous work on English /s, ʃ/ and Japanese /s, ɕ/ (Reidy, 2016b), the anterior sibilant showed a more dynamic spectro-temporal trajectory than posterior [ɕ].

The rise of M1 in the middle portion of the fricative could be a result of raising of the jaw, increased intensity (as seen in the intensity curves in *Paper I*) or changes in the linguo-palatal constriction area. In the articulatory-acoustic study of English /s/, Iskarous et al. (2011) observed that the increase in M1 was primarily related to jaw position, a pattern which might hold for these productions as well.

The GAMM also revealed differences in the final time points of the trajectory, related to lip position in the following vowel (neutral, spread or rounded). Given previous knowledge concerning anticipatory coarticulation (see e.g., Johnson 2012), it was expected that the rounded vowel context should elicited lower M1. Visual inspection of the plots revealed most anticipatory coarticulation in M1 trajectories of the velar [ɧ], and least in [s] (similar to the observations made by Iskarous et al. 2011 re coarticulatory resistance, and Zharkova et al. (2018) re less coarticulation in /s/ than /ʃ/). Notwithstanding the observed segment-specific differences in context-dependent variation in M1, my impression is that anticipatory coarticulation is perceptually discernible for all lingual fricatives, at least in some contexts. Looking further at speaker-specific patterns in both spectral dynamics (as proxy for coarticulation) and listeners perceptions of said patterns would be valuable contributions to the literature.

Individual differences

In *Paper II*, individual differences with respect to the structure of the voiceless fricative system (i.e., the relationship between [f, s, ɕ, ɧ]) in individual speakers was explored. M1 was highly discriminatory for lingual fricatives across all speakers, and the addition of M2 led to good separation of /f/ as well.

For most speakers, [ʂ̻] and [ɕ] overlapped almost entirely in the M1-M2 dimension (also noted by Shosted 2008), but for a few (n=5), there was little to no visible overlap. M1 for the retroflex [ʂ̻] was lower than for [ɕ] for most speakers, regardless of degree of category separation. Lindblad describes the sublingual cavity as larger in [ʂ] than in [ɕ], despite similar place of articulation, which is consistent with the present finding of lower M1 for the speakers that did differentiate between the two sounds. The differentiation between the retroflex [ʂ] and alveolo-palatal [ɕ] in some speakers suggests socio-phonetic or idiolectal variation, but further study is needed to delineate the source of variability.

Individual differences were noted for category dispersion and between-category distance (in the M1-M2 dimension). To appraise contrast between the sibilant pairs [s-ɕ] and [ɕ-ʂ], a category discriminability measure was calculated (taken from Romeo et al., 2013). Overall discriminability was higher for the [s-ɕ] contrast than for [ɕ-ʂ] but there were large individual differences with respect to contrast robustness.

A note on gender differences

Gender differences in the realisation of voiceless fricatives have been described in other languages and cultural contexts (see Sections 4.1.3 and 3.2.6). For the purposes of the present thesis, gender differences were not explored in depth, partly due the low number of participants and partly because the focus was on general characteristics of Swedish voiceless fricatives. Nonetheless, *Paper I* revealed that female speakers, on average had higher M1 values for [s] than male speakers. Visualisations of speaker-specific patterns in fricative contrast (*Paper II*), showed no clear gender differences although the overall discriminability between the sibilant pairs [s-ɕ] and [ɕ-ʂ] was higher for male than female speakers, in contrast to previous findings (e.g., Romeo et al., 2013).

With respect to the velar, older descriptions of Swedish report that the light allophone [ʂ] is associated with higher levels of education and with female gender, while the perceptually darker [ɧ] is denoted as "vulgar" (see Lindblad 1980 and references in Shosted 2008). In the present work, all included speakers produced /ɧ/ as [ɧ], so no such variation was explored.

Whether gender differences in Swedish voiceless fricatives are perceptually salient, or conveyed more systematically through other cues than the ones described herein, remains to be seen.

7.1.2 Swedish children's voiceless fricatives

The third paper presented an analysis of variation in children's productions of Swedish voiceless fricatives on two levels; phonemic variation was assessed through transcription and fine-grained phonetic variation through acoustic analysis. The transcription revealed differences in OoA of /s, ɕ, ɧ/ for some children, as compared to norms from LINUS and SVANTE (Lohmander et al., 2017b; Lundeborg Hammarström, 2019, see also Section 4.2.5). Specifically, transcribed accuracy showed that many children between three and five years of age reached higher levels of accuracy for /s/ than /ɕ, ɧ/. Note that the above observation is based on visual inspection of trends in transcribed accuracy; OoA and AoA are not reported in *Paper III*, as no criterion for acquisition was specified. The speech material used in the third paper differs from LINUS and SVANTE in a number of ways, which could have influenced the results. For example, *Paper III* only includes word-initial fricatives, which might be expected to lower AoA for /s/ (recall that word-medial and word-final consonants were generally acquired before word-initial consonants in Lohmander et al. 2017b, although frequency effects would predict earlier acquisition of word-initial /s/ in Swedish; Witte and Köbler 2019). Moreover, more opportunities for /ɕ, ɧ/ were provided here as compared to previous studies, which could have exposed subtle difficulties that were not previously accounted for. Note also that the speech task used in *Paper III* diverged from previous studies of Swedish children's speech acquisition, which could have affected the children's speech accuracy.

In addition to transcribed accuracy, developmental speech errors were also summarized, revealing that most substitutions of word-initial fricatives involved the production of a fricative or an affricative with a more anterior place of articulation (e.g., [s] for /ɕ/). An exception was the velar /ɧ/ which was often realised as glottal [h]. Affrication errors for target /s/, /ɕ/ and /ɧ/ were also prevalent (realised as [ts], [tɕ] and [kɧ] respectively).

The error analysis revealed different trends to those previously described; both backing and fronting of /s/ were observed in three-year old children in Palo (2022), whereas the children in *Paper III* seldom substituted a more posterior fricative for /s/ (only two occurrences of backing were idenitified). Moreover, although the errors described for /ɧ/ in his work were similar to the patterns uncovered here ([f] or [h] for /ɧ/), they were far rarer in Palo (2022) as compared to *Paper III*. For the five-year-old children described in Göthlin (2022), [θ] for /s/ substitutions were highly prevalent, similar to the patterns observed in this work. Note that speech errors in *Paper III* were presented on a group-level (across all ages) but were most prevalent in younger children. Compared to developmental speech errors in other languages, the substitutions observed here (i.e., primarily fronting of sibilant fricatives), are more similar to the patterns observed in English than in Japanese (Li et al., 2009).

With respect to the acoustic description, seven spectral features (M1-M4, global spectral peak as well as low-mid and mid-high spectral balance) were presented for all correct productions for each age-group. Descriptive statistics, visualisations and statistical classification (through a random forest) showed subtle age-related differences in productions that had been transcribed as correct, albeit with large variance. The acoustic cues M1, spectral peak and low-mid spectral balance were found to be most important in classification of voiceless fricatives for all age groups, apart from three year-olds. Despite similar mean acoustic values, classification accuracy increased with age, such that perceptually correct tokens produced by seven-year-olds were accurately classified to a higher extent than those produced by three-year-olds.

The same seven spectral cues were used in an investigation of acoustic contrast between the sibilants /s/ and /ɕ/ in individual children. Contrast robustness was parameterised as the proportion of correctly identified targets (i.e., /s/ and /ɕ/) by a random forest classifier that had been trained on all other children's correct sibilant productions. Overall, robustness of contrast increased with age, in line with previous investigations of English-acquiring children (Holliday et al., 2010). A drawback of the contrast measure used in *Paper III* is that if a child were to convey contrast using other cues than their peers, the model could have underestimated the contrast produced by that child. Nevertheless, the measure incorporated more acoustic features than those used in previous studies of sibilant contrast (Holliday et al., 2015; Romeo et al., 2013) and did, therefore, not rely on any single cue (e.g., M1) for fricative identification. Future estimates of between-fricative contrast could likely be improved by including additional acoustic cues, such as duration and intensity of the fricative, and frequency of F2 in the following vowel.

The contrast measure was weakly related to overall speech accuracy (PCC) and parent reports of intelligibility (ISC-SE). As there is variation in the order in which children acquire sounds, the modest relationship with PCC is not surprising. Intelligibility showed a weak correlation with both sibilant contrast and age, which could indicate that caregivers related their child to same-age peers in their reports.

Despite contributing new information, *Paper III* has only begun to scratch the surface of typical variation in children's acquisition of Swedish. In my view, one of the most important aspects of these results is that they cast light upon the need for systematic studies of *what* children do, and *how* they reach the phonetic target (i.e., productions that are transcribed as correct and overt contrasts) - not only *when* they get there. Ideally, such descriptions should be diverse with respect to demographics (SES, multilingualism, age, gender etc.), and large enough to begin to discern the boundaries of typical variation in Swedish speech acquisition.

A juxtaposition of adult and child fricatives

The adult and child data presented herein were collected in much the same way (the speech task and the target words were similar, but not identical) and the same spectral features (M1-M4, global spectral peak) were extracted. However, the use of different spectral estimation methods and frequency ranges impede comparisons across the adult and child speakers. Nevertheless, cautious comparison of MT spectra (not described in *Paper I* or *II*) from adults and children show that three- to seven-year-old children's perceptually accurate productions of [s] - on a group level - have slightly less distinct peaks and more high frequency energy than adult productions of the same sound. For the more posterior sibilant [ɕ], spectral shapes are more similar; a trough is visible around 2-2.5 kHz, a peak around 3-5 kHz and a monotonic decrease in energy towards the higher frequencies for both groups (adults and children). However, the adult productions show a larger difference in amplitude between said trough and peak (which could be interpreted as corresponding to sibilance, see Shadle 2023) and children's peak and trough frequency locations are higher. The velar [ɧ] also shows a similar shape, with a clear peak around 1-1.5 kHz and a rapid decrease in energy for higher frequencies, with the exception of a second, smaller peak around 5 kHz.

I hope to make available comparable acoustic data from adult's and children's voiceless fricatives in the near future.

7.1.3 Effects of lexicality and experience in gradient rating of voiceless fricatives

The perceptual study presented in *Paper IV* shows that SLPs, as a group, are more reliable and consistent in VAS ratings of children's sibilant fricatives than inexperience listeners. The higher reliability of SLP ratings is akin to previous reports (Meyer and Munson, 2021; Munson et al., 2012a) and bodes well for clinical implementation of the VAS. The variance of intra-rater reliability for inexperienced listener ratings was large, and some laypeople provide reliable ratings (for an investigation of the validity and reliability of crowdsourced VAS ratings, see Harel et al., 2017). A similar effect of experience was found in the consistency of ratings (shown in the *phi* parameter of the Beta regression). That is, ratings provided by SLPs exhibited less variance than those provided by inexperienced listeners.

With respect to lexicality, effects were observed for both groups and all individual listeners - the initial fricatives in words were rated as more target-like than those in syllables. Nonetheless, the effect of lexicality was less pronounced for clinicians, indicating that the thesis put forth by Meyer and Munson (2021) and Munson et al. (2012a) is true; SLPs' VAS ratings do seem to be more resilient to lexical bias. It is plausible that the interaction between experience and lexicality was mediated by a task effect (such as the one observed in Yu et al. 2023). That is, laypeople might be expected to be less familiar with the task of listening to and rating CV-syllables, than SLPs. The fact that effects of experience were less pronounced for the whole-word ratings might also indicate that previously reported effects of clinical experience on VAS ratings of speech are overestimated.

The statistical model of *Paper IV* showed strong effects of lexicality and experience. However, that which is statistically significant is not always clinical relevant (see e.g., Sand et al., 2022). Moreover, the model estimates were presented in log and log-odds units in *Paper IV*, which might impede interpretation of the magnitude of effects. In order to make the results more accessible, the model predictions were displayed on the original (transposed) VAS in a supplemental figure. The figure shows that the predicted effects are visually salient, but whether the observed differences would impact clinical decision-making is, as of yet, unclear.

Finally, the perceptual ratings from *Paper IV* were not compared to the acoustic characteristics of the children's sibilant fricatives. As such, the *accuracy* of ratings was not directly investigated, and the relationship between acoustics and perception in the two listener groups remains opaque. In view of the complexity of fricative acoustics and the variability in child productions (see Section 3.2 and 4.1.2, and *Paper III*), such a comparison was not deemed to be within the purview of the paper.

In summary, *Paper IV* revealed effects of experience and lexicality, and the results bring us closer to understanding how previous VAS studies with lexical and non-lexical stimuli relate to each other. Although the clinical significance of lexical effects is uncharted, it may be better to err on the side of caution and consider lexical bias when implementing VAS in the SLP clinic.

7.2 Methodological discussion

The following discussion considers the present PhD project as a whole, and discusses methods on a general and specific level. First, considerations related to the participants are reviewed, followed by a discussion of speech data and recording procedure, acoustic analyses and finally, statistical analyses.

7.2.1 Participants

The participants in the first two studies consisted of 20 adult speakers of working age. The third study included 31 child speakers between three to eight years of age, and the fourth study included 45 adult listeners, 18 of whom were clinically active SLPs. All participants were native speakers of Swedish (defined as having Swedish as one of their first languages) and were recruited in the broader Stockholm area.

The size of the adult participant groups herein are comparable to previous studies on speech production and perception, and recruitment for adult participants was relatively fast and simple. However, recruitment of children in post-pandemic Stockholm was both time-consuming and arduous. Despite efforts to facilitate participation (e.g., the opportunity to record at home) and over 1000 letters being sent to families with children of the fitting age, as well as posters being put up at libraries, preschools and SLP clinics around Stockholm, few families expressed interest in participating. The low *N* in each age group of children led to a change in direction of the research, from general and normative, to a focus on patterns of individual variation.

Most of the families who participated in the recordings for *Paper III* and *IV* had high Socio-Economic Status (53 of 62 caregivers had a university education, of whom 43 had completed three years or more of university level studies), leading to a skewed representation of SES in the child group. As such, this project is essentially based on a W.E.I.R.D (Henrich et al., 2010) population. Moreover, many caregivers expressed an interest in speech-language development, which may have led to increased motivation to participate and, possibly, different communicative/early literacy behaviour compared to the average caregiver.

The inclusion criteria of the project were formulated such that multilingual adults and children were invited to participate if they had Swedish as one of their first languages. The choice to include multilingual children was met by scepticism from some article reviewers and indeed, multilingualism is often cited as influential in speech development. However, the definition of multilingualism is not clear-cut and can encompass low-to-high levels of proficiency, early-to-late age of acquisition and a range of expressive and/or receptive language abilities. In the present thesis, all caregivers reported that their child's strongest language was Swedish and all children were enrolled in preschool or school, and we chose to include all children in our description (*Paper III*).

7.2.2 Speech data and recording procedure

The adult audio was of high quality; recorded with state of the art equipment in an aneachoic chamber. However, it would have been advantageous to include additional speech tasks (e.g., target words in carrier phrases and connected speech) as well as calibration tones to enable studies of connected speech and a comparison of intensity across speakers.

The child recordings were conducted in different recording environments; some children were recorded in their homes (n=7) and others in a sound-treated room at the Department of Linguistics, Stockholm University (n=24). The home recordings were conducted in quiet rooms (e.g., the child's own room or a quiet kitchen), but background noise levels were not measured and could have influenced the spectral measures presented here to some degree. To mitigate the influence of background noise, a microphone with a small circumference was used and placed close to the mouth. Moreover, all tokens with audible disturbances, such as intermittent background noise, touching of the microphone or speech from the caregiver, were excluded from acoustic analysis.

Speech production tasks

The speech production task used in recordings of children included both a picture and an audio prompt, and was modelled after previous studies of the acquisition of voiceless fricatives (see Edwards and Beckman, 2008a, for a discussion). The use of such a task ensures that speech can be collected from young children who might not name pictures correctly, and would thence require speech prompts from the test leader - which might vary across productions. In view of differences in memory, endurance and speech-language proficiency between three- and seven-year-old children, the chosen task also secured a congruous procedure across the age range included. However, the productions elicited in *Paper III* may not be representative of everyday speech. Future work on the acoustics of children's fricatives, based on more spontaneous speech would be welcome. Nevertheless, previous research indicates that speech accuracy does not seem to be affected by imitation; in an exploration of preschool-aged English-speaking children (n=267), McLeod and Masso (2019) found no significant differences in accuracy for word-initial consonants in imitated and spontaneous production. As such, accuracy of the speech collected via the current speech task should be roughly comparable to that elicited through picture-naming (e.g., Lundeborg Hammarström 2019 and Lohmander et al. 2017b).

The adult speech task was created to mirror the child task in as many ways as possible. The adults saw an orthographic representation of a target word and heard an audio-prompt produced by the same speaker that was used for the children. The dual prompt was not mandated for the adults, and may have been perceived as infantile. Hence, the adult speakers were told that the task was purposefully made easy, so their speech would be elicited in a comparable fashion to the children we were planning to record.

Transcription, annotation, segmentation

Transcription of speech is difficult, as it demands the denotation of a highly variable speech signal by a fixed set of transcription symbols. In the present thesis I used the symbols available in the International Phonetic Alphabet (The

International Phonetic Association, 2014), as well as "in-between" categories from the outset (see Stoel-Gammon, 2001). As such, the transcriptions were categorical but allowed three categories instead of two for any contrast (e.g., /s-ɕ/). This categorisation suggests that the "correct" category used in the present work might be narrower than in previous work (e.g., Lohmander et al., 2017b; Lundeborg Hammarström, 2019).

Regarding segmentation, consistent annotation of fricative onset and offset is important for analysis of duration or intensity that depend on the placement of the segment borders. However, the segmentation of fricative-vowel boundaries is not a trivial task (Jesus and Shadle, 2002; Shadle, 2011). While the segmentation of adult fricatives used in *Paper I* and *II* was relatively simple, the annotation of child speech proved to be challenging. Some of the children included in *Paper III*, produced lengthened and diffuse fricative-vowel transitions and silences between segments as well as bursts, spectral variability and silences within the fricative segment (see Figure 4.3 for some illustrative examples). This variability demanded careful consideration when annotating the regions to include in the acoustic analysis. I chose to be conservative in setting segment boundaries (including less rather than more), and when the onset or offset of the fricative could not be reliably determined, or if bursts or silences were present in the segment, a stable fricative region of at least 30 ms was selected for use in analysis. As a result of these inconsistencies in child productions, neither duration nor intensity was used in the acoustic description of the children's speech in *Paper III*.

7.2.3 Acoustic analysis

Choice of tokens

For adults, all tokens were judged to be correct and the choice of what sounds to include in the acoustic analysis was straightforward (i.e., all tokens without disturbances). For children, high variability in fricative production is expected, and what tokens to incorporate in acoustic analysis demands more thought. On the one hand, if one aims to describe speech on a functional level (e.g., measures that relate to speech intelligibility), it is problematic to exclude tokens that are transcribed as incorrect because speech errors are part of that child's speech. On the other hand, analysing all productions (including substitutions) would inflate category variability and, if the speech error involved substantial changes in place or manner of articulation, acoustic analysis may lead to erroneous inferences. In *Paper III*, the variability of child speech is described on two levels; broad, perceptually salient variability through transcription, and fine phonetic variability via acoustic analysis. The age-group analysis of spectral measures included only correct tokens so as to avoid variability from the large number of incorrect realisations in younger age groups. As I used an in-between transcription category throughout, some tokens that might have been transcribed as correct without the intermediate category were excluded from acoustic analysis. For the contrast measure, all productions from each child that had been transcribed as a sibilant fricative or affricate (i.e., both correct and incorrect) were included, to increase the ecological validity of the measure.

Spectral analysis

As previously acknowledged, the spectral estimation methods and frequency ranges used, differed between *Paper I, II* and *Paper III, IV*. The analysis of adult fricatives preceded the study of child productions, and part of the motivation for *Paper I* was to create a resource for cross-linguistic research. As such, I tried to make settings comparable to previous studies (e.g., Jongman et al., 2000). Specifically, the recordings were sampled at 44.1 kHz, but were down-sampled to 22 kHz prior to analysis (for static spectral measures). The default method for spectral estimation in Praat (i.e, FFT) was used for adult fricatives, due to convenience. During the preparation of *Paper III*, I became aware of the importance of appropriate spectral estimation (e.g., through the publication of Shadles 2023 paper). Hence, a reduced-variance spectral estimation method (MTS) was used for the child fricatives in the third paper. Additionally, the frequency range for child productions was adjusted to match studies of children's fricatives. Hence, a larger frequency range was used for spectral moments analysis in *Paper III* (0.3-15 kHz), as compared to *I* and *II*. Visual inspection of child spectra confirmed a significant amount of energy over 10 kHz, making the 11 kHz upper cut-off used for adult productions substandard.

With respect to spectral measures, global spectral peak and spectral moments were used in *Paper I, II* and *III*. In the first two papers, analysis methods were partly chosen to be comparable with previous work and partly due to availability of methods. In the third paper, spectral balance measures were added to improve the spectral description. In an effort to be comprehensive, I also planned to incorporate the F_M parameter (see Section 3.2) in the analysis, as it relates to front cavity size and could thus provide proxy for children's speech motor behaviour. However, despite trying different frequency ranges, the F_M algorithm did not reliably identify the lowest uncancelled spectral peak. In fact, some children did not have any peaks within the tested ranges and several children had two peaks of similar intensity within range. As such, the F_M parameter was not reported.

7.2.4 Statistical analysis

Summaries and visualisations of results in tables and figures are important to understand patterns in research data, and are an accessible way of presenting results to readers. Moreover, if an effect is robust, it is often visible to the naked eye. Nonetheless, with complex data, tables quickly become difficult to read and visualisations are insufficient to capture multidimensional variation. If one wants to understand multiplex and nested data, statistical modelling is often required.

Phonetic data is (almost) always hierarchical in nature, including clusters of data from specific participants, phonetic contexts and/or experimental conditions (e.g. fricative, vowel context and speaker in *Paper I*). To uncover such structures, multilevel, or mixed-effects, models are necessary. Including relevant nesting structure in the model will circumvent erroneous assumptions of independence (as repetitions of the same item from the same speaker cannot be regarded as independent from one another) and improve the validity of inferences for fixed effects (see e.g., Harel and McAllister, 2019)

In the analysis of dynamic speech phenomena, the time domain is introduced and auto-correlation of measures taken from adjacent time points needs to be handled (as acoustic measures from a number of adjacent time windows cannot be assumed to be independent of the neighbouring sample/s taken from the same sound).

The statistical methods of this thesis were chosen based on the research questions and the nature of the data; a generalised additive mixed model (GAMM) for the non-linear M1 trajectories in *Paper I*, random forests for the heteroskedastic and correlated spectral features in *Paper III* and Bayesian mixed-effects beta regression for the skewed and bounded VAS responses in *Paper IV*. Multilevel models with random effects were used to accommodate for participant and item effects, when applicable.

However, as a result of my status as a novice in statistics, the statistical models herein are not perfect. For example, as mentioned in *Paper I*, the inclusion of by-speaker trajectories in the GAMM would have been an interesting complement to the group-level analysis, to enable study of individual differences. However, we could not figure out how to include such smooths in the model without convergence issues. With regards to the Bayesian mixed-effects Beta regression in *Paper IV*, the model would likely have improved had I included minimal-word pair status as a predictor. It is possible that listeners' ratings, primarily of whole words, may have been influenced by minimal-word pair status. However, as the model was already highly parameterised and additional predictors might impede interpretation of the results, I decided not to add the predictor. I tried different priors and different scaling of VAS responses, which all yielded similar results, suggesting robust effects. The Bayesian implementation of the Beta regression produces a posterior distribution rather than point estimates, making quantification of uncertainty possible. For studies with potential clinical bearing, this is a particularly attractive feature.

With respect to communicating statistical results, model outputs that included p-values were reported, but I attempted to steer clear of phrases such as "statistically significant" in order to avoid over-reliance on this notorious value (Wagenmakers, 2007). Overall, my standpoint has been to embrace transparency and to be careful with the presentation of results, in order to avoid overly confident or conservative interpretations.

7.3 Ethical considerations

The data in *Paper III* and *IV* (i.e, child speech recordings and adult ratings of said recordings) involved young children. As children are particularly worthy of protection and as we planned to collect sensitive information concerning health (e.g., diagnosis of a speech-, language- or reading disorder), an ethical permit application was obtained prior to data collection. The registration number for the original approved ethical permit is 2019-02854[1]. However, during the COVID-19 pandemic, the Phonetics lab of the Department of Linguistics at Stockholm University closed down just as recruitment for the original project was about to start. Following discussions with my supervisors concerning time management (i.e., finishing on time), a number of changes were made to facilitate recruitment and recording. To accommodate for these changes, two amendments to the ethical permit application were created. The registration numbers for the two approved amendments are 2020-03306 and 2022-02168-02.

All children and their families were informed of the purpose of the study and that participation was voluntary. The children received a child version of the letter including an invitation to participate, descriptions of the tasks, a picture of the experiment leader (me) and explicit information that their participation was voluntary (e.g., "Du bestämmer själv om du vill vara med. Om du vill vara med men sen ändrar dig, är det helt okej." [You decide if you want to take part in the experiment. If you want to take part and then change your mind, that's perfectly fine.]). The caregivers received detailed information about the study and signed informed consent prior to recordings. The

[1] The original project aimed to explore the relationship between speech perception and - production in four- to six-year-old Swedish children with and without SSD (n= 60-120). A description of the original project can be found in Wikse Barrow et al. (2019).

families were given a cinema ticket for their time.

The data collected for *Paper I* and *II*, (i.e., adult speech recordings), was collected pending the approval of the ethical permit amendment for child recordings. The demographic questions and the speech material used in *Paper I* and *II* were designed to avoid eliciting any sensitive information, and was discussed extensively with my supervisors before data collection began. Prior to participation, the adult participants were given written and oral information concerning the study procedure as well as what personal data (e.g., age, gender) would be processed in the project. The information letters were modelled after the Swedish Ethical Review Authorities template. Participants were further informed that they could withdraw from the study at any time, without any consequences. All participants provided written informed consent prior to participation and received one cinema ticket as recompense for their time.

Data processing and storage

All data was pseudonymised early in the analysis pipeline and was treated with utmost care in order to protect the integrity of the participants. Processing and analysis of the data was performed on an encrypted and password protected computer in pseudonymised form, and the data was stored with high security, following the procedure described in the approved ethical permit.

7.3.1 Open science

To increase the transparency and reproducibility of the work presented herein, all steps in the analysis pipeline were documented, and shared via open source repositories, when possible. All papers were published Open Access so that stakeholders and other interested parties, as well as the research community, could gain access to the results.

In a broader context, open science practice can improve the quality of research through greater transparency, and thus contribute to increased faith in research (Allen and Mehler, 2019). At Stockholm University, the open science policy is very clear:

> Stockholm University strives for an open scientific system, where everyone has free and open access to scholarly texts, research results and research data. (Stockholm University Library, 2023)

However, to truly make code and data useful, one must not only share them, but also take efforts to make them as accessible as possible, for example by providing explicit and detailed comments and instructions for the code.

Notwithstanding the advantages of open science, ethical considerations demand that one be very careful about what information one shares, so as to protect the integrity of the participants. Within the scope of this thesis, I have only shared anonymised ratings and acoustic data (i.e. no audio or video whatsoever), with very sparse demographic information (e.g., only gender and age for children), to ensure that individual speakers cannot be identified from the data. The code to generate all results (models, tables, figures) in *Paper I, III* and *IV* has been uploaded to open-access repositories (see links in Chapter 6).

A note on open source software and code (credit where credit is due)

Publicly available software like Praat (Boersma and Weenink) has facilitated open science practice in phonetics by increasing the accessibility and reproducibility of results. In the present thesis, I exclusively use open source software and publicly available code for analyses. Praat (Boersma and Weenink) was used for annotation and segmentation of audio and for the acoustic analysis in *Paper I* and *II*. Statistical analyses were performed with R (R Core Team, 2013), using R Studio (RStudio Team, 2020). The `mgcv` (Wood, 2011) and `itsadug` (van Rij et al., 2020) packages were used to fit and evaluate the Generalised Additive Mixed Model in *Paper I*. In *Paper III*, the `spectRum` (Reidy, 2016a) package was used to create Multitaper spectra, and the `ranger` library (Wright and Ziegler, 2017) to grow Random forests. For *Paper IV*, Psychopy (Peirce, 2007) was used to implement the speech perception experiment and the `brms` library (Bürkner, 2017) was employed for the Bayesian mixed-effects Beta regression. Result visualisations and figures for publications were created with base R and the graphical library `ggplot` (Wickham, 2016).

7.4 Directions for future research

As relatively few phonetic studies have explored the production and perception of Swedish fricatives, there are a great many engaging directions for future research on the subject. In the interest of space, I briefly discuss a number of inquiries that I find particularly interesting below.

Studies on fricative coarticulation in adults and children, by means of spectral dynamics and fricative-to-vowel transitions would be highly informative regarding speech motor behaviour. Articulatory studies of voiceless fricatives would also be illuminating, with reference to individual differences in articulation and to the relationship between acoustics and articulation.

Additionally, exploring how acoustic characteristics of Swedish voiceless fricatives relate to listener evaluations would contribute to increased knowledge concerning functional speech communication (e.g., acoustic contrast in relation to intelligibility), and cross-language trends in speech perception.

With respect to dialectal, idiolectal and sociolectal variation, investigations of diachronic change in dialectal realisations of fricatives (e.g. in Northern Swedish, Finnish Swedish and Southern Swedish), as well as explorations of whether possible gender effects manifest differently in older and younger speakers, are fascinating directions for future research. Moreover, increased knowledge about socio-phonetic variation related to gender and multilingualism (e.g., in the realisation of /ɧ/ and /s/) would be a valuable contribution to the current state of knowledge.

Further explorations of the realisation of fricatives in Swedish children with SSD and adults with speech disorders would also be welcome, as would studies of the relationship between speech perception and production (e.g., regarding fricatives) in Swedish children. Note that this branch of inquiry necessitates the development of speech perception tests for Swedish, which are currently scarce.

In a broader context, explorations of the boundaries of variation and diversity in speech acquisition are important, for example through quantification of inconsistency and speech sound contrast in children with typical speech development.

7.5 Concluding remarks

This dissertation provides an acoustic description of Swedish voiceless fricative produced by adults and children, and investigates the influence of stimuli lexicality on category goodness judgements of children's sibilants.

Overall, spectral characteristics of the voiceless fricatives differed across place of articulation and between sibilants and non-sibilants. Substantial individual variation in the acoustics of Swedish voiceless fricatives was uncovered, both in adult and child speakers. Whether this variation is perceptible, and whether it conveys any socio-phonetic cues, remains to be seen. With regards to the perceptual investigation, an effect of lexicality was found for all listeners (both with and without clinical experience), in such a manner that words were rated as more prototypical than syllables. The results could indicate that the influence of experience cited in previous work is overstated.

Although this work contributes to an increased understanding of these complex sounds, many questioned regarding acoustic and perceptual characteristics of these voiceless fricatives remain unanswered - and even more questions have emerged through this work, which might inspire more research.

Svensk sammanfattning

Tonlösa frikativor är artikulatoriskt och akustiskt komplexa ljud som tillägnas relativt sent av barn. Det finns beskrivningar av tonlösa frikativor i andra språk, men för svenska är tidigare beskrivningar begränsade. Även kunskapen om svenska barns tillägnande av dessa komplexa ljud är något limiterad. Denna doktorsavhandling består av fyra delstudier som undersöker de svenska tonlösa frikativor /f, s, ɕ, ɧ/ från olika perspektiv. I studie I och II undersöks vuxna talares frikativor och deras statiska och dynamiska spektrala egenskaper. Studie III behandlar barns tonlösa frikativor och beskriver akustiska egenskaper, kontrast och individuell variation. I den fjärde studien undersöks effekten av lyssnarerfarenhet (logopeder vs. naiva lyssnare) och lexikalitet på bedömningar av barns tonlösa frikativor. Studien jämför skattningar av frikativor i olika lexikala kontexter (ord eller stavelse) på en Visuell Analog Skala (VAS).

De viktigaste bidragen kan i korthet beskrivas som 1) en beskrivning av de akustiska egenskaperna av svenska barns och vuxnas tonlösa frikativor och 2) en redogörelse för effekten av lexikalitet i VAS-bedömningar av barns tal. I förlängningen kan resultaten från denna avhandling bidra till ökad förståelse för den variation som kännetecknar typisk talutveckling. Den fjärde studiens resultat kan även vara av relevans för kliniska bedömningar av tal. För att ytterligare generalisera resultaten i denna avhandling behövs fler, mer omfattande studier som undersöker fler språkljudskontraster.

References

Allen, C. and Mehler, D. M. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS biology*, 17(5):e3000246. 69

Ancel, E. E., Smith, M. L., Rao, V. V., and Munson, B. (2023). Relating acoustic measures to listener ratings of children's productions of word-initial//and/w. *Journal of Speech, Language, and Hearing Research*, 66(9):3413–3427. 45

Apfelbaum, K. S., Kutlu, E., McMurray, B., and Kapnoula, E. C. (2022). Don't force it! gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America*, 152(6):3728–3745. 44

Babel, M. and McGuire, G. (2013). Listener expectations and gender bias in nonsibilant fricative perception. *Phonetica*, 70(1-2):117–151. 42

Bang, H. Y., Clayards, M., and Goad, H. (2017). Compensatory strategies in the developmental patterns of english /s/: Gender and vowel context effects. *Journal of Speech Language and Hearing Research*, 60(3):571–591. 34, 35

Baudonck, N., Buekers, R., Gillebert, S., and Van Lierde, K. (2009). Speech intelligibility of flemish children as judged by their parents. *Folia Phoniatrica et Logopaedica*, 61(5):288–295. 43

Baum, S. R. and McNutt, J. C. (1990). An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics*, 18(1):51–63. 25, 33

Behrens, S. and Blumstein, S. E. (1988a). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *The Journal of the Acoustical Society of America*, 84(3):861–867. 41

Behrens, S. J. and Blumstein, S. E. (1988b). Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics*, 16(3):295–298. 8, 47

Bergelson, E., Soderstrom, M., Schwarz, I.-C., Rowland, C. F., Ramírez-Esparza, N., R. Hamrick, L., Marklund, E., Kalashnikova, M., Guez, A., Casillas, M., et al. (2023). Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences*, 120(52):e2300671120. 28

Berk, S. and Lillo-Martin, D. (2012). The two-word stage: Motivated by linguistic or cognitive constraints? *Cognitive Psychology*, 65(1):118–140. 23

Blacklock, O. (2004). *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods.* PhD thesis, University of Southampton. 10

Boersma, P. and Weenink, D. Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/, year = 1994-2021. 70

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32. 51

Bruce, G. (2010). *Vår fonetiska geografi : om svenskans accenter, melodi och uttal*. Studentlitteratur, Lund, 1st edition. 3

Bruce, G. and Engstrand, O. (2006). The phonetic profile of Swedish. *Sprachtypologie und Universalienforschung*. 3

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28. 70

Chappell, W., García, C., and Davidson, J. (2023). Sociophonetics and fricatives. In Stelluf, C., editor, *The Routledge Handbook of Sociophonetics*, pages 176–194. Routledge. 16

Child Speech Disorder Research Network (2017a). Good practice guidelines for the analysis of child speech. Published on RCSLT members webpage (www.rcslt.org) and Bristol Speech and Language Therapy Research Unit webpage (www.speech-therapy.org.uk). 45

Child Speech Disorder Research Network (2017b). Good practice guidelines for the transcription of children's speech in clinical practice and research. Published on RCSLT members webpage (www.rcslt.org) and Bristol Speech and Language Therapy Research Unit webpage (www.speech-therapy.org.uk). 45

Coniglio, E. A., Chung, H., and Schellinger, S. K. (2022). Perception of children's productions of /l/: Acoustic correlates and effects of listener experience. *Folia Phoniatrica et Logopaedica*, 74(6):392–406. 45

Coplan, J. and Gleason, J. R. (1988). Unclear speech: Recognition and significance of unintelligible speech in preschool children. *Pediatrics*, 82(3):447–452. 23

Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., Scaff, C., Yankowitz, L., and Seidl, A. (2021a). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, 24(5):e13090. 22

Cychosz, M., Munson, B., and Edwards, J. R. (2021b). Practice and experience predict coarticulation in child speech. *Language Learning and Development*, 17(4):366–396. 26

Daniel, G. R. and McLeod, S. (2017). Children with speech sound disorders at school: Challenges for children, parents and teachers. *Australian Journal of Teacher Education (Online)*, 42(2):81–101. 30

de Boysson-Bardies, B. and Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67(2):297–319. 24

Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports*, 1(3):189–196. 29

Dodd, B., McIntosh, B., Crosbie, S., and Holm, A. (2023). Diagnosing inconsistent phonological disorder: quantitative and qualitative measures. *Clinical Linguistics & Phonetics*, pages 1–24. 29

Edwards, J. and Beckman, M. E. (2008a). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics*, 22(12):937–56. 46, 64

Edwards, J. and Beckman, M. E. (2008b). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development*, 4(2):122–156. 24

Edwards, J., Munson, B., and Beckman, M. E. (2011). Lexicon–phonology relationships and dynamics of early language development–a commentary on stoel-gammon's 'relationships between lexical and phonological development in young children'. *Journal of Child Language*, 38(1):35–40. 25

Engstrand, O. (1999). Swedish. In The International Phonetic Association, editor, *Handbook of the International Phonetic Association*, pages 140–142. Cambridge University Press, Cambridge. xv, xvii, 4, 5, 17

Ethnologue (2024). Ethnologue: Swedish (online). `https://www.ethnologue.com/language/swe/`. Accessed: 2024-01-02. 3

Fabiano-Smith, L. and Goldstein, B. A. (2010). Phonological acquisition in bilingual spanish–english speaking children. *Journal of Speech, Language, Hearing Research*, 53:160–178. 27

Flipsen Jr, P. (1995). Speaker-listener familiarity: Parents as judges of delayed speech intelligibility. *Journal of Communication Disorders*, 28(1):3–19. 43

Flipsen Jr, P. (2015). Emergence and prevalence of persistent and residual speech errors. *Seminars in Speech and Language*, 36:217–223. 32

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, 84(1):115–123. 8, 11, 47, 51

Fox, R. A. and Nissen, S. L. (2005). Sex-related acoustic changes in voiceless English fricatives. *Journal of Speech, Language and Hearing Research*, 48:753–765. 16, 34

Fox-Boyer, A., Lavaggi, S., and Fricke, S. (2022). Phonological variations in typically-developing Italian-speaking children aged 3; 0-4; 11. *Clinical Linguistics & Phonetics*, 36(2-3):241–259. 26

Fung, P., Schertz, J., and Johnson, E. K. (2021). The development of gendered speech in children: Insights from adult L1 and L2 perceptions. *JASA Express Letters*, 1(1). 27

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1):110–125. 45, 53

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*, 42(2):382–397. 24

Goldstein, B. A. and McLeod, S. (2012). Typical and atypical multilingual speech acquisition. In McLeod, S. and Goldstein, B. A., editors, *Multilingual aspects of speech sound disorders in children*. Multilingual Matters. 27

Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, 43(1):239–255. 24

Göthlin, C. (2022). En kartläggning av talavvikelsemönster hos 5-åriga svensktalande barn med typisk tal- och språkutveckling [Master's Thesis, Department of Clinical Intervention, Technology and Science, Karolinska Institutet]. 31, 38, 59

Hambly, H., Wren, Y., McLeod, S., and Roulstone, S. (2013). The influence of bilingualism on speech production: A systematic review. *International Journal of Language & Communication Disorders*, 48(1):1–24. 27, 28

Harel, D., Hitchcock, E. R., Szeredi, D., Ortiz, J., and Byun, T. M. (2017). Finding the experts in the crowd: Validity and reliability of crowdsourced measures of children's gradient speech contrasts. *Clinical Linguistics & Phonetics*, 31(1):104–117. 44, 45, 61

Harel, D. and McAllister, T. (2019). Multilevel models for communication sciences and disorders. *Journal of Speech, Language, and Hearing Research*, 62(4):783–801. 67

Hearnshaw, S., Baker, E., Pomper, R., McGregor, K. K., Edwards, J., and Munro, N. (2023). The relationship between speech perception, speech production, and vocabulary abilities in children: Insights from by-group and continuous analyses. *Journal of Speech, Language, and Hearing Research*, 66(4):1173–1191. 21

Hedrick, M. S. and Ohde, R. N. (1993). Effect of relative amplitude of frication on perception of place of articulation. *The Journal of the Acoustical Society of America*, 94(4):2005–2026. 41

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83. 28, 63

Holliday, J. J., Beckman, M. E., and Mays, C. (2010). Did you say susi or shushi? Measuring the emergence of robust fricative contrasts in English-and Japanese-acquiring children. In *Eleventh Annual Conference of the International Speech Communication Association*. 11, 15, 33, 34, 38, 60

Holliday, J. J., Reidy, P. F., Beckman, M. E., and Edwards, J. (2015). Quantifying the Robustness of the English Sibilant Fricative Contrast in Children. *Journal of Speech, Language and Hearing Research*, 58(3):622–637. 33, 35, 38, 42, 45, 53, 60

Holm, A., Van Reyk, O., Crosbie, S., De Bono, S., Morgan, A., and Dodd, B. (2023). Preschool children's consistency of word production. *Clinical Linguistics & Phonetics*, 37(3):223–241. 29

Howson, P. J. and Redford, M. A. (2022). A Cross-Sectional Age Group Study of Coarticulatory Resistance: The Case of Late-Acquired Voiceless Fricatives in English. *Journal of Speech, Language, and Hearing Research*, 65(9):3316–3336. 7, 8, 26, 35

Hughes, G. W. and Halle, M. (1956). Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America*, 28(2):303–310. 47

Hustad, K. C., Mahr, T. J., Natzke, P., and Rathouz, P. J. (2021). Speech development between 30 and 119 months in typical children I: Intelligibility growth curves for single-word and multiword productions. *Journal of Speech, Language, and Hearing Research*, 64(10):3707–3719. 23, 27

International Expert Panel on Multilingual Children's Speech (2012). Multilingual children with speech sound disorders: Position paper. Bathurst, NSW, Australia: Research Institute for Professional Practice, Learning and Education (RIPPLE), Charles Sturt University. Retrieved from http://www.csu.edu.au/research/multilingual-speech/position-paper. 24

Iskarous, K., Shadle, C. H., and Proctor, M. I. (2011). Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America*, 129(2):944–954. 15, 47, 57

Iuzzini-Seigel, J., Hogan, T. P., and Green, J. R. (2017). Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language, and Hearing Research*, 60(5):1194–1210. 29

Jesus, L. M. T. and Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, 30(3):437–464. 14, 47, 65

Joffe, V. and Pring, T. (2008). Children with phonological problems: a survey of clinical practice. *International Journal of Language & Communication Disorders*, 43(2):154–64. 45

Johnson, K. (2012). *Acoustic and Auditory Phonetics*. Wiley-Blackwell, Malden, MA, 3rd edition. 7, 8, 9, 57

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of english fricatives. *The Journal of the Acoustical Society of America*, 108(3):1252–1263. 8, 11, 14, 16, 47, 51, 56, 66

Julien, H. M. and Munson, B. (2012). Modifying speech to children based on their perceived phonetic accuracy. *Journal of Speech, Language and Hearing Research*, 55(6):1836–1849. 41, 44

Karlsson, F. (2006). *The acquisition of contrast: a longitudinal investigation of initial s + plosive cluster development in Swedish children*. PhD thesis, Umeå University, Faculty of Arts, Philosophy and Linguistics. 39, 51

Kehoe, M. and Philippart de Foy, M. (2023). The Development of Alveolar and Alveopalatal Fricatives in French-Speaking Monolingual and Bilingual Children. *Journal of Speech, Language, and Hearing Research*, 66(2):475–502. 35

Kent, R. D. (1992). The biology of phonological development. In Ferguson, C. A., Menn, L., and toel Gammon, C. S., editors, *Phonological development: Models, research, implications*, pages 65–90. New York Press. 24

Kidd, E. and Donnelly, S. (2020). Individual differences in first language acquisition. *Annual Review of Linguistics*, 6:319–340. 26

Klintö, K. and Lohmander, A. (2023). Perceptual assessment of cleft palate speech—Bridging the gap from research to clinical practice—the Swedish perspective. *Perspectives of the ASHA Special Interest Groups*, 8(5):986–1002. 46

Knight, R. A., Bandali, C., Woodhead, C., and Vansadia, P. (2018). Clinicians' views of the training, use and maintenance of phonetic transcription in speech and language therapy. *International Journal of Language & Communication Disorders*, 53(4):776–787. 46

Kochetov, A. (2017). Acoustics of Russian voiceless sibilant fricatives. *Journal of the International Phonetic Association*, 47(3):321–348. 11, 15, 56

Koenig, L. L., Lucero, J. C., and Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *The Journal of the Acoustical Society of America*, 124(5):3158–3170. 35

Koenig, L. L., Shadle, C. H., Preston, J. L., and Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, Hearing Research*, 56:1175–1189. 10, 11, 14, 15, 38, 47, 51

Kong, E. and Edwards, J. (2011). Individual Differences in Speech Perception: Evidence from Visual Analogue Scaling and Eye-Tracking. In *Proceedings of the 17th International Congress of Phonetic Sciences*, pages 1126–1129. 42

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5(11):831–43. 22

Lagerberg, T. B., Anrep-Nordin, E., Emanuelsson, H., and Strömbergsson, S. (2021). Parent rating of intelligibility: A discussion of the construct validity of the Intelligibility in Context Scale (ICS) and normative data of the Swedish version of the ICS. *International Journal of Language & Communication Disorders*, 56(4):873–886. 23

Lee, A., Whitehill, T. L., and Ciocca, V. (2009). Effect of listener training on perceptual judgement of hypernasality. *Clinical Linguistics & Phonetics*, 23(5):319–334. 46

Lee, S., Potamianos, A., and Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105(3):1455–1468. 35

Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese*. PhD thesis, The Ohio State University. 53

Li, F. (2017). The development of gender-specific patterns in the production of voiceless sibilant fricatives in Mandarin Chinese. *Linguistics*, 55(5):1021–1044. 34

Li, F., Edwards, J., and Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *J Phon*, 37(1):111–124. 25, 32, 33, 35, 59

Li, F. and Munson, B. (2016). The Development of Voiceless Sibilant Fricatives in Putonghua-Speaking Children. *Journal of Speech Language and Hearing Research*, 59(4):699–712. 32, 35

Li, F., Munson, B., Edwards, J., Yoneyama, K., and Hall, K. (2011). Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound. *The Journal of the Acoustical Society of America*, 129(2):999–1011. 43

Li, F., Rendall, D., Vasey, P. L., Kinsman, M., Ward-Sutherland, A., and Diano, G. (2016). The development of sex/gender-specific /s/ and its relationship to gender identity in children and adolescents. *Journal of Phonetics*, 57:59–70. 34

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5):358. 44

Lindblad, P. (1980). *Svenskans sje- och tje-ljud i ett allmänfonetiskt perspektiv*. PhD thesis, Lund University. 3, 4, 8, 16, 17, 18, 47, 56, 58

Littlejohn, M. and Maas, E. (2023). How to cut the pie is no piece of cake: Toward a process-oriented approach to assessment and diagnosis of speech sound disorders. *International Journal of Language & Communication Disorders*. 28

Lohmander, A., Holm, K., Eriksson, S., and Lieberman, M. (2017a). Observation method identifies that a lack of canonical babbling can indicate future speech and language problems. *Acta Paediatrica*, 106(6):935–943. 22

Lohmander, A., Lundeborg, I., and Persson, C. (2017b). SVANTE - The Swedish Articulation and Nasality Test - Normative data and a minimum standard set for cross-linguistic comparison. *Clinical Linguistics & Phonetics*, 31(2):137–154. xvii, 26, 30, 31, 38, 51, 58, 59, 64, 65

Lundeborg Hammarström, I. (2019). Manual till LINUS 2.0. LINköpingsUnderSökningen 2.0 - Ett fonologiskt bedömningsmaterial för barn från 3 år. *Linköping University*. xvii, 30, 31, 38, 51, 58, 64, 65

Luo, S., Min, Y., Meng, Z., and Ji, R. (2023). Acoustic and perceptual categorization of sibilants for mandarin children with ankyloglossia. *American Journal of Speech-Language Pathology*, 32(4):1489–1500. 35

Macken, M. A. and 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*, 7:41–74. 25

Maniwa, K., Jongman, A., and Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, 125(6):3962–3973. 15, 16

McCune, L. and Vihman, M. M. (2001). Early phonetic and lexical development. *Journal of Speech, Language, Hearing Research*, 44. 22

McLeod, S. and Baker, E. (2014). Speech- language pathologists practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics & Phonetics*, 28(7-8):508–531. 45

McLeod, S. and Crowe, K. (2018). Children's Consonant Acquisition in 27 Languages: A Cross-Linguistic Review. *American Journal of Speech-Language Pathology*, 27(4):1546–1571. 24, 25, 27, 32, 53

McLeod, S., Daniel, G., and Barr, J. (2013). "when he's around his brothers… he's not so quiet": The private and public worlds of school-aged children with speech sound disorder. *Journal of communication disorders*, 46(1):70–83. 30

McLeod, S., Harrison, L. J., and 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. 23

McLeod, S. and Masso, S. (2019). Screening children's speech: The impact of imitated elicitation and word position. *Language, Speech, and Hearing Services in Schools*, 50(1):71–82. 26, 64

McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, 152(6):3819–3842. 44

Meyer, M. K. and Munson, B. (2021). Clinical experience and categorical perception of children's speech. *International Journal of Language & Communication Disorders*, 56(2):374–388. 43, 45, 61

Miodonska, Z., Badura, P., and Mocko, N. (2022). Noise-based acoustic features of polish retroflex fricatives in children with normal pronunciation and speech disorder. *Journal of Phonetics*, 92:101149. 35

Morgan, L. and Wren, Y. E. (2018). A systematic review of the literature on early vocalizations and babbling patterns in young children. *Communication Disorders Quarterly*, 40(1):3–14. 22

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language & Hearing Research*. 24

Munson, B. (2004). Variability in /s/ production in children and adults. *Journal of Speech, Language, Hearing Research*, 47:58–69. 15, 35, 47

Munson, B., Crocker, L., Pierrehumbert, J. B., Owen-Anderson, A., and Zucker, K. J. (2015). Gender typicality in children's speech: A comparison of boys with and without gender identity disorder. *The Journal of the Acoustical Society of America*, 137(4):1995–2003. 27

Munson, B., Edwards, J., Schellinger, S. K., Beckman, M. E., and Meyer, M. K. (2010). Deconstructing phonetic transcription: covert contrast, perceptual bias, and an extraterrestrial view of vox humana. *Clinical Linguistics & Phonetics*, 24(4-5):245–60. 25, 33, 43, 53

Munson, B., Johnson, J. M., and Edwards, J. (2012a). 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. 43, 45, 46, 53, 61

Munson, B., Lackas, N., and Koeppe, K. (2022). Individual differences in the development of gendered speech in preschool children: Evidence from a longitudinal study. *Journal of Speech, Language, and Hearing Research*, 65(4):1311–1330. 27

Munson, B., Ryherd, K., and Kemper, S. (2017a). Implicit and explicit gender priming in English lingual sibilant fricative perception. *Linguistics*, 55(5):1073–1107. 43

Munson, B., Schellinger, S. K., and Carlson, K. U. (2012b). Measuring Speech-Sound Learning Using Visual Analog Scaling. *Perspectives on Language Learning and Education*, 19(1). 53

Munson, B., Schellinger, S. K., and Edwards, J. (2017b). Bias in the perception of phonetic detail in children's speech: A comparison of categorical and continuous rating scales. *Clinical Linguistics & Phonetics*, 31(1):56–79. 41, 44

Munson, B. and Urberg Carlson, K. (2016). An Exploration of Methods for Rating Children's Productions of Sibilant Fricatives. *Speech, Language and Hearing*, 19(1):36–45. 44, 45, 53

Namasivayam, A. K., Coleman, D., O'Dwyer, A., and Van Lieshout, P. (2020). Speech sound disorders in children: An articulatory phonology perspective. *Frontiers in Psychology*, 10:2998. 24

Narayanan, S. S., Alwan, A. A., and Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 98(3):1325–1347. 7

Nayeb, L., Wallby, T., Westerlund, M., Salameh, E.-K., and Sarkadi, A. (2015). Child healthcare nurses believe that bilingual children show slower language development, simplify screening procedures and delay referrals. *Acta Paediatrica*, 104(2):198–205. 28

Nettelbladt, U. and Salameh, E.-K. (2007). *Språkutveckling och språkstörning hos barn. Del 1, Fonologi, grammatik, lexikon*. Studentlitteratur, Lund. 29

Newman, R. S., Clouse, S. A., and Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, 109(3):1181–1196. 42

Nirgianaki, E. (2014). Acoustic characteristics of Greek fricatives. *The Journal of the Acoustical Society of America*, 135(5):2964–2976. 8, 11, 15, 47, 56

Nissen, S. L. and Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America*, 118(4):2570–2578. 11, 35

Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language, and Hearing Research*, 32(1):120–132. 16, 26

Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., and Tiede, M. (2018). How do children organize their speech in the first years of life? insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research*, 61(6):1355–1368. 26

Oller, D. K., Eilers, R. E., Neal, A. R., and Schwartz, H. K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of Communication Disorders*, 32(4):223–245. 22

Pace, A., Luo, R., Hirsh-Pasek, K., and Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, 3(2017):285–308. 27

Palo, J. (2022). En kartläggning av talavvikelser hos svenska 3-åriga barn med typisk tal- och språkutveckling [Master's Thesis, Department of Clinical Intervention, Technology and Science, Karolinska Institutet]. 31, 38, 59

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2):8–13. 70

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 70

Reidy, P. (2016a). spectRum [code]. https://github.com/patrickreidy/spectRum.git. 70

Reidy, P. F. (2015). A comparison of spectral estimation methods for the analysis of sibilant fricatives. *The Journal of the Acoustical Society of America*, 137(4):EL248–EL254. 10

Reidy, P. F. (2016b). Spectral dynamics of sibilant fricatives are contrastive and language specific. *The Journal of the Acoustical Society of America*, 140(4):2518–2529. 15, 47, 57

Riad, T. (2014). *The phonology of Swedish*. Oxford University Press. 3, 4, 17, 49

Rinaldi, P., Pasqualetti, P., Volterra, V., and Caselli, M. C. (2023). Gender differences in early stages of language development. some evidence and possible explanations. *Journal of Neuroscience Research*, 101(5):643–653. 27

Romeo, R., Hazan, V., and Pettinato, M. (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *The Journal of the Acoustical Society of America*, 134(5):3781–3792. 33, 34, 38, 47, 49, 50, 58, 60

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA. 70

Sand, A., Hagberg, E., and Lohmander, A. (2022). On the benefits of speech-language therapy for individuals born with cleft palate: A systematic review and meta-analysis of individual participant data. *Journal of Speech, Language, and Hearing Research*, 65(2):555–573. 62

Schellinger, S. K., Munson, B., and Edwards, J. (2017). Gradient perception of children's productions of /s/ and /theta/: A comparative study of rating methods. *Clinical Linguistics & Phonetics*, 31(1):80–103. 44, 45, 53

Schouten, B., Gerrits, E., and Van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41(1):71–80. 44

Schwab, J. F. and Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(4):264–275. 27

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., and Fletcher, P. (1996). Covert contrast as a stage in the acquisition of phonetics and phonology. In Broe, M. B. and Pierrehumbert, J. B., editors, *Papers in laboratory phonology V: Acquisition and the lexicon*, pages 194–207. Cambridge University Press. 25

Shadle, C. H. (2011). Acoustics and aerodynamics of fricatives. In Cohn, A., Fougeron, C., and Huffman, M., editors, *The Oxford Handbook of Laboratory Phonology*, pages 512–525. Oxford University Press Oxford. 8, 9, 10, 16, 65

Shadle, C. H. (2023). Alternatives to moments for characterizing fricatives: Reconsidering Forrest et al.(1988). *The Journal of the Acoustical Society of America*, 153(2):1412–1426. 10, 11, 14, 15, 51, 61, 66

Shadle, C. H., Chen, W.-R., Koenig, L. L., and Preston, J. L. (2023). Refining and extending measures for fricative spectra, with special attention to the high-frequency range. *The Journal of the Acoustical Society of America*, 154(3):1932–1944. 11, 14, 56

Shields, R. and Hopf, S. C. (2023). Intervention for residual speech errors in adolescents and adults: A systematised review. *Clinical Linguistics & Phonetics*, pages 1–24. 32

Shosted, R. (2008). Acoustic characteristics of Swedish dorsal fricatives. *Journal of the Acoustical Society of America*, 123(5):3888. 4, 17, 18, 47, 49, 56, 57, 58

Shriberg, L. D. and Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders*, 47(3):256–270. 30

Skahan, S. M., Watson, M., and Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology*, 16(3):246–259. 45

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., and Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55(4):779–798. 24

Souza, P., Gehani, N., Wright, R., and McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology*, 24(08):689–700. 43

Stockholm University Library (2023). Open science - stockholm univeristy library (online). `https://www.su.se/stockholm-university-library/about-the-library/open-science`. Accessed: 2023-12-07. 69

Stoel-Gammon, C. (2001). Transcribing the Speech of Young Children. *Topics in Language Disorders*, 21(4):12–21. 65

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, 38(1):1–34. 21, 22, 23

Stoel-Gammon, C. and Pollock, K. (2008). Vowel development and disorders. In Ball, Martin, J., Perkins, Michael, R., Müller, N., and Howard, S., editors, *The Handbook of Clinical Linguistics*, chapter 33, pages 525–548. John Wiley Sons, Ltd. 24

Stokes, S. F. and Surendran, D. (2005). Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children. *Journal of Speech, Language, and Hearing Research*, 48:577–591. 24

Stringer, H., Cleland, J., Wren, Y., Rees, R., and Williams, P. (2023). Speech sound disorder or DLD (phonology)? Towards a consensus agreement on terminology. *International Journal of Language & Communication Disorders*. 29

Strömbergsson, S., Salvi, G., and House, D. (2015). Acoustic and perceptual evaluation of category goodness of /t/ and /k/ in typical and misarticulated children's speech. *The Journal of the Acoustical Society of America*, 137(6):3422–35. 25, 44, 45

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/ in glaswegian. In Cole, J. and Hualde, J. I., editors, *Laboratory Phonology 9*, pages 65–86. Gruyter Mouton. 16

Swedish monophthongs chart (2009). Swedish monophthongs chart — Wikipedia, the free encyclopedia. [Online; accessed 1-July-2024]. xv, 4

Tambyraja, S. R., Farquharson, K., and Justice, L. (2020). Reading risk in children with speech sound disorder: Prevalence, persistence, and predictors. *Journal of Speech, Language, and Hearing Research*, 63(11):3714–3726. 30

The International Phonetic Association (2014). *Handbook of the International Phonetic Association*. Cambridge University Press, Cambridge. 64

Unsworth, S. (2013). Current issues in multilingual first language acquisition. *Annual Review of Applied Linguistics*, 33:21–50. 27, 28

Urberg-Carlson, K., Munson, B., and Kaiser, E. (2009). Gradient measures of children's speech production: Visual analog scale and equal appearing interval scale measures of fricative goodness. *The Journal of the Acoustical Society of America*, 125(4_Supplement):2529–2529. 44

van Rij, J., Wieling, M., Baayen, R. H., and van Rijn, H. (2020). itsadug: Interpreting time series and autocorrelated data using gamms. R package version 2.4. 70

Veríssimo, A., van Borsel, J., and de Britto Pereira, M. (2012). Residual /s/ and /r/ distortions: The perspective of the speaker. *International Journal of Speech-Language Pathology*, 14(2):183–186. 32

Vihman, M. and Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics*, 45. 29

Vihman, M. M., Ferguson, C. A., and Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*, 7(1):3–40. 26

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5):779–804. 68

Waring, R. and 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 & Communication Disorders*, 48(1):25–40. 29

Weismer, G. (2008). Speech intelligibility. In Ball, M. J., Perkins, M. R., Müller, N., and Howard, S., editors, *The Handbook of Clinical Linguistics*, chapter 35, pages 568–582. John Wiley & Sons, Incorporated. 23

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. 70

Wikse Barrow, C., Körner, K., and Strömbergsson, S. (2023a). A survey of swedish speech-language pathologists' practices regarding assessment of speech sound disorders. *Logopedics Phoniatrics Vocology*, 48(1):23–34. 28, 45, 46

Wikse Barrow, C., Strömbergsson, S., and Heldner, M. (2019). A multidimensional investigation of covert contrast in swedish acquiring childrenś speech - a project description. In *Proceedings of FONETIK 2019*. Zenodo. 68

Wikse Barrow, C., Włodarczak, M., Thörn, L., and Heldner, M. (2023b). Erratum: Static and dynamic spectral characteristics of swedish voiceless fricatives [j. acoust. soc. am. 152 (5), 2588–2600 (2022)]. *The Journal of the Acoustical Society of America*, 153(3):1933–1933. 48

Witte, E. and Köbler, S. (2019). Linguistic materials and metrics for the creation of well-controlled Swedish speech perception tests. *Journal of Speech, Language, and Hearing Research*, 62(7):2280–2294. 38, 59

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36. 70

Wren, Y., Miller, L. L., Peters, T. J., Emond, A., and Roulstone, S. (2016). Prevalence and predictors of persistent speech sound disorder at eight years old: Findings from a population cohort study. *Journal of Speech, Language, and Hearing Research*, 59(4):647–673. 27

Wren, Y., Pagnamenta, E., Orchard, F., Peters, T. J., Emond, A., Northstone, K., Miller, L. L., and Roulstone, S. (2023). Social, emotional and behavioural difficulties associated with persistent speech disorder in children: A prospective population study. *JCPP advances*, 3(1):e12126. 30

Wren, Y., Pagnamenta, E., Peters, T. J., Emond, A., Northstone, K., Miller, L. L., and Roulstone, S. (2021). Educational outcomes associated with persistent speech disorder. *International Journal of Language & Communication Disorders*, 56(2):299–312. 30

Wright, M. N. and Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1):1–17. 70

Yu, M. E., Cooper, A., and Johnson, E. K. (2023). Who speaks "kid?" how experience with children does (and does not) shape the intelligibility of child speech. *Journal of Experimental Psychology: Human Perception and Performance*. 23, 43, 61

Zharkova, N. (2018). An ultrasound study of the development of lingual coarticulation during childhood. *Phonetica*, 75(3):245–271. 26

Zharkova, N. (2021). Development of the voiceless sibilant fricative contrast in three-year-olds: an ultrasound and acoustic study. *Journal of Child Language*, 48(6):1126–1149. 33, 35

Zharkova, N., Hardcastle, W. J., and Gibbon, F. E. (2018). The dynamics of voiceless sibilant fricative production in children between 7 and 13 years old: An ultrasound and acoustic study. *The Journal of the Acoustical Society of America*, 144(3):1454–1466. 47, 57

Zharkova, N., Hewlett, N., and Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, 15(1):118–140. 26

Żygis, M., Pape, D., Jaskuła, M., and Koenig, L. L. (2023). Do children better understand adults or themselves? an acoustic and perceptual study of the complex sibilant system of polish. *Journal of Phonetics*, 100:101227. 32, 35, 41, 42