UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS

Acquisition of English fricatives by Vietnamese users of the ELSA app: An Acoustic Study

Mestrado Bolonha em Linguística

Filipa Isabel Filipe Pinto

2023

Relatório de Estágio especialmente elaborado para a obtenção do grau de Mestre, orientado pela Professora Doutora Helena Gorete Silva Moniz e pelo Mestre Raphael Girard

Agradecimentos

Existem muitas pessoas às quais poderia agradecer, no entanto, o número de páginas não seria suficiente e por isso tentarei inserir todos os grupos neste pequeno espaço.

Sinto que devo começar por agradecer à minha família, tanto aos meus pais como ao meu irmão por todo o apoio e força que me deram para seguir o que mais gostava. Julgo que tudo começou em criança, quando me incentivaram a estudar e principalmente a ganhar o bichinho da curiosidade para conhecer novas áreas e saber mais, apenas para saber. Mais do que um incentivo, deram-me a possibilidade de seguir os meus estudos e fazer algo que me deixa feliz, o que nem sempre foi fácil e obrigou a muito esforço da parte deles.

Não poderia deixar de agradecer aos meus amigos que sempre me apoiaram e tornaram esta jornada muito mais feliz e divertida e com um especial agradecimento às minhas amigas ruemianas, que me aturaram nos dias mais difíceis e me confortaram nas nossas longas noites de chá.

Em particular, gostaria de agradecer ao meu supervisor na ELSA, Mestre Raphael Girard, pela sua ajuda e tempo. Não seria possível concluir este trabalho sem o seu apoio, nem sem a sua paciência. Para além de um fantástico mentor que me proporcionou oportunidades de aprendizagem que nunca esperei ter, tornou-se um colega e estimado amigo.

Não poderia também deixar de agradecer ao doutor Xavier Anguera pela possibilidade de realizar este estágio numa empresa tão notável como a ELSA Speak e a permitir que eu aprendesse e desenvolvesse as minhas competências e capacidades num ambiente tão agradável como o que existe em toda a equipa de pesquisa.

Por fim, mas não menos importante, tenho de agradecer à professora doutora Helena Moniz, minha orientadora, pela sua disponibilidade e paciência neste que foi um longo caminho.

Resumo

O principal objetivo desta tese é analisar a produção de fricativas não vozeadas, por utilizadores da aplicação ELSA[1] Speak. Esta tese teve por base o estágio realizado na empresa para obtenção do grau de mestre em Linguística.

O trabalho realizado ao longo do estágio teve como finalidade ajudar e melhorar a aprendizagem dos falantes de uma segunda língua, neste caso a língua vietnamita. Ao entender as produções dos falantes, torna-se mais simples proporcionar um melhor *feedback*, o que consequentemente, deixa o utilizador mais satisfeito e oferece uma melhor experiência ao utilizador da aplicação. De forma a melhorar a deteção de erros da aplicação, a cobertura e precisão de exemplos e estruturas a anotar tinha de ser melhorada, assim como as anotações necessitavam de ser mais específicas e de terem rótulos que fossem para além de "OK" ou "ERRO", inicialmente utilizadas. Assim sendo, a anotação fonética foi a principal tarefa deste estágio. Outro dos motivos para esta análise, foi a relevância académica sobre o assunto, uma vez que existem alguns estudos conhecidos sobre as características acústicas de fricativas (Stevens, 1999; Jongman *et al.*, 2000), porém há menos estudos sobre aquisição de L2 de fricativas de inglês (Kitikanan, 2016), particularmente sobre a língua vietnamita.

Deste modo, este trabalho pode ser dividido em três grandes partes.

A primeira parte é uma parte teórico-prática (capítulo 2) aborda a empresa em que o estágio foi realizado, a ELSA Speak. Nesta parte analisam-se tanto as equipas que constituem a empresa, bem como a aplicação desenvolvida pela mesma. Esta parte também se destaca por uma explicação dos vários trabalhos desenvolvidos ao longo do estágio e observações quanto ao que foi realizado durante esse período. De modo a realizar esta análise foram ouvidos e anotados vários áudios. Primeiramente e numa fase de aprendizagem, foram analisados vários fones, desde oclusivas a ditongos, e posteriormente o grupo de fones para análise mais aprofundada, as fricativas não vozeadas. Estes fones foram escolhidos para uma maior análise

[1] ELSA significa *English Language Speech Assistant (Assistente de Fala da Língua Inglesa).*

devido à dificuldade de aquisição que os falantes vietnamitas demonstram para a sua produção.

A segunda parte é teórica e aborda a revisão da literatura (capítulo 3). Primeiramente, é realizada uma análise sobre as fricativas em contraste com outras classes de sons, seguido de uma breve revisão das fricativas na língua inglesa, variante de inglês americano padrão. De seguida, é efetuada uma pesquisa sobre análise acústica e aprofunda-se o que as várias medidas acústicas podem revelar sobre as fricativas, sendo que este aprofundamento tem como base os dados encontrados na língua inglesa. Consequentemente, é efetuada uma análise relativamente à língua vietnamita: começando com uma breve explicação do que constitui a língua, ou seja, o inventário fonético correspondente a consoantes e vogais, uma breve explicação dos tons existentes em vietnamita, passando depois para uma análise fonotática da língua e terminando com uma revisão sobre as dificuldades até hoje encontradas pelos falantes nativos de vietnamita, quando produzem as fricativas inglesas. Finalmente, esta segunda parte termina com um olhar para três modelos de aquisição de segunda língua: a Hipótese da Análise Contrastiva (HAC), (Lado, 1957); o Modelo de Aprendizagem da Fala (Speech Learning Model: SLM), (Flege, 1987); e o Modelo de Assimilação Perceptual (Perceptual Assimilation Model: PAN-2), (Best & Tyler, 2007). Após a revisão da literatura, foram elaboradas as hipóteses e metodologias de trabalho (capítulo 4 e 5, respetivamente).

A terceira parte deste relatório (capítulo 6 e 7) é prática e debruça-se sobre as fricativas não vozeadas e sobre a análise que é realizada, alinhado com o que foi anteriormente explorado na revisão da literatura. Os áudios analisados são de falantes autodeclarados nativos vietnamitas e foram analisadas apenas fricativas em posição inicial de palavra. Para este estudo, o contexto vocálico escolhido foi a vogal /æ/, que existe na língua inglesa, no entanto, não faz parte do inventário vietnamita. As anotações estavam divididas em três categorias diferentes (“OK”, “ERRO”, “FORA”). Na categoria “ERRO” era possível escolher uma subcategoria que consistia numa classificação de fones específicos a partir do Alfabeto fonético

internacional[2]. Os fones foram categorizados e posteriormente anotados por um segundo linguista. Cada fone tinha um total de 1000 áudios para anotar, de forma a criar um conjunto de dados equivalente. No total, 5000 áudios foram anotados pelos dois anotadores. Depois de anotados os áudios, estes foram analisados acusticamente com diferentes medidas: quatro pistas espectrais (centróide, variância, assimetria e curtose), localização do pico, medidas de transição da informação (frequência de F2 em onset, intercetação, inclinação) e medidas de amplitude (amplitude normalizada e relativa). De seguida, foi realizada uma análise de aquisição de L2 em que foram analisadas as produções dos falantes vietnamitas e possíveis explicações para essas produções. Esta análise foi efetuada nas fricativas que foram anotadas manualmente.

Esta tese testa duas principais hipóteses sobre a deteção acústica. O centróide, a localização do pico, a frequência de F2 em onset, a intercetação e a inclinação devem distinguir fricativas em termos de modo e ponto de articulação relativamente à localização da constrição como anterior vs. posterior. Portanto, devem distinguir algumas fricativas produzidas na frente da cavidade oral (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) de algumas fricativas produzidas na parte posterior da cavidade oral (ç, x, χ, ħ, h) (hipótese 1 e 4). Variância, assimetria, curtose, amplitudes normalizadas e relativas devem distinguir fricativas quanto ao grau de constrição. Portanto, devem distinguir entre uma constrição mais estreita (s̪, s̺, s, ʃ, ʂ) e uma constrição mais ampla (ɸ, f, θ, ç, x, χ, ħ, h) (hipótese 2 e 6). Consequentemente, também são testadas quais as medidas acústicas mais eficientes para distinguir as diferentes fricativas.

Este relatório testa também duas hipóteses nas fricativas analisadas quanto à aquisição de segunda língua. Primeiro, os falantes da língua vietnamita não devem ter dificuldades ao pronunciar os fonemas /f, h/ na posição de *onset* (início de sílaba), uma vez que estes fones existem no inventário da língua e podem ser produzidos na posição de onset (hipótese LA_1). Também se prevê que os falantes da língua vietnamita terão dificuldades com /θ, s, ʃ/ (hipótese LA_2) na posição de onset, já que estes não existem no inventário vietnamita. Logo, prevê-se

[2] IPA Chart, http://www.internationalphoneticassociation.org/content/ipa-chart, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2018 International Phonetic Association.

que os falantes irão mapear estes fonemas para o modo articulatório e propriedades acústicas mais próximos permitidos na língua vietnamita. A fricativa interdental /θ/ deverá ser substituída pelos fonemas /t, tʰ, t̪, t̪ʰ/ ou /s̪/. No caso da fricativa alveolar /s/, prevê-se que o fone será pronunciado como /s̪/ e /ʂ/. No caso de /ʃ/, prevê-se que o fone será pronunciado como /s/ ou /ʂ/.

Neste estudo, foi possível provar que existem várias medidas acústicas robustas e que fornecem informação sobre as fricativas e as diversas formas de articulação, independentemente da variação por falante. Foram ainda confirmadas várias das hipóteses sobre a deteção acústica e as hipóteses de aquisição da língua.

A tese mostra que as propriedades espectrais e medidas de amplitude distinguem com sucesso uma ampla variedade de fricativas, o que é confirmado por estudos anteriores. No entanto, estes estudos anteriores concentraram-se num conjunto limitado de fricativas (e.g., Jongman *et al.*, 2000; Nirgianaki, 2014; Wikse Barrow *et al.*, 2022). As medidas de transição de informação (frequência de F2 em onset, intercetação e inclinação) foram as medidas em que houve menos diferenças significativas entre fricativas não vozeadas.

Os falantes vietnamitas tendem a utilizar estratégias percetivas e articulatórias da sua própria língua ao produzir tanto os sons conhecidos quanto os sons desconhecidos, do inventário da primeira língua. A influência do alfabeto vietnamita é visível na produção de fricativas como /f/ e /θ/. Como era esperado, as fricativas interdental (θ) e pós-alveolar (ʃ) representam os maiores desafios para os falantes. Algumas das substituições mais comuns foram por fones do inventário fonético vietnamita, tais como /s̪, tʰ, ʂ, x/.

Palavras-chave:

Fricativas não vozeadas; aquisição de L2; análise acústica; vietnamita; ELSA Speak.

Abstract

The main objective of this thesis was to analyze the production of voiceless fricatives by ELSA users. This thesis was created while doing an internship at ELSA[3] Speak. The study focuses on the acoustic properties of fricatives, which have been extensively studied but often with a limited number of speakers (Stevens, 1999; Jongman *et al.*, 2000). In contrast, this thesis benefits from analyzing hundreds of speakers, providing valuable insights into fricative acquisition. Furthermore, academically speaking there are some studies on L2 acquisition of fricatives (Kitikanan, 2016). However, fricatives are hard to produce and analyze, therefore they have earned attention in the literature.

The analysis focused on fricatives in the onset position. To perform this analysis, 1000 audio files from self-declared native Vietnamese speakers were listened to and annotated. Annotations were separated into different categories ("*OK", "ERROR", "OUT"*) and the error tokens were further classified into specific phones (IPA). After the annotation process was completed, the audios were analyzed acoustically (spectral moments, transition information, and amplitude). Alongside, there was a comparison of the results of the annotations with what is found in the literature in terms of perceptual and articulatory differences, in the acquisition of these fricatives by Vietnamese speakers.

The thesis reveals that spectral properties and amplitude measurements successfully distinguish a wide range of fricatives, while previous studies were successful but focused only on a limited subset of fricatives (e.g., Jongman *et al.*, 2000; Nirgianaki, 2014; Wikse Barrow *et al.*, 2022). Vietnamese speakers tend to use phones of their language, or phones that are somewhat similar to them, when producing both known and unknown sounds. As anticipated, the interdental and postalveolar fricatives pose the greatest challenges for the speakers. The influence of the Vietnamese alphabet is probable in the production of fricatives such as /f/ and /θ/.

[3] ELSA stands for *English Language Speech Assistant.*

Keywords:

Voiceless fricatives; second language acquisition; acoustic analysis; Vietnamese; ELSA Speak.

List of abbreviations

App. - Application

CEO - Chief Executive Officer

CTO - Chief Technology Officer

IPA - International Phonetic Alphabet

NLP - Natural Language Processing

ASR - Automatic Speech Recognition

B2C - Business to Consumer

B2B - Business to Business

B6 - Benchmark6

B7 - Benchmark7

AI - Artificial Intelligence

M1 - First spectral moment

M2 - Second spectral moment

M3 - Third spectral moment

M4 - Fourth spectral moment

SD - Standard deviation

SLA - Second language acquisition

L1 - First Language

L2 - Second Language

NT - Native language

TL - Target language

CAH - Contrastive Analysis Hypothesis

SLM - Speech Learning Model

PAM - Perceptual Assimilation Model

PAM - L2 - Perceptual Assimilation Model - L2

SS - SpeechServer

VI - Vietnamese

Table of Contents

List of Figures

List of Tables

1. Introduction

After starting the master's program and completing the first year, it became evident that the best way to learn was through experience. Once presented with the opportunity to undertake an internship at ELSA Speak[4], it was undoubtedly one of my top choices. This decision was primarily driven by the work developed in this company. The love for languages has always been present in my life, both for learning and teaching. Therefore, when looking at a state-of-the-art app. that teaches English pronunciation, there was no hesitation in my resolve to work towards the goal of helping someone speak English with a better pronunciation.

In this internship, it was possible to look further into the detection of errors in English phones and extend my linguistic knowledge in doing so. This is the first reason to conduct this analysis. The first step to do this thesis was to discover which phones would be analyzed more specifically, since it was not possible to analyze all of them. This choice was based on two different aspects: providing more information and improving ELSA's capacities in a specific issue; conducting an academically speaking interesting study. After considering the two, the group of voiceless fricatives was selected.

In order to achieve the goal of performing at the best during the internship and attempting to create an interesting study, it was necessary to combine the knowledge from the internship and from the entire company, with the literature.

This report starts by introducing the Elsa company, both teams within the company, and also the app. and the first steps to use it, as well as an explanation of some of the linguistic tasks performed in the company. More importantly, there is also an explanation and discussion of what my work was during my internship. After this introduction, there will be a chapter with state of the art, which will explore English and Vietnamese fricatives further on; a section with acoustic measurements and information found in previous studies about the English fricatives selected for the study (e.g., Shadle, 1990; Jongman *et al.*, 2000; Gordon *et al.*, 2002); and

[4] Henceforth, ELSA Speak will be referred to as ELSA.

finally, three L2 acquisition models, Contrastive Analysis Hypothesis (Lado, 1957), Speech Learning Model (Flege, 1987) and Perceptual Assimilation Model (Best & Tyler, 2007). Next the research questions of this thesis and the methodology employed to obtain the data and perform the analysis are addressed. The hypotheses were built with two main questions in mind: how do native speakers of Vietnamese produce the voiceless English fricatives and how well can these fricatives and different productions be detected. Subsequently, the results were analyzed: the human annotated data according to the acoustic measurements and second language acquisition models.

2. ELSA: A look at the company

This chapter will introduce the ELSA company, starting with a brief description of the teams to which the internship was closely related to, and a brief demonstration and explanation of the app. itself. Afterwards, there will be an explanation of what linguistic tasks are performed at Elsa and, more specifically, what the internship work was dedicated to.

2.1. ELSA

ELSA is a start-up founded in San Francisco in the year of 2015. This startup was founded by Vu Van and Xavier Anguera, the Chief Executive Officer (CEO) and Chief Technology Officer (CTO) of the company, respectively. The company also has offices in Lisbon, Vietnam and India. Since its creation, the ELSA application (app.) has gathered users all around the world and already counts with over 13 million users, being available for Android and iOS systems. The company also has numerous partnerships with other companies and educational institutions such as schools, universities, and language centers, particularly in Asia and Latin America.

ELSA stands for *English Language Speech Assistant,* which was created to help people improve their pronunciation in English. The pronunciation learned in the application is Standard American English, and the phonetic alphabet used to teach is the International Phonetic Alphabet (IPA)[5].

ELSA is a virtual teacher and speaking coach that helps users improve their pronunciation through various exercises which allow the user to improve the most problematic sounds of its own pronunciation. All of this is done using automatic speech recognition (ASR).

[5] IPA Chart, http://www.internationalphoneticassociation.org/content/ipa-chart, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2018 International Phonetic Association.

2.2. ELSA teams

ELSA has different teams working together in different areas. Some of these teams are more directly connected to the work undertaken during the internship and the team in which the work was developed, the Research Team. There are three main teams linked to the Research Team: Product Management, Business Development and Content. In the subsection 2.2., it will provide an insight to what each of these teams does and how they are connected. The engineering team provides support to all the teams, however, since the internship work was not in close collaboration, this thesis will not detail the work of this team.

2.2.1. Research Team

ELSA has different teams developing different areas and these teams complement each other. The work developed along the internship was in the Research Team. Within this team different main areas are covered, such as speech and signal processing, natural language processing (NLP), linguistics, among others. This team's main goals are to improve the speech recognition models, to enhance the users' assessment system, to improve and adapt the recommendation of content for each user according to the difficulty of the lessons and level of proficiency of the user, and many others.

One of the areas this team focuses on is NLP, which can be used in search engines, machine translation, chatbots, text summarization or simplification, inter alia. Nowadays, NLP is a synonym of working with neural networks. In this modern NLP era, there are many (un)labeled datasets that create pre-trained base models, which are then used and tailored to niche tasks. In order to do this work, there are language models that predict certain aspects and characteristics, such as the transformer language model. This neural network model allows very efficient computation and builds hierarchical representations of text, for instance, this model can predict words based on their context in a sentence.

At ELSA, NLP is used for:

- Punctuation for output from Automatic Speech Recognition (ASR).
- Grammar error correction.
- Relevance prediction in context.
- Word prominence.
- Synonym prediction in context.

This is an important research work done mainly by speech scientists in the team, and it is a very important job to improve ELSA's predictions for specific tasks.

Another area the Research Team focuses on is pronunciation, which is the area on which the internship was mostly focused on. This specific task involves an evaluation of the quality of the content of the app. by looking at the quality of audios and by the creation and evaluation of resources, as well as by annotating audio and text, which allows the evaluation and training of the technology. The annotation of audios is what the internship was mainly focused on. These annotations detect existing errors for each phone, which are then used to train and assess the system, as well as improve the feedback hints in the app. Figure 1 shows the process from the annotations to the ending results of the task.

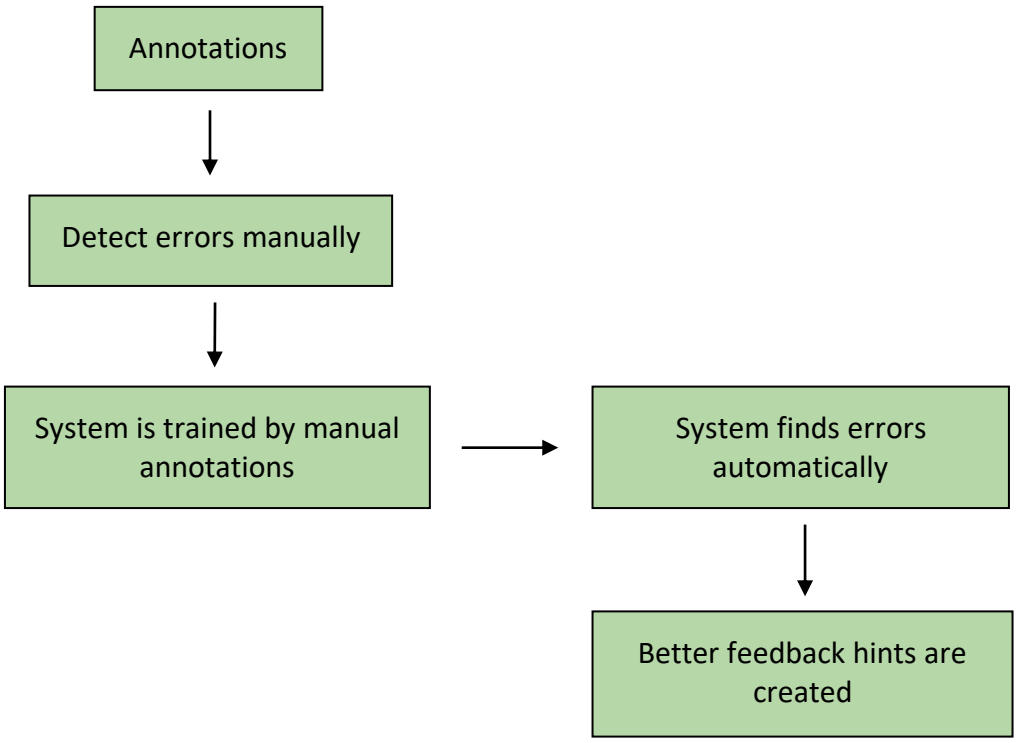

Figure 1 - From annotations to ending results

There are also other areas in the speech team where linguists are more involved, such as verifying phonetic transcriptions and examining the quality of translations.

2.2.2. Product management

As any system, the ELSA app. needs improvements over the course of time. In order to make them, it is necessary to have a team that can check what is not working well, and why users are quitting the app. at a certain stage of the process, if the user's journey is adequate, what customers are complaining about, if there are bugs happening. All of this needs to be analyzed and looked at carefully, hence there is a **product management team** taking care of all the mentioned tasks.

This team must **plan, research, design, adapt, and measure the user experience and lifecycle**. What does this mean? They need to see what needs to be improved, either because the user is quitting the app. in the first few steps, or it's not upgrading to a subscription. Thus, **plans** must be made to see what can be changed in order to improve. **Research** is conducted to assess ELSA's health. It is necessary to gather a comprehensive understanding of where the app. is and what are the major areas that are doing well and are not doing well from a product perspective and ultimately from a user perspective. **Design** is to turn a vision into a product by creating an appealing and intuitive design that improves the application, since the visual part of the app. is extremely important. As a consumer, it is very important to understand how the visual part of an app. is to either make me like it or to never use it. Therefore, this part is crucial. The product team needs to **adapt** the changes to the needs and requests of the customers, so they talk to users, research competition, test previous developments and file bug reports, after which having made modifications, they conduct data analysis and **measure** the success rate of the changes to see if they were good changes. These changes can bring in more or less customers and they can change the revenue. If they are good changes, they stay and eventually are improved. If they are bad changes, it is necessary to understand what went wrong, maybe modify them, or go back to what was prior to the changes. Therefore, this team updates versions of the app. to consider customization and main issues in a constant way.

Here are some of the tasks the Product Management team does:

- Sync up with Product team.
- Review latest designs.
- Groom with engineers.
- Test previous development and file bug reports.
- Write specs for upcoming development.
- Talk to users.
- Conduct data analysis.
- Research competition.
- Write down ideas.
- Let team review ideas.
- Host a product inception forum.

2.2.3. Business Development

Business development team develops the business. It started working with a business-to-consumer (B2C) marketing model, which works directly with the consumer, and evolved into a business-to-business (B2B) marketing model. This second model means that the team works directly with companies that have local employees who need to know how to talk and be understood in English, or with specific groups such as professors, tutors, universities, and linguistic centers that use ELSA to help them teach.

This means that the Business Development team contacts both consumers and companies to sell the product. Since this team has direct contact with the consumers of ELSA, they clarify some of the questions consumers or potential consumers might have and align with other teams on how to solve the issues in a technologically oriented way.

In summary, this team is responsible for finding different clients and partners and selling the ELSA app. to them, as well as giving them feedback and answering their questions. In order

to answer many of the questions and provide good feedback to clients, it is important to be in contact with the Research Team and obtain specific answers for each question.

2.2.4. Content

Content is also an important part of ELSA. This team oversees writing the content, which means writing everything that is in the app. There are four stages in building content. The first stage is the project proposal, in which gaps are found and content is generated based on user research. The second stage is the writing process for each main game in the app. - listening, pronunciation, word stress, and conversation exercises plus intonation, which is very minimal. The third stage is to normalize the process, which means adding more information to the content, both at a normal level (phone) and higher level (word, sentence), and ensuring that they can be used in the system. The last stage is to obtain the content released in three directions: voicing, content check and upload, and release. The content team also writes and creates original video content for the application.

In addition, the content team also works with influencers and secures partnerships which secure users, also known as the B2C model. However, ELSA has been working on a new frontier by gathering data from B2B and sales teams, which provides a list of companies and creates a need for career-focused modules.

The content team and the research teams work together by writing assessments (mini assessments) and testing/verifying vocabulary.

2.3. Inside the app.

Now that it has been explained how the company works, the first few steps when using the app. and the user's journey will be illustrated.

When the app. is downloaded to a device, four onboarding questions allow users to define their profile:

1) Native language: there are 58 languages to choose from and an option for users who do not speak any of the languages.
2) Topics of interest: it is asked why the user is practicing English and there are six different options to choose from: Travel; Job opportunities; Education; Live & Work Abroad; Culture & Entertainment; or Other.
3) Level of English: there are three possible choices: Beginner; Intermediate and Advanced.
4) Practice time: it is asked what our practice goal is, which means the amount of time we want to spend practicing.

Once these questions are answered, the user can choose to make a plan and use the app. or perform a **mini assessment test** to identify his/her speaking issues. In this mini assessment test, there are ten sentences to repeat, in order to discover what the user is good at and what should be improved. After the user has taken the test, there is a specific training program made for him/her. Once all of this is completed, the user can begin completing the training program. Figures 2, 3 and 4 show some screenshots with a few steps when entering the app.:

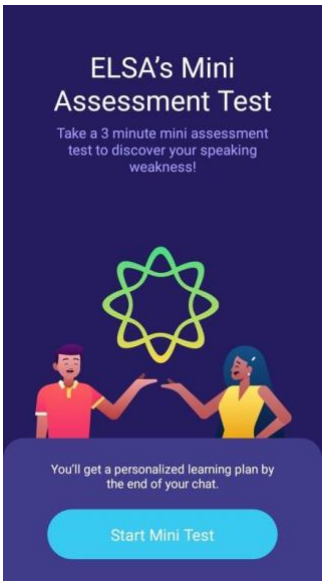

Figure 2 - Mini assessment test

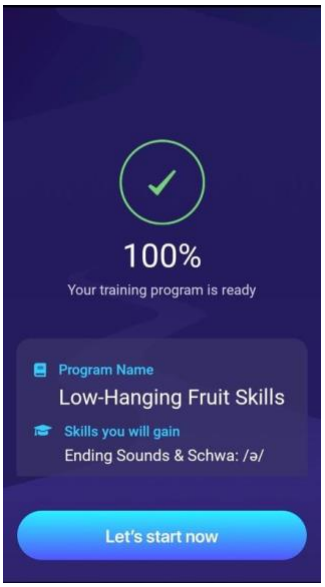

Figure 3 - Name of our program

Figure 4 - Laid out training program

When using the app. there are five different types of learning materials and content:

1) Pronunciation: this corresponds to the ability to pronounce the sounds clearly, which include consonants and vowels.
2) Listening: Corresponds to the ability to hear the difference between similar words, which helps with the comprehension and pronunciation.
3) Fluency: Represents the ability to speak swimmingly by working on natural rhythms and pausing time.
4) Word Stress: Represents the ability to correctly stress syllables.
5) Intonation: Corresponds to the ability to stress the correct words in a sentence.

The content materials are organized in the following manner:

Planets → Modules → Lessons → Exercises

In order to record the exercises in the app., it is necessary to click on the "microphone button" and speak. After the recording is made, feedback is provided to the user. At this point, the user can perform the recording again if him/her wants to repeat it and/or listen to his/her own recording. To track his/her performance and check his/her scores, the user can check his/her profile.

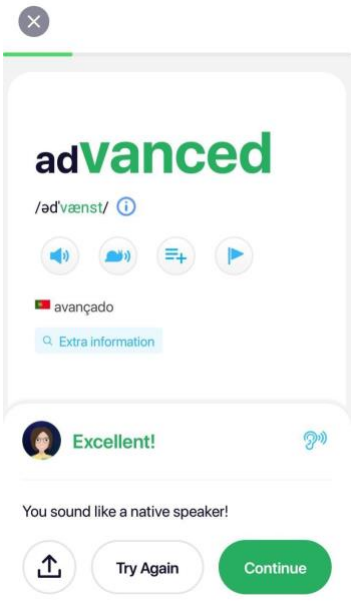

Figure 5 - Performance tracking

Figure 6 - Recording and feedback

In addition to exploring the modules and following the program, it is possible to navigate through different tabs, watch video conversations, explore study sets, use the dictionary, accompany the user evolution and much more.

2.4. Linguistics at ELSA

This section of the second chapter focuses on the main linguistic tasks developed in the Research Team. It begins with an explanation of what annotations are, as well as the tools and criteria to do it; a brief description of what content check and localization are.

2.4.1. Annotations

The internship at ELSA was made throughout four months and there was one main task to complete: annotating audio files in order to augment the coverage of annotated data and improve error detection. This task was completed by listening to a large number of audio files focusing on specific phones (especially focusing on the phones more relevant to the project) and by annotating them, which consisted of labeling the audios. These annotated data are necessary for training and benchmarking.

The annotated phones were selected and prioritized from a previously made benchmark (*Benchmark6*) with the results collected from already available annotations. This task was inserted into a project called semi-supervised-annotations which is a project of pronunciation. The annotations in ***Benchmark 6*** (B6) were sometimes repetitive in more than one way. There could be a set of tokens to annotate, and all of them or many could have the same word, hence, there was no variety in the annotations. In addition, the same audio from the same person sometimes came up more than once because users just repeated it. There were also a few audios in which the phone to be analyzed was incorrectly selected by the computer, which created more annotations than necessary.

Since B6 had all the mentioned issues, the data were not as robust as the team wanted, therefore, it was necessary to create B7, in order for these things to not take place and as a way

to find out why they did, and also in order to have better labeled data as a way for the system to have greater results. The main objective of this study is to create a new benchmark based on quality data selection and annotation for robust testing.

2.4.1.1. Tools

The entire annotation process was performed using a customized version of the web annotation tool, *wavesurfer.js.*[6] This tool was adapted specifically for the semi-supervised-annotations project and was created to facilitate the process of annotation in a simple and fast way, enabling the annotator to hear complete audios or just a target phone. While at the company, the tool suffered some changes, and is constantly being improved to turn it into a better tool. When fetching data for the annotation tool, there is a process to filter and prepare the information. This process allows the data to be chosen according to various features, such as the pre- and post-target phones or the mother tongue of the users. The system chooses the data and attempts to avoid the repetition of the same words and users. After running the search with all the features mentioned previously, the data is downloaded and prepared for annotations. All data is anonymous.

The annotations interface includes a waveform which is displayed in the center of the screen, a play button, two arrow buttons, four category buttons, and an annotation text box. There is also a count of annotations with the number of "*OK's*", "*ERR's*" and "*OUT's*", as well as the written discourse that the user said. The waveform is where the audio plays, where we can play the target phone, which is aligned by the computer, or where we can select a specific segment of the audio to play. In order for the audio to run, it is necessary to press the Play button or press P as a shortcut, and it is also possible to listen to the neighboring phones by pressing S, also as a shortcut.

[6] https://wavesurfer-js.org/ and/or https://github.com/katspaugh/wavesurfer.js

After listening to the audio, the annotations can be labeled into four different categories: ok, warning, error and out (OK, WARN, ERR, OUT)[7]. Once the category is chosen, the category button will turn green. If the annotator considers it necessary to add more information about the phone or audio, he can type an optional subcategory into a field. When the annotation is complete, one of the two arrows is used to either go to the next audio or to the previous if necessary.

2.4.1.2. Annotations and criteria

As mentioned previously, the annotations can be placed in one of four different categories: "*OK", "WARN", "ERR", "OUT"*. For each category, there are specific criteria so that every annotator can do the work with the same guidelines. This is very important because the annotations must be consistent, even if the opinions of the annotators differ. It is for this reason very important to have an inter-annotator agreement (Mella *et al.*, 2015).

Regarding the criteria for the annotation process, these were the more general guidelines for the inter annotator agreement:

- When there is no problem in the pronunciation and the token sounds like a native production, it is placed in the OK category (as displayed in figure 7 by *¹).
- If there is an error, the audio must be placed in the ERR category and then explain what the error is. In these cases, the explanation depends on the analyzed phone (as displayed in figure 7 by*²).
- While annotating, it was decided that the WARN category would be removed and only in the future will be revisited, for a finer partition within the NON-OK set (as displayed in figure 7 by *³).
- Finally, any audio that is imperceptible should be placed in the OUT category. The audio can be incomprehensible, either because the speaker uttered nonperceived segments,

[7] Each category is described in 2.4.1.2 Annotations and criteria section of this chapter.

because of the microphone used, or because of the noise in the recording (as displayed in figure 7 by *4).

These guidelines were created by linguists on the research team and improved as the internship took place. The guidelines can also be displayed in a decision tree, as illustrated below.

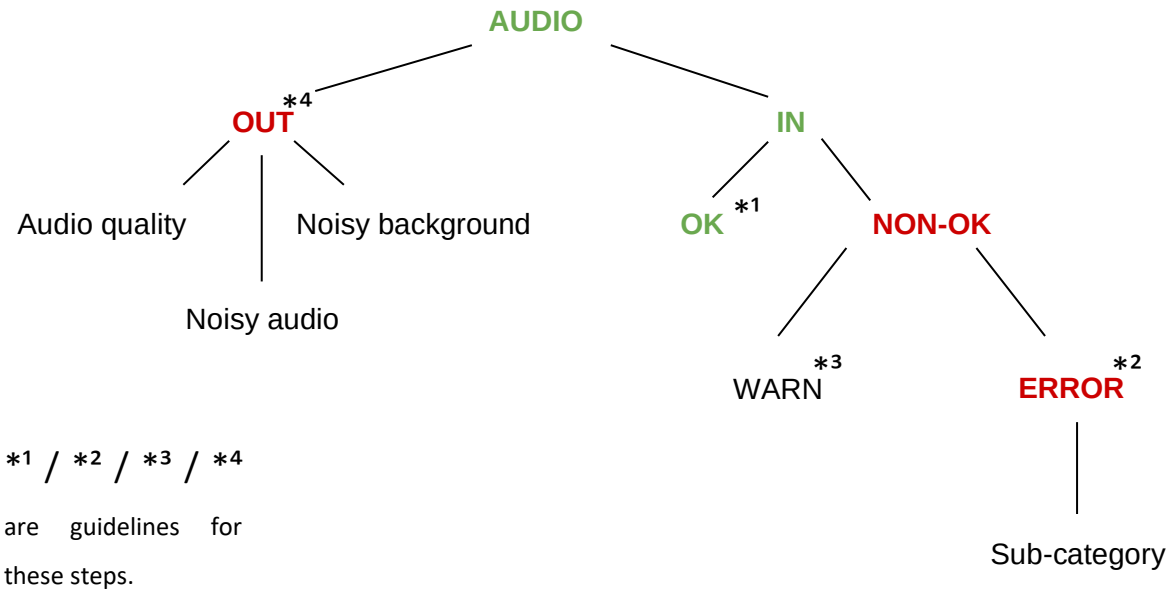

Figure 7 - Decision tree

2.4.2. Content check

Content check is a task conducted by linguists in collaboration with the content team. This task consists in verifying the already existing and the newly made content. In order to execute this task, the linguists need to check the lexicon's transcriptions and label them. These transcriptions are verified in terms of pronunciation, syllabification and alignment.

In order to do this verification, it is necessary to check the source of the transcription and provide linguistic information about the transcription and the pronunciation, using labels to trace the method used to obtain the transcription. The transcriptions can occur textually in a particular dictionary or in the case of inflected forms, it can be inferred from related entries in

that dictionary. The labels used provide meaningful linguistic information and can define the relevance of a particular variant to the users. Some labels are automatically added in the code in decoding graph creation or automatically added for specific code support.

2.4.3. Localization

Localization is the translation and adaptation of the translation to a specific country and culture. This area of translation is useful to make the app. available to more users, by making sure that the content is more suited for their language and culture. The Localization team at ELSA is in charge of translating and creating translation orders to help the other teams at ELSA ensure that the translations are as accurate as possible, throughout the different content types in the company. This way, the users will have the app. and website that is adapted to their culture and their language, making the experience of learning a new language easier.

The Localization and Translation quality team at ELSA takes care of the translation requests and helps with the languages that are supported in the app: Vietnamese, Korean, Hindi, Brazilian Portuguese, Indonesian, Thai, Japanese, French and Spanish. The team also helps by adding screenshots of the content and reviewing the items that are in the tool Lokalise, a localization and translation management tool. This ensures that all the translators can have as much context as possible and helps keeping the glossaries in the tool up to date to maintain consistency throughout all the content types of the company. The team also orders the translation through Lokalise for the languages that do not have an in-house translator, so that the app. is still updated with all the supported languages. The team also finds translators and assesses them to check if their background aligns with what the company is looking for.

2.5. Internship work

This section describes in more detail the annotations done during the internship and a few brief observations about the work developed.

2.5.1. Annotation effort: from B6 to B7

To update benchmarking according to specific criteria, a detailed annotation process is required. This process began by annotating data that no one else had annotated, neither human, nor machine. To annotate this data, it was necessary to find labels for each phone, and only upon listening to several audios it was possible to observe which errors were being made. Subsequently, it became more evident which errors occurred and the best way to label them. After doing this, the process was continued by annotating Benchmark6. These annotations could be a check of annotations already done by someone or by a computer. Once these two tasks were completed, both annotations were merged and B7 annotations were created. To clarify the process, some of the annotated and categorized phones are described in this section.

One of the first phones analyzed was the diphthong /**ɔɪ**/. This phone was annotated in **coda** position (e.g., /əˈn**ɔɪ**/) and productions were from different languages. In the first phase of the process, it was noticeable that /ɔɪ/ final can be pronounced as an error both onset or/and offset, which means the error can be at the beginning or end of the diphthong. Several errors were made, and the following were found: sometimes speakers pronounced this diphthong as a bisyllabic sound, meaning two syllables were produced where only one should have been; often the users produced schwa (e.g. /əˈnə/) and /ɪ/ (e.g. /əˈnɪ/). After reaching these conclusions, this phone's errors were categorized as ERR: onset, ERR: offset, ERR: bisyllabic and in some cases, specifically as ERR: onset: schwa and ERR: offset: ɪ. Furthermore, there were moments when the error occurred both onset and offset, thereby making it useful to have a specific label for this. Thus, the option at the time was to name it ONSET. To decide the annotation process, a decision tree was used. To determine the labeling criteria for the errors, it was necessary to hear the wrong pronunciations and discern the phonetic and phonological characteristics of the mispronunciations. Based on the findings, an appropriate classification system was established for accurate labeling. Sometimes this label was the mispronounced phone, or a broader characteristic if it is not possible to clearly define the errors as another phone. Once this step was finalized, the data was ready to be annotated. After these first annotations, the B6 dataset was also annotated and once both were done, everything was merged and B7 created.

One other consonant analyzed was the voiced labiodental fricative /**v**/. This phone was annotated in **onset** position (e.g., /**v**ɛɪg/) and productions were from Spanish and Vietnamese speakers. These two languages were chosen as a trial to choose one language for the fricatives which are analyzed in depth ahead in this thesis. In order to do this, two hundred annotations from Spanish and two hundred annotations from Vietnamese were extracted, which were then merged in a single group. When analyzing the "*ERR*" categories for this phone it was not easy, as it was difficult to distinguish some of the sounds being produced, especially among the bilabial sounds. Thus, it was decided to separate the bilabial error category into two categories - bilabial stop - /b/ and bilabial fricative - /β/. It was also decided to have an approximant category which was labeled labiodental approximant - /ʋ/. Again, after the annotations of the first dataset, B6 was annotated and then B7 created.

Another fricative that was analyzed is the voiceless alveolar fricative /**s**/. This phone was annotated in **onset** position (e.g., /ˈ**s**ɛl.ə.bɹeɪt/) and productions were also from Spanish and Vietnamese speakers. In /s/ initial there were many different possible "*ERR*" categories. In the first annotations, there were many "*OUT*" audios, either because the beep overlapped, audio quality wasn't good or there was a noisy background. This is particularly important for fricatives, because they are basically noise themselves, unlike, say, vowels, which are much more robust to noisy backgrounds. This excluded a lot of recordings. In the Benchmark6 annotations, there were not as many "*OUT*" audios as before. Several errors were made in the audios. The following were found: speakers made an insertion on the left of the consonant and a few times an insertion on the right of the consonant; speakers retracted the consonant making 'ʃ like sounds' instead of /s/; other times speakers geminated the consonant doing a longer version of it; speakers also advanced the consonant, hence /s/ was similar to /θ/ or at least sounded more dental than the expected target; speakers also voiced the consonant at times producing a /z/ instead of /s/; and there was also deletions, therefore, it wasn't produced. In the B6 annotations some new errors were found, such as the occurrence of /t/ and the production of /tʃ/. After reaching these conclusions, the errors of this phone were labeled as: insertion_right, retracted, advanced, voiced, deletion, sub:t, and sub:tʃ. This

distinction and knowledge of different categories is very important, as it is necessary to maintain the same name throughout the annotations. Having a specific phone as a category (e.g. ERR-sub:ʃ) can be very interesting since it is a detailed category, but in some situations it is not possible to define a phone within one of these categories, hence having more general labels, such as retracted or advanced might be safer.

There was also another fricative which was analyzed, the voiceless labiodental fricative /**f**/. This phone was annotated in **coda** position (e.g., /ʃɛ**f**/) and productions were also from Spanish and Vietnamese speakers. It was concluded that Vietnamese speakers produce /f/ in coda mainly as a bilabial stop /p/, which is something they allow in coda, or do various insertions on the right side of the phone. There were also several deletions, one alveolar implosive /ɗ/, which might be an exception and also the voiceless alveolar stop /t/ sound, which might be recurrent or not. After reaching these conclusions, it was decided the errors would be labeled as: bilabial stop, deletion, alveolar implosive and alveolar stop. There was also another type of error that was unexpected, which was named 'intrusive_s' and it consisted in an attempt to produce a fricative, but what was produced was something like: /sf/, /sp/, /ps/… VI speakers also turned /f/ into its counterpart, the voiced labiodental fricative /v/. After reaching these conclusions, it was decided errors would be labeled as: bilabial stop, deletion, alveolar implosive, alveolar stop, intrusive_s and voiced. As for the Spanish speakers, besides the errors also produced by the Vietnamese speakers, they also seem to do a voiced labial stop (b). Again, after annotating the first set of data, B6 was annotated and then B7 created. In addition, /**f**/ in **onset** position (e.g., /'**f**ɹɛnd/) was also annotated, in the same conditions as /f/ in coda position and both Spanish and Vietnamese speakers apparently don't have major problems producing it. In fact, it is very complicated to find any mispronounced productions. There were some productions of bilabial stops, some deletions and some insertions on the left.

The last phone being discussed in this section is the voiceless interdental fricative /**θ**/ in the **onset** position (e.g., /**θ**æŋk/)**.** Only Vietnamese speakers' productions were analyzed. Vietnamese speakers tend to produce /θ/ as the alveolar stop /t/ or as the aspirated dental stop /t^h/. Less commonly they produce the voiced interdental fricative /ð/ and the postalveolar

affricate. They also produce it as the labiodental fricative /f/. Given the data extracted and the Benchmark6 data the errors were labeled as: alveolar stops, aspirated dental stops, alveolar fricatives, labiodental fricatives and then voiced, deletion, voiced alveolar fricative, bilabial stop, one insertion left and one sub:K which were certainly exceptions. Afterwards this data was merged and B7 created.

These phones are just a brief example of what was done in the process of annotating and some of the conclusions made from those annotations. The same process was done for the experiment of voiceless fricatives which is the main focus of this thesis. Since this group of phones is analyzed in more depth, it will not be discussed thoroughly in this chapter. Instead, there will be an analysis only after explaining acoustic and language acquisition characteristics observed in the annotated data.

2.5.2. Internship project: learning outcomes

The internship at ELSA was a very significant step in my personal path. This experience allowed previously learned knowledge to be used and new skills to be gained and developed. The project permitted mainly the use of phonological and phonetic knowledge from the previous academic trajectory, but it also allowed to gain the capacity of an annotator. This was the main role in the company: labeling information so that the machine and system can use it, in this case, information about phones. The system needs these annotated labeled datasets to process, comprehend and learn, in order to make it functionable for language-based AI models and to have quality outputs. In doing this work the coverage of annotated data was augmented, which helps the system by improving error detection and it is also information that can be used to improve the feedback in the app.

The entire process of annotating allowed us to learn more about each phone which was analyzed. In trying to label the data, it was necessary not only to listen to the audios, but also to discriminate what productions were being made, which trained the ear at every step of the way improving the analysis. Furthermore, it was necessary to figure out what was the best way of

labeling them, taking into account phonetic and phonological aspects, but also how the system would handle it.

In every phone annotated it was possible to observe what mispronunciations were being made. When possible, it was important to spot patterns or to figure out the reasons why users were mispronouncing the phones, in order to understand how to label them, how to give better feedback, to know both the English language better and how the acquisition of it by other languages work and what difficulties users have when learning.

The process of annotating also helped improve the annotation tools, made this process faster as a team and helped giving feedback over the annotations results.

The choice of analyzing fricatives more in depth was based on these phones needing to be upgraded with respect to the annotation coverage, which would provide better error detection. Furthermore, academically speaking there are some studies on L2 acquisition of fricatives (Kitikanan, 2016). However, fricatives are hard to produce and analyze, therefore they have earned attention in the literature.

When looking deeper into fricatives, it was possible to observe that an even shorter group of phones was needed, due to the time this thesis had to be turned in. After conducting some research, the decision was to analyze voiceless fricatives in the onset position and analyze them using acoustic measurements to compare L2 productions and native productions.

3. State of the art

Since the main aim of this study is to analyze the productions of English voiceless fricatives by Vietnamese learners, this chapter starts by giving a definition of what fricatives are in contrast to other classes of sounds like stops or sonorants and also by giving an overview of their typology: how many fricatives languages typically have, what fricatives are more/less frequent (3.1. General aspects of fricatives). Afterwards, literature about General American English is reviewed (3.1.1. General American English), followed by a look at acoustic measurements and articulation of the voiceless fricatives (3.2. The acoustics and articulation of fricatives). Subsequently, the Vietnamese language is briefly described, as well as difficulties that Vietnamese speakers encounter when trying to produce voiceless fricatives (3.3. Vietnamese language). Finally, literature about L2 acquisition is reviewed and more specifically three models, the Contrastive Analysis Hypothesis (Lado, 1957), the Speech Learning Model (Flege, 1987), and Perceptual Assimilation Model (Best & Tyler, 2007), (3.4. The L2 acquisition of fricatives).

3.1. General aspects of fricatives

Consonants are divided according to their manner of articulation - stops, fricatives, approximants, trills and flaps - as well as by their place of articulation (Ladefoged & Johnson, 2015). Speech sounds can also be distinguished on the basis of the degree of constriction or the degree to which they let the air flow through the oral cavity - obstruents and sonorants. Obstruents are characterized by a considerable constriction in the air flow and sonorants are characterized by a continuous air flow.

According to Maddison (1984) almost all languages have fricatives. All fricatives can occur at all places of articulation; however languages only pick a subset of these (Roach, 1984). Fricatives are produced by letting/forcing the air flow through a narrow constriction. This constriction which is different according to each fricative will cause an audible friction (*e.g.*, Shadle, 1990; Stevens 1999; Jongman *et al.*, 2000).

Fricatives can be sibilants - /s, z, ʃ, ʒ/ or non-sibilants - /f, v, θ, ð/. Sibilants are pronounced with a characteristic hissing sound. Some sibilants can be produced in the same place of articulation as the non-sibilants and only be distinguished by a variety of subtle tongue shapes. (Maddieson, 1984)

> In phonetics, sibilants are consonants in which the tip or blade of the tongue goes near the roof of the mouth through a narrow channel in the oral cavity, while air stream is pushed past the tongue to make a characteristic high pitch hissing or hushing sound. (Demirezen, 2016: p. 751)

According to the L1-L2*map*[8], which shows a contrastive analysis of the phonetic inventories of more than 500 languages, the English and Vietnamese fricatives (languages chosen for this thesis) occur in different languages. The following table will give an overview of the percentages and numbers of each fricative in all languages, as well as some examples of languages they occur in:

Fricative	How many languages?	Some of languages it occurs	Percentages
f	211 languages	Armenian, Norwegian, Russian, Vietnamese, Mandarin.	43.15%
v	106 languages	Bulgarian, Finnish, Turkish, Vietnamese	21.68%
θ	23 languages	Albanian, Arabic or Greek	4.70%
ð	28 languages	Danish, Greek	5.73%
s	231 languages	Moroccan, Polish, Vietnamese	47.24%
z	64 languages	German, Lithuanian or Vietnamese	13.09%
ʃ	211 languages	Romanian, Polish, Ukrainian	43.15%
ʒ	72 languages	Arabic, French	14.72%

[8] L1-L2*map* can be found at https://l1-l2map.hf.ntnu.no (Koreman, J., Bech, Ø., Husby, O. & Wik, P.). This is a tool created by the Norwegian University of Science and Technology which contains phonetic inventories of more than 500 languages.

ʂ	25 languages	Basque, Swedish	5.11%
x	98 languages	German, Greek, Russian, Vietnamese	20.04%
ɣ	56 languages	Irish, Turkish, Vietnamese	11.45%
h	305 languages	Indonesian, Japanese, Turkish, Vietnamese	62.37%

Table 1 - Overview of the percentages and numbers of each English and Vietnamese fricative in all languages, according to the L1-L2map tool

This thesis will focus on the analysis of voiceless fricatives, therefore the core analysis will be targeting them. Aside from the pair θ - ð, all voiceless fricatives occur considerably more than voiced fricatives in the languages. Furthermore, both interdental fricatives seem to occur significantly less than other fricatives. The voiceless retroflex fricative is also one of the fricatives with lower occurrences in the languages. On the other hand, /f, s, ʃ, h/ occur in a significant percentage of all languages.

3.1.1. General American English

In the case of English (General American English), there are five groups of sounds and eight places of articulation - bilabial, labiodental, dental, alveolar, postalveolar, palatal, velar and glottal. The phonetic inventory for this language can be found below:

	Labial	**Dental**	**Alveolar**	**Post-alveolar**	**Palatal**	**Velar**	**Glottal**
Nasal	m		n			ŋ	
Stop	p b		t d			k g	
Affricate				tʃ dʒ			
Fricative	f v	θ ð	s z	ʃ ʒ			h
Approximant			l	ɹ	j	w	

Table 2 - General American English inventory, according to the L1-L2map tool[9]

[9] Idem.

In English there are nine fricatives, four voiced fricatives and their counterparts, plus the glottal h.

	Labiodental	Dental	Alveolar	Palato-alveolar	Glottal
Voiceless	f	θ	s	ʃ	h
Voiced	v	ð	z	ʒ	

Table 3 - Place of articulation of the English fricatives[10]

The voiced fricatives have either no voicing or almost none in initial and final position, but in between voiced sounds they might be voiced. The glottal is neither voiced nor voiceless. (Roach, 1984) All these fricatives with the exception of the glottal can be found in initial, medial and final position. The glottal can be found only in initial and medial position.

3.2. The acoustics and articulation of fricatives

This section describes the articulation of voiceless fricatives and presents information about common acoustic measurements frequently employed for fricative analysis.

3.2.1. Articulation of fricatives

As it was stated before, fricatives are produced with a slight constriction when forcing the air out. This is different from stops or approximants, because stops are produced with a complete constriction of the air flow and approximants are produced with a continuous airflow, since the constriction is not sufficiently narrow to create a turbulent airflow (Maddieson, 1984). Each English fricative has the following articulation characteristics (Roach, 1984; Kitikanan, 2016):

[10] Roach, Peter, 1984, *English phonetics and phonology: a practical course*. Cambridge: Cambridge University Press, 1983. Pp. x, 212. *RELC Journal*. 1984;15(1):117-118. Page 48

- The labiodental fricatives (f, v) are produced with the upper teeth and lower lip in contact with each other.
- As for the interdental (θ, ð), the tip of the tongue touches against the teeth and the air exits through a little gap between them. Both labiodental and interdental fricatives are characterized by having a weak noise.
- The alveolar fricatives (s, z) are produced with the tip of the tongue against the roof of the mouth (alveolar-ridge) and the sound exits through the center of the tongue.
- As for the palato-alveolar (ʃ, ʒ), are articulated with the tongue placed further back than the alveolar sounds and the sound exits also like the alveolar, but the passage is wider.
- The glottal (h) is produced in the vocal folds and the lip and tongue positions are according to the vowel next to the glottal.

When a fricative is produced and the air is forced through a constriction, the front part of the oral cavity of each constriction is what defines the spectral shape of the fricative. (Jongman *et al.*, 2000).

3.2.2. Acoustic analysis

In order to perform an acoustic analysis, there are different manners of measuring data. There are many different measurements, but according to Stevens (1999), Jongman *et al.* (2000) and Kitikanan (2016), these are the core measurements which can be used:

- Spectral properties: usually four moments (centroid, standard deviation (SD), skewness and kurtosis) and peak location.
- Transition information: onset F2 frequency, intercept and slope.
- Amplitude: normalized amplitude and relative amplitude.

Spectrum analysis determines how the acoustic energy is distributed across frequencies in a sound or speech signal. The four spectral moments are a mathematical way of describing the shape of a distribution or curve. In the next figures it is possible to observe the four spectral moments, as well as peak location and normalized amplitude.

The first moment (M1 or **centroid**) is used to find the average of the overall distribution, which means the mean of all the peaks of the spectrum (Jongman *et al.*, 2000; Kitikanan, 2016). This measure is generally assumed to correlate with the place of articulation where the fricative is being produced. The shorter the cavity in front of the constriction, the higher the centroid frequency. In case the centroid is higher, then it means the fricative is more fronted. The centroid should show differences between fricatives produced in the front of the oral cavity and fricatives produced in the back of the oral cavity (Stevens, 1999; Jongman *et al.*, 2000; Nirgianaki, 2014; Kitikanan, 2016)

For example, the frequency range of /ʃ/ is lower than /s/ and /s/ should have the highest centroid (Jongman *et al.*, 2000; Kitikanan, 2016). However, Fu *et al.* (1999) and Jongman *et al.* (2000) report that /ʃ/ generally has a lower mean than /s/. Additionally, differences in frequency should be easily noted between /ʃ/ and the rest of the fricatives and between /s/ and the rest of the fricatives. (Fu *et al.*, 1999; Gordon *et al.*, 2002) The fricatives /θ/ and /f/ should be hard to distinguish (Behrens & Blumstein, 1998a; Jongman *et al.*, 2000). In sum, the voiceless fricatives should appear in the following order, starting with the higher frequencies: f > θ > s > ʃ > h. /h/ is not analyzed in most studies (Jongman *et al.*, 2000), since it is sometimes considered as a voiceless vowel. However, it is a consonant produced further back in the vocal tract and it is influenced by the following vowel (in this thesis /æ/). This means that the oral cavity in front of the constriction is longer therefore this consonant should have a lower centroid. In previous research (Shadle, 1990; Behrens & Blumstein, 1988a; Jongman *et al.*, 2000), the non-sibilants (f, θ) were distinguishable from the two sibilants (s, ʃ) in average centroid values.

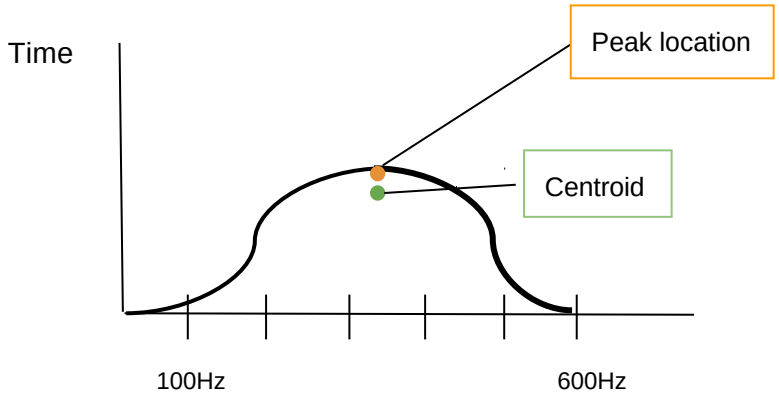

Figure 8 - Centroid and peak location (explained above)

The second moment (M2 or **standard deviation**) is the average square distance from the centroid. This shows the dispersion of the spectrum based on the mean frequency. Hence, if this square (SD) is bigger, then it is corresponding to a non-sibilant. If it is a smaller SD, which means the spectrum is less dispersed, then it is generally of a sibilant. The standard deviation for sibilant fricatives is expected to be low and for non-sibilants high. Several studies (Shadle & Mair, 1996; Jongman *et al.* ,2000; Kitikanan, 2016; Nguyen *et al.* 2022) show that between the fricatives /f, s, ʃ/, the lowest standard deviation is for /ʃ/ and the highest is for /f/. As for /s/ it is between the two fricatives, and it is similar to /ʃ/. The fricatives /f, θ, h/ should have a low energy and a dispersed spectrum which translates into larger values in the standard deviation. (Fu *et al.*, 1999)

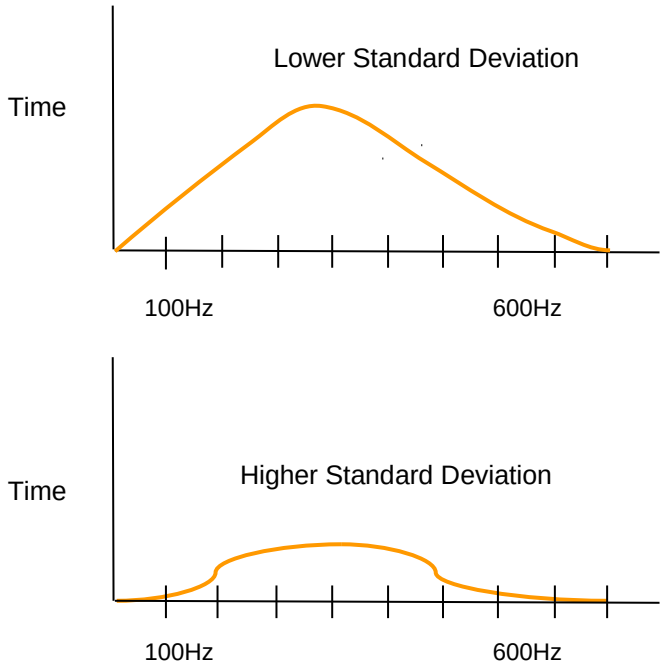

Figure 9 - Standard deviation

Third moment (M3 or **Skewness**) provides insights into the asymmetry of a distribution around the mean. For the distribution to be symmetric it must be zero. If the distribution is more to the right than to the left of the mean, the skewness is positive. Positive skewness suggests an energy distribution and concentration more in the lower frequencies. Positive skewness indicates a fricative is produced farther back. If the distribution is more to the left than to the right of the mean, the skewness is negative. Negative skewness suggests an energy

distribution and concentration more in the higher frequencies (Jongman *et al.*, 2000; Kitikanan, 2016). The alveolar fricative is expected to have a negative skewness and lower values than the postalveolar fricative which is expected to have positive skewness. (McFarland, 1996; Jongman *et al.*, 2000)

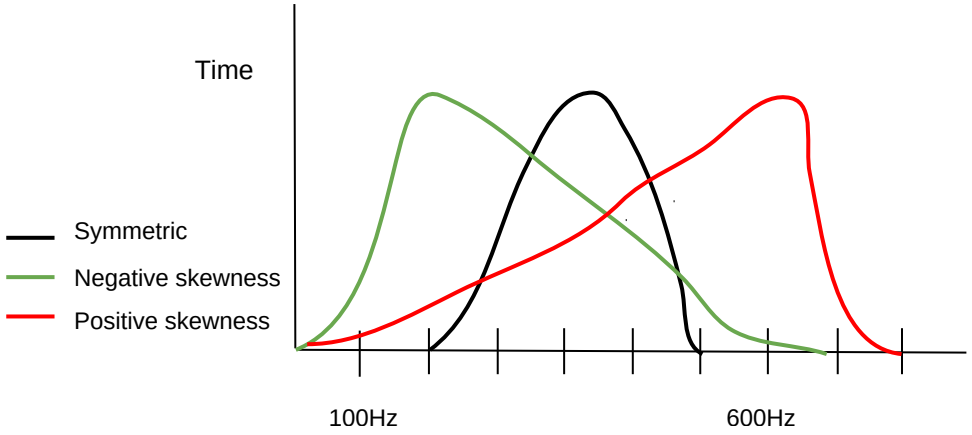

Figure 10 - Skewness

Fourth moment (M4 or **Kurtosis**) shows the spectrum distribution in relation to its flatness or peakedness. The kurtosis is positive when the peakedness is high, which means it has well-defined peaks. The higher the kurtosis value, the more peaks it has. When kurtosis is negative, it means that the distribution is more or less flat, which means it does not have well-defined peaks. If the peakiness is higher, it is likely a sibilant (Jongman *et al.*, 2000; Kitikanan, 2016). According to Tomiak (1990) (cited in Jongman *et al.*, 2000) and McFarland, (1996), /s/ has a significantly positive kurtosis and /ʃ/ has a smaller positive kurtosis than /s/.

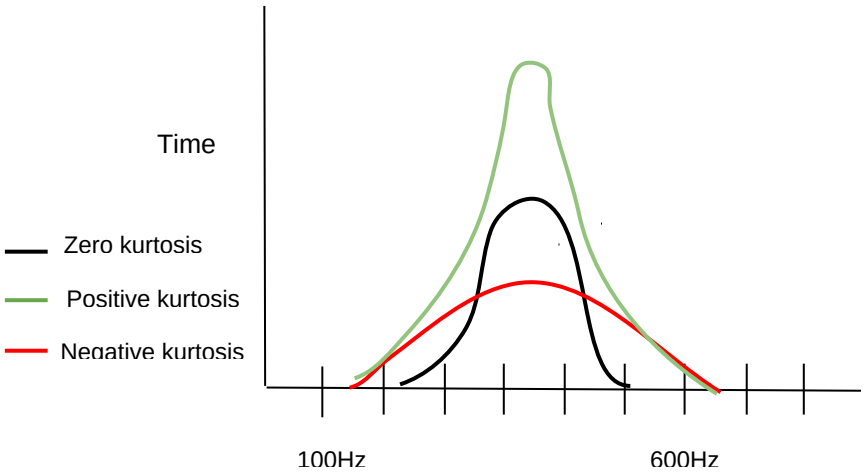

Figure 11 - Kurtosis

These four moments have been analyzed by Jongman *et al.* (2000) and Nirgianaki (2014) regarding some fricatives. In the table below some of the results are shown as reference.

	Jongman *et al.* (2000)				Nirgianaki (2014)				
	f	θ	s	ʃ	f	θ	s	ç	x
Centroid	5108Hz	5137Hz	6133Hz	4229Hz	4931Hz	5230Hz	5482Hz	4625Hz	3397Hz
SD	6.37	6.19	2.92	3.38	1.64	1.63	1.21	1.25	1.38
Skew.	0.077	-0.083	0.229	0.693	0.615	0.519	0.629	1.486	1.962
Kurt.	2.11	1.27	2.36	0.42	0.44	0.81	1.79	3.59	7.92
Peak	7733Hz	7470Hz	6839Hz	3820Hz	4492Hz	5178Hz	5145Hz	4149Hz	2377Hz

Table 4 - Mean spectral moment values for each place of articulation, averaged across speakers, window location, voiced and voiceless tokens, and vowel context[11]

According to Jongman *et al.* (2000), some spectral moments differentiate the two sibilants well, however it is not as easy to distinguish the non-sibilants. Furthermore, standard deviation and skewness seem to be the two better performing spectral properties when distinguishing places of articulation.

Peak location is the measurement that indicates the highest peak in fricative noise. It can distinguish fricatives according to place of articulation. If the place of articulation is more forward in the oral cavity, then the value of the peak is higher. Jongman *et al.* (2000) reports that postalveolars are normally around 2.5-3Hz and alveolar fricatives are normally around 4-5Hz. In the study's results, there are the following values for peak location in these fricatives: s, z - 6839 Hz; ʃ, ʒ - 3820 Hz. These values demonstrate that the further back is the place of articulation the lower is the frequency. The postalveolar and alveolar are defined by clear and distinct shapes in the spectrum and the labiodental and interdental are defined by a flatter

[11] Jongman, Allard & Wayland, Ratree & Wong, Serena. (2000). Acoustic characteristics of English fricatives. The Journal of the Acoustical Society of America. 108. 1252-63. 10.1121/1.1288413. Page 1257.
Nirgianaki E. (2014). *Acoustic characteristics of Greek fricatives*. *The Journal of the Acoustical Society of America*, *135*(5), 2964–2976. https://doi.org/10.1121/1.4870487. Page 2969.

spectrum. The spectral peaks can be different according to the speaker. In previous research (Shadle, 1990; Behrens & Blumstein, 1988a; Jongman *et al.*, 2000), the non-sibilants (f, θ) were distinguishable from the two sibilants (s, ʃ) in average peak location. The fricatives /θ/ and /f/ should be hard to distinguish (Behrens & Blumstein, 1988a; Jongman *et al.*,2000). Additionally, differences in frequency should be easily noted between /ʃ/ and the rest of the fricatives and also between /s/ and the rest of the fricatives. (Fu *et al.*, 1999; Gordon *et al.*, 2002)

Onset F2 frequency is the value of transition between the transition of fricative and vowel. When preceded by a fricative, a vowel can give information about the place of articulation of the specific fricative that is being produced alongside the vowel, specifically through F2 frequency. This measure increases with high vowels. (Jongman *et al.*, 2000) The further back is the constriction in the oral cavity, the higher is the F2.

Slope is measured in the F2 of the vowel, at onset and at vowel midpoint (Jongman et al. 2000) and it expresses the information of the F2 in different points. A fricative plus the F2 of the vowel gives information about the articulatory position of the fricative.

In table 5, it is possible to see Jongman *et al.* (2000) and Nirgianaki (2014) values averaged across voiced and voiceless tokens, as well as vowels, for the F2 onset frequency measure.

Jongman *et al.* (2000)					Nirgianaki (2014)				
	/f,v/	/θ,ð/	/s,z/	/ʃ,ʒ/	f	θ	s	ç	x
Mean F2 onset	1661Hz	1833Hz	1832Hz	1982Hz	1371Hz	1514Hz	1629Hz	2087Hz	1074Hz
Slope	0.738	0.530	0.517	0.505	0.662	0.687	0.585	0.442	1.087

Table 5 - Mean F2-onset (Hz) and slope values[12]

[12]Ibid.

Normalized amplitude shows the difference in amplitude between the target fricative and surrounding vowels. It is the amplitude measured within the spectrum as one (specifically fricative and surrounding vowel). If the fricative has a higher normalized amplitude, then it is a sibilant (first example). If the fricative has a lower normalized amplitude, then it is a non-sibilant (second example). The sibilants should have higher values than non-sibilants (Behrens & Blumstein, 1988a).

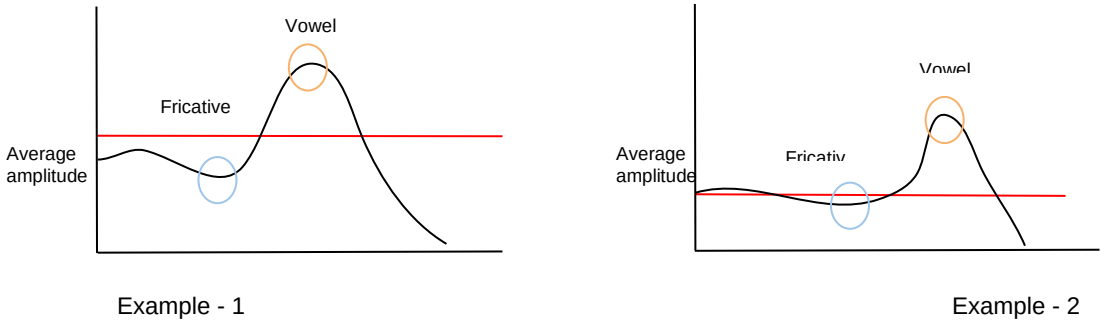

Figure 12 - Normalized amplitude

Relative amplitude shows the difference between fricative and vowel amplitude at F3 and/or F5. This measurement is ought to distinguish sibilants in terms of place of articulation.

Table 6 shows the average spectral mean values for different variants of English in Gordon *et al.* (2002)[13]:

	f	θ	*s*	ʃ	ʂ	ʂ̺
Chickasaw (North America (Oklahoma)	4562Hz	-	5163Hz	4679Hz	-	-
Apache (Western) - North America (Arizona)	-	-	5461Hz	4859Hz	-	-
Aleut (Western) - North America (Aleutian Islands)	-	-	5219Hz	-	-	-
Montana Salish - North America (Montana)	-	-	4601Hz	4134Hz	-	-
Hupa - North America (California)	-	-	4797Hz	4440Hz	-	-
Gaelic (Scottish) - Europe (Scotland)	4415Hz	-	4884Hz	4396Hz	-	-
Toda - Asia (India)	4268Hz	4111Hz	4529Hz	4704Hz	4535Hz	5027Hz

Table 6 - Average spectral mean in Gordon et al. (2002), table adapted from different pages

[13]Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association, 32*, 1-36.

3.3. Vietnamese language

3.3.1. Inventory and phonetics

Vietnamese is a tonal language, which means that this prosodic feature allows the distinction of meaning of the words of the language. There are six tones in Vietnamese: (-), (ã), (á), (ả), (à), (ạ), (Emerich, 2012). The tones are generally defined by the nucleus.

Tones	Name	Description
(-)	level	Mid or high-mid trailing pitch, nearly level when syllable is not final in pause group; falls to low range in final syllables
(ã)	broken	High rising pitch
(á)	rising	High rising pitch, might be heard as high level in rapid speech
(ả)	curve	Mid-low dropping pitch, not too abrupt, with a rise at the end
(à)	falling	Low trailing pitch
(ạ)	drop	Low dropping pitch with an abrupt falling

Table 7 - Vietnamese tones. Adapted from Emerich (2012) [14]

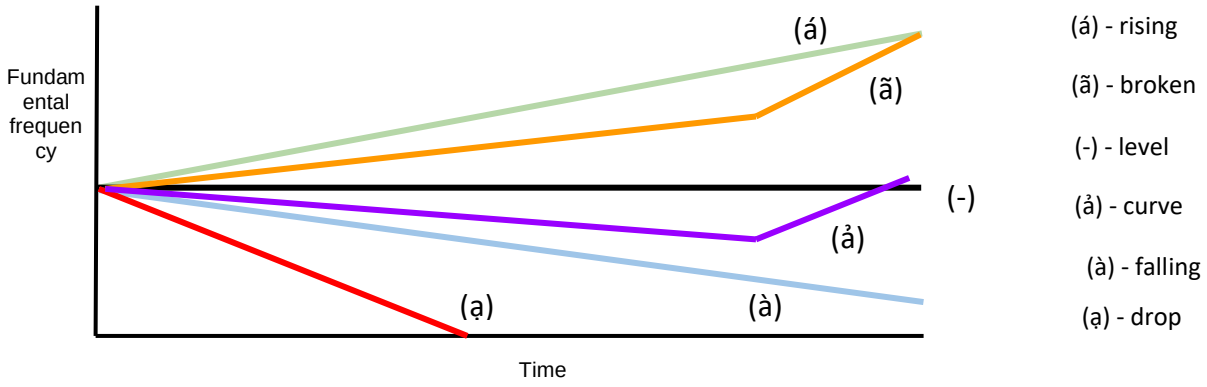

Figure 13 - Vietnamese tones in visualization

[14] Emerich, Giang Huong, "The Vietnamese Vowel System" (2012). *Publicly Accessible Penn Dissertations*. 632. p. 31

The Vietnamese language contains in its inventory (Hoang, 1965; Emerich, 2012):

- 11 vowels: a / ɐ / ɤ / ʌ / i / e / ɛ / ɯ / u / o / ɔ;
- 18 diphthongs: aj / ɐj / ʌj / ɔj / oj / ɤj / ɯj / aw / ɐw / ʌw / ɛw / ew / iw / ɯw / ie / ɯɤ / uo

According to the Hanoi dialect shown in Emerich (2012), and the Saigon dialect shown in Hoang (1965), VI has 25 consonants, which are displayed in the table below:

			Labial	**Dental**	**Alveolar**	**Retroflex**	**Palatal**	**Velar**	**Glottal**
Nasal			m	n			ɲ	ŋ	
Stop/ Affricate	Voiceless	Unaspirated	p[15]	t		ʈ	c	k g	ʔ
		Aspirated		tʰ					
	Voiced	Unaspirated	ɓ		ɗ				
Fricative	Voiceless		f		s	ʂ		x	h
	Voiced		v		z			ɣ	
Approximant				l			j	w	

Table 8[16] - Vietnamese inventory, with consonants in the Hanoi dialect and the Saigon dialect

When it comes to loan words, Vietnamese speakers can also employ the consonant /r/ in the onset position. The consonants /p j r/, in the onset position, occur mostly in loans. (Kirby, 2011)

[15] According to (Hoang, 1965; Nguyen, 2014), /p/ does not occur in word-initial position, except in loan words.
[16] In blue are the consonants which only exist in the Saigon dialect - /p, ʈ, ʂ, j, g/. In orange are the consonants which only exist in the Hanoi dialect - /z, w, ʔ/.

3.3.2. Vietnamese phonotactics

In Vietnamese, words are normally single-syllable words. These words are composed either by a vowel, a diphthong or a vowel preceded or followed by a consonant (Hoang, 1965). Only vowels can be the nucleus. Subsequently, Vietnamese's syllable structure is C_1VC_2 (Emerich, 2012). The consonants in onset (C_1), can be any of the consonant's existent in the Vietnamese language and consonants in coda (C_2) can be the voiceless stops - /p / t / k / - the voiced nasals - / m / n / ŋ / - and the voiced approximants / w / j /. It is not possible for a word to have more than three phones in Vietnamese. Whenever a loan word has multiple syllables, the word is adjusted to the structure of Vietnamese.

The Vietnamese language has the following consonant distribution:

- Every consonant can occur in onset position.
- There are no medial consonants.
- Only eight consonants are allowed in coda position: /p /, /t /, / k /, / m /, / n /, / ŋ /, / w/ and /j/ (Emerich, 2012).

As established before, in Vietnamese every consonant can occur in onset position and after this consonant there is always a vowel or a diphthong. Therefore, Vietnamese does not have complex onsets.

When observing the language, it is possible to see there is one type of consonant clusters permitted in the onset, which is C+w. All consonants except for /f/ can be followed by /w/. However, if the /w/ is considered labialized /^{w}/, then Vietnamese does not have consonant clusters in the onset position. Vietnamese does not have (initial) medial or final clusters. This leads to many difficulties when trying to pronounce words in other languages, especially in English, which is a language with multiple clusters. This makes it particularly complicated in the acquisition of this language. (Nguyen, 2007) Also, according to Tuan (2011), some groups of sounds might be difficult to pronounce, despite the fact that they might exist in the Vietnamese language.

3.3.3. Vietnamese fricatives

In Vietnamese there are nine fricatives if both dialects (Hanoi and Saigon) are taken into account, though the voiceless retroflex /ʂ/ only exists in the Hanoi dialect and the voiced alveolar /z/ only exists in the Saigon dialect. Despite the number being the same as in English, these are not the same fricatives. Nonetheless, the Vietnamese language also has four voiced fricatives and their counterparts, plus the glottal h.

	Labial	Dental/Alveolar	Retroflex	Velar	Glottal
Voiceless	f	s	ʂ	x	h
Voiced	v	z	-	ɣ	

Table 9 - Vietnamese fricatives

3.3.3.1. Difficulties in pronouncing voiceless English fricatives

In Vietnamese most consonants occur in the onset position, but not in the coda position. In English most consonants occur in all positions. For this, Vietnamese speakers have more difficulties in coda than with the onset. Coda position consonants tend to not be released by Vietnamese learners, since in the Vietnamese language final sounds are not released. (Tuan, 2011). Fricatives are not allowed in coda position in Vietnamese, thus when learning English there is a tendency to drop the final consonant. In Nguyen (2007), it is stated that Vietnamese speakers learning English tend to produce sounds which are peculiar by changing them to more similar sounds of their L1, and when sounds are particularly complicated, as in complex consonant clusters, Vietnamese learners tend to omit these sounds.

When it comes to the labiodental fricative /f/, it is often produced as /p/, an allowed sound in the Vietnamese language (Nguyen, 2007).

Since the Vietnamese language does not have /θ/ in the inventory, learners have

difficulties articulating this sound (Bui, 2016). Fricatives are also not allowed in coda positions in Vietnamese and are never released. The fact that /θ/ is represented orthographically by "th" generates confusion, because the same representation in Vietnamese is the equivalent to /t^h/. According to Tam (2005) and Bui (2016), /θ/ is mispronounced mostly as /t^h/, but also as /t/, /s/, /d/, /z/ and /ð/. According to Tuan (2011), /θ/ and /ʃ/ are some of the most difficult sounds to pronounce for Vietnamese speakers, because they are not in the Vietnamese language inventory.

In English, the voiceless alveolar fricative /s/ is a particularly common and frequent sound. This leads to an extra attention and consciousness of this sound for foreign speakers, which contributes to a smaller percentage of errors with word-final /s/. Nevertheless, since this phone is an unknown or foreign phone in coda position, speakers unavoidably make mistakes when producing it. Among the most common errors in coda position is the omission of /s/. There is also an over pronunciation of /s/ in ending sounds (Nguyen, 2007). According to Tam (2005), /s/ is mispronounced as /ʃ/.

According to Tuan (2011), /ʃ/ is not very difficult to pronounce for Vietnamese speakers in the onset position, however it is much harder to pronounce in the coda position. This phone is often produced as /s/ or as the retroflex /ʂ/ and it is also frequently omitted in coda position. In Tam (2005), it is stated that /ʃ/ is also mispronounced as /z/.

A final note, for /h/ no information was found in the literature.

In sum, studies show the following substitutions.

Fricatives	Substitutions
f	p
θ	t^h / t / s / d / z / ð
s	ʃ
ʃ	s / ʂ
h	-

Table 10 - Vietnamese speakers' substitutions for fricative sounds

3.4. The L2 acquisition of fricatives

While the models addressed in the next section are not the main focus of this thesis, it is worth considering some aspects that align with the analysis. In second language acquisition (SLA), several aspects are taken into account, particularly perception and production. When it comes to perception and production of non-native sounds, these two aspects can be linked or unrelated (Cheon, 2005). Studies have different perspectives, some more oriented towards production, others towards perception, and less frequently towards the combination of both. Several studies (Polivanov, 1931; Ladefoged, 1967; Goto, 1971; Baddeley & Hitch, 1974; Flege, 1987; Major, 2001) have tried to confirm theories about perception and production.

One of the main theories is perception is followed by production (Polivanov, 1931). In this theory, it is defended that perception can affect the production of L2 learners, since a deficient perception can lead to a poor production (Flege, 1987; Kitikanan, 2016). Sometimes, an L2 learner might perceive one sound as similar to one he/she has in his/her own language, producing it inaccurately. In a study about production and perception of English vowels with L1 Italian speakers, conducted by Flege *et al.* (1999), evidence was found to confirm the perception-production theory. Furthermore, according to Tuan (2011), someone trying to learn a second language will listen to the L2 sounds and may create some parallel/ transfer with the already known sounds. This also supports the idea that perception influences production. Other L2 learners are also not capable of producing certain non-native sounds, even though they are able to perceive them correctly (Cheon, 2005).

Various factors impact the perception and production abilities of a second language (L2) learner. These factors include linguistic proficiency, exposure to the language which influences the production to be more or less native-like, sociolinguistic circumstances, the amount of time spent listening to certain words and sounds, the preceding or following contexts that can influence the production of sounds due to coarticulation effects, difficulties in perceiving contrasts between sounds which are allophonic in their mother tongue, *inter alia*. (Kitikanan, 2016). Additionally, errors can occur due to articulatory difficulties (Cheon, 2005). When

speakers try to produce a second language, there are pronunciation patterns that reflect their first language.

Below three language acquisition models will be contemplated, and conclusions will be derived from them. These models are the Contrastive Analysis Hypothesis, Speech Learning Model and the Perceptual Assimilation Model - L2.

3.4.1. Contrastive Analysis Hypothesis

The Contrastive Analysis Hypothesis (CAH) was presented by Lado (1957) and it is a model which compares the native language and the target language while anticipating the speakers' difficulties in the second language. This model focuses on language inventory rather than seeking a connection between production and perception. This model proposes that L2 learners transfer the phonological system of their native language and assumes that the errors are due to this transfer or to the interference from the L1.

According to CAH, it is predicted that L2 learners **do not have many difficulties** when three of these things occur:

- If a phone exists in both the native language and the second language (phonemic existence). In the case of Vietnamese, an example would be the English /f/, which also exists in Vietnamese. This is expected because there should be a positive transfer from the first to the second language.
- An allophone exists and it is different in L1 and L2. For example, the English /s/ is a voiceless alveolar fricative, however, in Vietnamese it is a voiceless dental fricative.
- When a phone exists in both L1 and L2 but is constrained by different phonotactics. For example, Vietnamese speakers may find difficult to pronounce the English /f/ in coda position, because it is forbidden in Vietnamese.

Difficulties arise for speakers when a phone is present in the L2 but not in their native language, resulting in a distinct element (Lado, 1957; Cheon, 2005). For Vietnamese speakers learning English it would be the case of /ʃ/ and /θ/.

This model is considered to have numerous problems, since comparing only the phonological inventory might not be sufficient (Kitikanan, 2016). Regardless, this model is still quite useful to help predict the difficulties Vietnamese speakers can potentially face with English fricatives. On one side, English has /f, v, θ, ð, s, z, ʃ, ʒ, h/ as fricatives and on the other side, Vietnamese has /f, v, s, x, ɣ, h/ and also /z, ʂ/ in specific dialects (dialects that are spoken by many Vietnamese speakers), which means the two languages have a different inventory that can be compared.

It is important to add that it is extremely difficult to say two languages are similar and have similar sounds. Also, it cannot be assumed that because two languages are described as having a given phone, that this phone is necessarily exactly the same in both languages. This does not mean they are equal. (Stevens, 1999)

3.4.2. Speech Learning Model

The Speech Learning Model (SLM) (Flege, 1987) establishes a link between perception and production, in the sense that listeners which don't have an accurate perception will not be able to correctly produce it (Kitikanan, 2016). Nonetheless, this does not mean that a poor production comes from a poor perception directly. There can be other factors at play, such as articulatory difficulties, that contribute to inaccuracies in producing a phone.

SLM is helpful in clarifying differences between sounds: *new*, *similar* or *identical*. This terminology was adopted by Flege (1987) and it is crucial to understand for the work conducted in this thesis. This distinction is built according to specific criteria, such as acoustic/auditory judgements and the IPA, which aid in predicting similarity (Cheon, 2005). According to the SLM, the L1 processes are kept in the L2, and the context (onset or coda) is important in L2 perception. In Flege's terminology the differences between L1 and L2 sounds are:

- *New* L2 sound - no equivalent sound exists in L1, both acoustically and perceptually. In Vietnamese it would be the case of the English fricatives /θ/ and /ʃ/.

- *Similar* L2 sound - L1 and L2 have the same IPA symbol, however the two are phonetically different (allophones). In Vietnamese it would be the case of the English fricative /s/, which is an alveolar fricative in English and a dental fricative in Vietnamese.

- *Identical* L2 sound - L1 and L2 have the same IPA symbol, however, there aren't any phonetic or phonemic differences noticed between the two sounds. In Vietnamese it would be the case of the English fricative /f/ and /h/.

This model has a few predictions according to what was said previously. In the SLM, a sound which is *identical* in L2 and L1 is perceived and produced accurately - positive transfer. If a sound is *similar* in L2 and L1, it is not perceived and produced as a native speaker would, because the two sounds are considered equivalent. Thus, this sound is substituted by an L1 sound which resembles more the L2 sound. Similar sounds are a complicated combination for L2 learners, since it might be difficult to distinguish differences in a very similar phone. This leads to a transfer from L1, of a sound the learner knows, to the sound they believe is the same in L2. The *new* L2 sound does not have any equivalent, therefore it can be perceived better than a similar or identical sound, since new sounds are acoustically robust and prominent to L2 listeners.

According to this model, L2 shared fricatives are difficult for Vietnamese learners, while fricatives that are not shared will be easier. In an acoustic analysis, there are some acoustic measurements in which it is difficult to identify if sounds are phonetically similar or different in an L1 and L2, depending on the characteristics.

Hence, if the predictions and hypothesis of SLM are confirmed in Vietnamese, it is expected that the *new* sounds /ʃ, θ/ will be produced with identical phonetic characteristics to those of native speakers of English. The same principle applies to the English sounds /f/ and /h/ in onset position, as they are considered identical sounds. This does not apply for /s/ in the

onset position, as it falls under the category of similar sounds, thus making it more difficult to pronounce.

3.4.3. Perceptual Assimilation Model - L2

The Perceptual Assimilation Model - L2 (PAM-L2) was developed by Best & Tyler (2007) and it derives from the Perceptual Assimilation Model (PAM), designed by Best (1995).

According to this model, learners are capable of perceiving differences between sounds produced by native and non-native speakers. The non-native sounds are likely going to be perceived as the closest articulatory phone in their native language. Therefore, an unfamiliar sound (phonetic segment) is perceived as being part of a good or bad phonological segment of their native language (categorized). It can also be perceived as not having any resemblance to any native sound (uncategorized) or even as a non-linguistic sound that is not speech (non-assimilated). The resemblance between the non-native sound and the native sound is anticipated in order to see what assimilations will be made between the sounds. (Cheon, 2005)

Additionally, it is believed that learners might have difficulties in distinguishing sounds which are not in their native language, when these sounds have similar articulatory gestures as one or more sounds of their native language (Kitikanan, 2016). PAM-L2 considers the phonetic, phonological and articulatory differences between L1 and L2, which leads to the following predictions: (Best & Tyler, 2007; Kitikanan, 2016)

Predictions	Explanation
Two category assimilation	Two sounds from the second language are perceived as equivalent to two L1 sounds.
Category goodness assimilation	Two sounds from the second language are perceived as one L1 sound, however one is more divergent than the other.
Single-category assimilation	Two sounds from the second language are perceived as one L1 sound, however these two sounds can be categorized as good or poor tokens of the one L1 sound.

Uncategorized-uncategorized assimilation	There is no match between L1 and L2 sounds, therefore they are different and are perceived as different.

Table 11 - Predictions based PAM-L2 model

4. Research questions and hypotheses

This chapter presents the research questions and hypotheses addressed in this study. The analysis primarily relies on predictions derived from these hypotheses, which guide our research methodology. The first section covers the hypotheses related to the acoustic analysis, with hypotheses for the spectral properties, the transition information and the amplitude. The second section concerns hypotheses related to the language acquisition perspective, which takes into account the native language of the speakers.

4.1. Detection hypotheses

From the **spectral properties** measurements:

1. The centroid and peak location should distinguish fricatives in terms of the length of the cavity which is in front of the constriction. Therefore, they should be successful in distinguishing fricatives produced in the front of the oral cavity /ɸ, f, θ, s̪, s̺, s, ʃ, ʂ/ from fricatives produced in the back of the oral cavity /ç, x, χ, ħ, h/. (Jongman *et al.*, 2000; Jones & Nolan, 2007; Nirgianaki, 2014; Wikse Barrow *et al.*, 2022) Consequently, within English target fricatives /f, θ, s, ʃ, h/, the labiodental and the interdental should be easy to distinguish from the glottal, since the length of the front cavity is much longer for the glottal than for the other two fricatives.

    a. Within the front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ), it should be easy to distinguish between the fricatives /ɸ, f/ and the fricatives /ʃ, ʂ/, since the length of the cavity in front of the constriction is relatively different.
    b. Within the back fricatives (ç, x, χ, ħ, h), it should be easy to distinguish between /ç, x/ and /h/, since the length of the cavity is relatively distant.
    c. The hardest phones to distinguish based on the length of the front cavity are phones at adjacent places of articulation.

i. (θ/s), (s/ʃ) are an exception and should be distinguishable (Behrens & Blumstein, 1988a; Fu *et al.*, 1999; Jongman *et al.*, 2000; Gordon *et al.*, 2002).

2. The standard deviation, skewness and kurtosis should distinguish between differences in size and shape of the constriction (Stevens, 1999; Jongman *et al.*, 2000). Therefore, they should distinguish between a narrower constriction (s̻, s̪, s, ʃ, ʂ) and a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h). Within English target fricatives (f, θ, s, ʃ, h), the labiodental, the interdental and the glottal have a wider constriction and therefore should be distinguishable from the alveolar and postalveolar, which have a narrower constriction:

    a. Within narrow fricatives it should be possible to make distinctions between the alveolar /s/ and some of the other narrower fricatives (s̻, s̪, ʃ, ʂ);
    b. Within wider fricatives it should be possible to make distinctions between the palatal /ç/ and the other wider fricatives (ɸ, f, θ, x, χ, ħ, h);

3. The most effective spectral properties measurements should be the centroid, skewness and peak location (Forrest *et al.*, 1988; Shadle & Mair, 1996; Jongman *et al.*, 2000; Nirgianaki, 2014).

From the **transition information** measurements:

4. Onset F2 frequency, intercept and slope[17] should distinguish fricatives in terms of the length of the cavity which is in front of the constriction. Therefore, they should be successful in distinguishing fricatives produced in the front of the oral cavity /ɸ, f, θ, s̻, s̪, s, ʃ, ʂ/ from fricatives produced in the back of the oral cavity /ç, x, χ, ħ, h/. Consequently, within English target fricatives (f, θ, s, ʃ, h), the labiodental and the interdental should be easy to distinguish from the glottal, since the length of the front cavity is much longer for the glottal than for the other two fricatives.

[17] These measurements are dependent on the vowels, and it being well pronounced.

a. Within the front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ), it should be possible to distinguish between the fricatives /ɸ, f/ and the fricatives /ʃ, ʂ/, since the length of the cavity in front of the constriction is relatively different.
b. Within the back fricatives (ç, x, χ, ħ, h), it should be possible to distinguish between /ç, x/ and /h/, since the length of the cavity is relatively distant.
c. The hardest phones to distinguish based on the length of the front cavity are phones at adjacent places of articulation.

5. The most effective transition information measurements should be the F2 onset frequency (Nirgianaki, 2014).

From the **amplitude** measurements:

6. Normalized amplitude and relative amplitude should distinguish between differences in size and shape of the constriction (Jongman *et al.*, 2000). Therefore, it should distinguish between a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h). Within English target fricatives (f, θ, s, ʃ, h), the labiodental, the interdental and the glottal have a wider constriction and therefore should be distinguishable from the alveolar and postalveolar, which have a narrower constriction.

    a. Within narrow fricatives it should be possible to make distinctions between the alveolar /s/ and the other narrower fricatives (s̪, s̺, ʃ, ʂ), since all are produced with different constriction shapes and sizes.
    b. Within wider fricatives it should be possible to make distinctions between the palatal /ç/ and the other wider fricatives (ɸ, f, θ, x, χ, ħ, h).

7. Both the normalized amplitude and relative amplitude should be effective measurements in distinguishing fricatives (Jongman *et al.,* 2000; Nirgianaki, 2014; Wikse Barrow *et al.*, 2022).

4.2. Language acquisition hypotheses

All language acquisition hypotheses will have an index of LA (language acquisition) to avoid confusion with the detection hypothesis.

Considering all the information addressed in the state-of-the-art chapter (chapter 3), these are the second acquisition hypothesis for this thesis:

1) Vietnamese speakers should not have problems pronouncing /f, h/ in onset position, because these phones exist in their inventory and can be produced in the onset position.

    a) Vietnamese spelling/orthography does not have the letter "f", it has the letter "ph" for the voiceless labiodental fricative sound. This may make the production of this fricative harder.
    b) Vietnamese spelling/orthography uses the "h" letter.

2) Vietnamese has two coronals in their fricative inventory (s̪, ʂ), while English has three (θ, s, ʃ). Therefore, Vietnamese speakers will have problems pronouncing all three English phones:

    a) The prediction is that Vietnamese speakers will map /θ/ to:

        i) The labiodental and the dental /s̪/, due to being the fricatives with the closest place of articulation in their L1 and also due to their acoustic similarity.
        ii) to the stops /t, tʰ, t̪, t̪ʰ/ due to the fact that Vietnamese uses the latin alphabet and "th" in its spelling/orthography.

    b) The prediction is that Vietnamese speakers will map /s/ as the dental /s̪/ or the retroflex /ʂ/ due to them being the fricatives with the closest place of articulation in their L1 and also due to their acoustic similarity.

c) The prediction is that Vietnamese speakers will map /ʃ/ to:

i) the retroflex /ʂ/, since it is the closest place of articulation in their L1 and also due to acoustic similarities.

ii) the dental /s̪/ due to perceived similarity (Flege, 1987).

5. Methodology

This chapter starts by explaining the sources of the data and its anonymity, followed by an explanation of the criteria employed to the various datasets: human annotated data, testing data and benchmark data (5.1.). Subsequently, the annotations and inter-annotator agreement are described (5.2.). Lastly, the process to measure the data acoustically is detailed. (5.3.).

5.1. Data collection

This section describes the criteria and different phases used to collect the human annotated data and finishes by describing the benchmark data used for comparison.

5.1.1. Human annotated data

The data for this analysis was obtained from the user's data of the ELSA app. All the data selected is anonymous. An ELSA internal tool[18] was created for linguists to have access to the data. However, in this tool there is no access to the identity of the users, which means it is impossible to know who the speakers are. This safeguards the identity of the users.

The human annotated data was acquired in two phases. Firstly, there was a pilot project created in order to get a sense of the existing data, and to create the guidelines for the annotators. In this pilot there were 1000 audios, of which the goal was to get 50 audios of good productions and 50 audios of each error type. This number was chosen in order to have a similar sample size for each type. After reaching these numbers, the remaining audios were excluded from the projects[19]. The second phase was the selection of the main experimental data. Each fricative project had 1000 audios, except /ʃ/ which had 758.[20] /ʃ/ has fewer data points given the difficulty in getting the data. All the audios were annotated by two linguists and only the audios with agreement between both linguists were included in the analysis.

[18] More information about the tool can be found in 2.4.1.1. Tools.
[19] Specific details about the data collection are explained ahead (section 5.2.)
[20] Number of OUT's can be found on page 69.

The last phase of this analysis is to include the reference values of the LibriSpeech[21] dataset.

Human annotated data was selected according to the following criteria:

- Proficiency level of the speakers.
- Mother tongue.
- Context vowels.
- Word position.
- Specific words to be avoided.

This study investigates the errors speakers produce when learning a second language. It is focused on the typical ELSA user, who is neither too advanced nor too beginner.

The selected language for this analysis was Vietnamese[22]. Furthermore, Vietnamese has different dialects, among these there are two main dialects, Saigon and Hanoi[23].

All the fricatives were analyzed in onset position and in word and sentence initial position. Some examples of words in onset position for each fricative are: /f/ - /ˈ**f**æməli/; /θ/ - /ˈ**θ**æŋks/; /s/ - /ˈ**s**æd/; /ʃ/ - /ˈ**ʃ**æl/; /h/ - /ˈ**h**æf/.

Since fricatives in certain vowel contexts were limited due to the limitation of content and the limitation of the English lexicon (some C+V pairs are scarce), for this report one context vowel was chosen - front low /æ/. This vowel is stressed in all fricatives, except in /ʃ/. This fricative includes both stressed and unstressed vowels to level the lack of data. This vowel was selected by observing a histogram with every combination of fricative+vowel diphthong

[21] Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). *Librispeech: An ASR corpus based on public domain audio books*. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5206-5210.
[22] The language is self-declared by the users when signing in the app.
[23] Dialectical differences were not taken into account in this study. Inventory available in section 3.3.1. Inventory and phonetics. It is possible to see the inventory taken into consideration in this thesis and the respective dialectal differences.

frequency, in one day's worth of logs in the app. The occurrence of the vowels with each fricative was analyzed.

No specific words were removed in the search, given that some contexts only occur in a few words with the chosen vowel, and since the search algorithm is made to avoid too many repetitions of the same word and of the same speaker whenever possible. Figure 14 shows all the words found in the data.

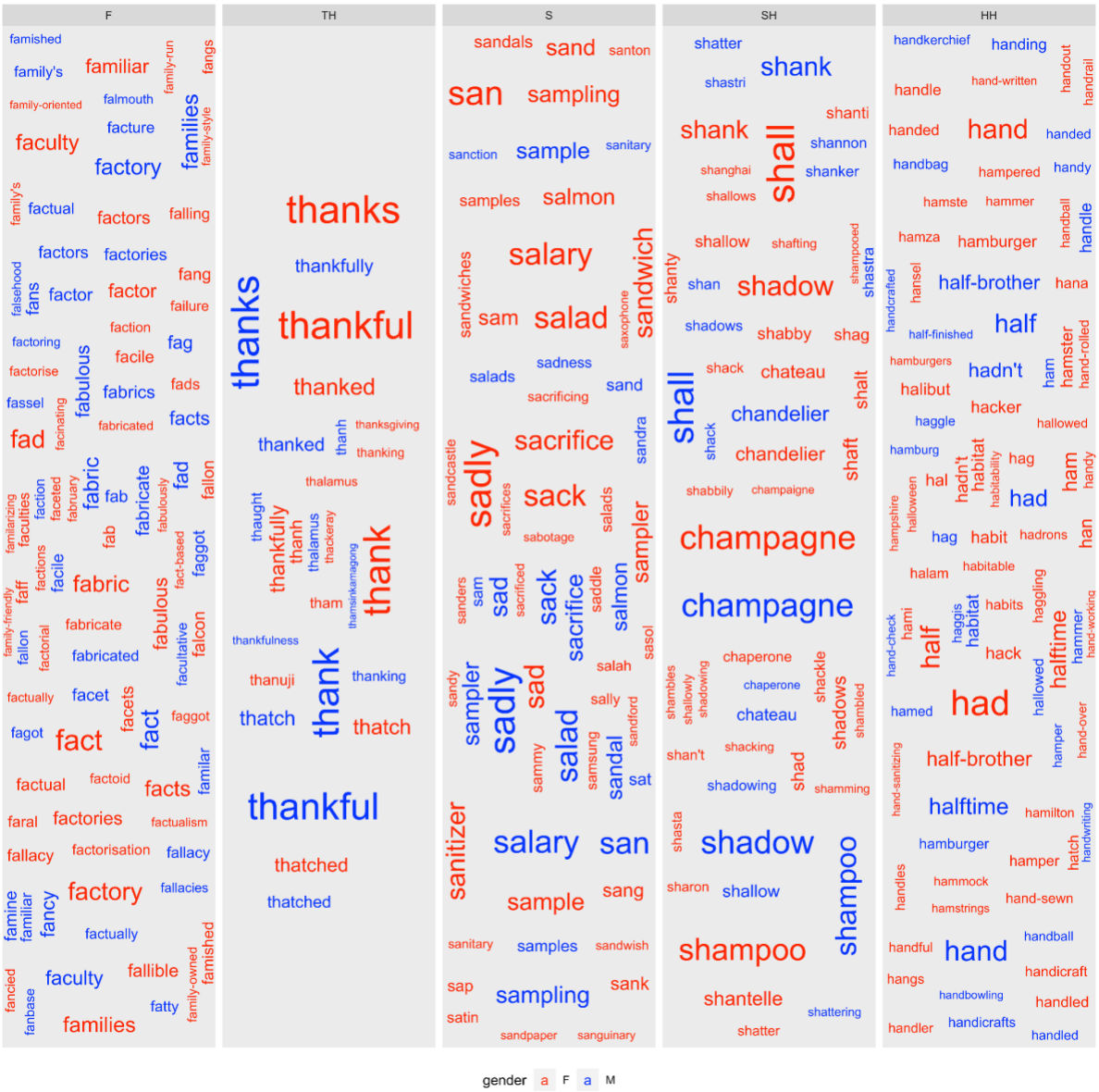

Figure 14 - Words found in the training data

5.1.2. Benchmark data

In order to compare the values from this experiment to reference values, the LibriSpeech[24] dataset was used. This corpus has around 1000 hours of read English speech which is generally suitable for training and evaluating speech recognition systems (Panayotov *et al.*, 2015). For comparison, it was only extracted a maximum of one token per speaker and the data is mostly balanced per gender[25].

5.2. Annotations and inter-annotator agreement

The annotation process was made using a customized version of the web annotations tool *wavesurfer.js*[26]. Pilot annotations were annotated by the author. The human annotated data was annotated by two seasoned linguists, which intends to validate the annotations. An inter-annotator agreement was created. This agreement intends to create clear guidelines which define a strategy to annotate accurately and consistently. It intends to identify ambiguities or difficulties. It also intends to attest the performance of the annotators, to understand the choices made and to ensure the reliability of the annotations. When the annotations process is valid and can be replicated, the annotations should be reliable. To measure the reliability, the kappa coefficient was applied (Artstein, 2017; Brants, 2000).

Annotations were separated into categories ("*OK*", "*ERROR*", "*OUT*"). When a token sounded like a native production of the target phone it was marked as "*OK*". A token which did not sound like a native production of the target phone was marked as "*ERR*". Within the "*ERR*" category there were subcategories which were outlined according to each phone. The annotator had to further classify them.

---

[24] Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). *Librispeech: An ASR corpus based on public domain audio books*. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5206-5210.

[25] Gender numbers: /f/: (M - 488; F - 511), /θ/: (M - 453; F - 547), /s/: (M - 508; F - 482), //ʃ/: (M - 502; F - 498), /h/: (M- 523; F - 477)

[26] https://wavesurfer-js.org/ and/or https://github.com/katspaugh/wavesurfer.js

In the annotations, a lot of "*OUT*" audios were found, either because of the beep of the app. was overlapping with the production, audio quality wasn't good or there was a noisy background. This is particularly important for fricatives, because they are basically noise themselves, unlike, say, vowels, which are much more robust to noisy backgrounds. This excluded the following number of recordings for each fricative: /f/ - 105; /θ/ - 112; /s/ - 129; /ʃ/ - 75; /h/ - 189.

5.3. Acoustic measurements

The acoustic measurements were applied to the human annotated data. This dataset is analyzed acoustically using state of the art metrics (Forrest *et al.*, 1988; Behrens & Blumstein, 1988a; Shadle & Mair, 1996; Jongman *et al.*, 2000;). Firstly, the spectral properties: peak location and the four spectral moments (centroid, standard deviation, skewness and kurtosis); secondly, the transition information: onset F2 frequency, intercept and slope; and thirdly, amplitude: normalized amplitude and relative amplitude.

To extract the measurements for the analysis of the human annotated data different internal Praat scripts[27] were used. The first script extracts the first four spectral moments from fricative spectrograms, as well as intensity and duration. The second script extracts the values of mean formants, the four spectral moments plus F0 and duration.

All measurements were computed following the methods from Jongman *et al.* (2000) as closely as possible. For the spectral moments, a Fast Fourier Transform (FFT) was calculated with a 40ms full Hamming window, which was calculated in four different parts of the fricative. These are the onset part, the middle part, the offset part and also the centered part of the offset of the fricative. The centered offset part included the last 20ms of the fricative and the first 20ms of the vowel.

[27] Modified versions, bits and pieces from Dicanion's Vowel acoustic script, Dicanio's Spectral Tilt Script for Praat, Dicanio's Spectral Envelope Script for Praat, Dicanio's Spectral Moments Script for Praat, Reetz's spectrum script, Kawahara's intensity script.

Onset F2 frequency was calculated also using a FFT with a 23.3ms window at the vowel's midpoint. The slope was calculated at F2 onset and midway in the vowel. F2 at vowel onset was calculated using FFT with a 23.3ms full Hamming window. In this study, the measurement of slope was employed instead of the locus equation methodology used by Jongman *et al.* (2000). This measurement is typically made by extracting several tokens from multiple speakers. However, the nature of this dataset did not allow for the utilization of multiple tokens from the same speaker. Therefore, two instances of a sound were taken into consideration for the analysis. This approach was chosen to adapt to the available data constraints while still allowing for meaningful analysis of the acoustic characteristics of the speech sounds. Since the data is from non-native English speech, vowels which are not the target vowel are excluded from these measurements. This may limit the results of the measurements.

Normalized amplitude (in dB) was calculated for the entire fricative segment. This measurement was done by taking three consecutive pitch periods where the vowel was at maximum intensity and then by calculating the difference between the entire fricative segment amplitude subtracted by the vowel amplitude. Relative amplitude (in dB) was calculated using a FFT at vowel onset with a 23.3ms Hamming window. The amplitude of spectral regions which correspond to F3 and F5 in the vowel was measured in the fricatives.[28] The fricative FFT was calculated at the center of the fricative with a 23.3ms Hamming window. (Jongman *et al.*, 2000; Hedrick & Ohde, 1993). Jongman *et al.* (2000) only considers relative amplitude at F3 for the target fricatives /s, ʃ/ and relative amplitude at F5 for the target fricatives /f, θ/. Given the focus on detection in this study, all fricatives were analyzed both at F3 and F5. This approach ensures that the analysis remains unbiased and allows for the examination of fricatives across various contexts and conditions.

For each spectral property, measurements were taken in four windows and an average of the four windows was also computed. An ANOVA analysis of the type III and Bonferroni post hoc tests were conducted to compare variances across the means of the different phones.

[28] If F3 and F5 were not possible to extract from the vowel the reference value in Praat was used.

6. Results and Discussion

This chapter will expose the results from the human annotated data to verify the hypothesis predictions, followed by a discussion regarding these results. Afterwards, an analysis of second language acquisition results will be presented and discussed.

6.1. Acoustic analysis

6.1.1. Human annotated data

Acoustic analysis was conducted only within the fricative sounds, not taking into account any other classes, such as stops or affricates. The analysis begins with the human-annotated data for the spectral properties: four spectral moments (centroid, standard deviation, skewness and kurtosis) and peak location. Then, the transition information is analyzed: F2 onset frequency, interception and slope. Finally, the amplitude measurements are observed: normalized amplitude and relative amplitude.

6.1.1.1 Spectral properties

This section is organized with the following structure: firstly, the general, averaged tendencies of the four windows and afterward, the specific information for the four window locations. The same structure will be used for all measurements. Within each average and window, the following aspects will be analyzed: general findings; contrasts between English target fricatives; contrasts between English targets and non-target fricatives; relation between annotated target and LibriSpeech benchmarks.

Figure 15 shows all spectral properties averaged across the four windows and figure 16 shows the four windows of the spectral properties can be found.

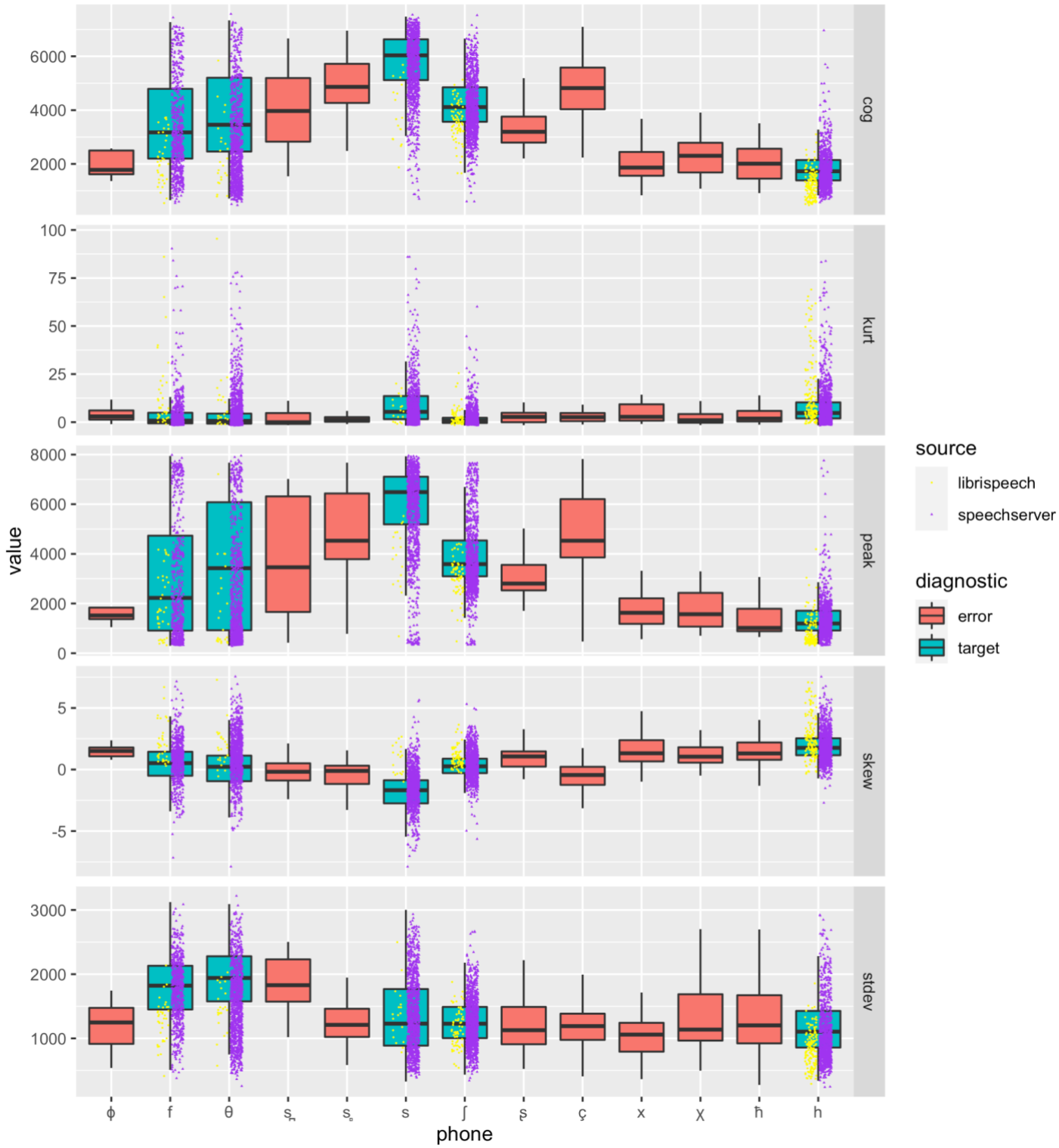

Figure 15 - Spectral properties averaged across the four windows

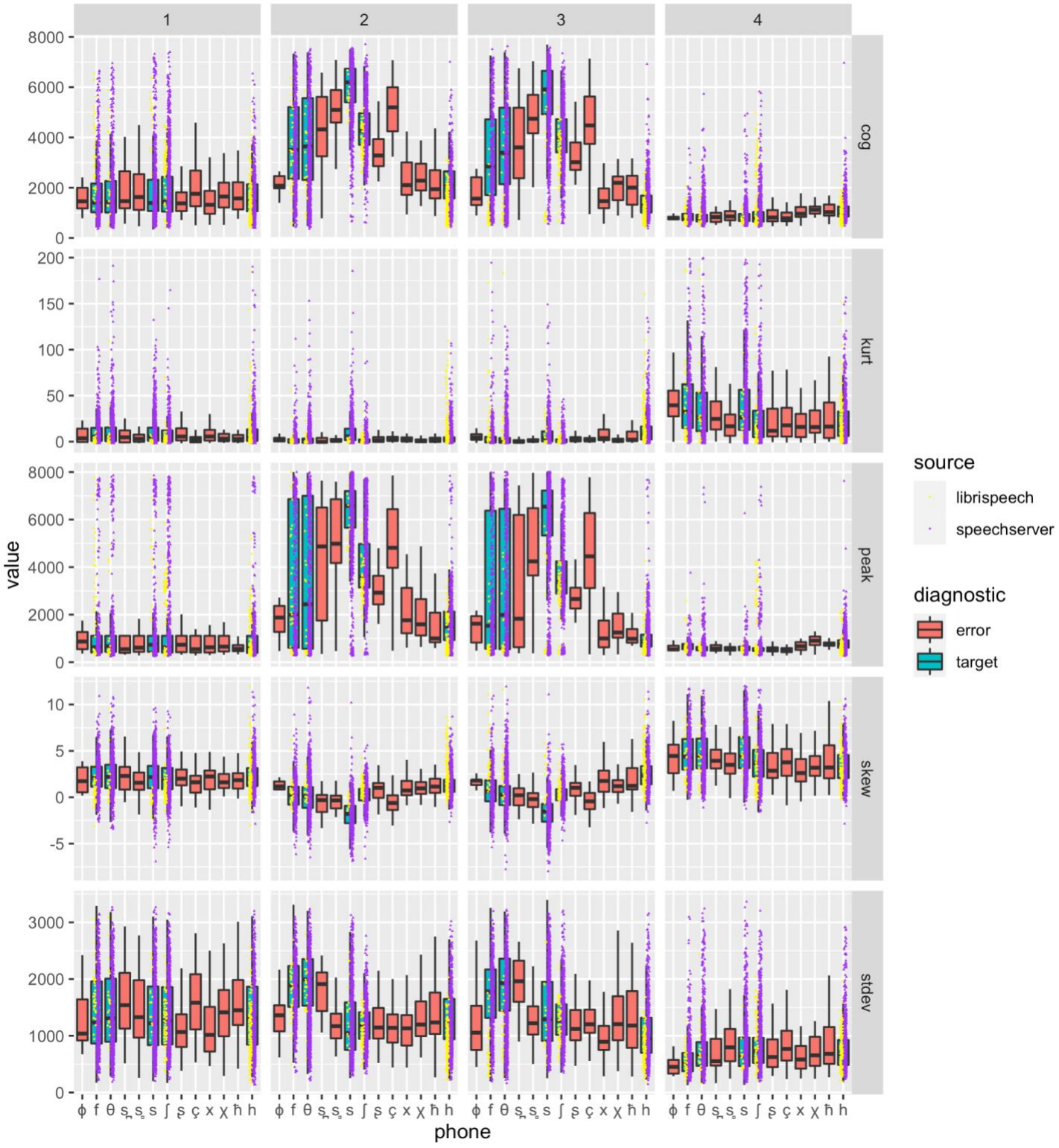

Figure 16 - Four windows of the spectral properties

6.1.1.1.1. Centroid

Below, there is a graphic with a visual representation of the statistical differences in all fricatives, and in all windows and average of the centroid. These differences are analyzed further in more detail.

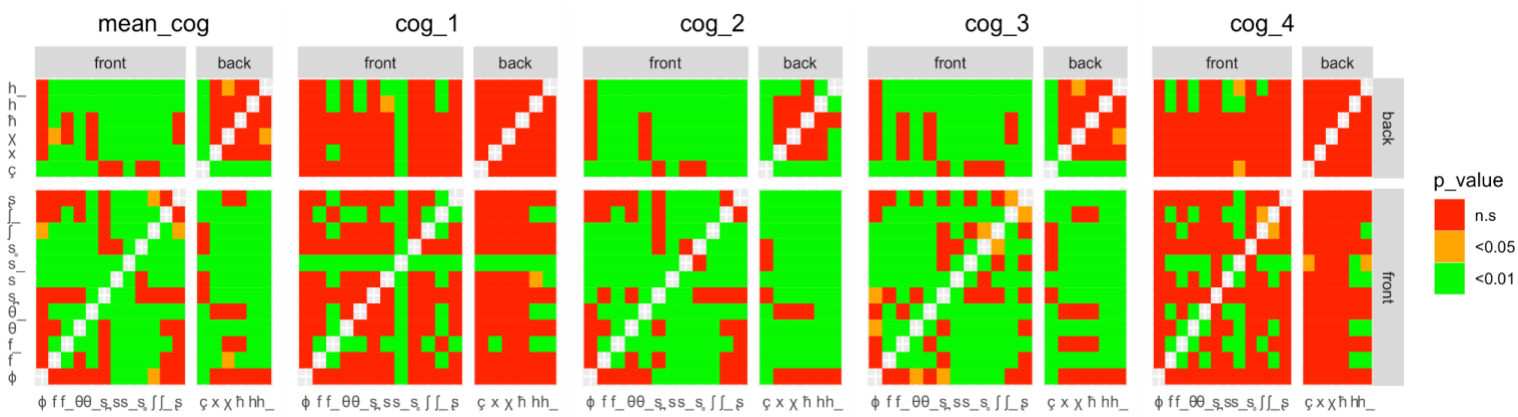

Figure 17 - Visualization of the statistical differences between fricatives in the centroid

Average

In the ANOVA analysis the averaged values of the four windows of the centroid showed a significant main effect of phone identity [(1,8205 F= 386.22), p< 0.01, η^2 = 0.42]. The results of the Bonferroni post hoc tests[29] revealed the following significant contrasts between some front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) and some back fricative (ç, x, χ, ħ, h), (hypothesis 1):

- The bilabial (ɸ) was significantly different (p<0.01) from the back palatal fricative (ç), but **not** from the other back fricatives;
- The labiodental (f) was significantly different (p<0.01) from all back fricatives /x, χ, ħ, h/ and marginally different (p<0.05) from /χ/;
- The interdental (θ) was significantly different (p<0.01) from the back fricatives /ç, h/ and **not** significantly different from /x, χ, ħ/;

[29] Bonferroni post hoc tests were performed with the same number of comparisons [18(18-1)/2 = 153] every time they were ran.

- The dental-alveolar (s̪) and the laminal (s̻) were significantly different ($p<0.01$) from the back fricatives /ħ, h/ and **not** significantly different from /ç, x, χ/;
- The alveolar (s) and the postalveolar (ʃ) were significantly different ($p<0.01$) from all the back fricatives /x, χ, ħ, h/, except /ç/;
- The retroflex (ʂ) was significantly different ($p<0.01$) from the back fricatives /ç, x, h/ and **not** significantly different from /χ, ħ/.

Within the front fricatives (ɸ, f, θ, s̪, s̻, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1a):

- The bilabial (ɸ) was significantly different ($p<0.01$) from /s̻, s/, marginally different ($p<0.05$) from /ʃ/ and **not** significantly different from /f, θ, s̪, ʂ/;
- The labiodental (f) was significantly different ($p<0.01$) from /s̻, s, ʃ/ and **not** significantly different from /ɸ, θ, s̪, ʂ/;
- The interdental (θ) was significantly different ($p<0.01$) from /s, ʃ/ and **not** significantly different from /ɸ, f, s̪, s̻, ʂ/;
- The dental-alveolar (s̪) was significantly different ($p<0.01$) from /s̻, s/, marginally different ($p<0.05$) from /ʃ/ and **not** significantly different from /ɸ, f, θ, ʂ/;
- The laminal (s̻) was significantly different ($p<0.01$) from /ɸ, f, s̪/, marginally different ($p<0.05$) from /ʃ/ and **not** significantly different from /θ, s, ʂ/;
- The alveolar (s) was significantly different ($p<0.01$) from /ɸ, f, θ, s̪, ʂ, ʃ/ and **not** significantly different from /s̻/;
- The postalveolar (ʃ) was significantly different ($p<0.01$) from /f, θ, s/ and marginally different ($p<0.05$) from /ɸ, s̪, s̻, ʂ/;
- The retroflex (ʂ) was significantly different ($p<0.01$) from /s/, marginally different ($p<0.05$) from /ʃ/ and **not** significantly different from /ɸ, f, θ, s̪, s̻/.

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1b):

- The palatal (ç) was significantly different ($p<0.01$) from all the back fricatives /x, χ, ħ, h/ and no other significant differences were found.

As for the English target fricatives (f, θ, s, ʃ, h) there were the following findings, (hypothesis 1):

- /s, ʃ, h/ were distinguishable from all English target fricatives ($p<0.01$);
- /f/ and /θ/ were not significantly different from each other.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[30] and not significantly different from the rest:

- /f/ was significantly different ($p<0.01$) from /s̪/ and **not** significantly different from /ɸ, s̪, ʂ/;
- /θ/ was significantly different ($p<0.01$) from /ç/ and **not** significantly different from /s̪, ʂ/;
- /s/ was significantly different ($p<0.01$) from /s̪, ʂ/ and **not** significantly different from /s̪, ç/;
- /ʃ/ was significantly different ($p<0.01$) from /s̪, ʂ/ and **not** significantly different from /ç/;
- /h/ was significantly different ($p<0.01$) from /ç/ and **not** significantly different from /x, χ, ħ/.

Among all fricatives, the spectral mean was highest for /s/ (5619 Hz) and lowest for /h/ (1732 Hz), as it is possible to observe in table 12, in page 121.

When looking at pairs of fricatives that have adjacent places of articulation (hypothesis 1c), it is possible to observe that:

- the bilabial (ɸ) and the labiodental (f) were **not** significantly different;

[30] These errors will be discussed further in section 6.2. Language acquisition perspective.

- the labiodental (f) and the interdental (θ) were **not** significantly different;
- the interdental (θ) and the dental-alveolar (s̪) were **not** significantly different;
- the dental-alveolar (s̪) and the laminal (s̻) were **not** significantly different;
- the laminal (s̻) and the alveolar (s) were **not** significantly different;
- the alveolar (s) and postalveolar (ʃ) were significantly different ($p<0.01$);
- the postalveolar (ʃ) and the retroflex (ʂ) were significantly different ($p<0.05$);
- the retroflex (ʂ) and the palatal (ç) were significantly different ($p<0.01$);
- the palatal (ç) and the velar (x) were significantly different ($p<0.01$);
- the velar (x) and the uvular (χ) were **not** significantly different;
- the uvular (χ) and the pharyngeal (ħ) were **not** significantly different;
- the pharyngeal (ħ) and the glottal were **not** significantly different;

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it was possible to observe:

- /f, θ, s, ʃ/ were significant different ($p<0.01$) from the LibriSpeech /f, θ, s, ʃ/;
- /h/ was not significantly different from the LibriSpeech /h/.

First window

As it is possible to visually observe in figure 18, in the first window of the centroid there was a marginal main effect of phone identity [(1,3112 F= 1.93), $p< 0.027$, $\eta^2 = 0.0072$]. The results of the Bonferroni post hoc tests revealed that the front fricative /s/ was significantly different ($p<0.01$) from the back fricative /h/, (hypothesis 1). No other significant differences were found in this window.

Second window

The ANOVA analysis performed in the second window of the centroid had a significant main effect of phone identity [(1,3112 F= 242.63), $p< 0.01$, $\eta^2 = 0.48$]. Bonferroni post hoc tests

indicated that there were significant contrasts between some front and some back fricatives, (hypothesis 1):

- The labiodental (f), the interdental (θ), the alveolar (s), the postalveolar (ʃ) and the retroflex (ʂ) were significantly different (p<0.01) from all the back fricatives (ç, x, χ, ħ, h);
- The bilabial (ɸ) was significantly different (p<0.01) from the palatal fricative (ç), but **not** from the other back fricatives;
- The dental-alveolar (s̪) and the laminal (s̻) were significantly different (p<0.01) from the back fricatives (x, χ, ħ, h) and **not** significantly different from /ç/.

Within the front fricatives (ɸ, f, θ, s̪, s̻, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1a):

- The bilabial (ɸ) was significantly different (p<0.01) from /s̻, s, ʃ/, marginally different (p<0.05) from /s̪/ and **not** significantly different from /f, θ, ʂ/;
- The labiodental (f) was significantly different (p<0.01) from /s̻, s, ʃ/ and **not** significantly different from /ɸ, θ, s̪, ʂ/;
- The interdental (θ) was significantly different (p<0.01) from /s̪, s̻, s, ʃ/ and **not** significantly different from /ɸ, f, ʂ/;
- The dental-alveolar (s̪) was significantly different (p<0.01) from /θ, s, ʃ, ʂ/, marginally different (p<0.05) from /ɸ/ and **not** significantly different from /f, s̪, s̻/;
- The laminal (s̻) was significantly different (p<0.01) from /ɸ, f, θ, s, ʃ, ʂ/ and **not** significantly different from /s̪/;
- The alveolar (s) was significantly different (p<0.01) from /ɸ, f, θ, s̪, s̻, ʂ, ʃ/ and **not** significantly different from /s̻/;
- The postalveolar (ʃ) was significantly different (p<0.01) from all the front fricatives;
- The retroflex (ʂ) was significantly different (p<0.01) from /s̪, s̻, s, ʃ/ and **not** significantly different from /ɸ, f, θ/.

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1b):

- The palatal (ç) was significantly different ($p<0.01$) from all the back fricatives /x, χ, ħ, h/ and no other significant differences were found.

As for the English target fricatives (f, θ, s, ʃ, h) there were the following findings, (hypothesis 1):

- /s, ʃ, h/ were distinguishable from all English target fricatives ($p<0.01$);
- /f/ and /θ/ were not significantly different from each other.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative **errors** that were found in the annotations[31] and not significantly different from the rest:

- /f/ was significantly different ($p<0.01$) from /s̪/ and **not** significantly different from /ɸ, s̪, ʂ/;
- /θ/ was significantly different ($p<0.01$) from /s̪, ç/ and **not** significantly different from /ʂ/;
- /s/ and /ʃ/ were significantly different from all non-English fricatives ($p<0.01$).
- /h/ was significantly different ($p<0.01$) from /ç/ and **not** significantly different from /x, χ, ħ/.

When looking at pairs of fricatives that have adjacent places of articulation (hypothesis 1c), it was possible to observe that:

- the bilabial (ɸ) and the labiodental (f) were **not** significantly different;
- the labiodental (f) and the interdental (θ) were **not** significantly different;
- the interdental (θ) and the dental-alveolar (s̪) were significantly different ($p<0.01$);

[31] These errors will be discussed further in section 6.2. Language acquisition perspective.

- the dental-alveolar (s̪) and the laminal (s̻) were significantly different (p<0.01);
- the laminal (s̻) and the alveolar (s) were significantly different (p<0.01);
- the alveolar (s) and postalveolar (ʃ) were significantly different (p<0.01);
- the postalveolar (ʃ) and the retroflex (ʂ) were significantly different (p<0.01);
- the retroflex (ʂ) and the palatal (ç) were significantly different (p<0.01);
- the palatal (ç) and the velar (x) were **not** significantly different;
- the velar (x) and the uvular (χ) were **not** significantly different;
- the uvular (χ) and the pharyngeal (ħ) were **not** significantly different;
- the pharyngeal (ħ) and the glottal (h) were **not** significantly different;

Third window

In the ANOVA analysis, the third window had a significant main effect of phone identity (1,3112 F= 214.95], p< 0.01, η2 = 0.45]. Bonferroni post hoc tests indicated that there were significant contrasts between some front and some back fricatives, (hypothesis 1):

- The labiodental (f), the interdental (θ), the alveolar (s) and the retroflex (ʂ) were significantly different (p<0.01) from all the back fricatives (ç, x, χ, ħ, h);
- The bilabial (ɸ) was significantly different (p<0.01) from the palatal fricative (ç), but **not** from the other back fricatives;
- The dental-alveolar (s̪), the laminal (s̻) and the postalveolar (ʃ) were significantly different (p<0.01) from the back fricatives (x, χ, ħ, h), but **not** significantly different from /ç/.

Within the front fricatives (ɸ, f, θ, s̪, s̻, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1a):

- The bilabial (ɸ) is significantly different (p<0.01) from /s̻, s, ʃ/, marginally different (p<0.05) from /θ, s̪/ and **not** significantly different from /f, ʂ/;

- The labiodental (f) is significantly different (p<0.01) from /θ, s̪, s, ʃ/ and **not** significantly different from /ɸ, s̻, ʂ/;
- The interdental (θ) is significantly different (p<0.01) from /f, s̻, s̪, s, ʃ/, marginally different (p<0.05) from /ɸ/ and **not** significantly different from /ʂ/;
- The dental-alveolar (s̻) is significantly different (p<0.01) from /θ, s, ʂ/, marginally different (p<0.05) from /ɸ, ʃ/ and **not** significantly different from /f, s̻, s̪/;
- The laminal (s̪) is significantly different (p<0.01) from /ɸ, f, θ, ʂ/, marginally different (p<0.05) from /ʃ/, and **not** significantly different from /s̻, s/;
- The alveolar (s) is significantly different (p<0.01) from /ɸ, f, θ, s̻, ʃ, ʂ/ and **not** significantly different from /s̪/;
- The postalveolar (ʃ) is significantly different (p<0.01) from /ɸ, f, θ, s/, marginally different (p<0.05) from /s̻, s̪/ and **not** significantly different from /ʂ/;
- The retroflex (ʂ) is significantly different (p<0.01) from /s̻, s̪, s/ and **not** significantly different from /ɸ, f, θ, ʃ/.

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1b):

- The palatal (ç) was significantly different (p<0.01) from all the back fricatives /x, χ, ħ, h/ and no other significant differences were found.

As for the English target fricatives (f, θ, s, ʃ, h), Bonferroni post hoc tests also indicated that all target English fricatives were significantly different from each other (p<0.01), (hypothesis 1).

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[32] and not significantly different from the rest:

[32] These errors will be discussed further in section 6.2. Language acquisition perspective.

- /f/ was significantly different (p<0.01) from /s̪/ and **not** significantly different from /ɸ, s̪, ʂ/;
- /θ/ was significantly different (p<0.01) from /s̪, ç/ and **not** significantly different from /ʂ/;
- /s/ was significantly different (p<0.01) from /s̪, ʂ, ç/ and **not** significantly different from /s̪/;
- /ʃ/ was marginally different (p<0.05) from /s̪/ and **not** significantly different from /ʂ, ç/;
- /h/ was significantly different (p<0.01) from /ç/ and **not** significantly different from /x, χ, ħ/.

When looking at pairs of fricatives that have adjacent places of articulation (hypothesis 1c), it was possible to observe that:

- the bilabial (ɸ) and the interdental (θ) were marginally different from each other (p<0.05);
- the labiodental (f) and interdental (θ) were significantly different from each other (p<0.01);
- the interdental (θ) and the dental-alveolar (s̪) were significantly different (p<0.01);
- the dental-alveolar (s̪) and the laminal (s̪) were **not** significantly different from each other;
- the laminal (s̪) and the alveolar (s) were **not** significantly different from each other;
- the alveolar (s) and postalveolar (ʃ) were significantly different (p<0.01);
- the postalveolar (ʃ) and the retroflex (ʂ) were **not** significantly different;
- the retroflex (ʂ) and the palatal (ç) were **not** significantly different;
- the palatal (ç) and the velar (x) were significantly different (p<0.01) from each other ;
- the velar (x) and the uvular (χ) were **not** significantly different;
- the uvular (χ) and the pharyngeal (ħ) were **not** significantly different;
- the pharyngeal (ħ) and the glottal were **not** significantly different.

Fourth window

In the fourth window of the centroid there was a marginal main effect of phone identity [(1,3112 F= 13.1), p< 0.01, η^2 = 0.05]. Bonferroni post hoc tests indicated that there were significant contrasts between some front and some back fricatives, (hypothesis 1):

- The labiodental (f) was significantly different (p<0.01) from the back fricatives /χ, ħ, h/, marginally different (p<0.05) from /x/ and **not** significantly different from /ç/;
- The interdental (θ) was significantly different (p<0.01) from /χ, h/, marginally different (p<0.05) from /ħ/ and **not** significantly different from /ç, x/;
- The alveolar (s) was significantly different (p<0.01) from all back fricatives, except /ç/;
- The postalveolar (ʃ) was significantly different (p<0.01) from /χ/ and **not** significantly different from /ç, x, ħ, h/;
- The retroflex /ʂ/ was significantly different (p<0.01) from /χ, h/ and **not** significantly different from /ç, x, ħ/;
- The rest of the front fricatives (ɸ, s̪, s̪ʲ) were **not** significantly different from any of the back fricatives.

Within the front fricatives (ɸ, f, θ, s̪, s̪ʲ, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1a):

- The labiodental (f) was significantly different (p<0.01) from /s̪ʲ/ and **not** significantly different from any of the other front fricatives;
- The interdental (θ) was significantly different (p<0.01) from /s̪, s̪ʲ/ and **not** significantly different from /ɸ, f, s, ʃ, ʂ/;
- The dental-alveolar (s̪) was significantly different (p<0.01) from /θ/ and **not** significantly different from any of the other front fricatives;
- The laminal (s̪ʲ) was significantly different (p<0.01) from /f, θ/, marginally different (p<0.05) from /s/ and **not** significantly different from /ɸ, s̪, s̪ʲ, ʃ, ʂ/;

- The alveolar (s) was marginally different ($p<0.05$) from /s̪/ and **not** significantly different from all the other front fricatives;
- The bilabial (ɸ), the postalveolar (ʃ) and the retroflex (ʂ) were **not** significantly different from any of the front fricatives.

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated that none of the back fricatives were significantly different, (hypothesis 1b).

As for the English target fricatives (f, θ, s, ʃ, h) there were the following findings, (hypothesis 1d, 1e):

- /h/ was significantly different ($p<0.01$) from all English target fricatives;
- All the other fricatives were not significantly different.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[33] and not significantly different from the rest:

- /f/ was significantly different ($p<0.01$) from /s̪/ and **not** significantly different from /ɸ, s̪, ʂ/;
- /θ/ was significantly different ($p<0.01$) from /s̪, ç/ and **not** significantly different from /ʂ, ç/;
- /s/ was marginally different ($p<0.05$) from /s̪/ and **not** significantly different from /s̪, ʂ, ç/;
- /ʃ/ was not significantly different from any non-English fricative errors;
- /h/ was not significantly different from any non-English fricative errors;

When looking at pairs of fricatives that have adjacent places of articulation (hypothesis 1c), it was possible to observe that:

- the bilabial (ɸ) and the labiodental (f) were **not** significantly different;

[33] These errors will be discussed further in section 6.2. Language acquisition perspective.

- ❖ the labiodental (f) and the interdental (θ) were **not** significantly different;
- ❖ the interdental (θ) and the dental-alveolar (s̪) were significantly different (p<0.01);
- ❖ the dental-alveolar (s̪) and the laminal (s̻) were **not** significantly different;
- ❖ the laminal (s̻) and the alveolar (s) were marginally different (p<0.05);
- ❖ the alveolar (s) and postalveolar (ʃ) were **not** significantly different;
- ❖ the postalveolar (ʃ) and the retroflex (ʂ) were **not** significantly different;
- ❖ the retroflex (ʂ) and the palatal (ç) were **not** significantly different;
- ❖ the palatal (ç) and the velar (x) were **not** significantly different;
- ❖ the velar (x) and the uvular (χ) were **not** significantly different;
- ❖ the uvular (χ) and the pharyngeal (ħ) were **not** significantly different;
- ❖ the pharyngeal (ħ) and the glottal were **not** significantly different;

6.1.1.1.2. Standard deviation

Below, there is a graphic with a visual representation of the statistical differences in all fricatives and in all windows and average of the standard deviation. These differences are analyzed further in more detail.

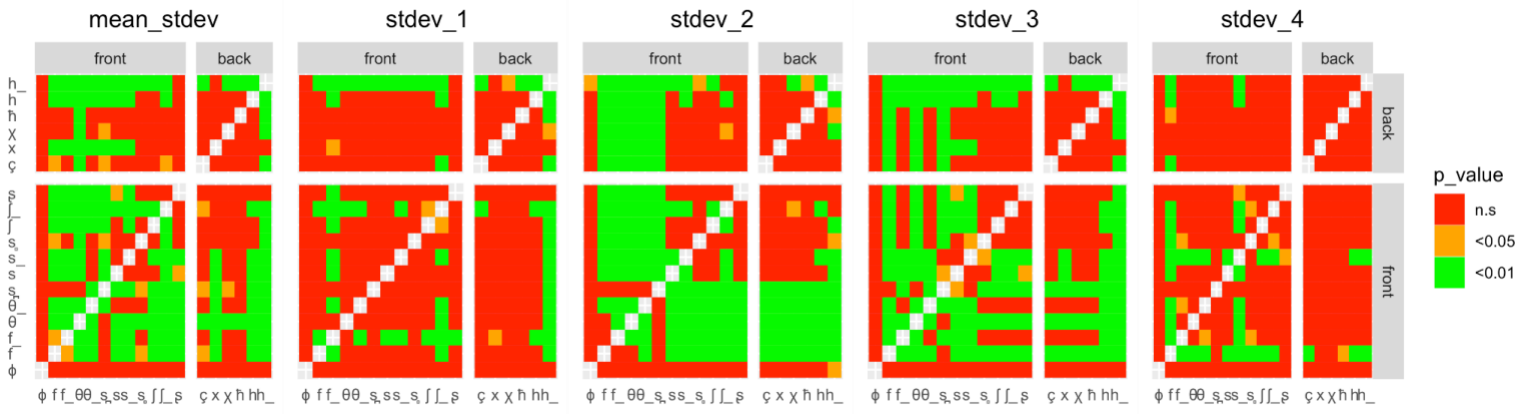

Figure 18 - Visualization of the statistical differences between fricatives in standard deviation

Average

In the ANOVA analysis, the values of the four windows of the standard deviation were averaged and there was a significant main effect of phone identity [(1,8205 F= 102.68), p< 0.01, η^2 = 0.16]. Bonferroni post hoc tests indicated that there were significant contrasts between some phones with a more narrow constriction (s̪, ʂ, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The dental-alveolar (s̪) was significantly different (p<0.01) from /ç, h/ and **not** significantly different from /ɸ, f, θ, x, χ, ħ/;
- The laminal (ʂ) was marginally different (p<0.05) from /f/ and **not** significantly different from /ɸ, θ, ç, x, χ, ħ, h/;
- The alveolar (s) was significantly different (p<0.01) from /f, θ, x, h/ and **not** significantly different from /ɸ, ç, χ, ħ/;
- The postalveolar (ʃ) was significantly different (p<0.01) from /f, θ/ and **not** significantly different from /ɸ, ç, x, χ, ħ, h/;
- The retroflex (ʂ) was significantly different (p<0.01) from /f, θ, x/ and **not** significantly different from /ɸ, ç, χ, ħ, h/;

Within the narrower fricatives (s̪, ʂ, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪) was significantly different (p<0.01) from /s/ and **not** significantly different from /ʂ, ʃ, ʂ/;
- The laminal (ʂ) was **not** significantly different from any of the narrower fricatives;
- The alveolar (s) was significantly different (p<0.01) from /s̪/, marginally different (p<0.05) from /ʂ/ and **not** significantly different from /ʂ, ʃ/;
- The postalveolar (ʃ) was not significantly different from any of the narrower fricatives;
- The retroflex (ʂ) was significantly different (p<0.05) from /s/ and **not** significantly different from /s̪, ʂ, ʃ/.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2b):

- The bilabial (ɸ) was not significantly different from any of the other wider fricatives;
- The labiodental (f) was significantly different ($p<0.01$) from /θ, x, h/, marginally different ($p<0.05$) from /ç/ and **not** significantly different from /ɸ, χ, ħ/;
- The interdental (θ) was significantly different ($p<0.01$) from /f, ç, x, ħ, h/ and **not** significantly different from /ɸ, χ/;
- The palatal (ç) was significantly different ($p<0.01$) from /θ/, marginally different ($p<0.05$) from /f/ and **not** significantly different from /ɸ, x, χ, ħ, h/;
- The velar (x) and the glottal (h) were significantly different ($p<0.01$) from /f, θ/ and **not** significantly different from the rest of the wider fricatives;
- The uvular (χ) and the pharyngeal (ħ) were significantly different ($p<0.01$) from /θ/, but **not** significantly different from the rest of the wider fricatives;

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 2):

- /f/ and /θ/ were significantly different ($p<0.01$) from all other English target fricatives;
- /s/ and /h/ were not significantly different from /ʃ/.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[34] and not significantly different from the rest:

- /f/ was significantly different ($p<0.01$) from /ʂ/, marginally different ($p<0.05$) from /s̪/ and **not** significantly different from /ɸ, s̪/;
- /θ/ was significantly different ($p<0.01$) from /ʂ, ç/ and **not** significantly different from /s̪/;

[34] These errors will be discussed further in section 6.2. Language acquisition perspective.

- /s/ was significantly different (p<0.01) from /s̪/, marginally different (p<0.05) from /ʂ/ and **not** significantly different from /s̺, ç/;
- /ʃ/ was not significantly different from any non-English fricative errors;
- /h/ was not significantly different from any non-English fricative errors.

In this dataset, the standard deviation for sibilant fricatives was lower and for non-sibilants it was higher[35]. Among all fricatives, the standard deviation was highest for /s̪/ (1.35) and lowest for /x, h/ (1.03). If only the English target fricatives are considered, the highest was /f/ (1.32) and the lowest was still /h/ (1.03).

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it is possible to observe:

- /f/ was significantly different (p<0.05) from the LibriSpeech /f/;
- /θ, ʃ, h/ were significant different (p<0.01) from the LibriSpeech /θ, ʃ, h/, respectively;
- /s/ was not significantly different from the LibriSpeech /s/.

First window

In the first window there was not a significant main effect of phone identity [(1,3112 F= 2.13), p< 0.013, η^2 = 0.008]. The results of the Bonferroni post hoc tests revealed that no significant differences were found in this window, (hypothesis 2).

Second window

In the ANOVA analysis, the second window has a significant main effect of phone identity [(1,3112 F= 76.19), p< 0.01, n^2 = 0.22]. Bonferroni post hoc tests indicated that there were significant contrasts between some phones with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

[35] Numbers can be found in table 12, in page 120.

- The dental-alveolar (s̪) was significantly different (p<0.01) from /θ, ç, h/, marginally different from (p<0.05) from /ħ/ and **not** significantly different from /ɸ, f, x, χ/;
- The laminal (s̻) was marginally different (p<0.05) from /f, θ/ and **not** significantly different from the rest of the wider fricatives;
- The alveolar (s), the postalveolar (ʃ) and the retroflex (ʂ) were significantly different (p<0.01) from /f, θ/ and **not** significantly different from the rest of the wider fricatives;

Within the narrower fricatives (s̪, s̻, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪) was significantly different (p<0.01) from /s̻, s/ and **not** significantly different from /ʃ, ʂ/;
- The laminal (s̻) was significantly different (p<0.01) from /s̪/ and **not** significantly different from the rest of the more narrow fricatives;
- The alveolar (s) was significantly different (p<0.01) from /s̪/ and **not** significantly different from the rest of the more narrow fricatives;
- The postalveolar (ʃ) and the retroflex (ʂ) were **not** significantly different from any of the narrower fricatives;

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2b):

- The bilabial (ɸ) was **not** significantly different from any of the wider fricatives;
- The labiodental (f) was significantly different (p<0.01) from /ç, x, χ, ħ, h/ and **not** significantly different from /ɸ, θ/;
- The interdental (θ) was significantly different (p<0.01) from /ç, x, χ, ħ, h/ and **not** significantly different from /ɸ, f/;
- The palatal (ç), the velar (x), the uvular (χ), the pharyngeal (ħ) and the glottal (h) were significantly different (p<0.01) from /f, θ/ and **not** significantly different from the rest of the wider fricatives.

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 2):

- /f/ and /θ/ were significantly different ($p<0.01$) from /s, ʃ, h/;
- /f, θ/ were not significantly different from each other;
- /s, ʃ, h/ were not significantly different from each other.

If only the English target fricatives are considered, then the narrower fricatives (s, ʃ) were significantly different ($p<0.01$) from the wider fricatives (f, θ, h).

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[36] and not significantly different from the rest:

- /f/ was significantly different ($p<0.01$) from /s̪, ʂ/ and **not** significantly different from /ɸ, s̻/;
- /θ/ was significantly different ($p<0.01$) from all non-English fricative errors;
- /s/ was significantly different ($p<0.01$) from /s̻/ and **not** significantly different from /s̪, ʂ, ç/;
- /ʃ/ was not significantly different from any non-English fricative errors;
- /h/ was not significantly different from any non-English fricative errors.

Third window

In the ANOVA analysis, the third window had a significant main effect of phone identity [(1,3112 F= 58.98), $p< 0.01$, $\eta^2 = 0.18$]. Bonferroni post hoc tests indicated that there were significant contrasts between some phones with a more narrow constriction (s̻, s̪, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The dental-alveolar (s̻) was significantly different ($p<0.01$) from /θ, ç, ħ, h/ and **not** significantly different from /ɸ, f, x, χ/;

[36] These errors will be discussed further in section 6.2. Language acquisition perspective.

- The laminal (s̪) and the retroflex (ʂ) were significantly different ($p<0.01$) from /f, θ/ and **not** significantly different from /ɸ, ç, x, χ, ħ, h/;
- The alveolar (s) was significantly different ($p<0.01$) from /f, θ, x, h/ and **not** significantly different from /ɸ, ç, χ, ħ/;
- The postalveolar (ʃ) was significantly different ($p<0.01$) from /f, θ, h/ and **not** significantly different from /ɸ, ç, x, χ, ħ/.

Within the narrower fricatives (s̪, s̪, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪) was significantly different ($p<0.01$) from /s̪/, marginally different ($p<0.05$) from /s/ and **not** significantly different from /ʃ, ʂ/;
- The laminal (s̪) was significantly different ($p<0.01$) from /s̪/ and **not** significantly different from /s, ʃ, ʂ/;
- The alveolar (s) was significantly different ($p<0.01$) from /ʃ/, marginally different ($p<0.05$) from /s̪, ʂ/ and **not** significantly different from /s̪/;
- The postalveolar (ʃ) was significantly different ($p<0.01$) from /s/ and **not** significantly different from /s̪, s̪, ʂ/;
- The retroflex (ʂ) was significantly different ($p<0.05$) from /s/ and **not** significantly different from /s̪, s̪, ʃ/.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2b):

- The bilabial (ɸ) was **not** significantly different from any of the other wider fricatives;
- The labiodental (f) was significantly different ($p<0.01$) from /θ, ç, x, χ, ħ, h/ and **not** significantly different from /ɸ/;
- The interdental (θ) was significantly different ($p<0.01$) from /f, ç, x, χ, ħ, h/ and **not** significantly different from /ɸ/;

- The palatal (ç), the velar (x), the uvular (χ), the pharyngeal (ħ) and the glottal (h) were significantly different (p<0.01) from /f, θ/ and **not** significantly different from the other wider fricatives.

As for the English target fricatives (f, θ, s, ʃ, h), Bonferroni post hoc tests indicated that all English target fricatives were significantly different from each other (p<0.01), (hypothesis 2).

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[37] and not significantly different from the rest:

- /f/ was significantly different (p<0.01) from /s̪, ʂ/ and **not** significantly different from /ɸ, s̻/;
- /θ/ was significantly different (p<0.01) from all non-English fricative errors;
- /s/ was significantly different (p<0.01) from /s̻/ and **not** significantly different from /s̪, ʂ, ç/;
- /ʃ/ was not significantly different from any non-English fricative errors;
- /h/ was not significantly different from any non-English fricative errors.

Fourth window

In the ANOVA analysis, the fourth window had a marginal main effect of phone identity [(1,3112 F= 9.91), p< 0.01, η^2 = 0.04]. Bonferroni post hoc tests indicated that there were significant contrasts between some phones with a more narrow constriction (s̻, s̪, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The laminal (s̪), the alveolar (s) and the postalveolar (ʃ) were significantly different (p<0.01) from /f/, but **not** significantly different from any of the wider fricatives;
- The dental-alveolar (s̻) and the retroflex (ʂ) were **not** significantly different from any of the wider fricatives;

[37] These errors will be discussed further in section 6.2. Language acquisition perspective.

Within the narrower fricatives (ʂ̈, ʂ̟, s, ʃ, ʂ) the Bonferroni post hoc tests indicated that none of the fricatives were significantly different (hypothesis 2a).

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated that these fricatives were not significantly different, except for the labiodental that was significantly different ($p<0.01$) from /θ, ç, h/ and marginally different ($p<0.05$) from /ħ/, (hypothesis 2b).

As for the English target fricatives (f, θ, s, ʃ, h) there were the following findings (hypothesis 2):

- /f/ was significantly different ($p<0.01$) from all English target fricatives;
- All the other fricatives were **not** significantly different.

The English target fricatives were also not significantly different from non-English fricative errors, with the exception of /f/ that was significantly different ($p<0.01$) from /ʂ̟, ç/ and marginally different ($p<0.05$) from /ħ/.

6.1.1.1.3. Skewness

Below, there is a graphic with a visual representation of the statistical differences in all fricatives and in all windows and average of skewness. These differences are analyzed further in more detail.

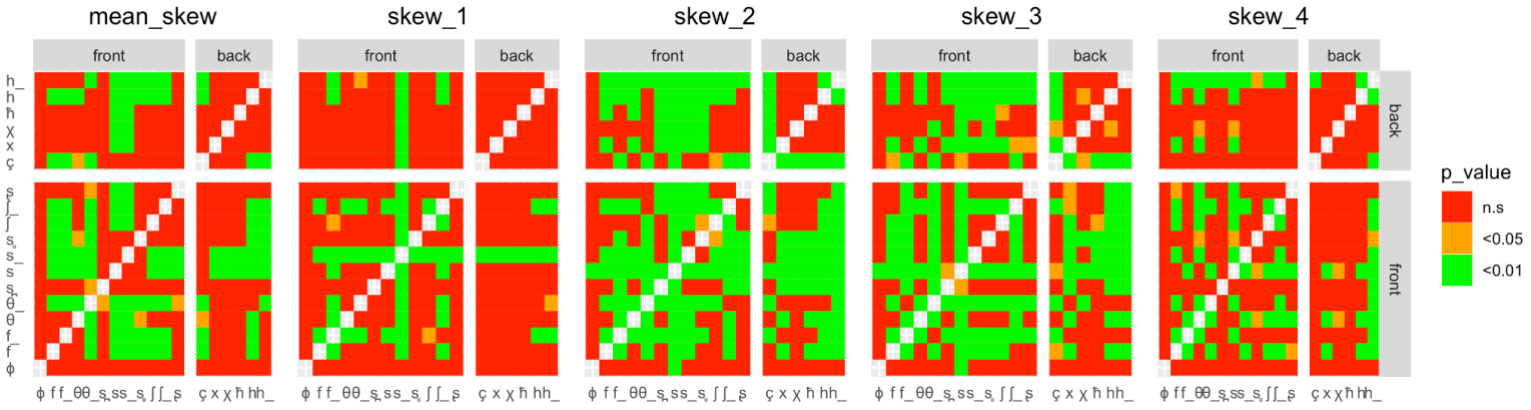

Figure 19 - Visualization of the statistical differences between fricatives in skewness

Average

In the ANOVA analysis, the values of the four windows of skewness were averaged and there was a significant main effect of phone identity [(1,8205 F= 70.23), p< 0.01, η^2 = 0.13]. Bonferroni post hoc tests indicated that there were significant contrasts between some phones with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The dental-alveolar (s̪) was not significantly different from any of the wider fricatives;
- The laminal (s̺) was significantly different (p<0.01) from /f, h/ and **not** significantly different from /ɸ, θ, ç, x, χ, ħ/;
- The alveolar (s) was significantly different (p<0.01) from /f, θ, x, χ, ħ, h/ and **not** significantly different from /ɸ, ç/;
- The postalveolar (ʃ) was significantly different (p<0.01) from /f, h/ and **not** significantly different from /ɸ, θ, ç, x, χ, ħ/;
- The retroflex (ʂ) was **not** significantly different from any of the wider fricatives;

Within the narrower fricatives (s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪) was **not** significantly different from any of the narrower fricatives;
- The laminal (s̺) was **not** significantly different from any of the narrower fricatives;
- The alveolar (s) was significantly different (p<0.01) from /ʃ, ʂ/ and **not** significantly different from /s̪, s̺/;
- The postalveolar (ʃ) was significantly different (p<0.01) from /s/ and **not** significantly different from /s̪, s̺, ʂ/;
- The retroflex (ʂ) was marginally different (p<0.05) from /s/ and **not** significantly different from /s̪, s̺, ʃ/.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2b):

- The bilabial (ɸ) was **not** significantly different from any of the wider fricatives;
- The labiodental (f) was significantly different ($p<0.01$) from /ç, h/ and **not** significantly different from /ɸ, θ, x, χ, ħ/;
- The interdental (θ) was significantly different ($p<0.01$) from /h/, marginally different ($p<0.05$) from /ç/ and **not** significantly different from /ɸ, f, x, χ, ħ/;
- The palatal (ç) was significantly different ($p<0.01$) from /f, h/, marginally different ($p<0.05$) from /θ/ and **not** significantly different from /ɸ, x, χ, ħ/;
- The velar (x), the uvular (χ) and the pharyngeal (ħ) were **not** significantly different from any of the wider fricatives;
- The glottal was significantly different ($p<0.01$) from /f, θ, ç/ and **not** significantly different from /ɸ, x, χ, ħ/.

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 2):

- /s/ and /h/ were significantly different ($p<0.01$) from all other English target fricatives;
- /f/ was significantly different ($p<0.01$) from /s, ʃ, h/, but not significantly different /θ/;
- /ʃ/ was significantly different ($p<0.01$) from /f, s, h/, but not significantly different /θ/;
- /θ/ was significantly different ($p<0.01$) from /s, h/.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[38] and not significantly different from the rest:

- /f/ was significantly different ($p<0.01$) from /s̪/ and **not** significantly different from /ɸ, s̪, ʂ/;

[38] These errors will be discussed further in section 6.2. Language acquisition perspective.

- /θ/ was marginally different (p<0.05) from /ç/ and **not** significantly different from /s̪, ʂ/;
- /s/ was significantly different (p<0.01) from /ʂ/ and **not** significantly different from /s̪, s̺, ç/;
- /ʃ/ was not significantly different from any non-English fricative errors;
- /h/ was significantly different (p<0.01) from /ç/ and **not** significantly different from /x, χ, ħ/.

Among all fricatives, skewness was positive and highest for /h/ (1.950) and negative and lowest for /s/ (-1.560), as shown in table 12, in page 121.

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it is possible to observe:

- /θ/ was significantly different (p<0.01) from the LibriSpeech /θ/;
- /f, s, ʃ, h/ were not significantly different from the LibriSpeech respective /f, s, ʃ, h/.

First window

In the ANOVA analysis the first window did not have a significant main effect of phone identity [(1,3112 F= 1.46), p< 0.01, η^2 = 0.005]. The results of the Bonferroni post hoc tests revealed that no significant differences were found in this window, (hypothesis 2).

Second window

In the ANOVA analysis the second window had a significant main effect of phone identity [(1,3112 F= 129.71), p< 0.01, η2 = 0.33]. Bonferroni post hoc tests indicated that there were significant contrasts between some phones with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The dental-alveolar (s̪) was significantly different (p<0.01) from /x, χ, ħ, h/ and **not** significantly different from /ɸ, f, θ, ç/;

- The laminal (s̪) was significantly different (p<0.01) from /x, χ, ħ, h/ and **not** significantly different from /ɸ, f, θ, ç/;
- The alveolar (s) was significantly different (p<0.01) from all the wider fricatives;
- The postalveolar (ʃ) was significantly different (p<0.01) from /ç, h/, marginally different (p<0.05) from /x, ħ/ and **not** significantly different from /ɸ, f, θ, χ/;
- The retroflex (ʂ) was significantly different (p<0.01) from /ç/, marginally different (p<0.05) from /f, θ/ and **not** significantly different from /ɸ, x, χ, ħ, h/;

Within the narrower fricatives (s̪, s̪, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪) was significantly different (p<0.01) from /s, ʂ/, marginally significantly different (p<0.05) from /ʃ/ and **not** significantly different from /s̪/;
- The laminal (s̪) was significantly different (p<0.01) from /s/, marginally different (p<0.05) from /ʃ/ and **not** significantly different from /s̪, ʂ/;
- The alveolar (s) was significantly different (p<0.01) from all the narrower fricatives;
- The postalveolar (ʃ) was significantly different (p<0.01) from /s/, marginally different (p<0.05) from /s̪, s̪/ and **not** significantly different from /ʂ/;
- The retroflex (ʂ) was marginally different (p<0.05) from /s̪, s̪, s/ and **not** significantly different from /ʃ/.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2b):

- The bilabial (ɸ) was not significantly different from any of the wider fricatives;
- The labiodental (f) was significantly different (p<0.01) from /x, ħ, h/, marginally different (p<0.05) from /ç/ and **not** significantly different from /ɸ, θ, χ/;
- The interdental (θ) was significantly different (p<0.01) from /x, ħ, h/, marginally different (p<0.05) from /χ/ and **not** significantly different from /ɸ, f, ç/;

- The palatal (ç) was significantly different ($p<0.01$) from /x, χ, ħ, h/, marginally different ($p<0.05$) from /f/ and **not** significantly different from /ɸ, θ/;
- The uvular (χ) was significantly different ($p<0.01$) from /ç/, marginally different ($p<0.05$) from /θ/ and **not** significantly different from /ɸ, f, x, ħ, h/;
- The velar (x), the pharyngeal (ħ) and the glottal (h) were significantly different ($p<0.01$) from /f, θ, ç/, but not from the rest of the wider fricatives;

As for the English target fricatives (f, θ, s, ʃ, h) there were the following findings (hypothesis 2):

- /s, h/ were significantly different ($p<0.01$) from all English target fricatives;
- /f, θ, ʃ/ were not significantly different from each other.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[39] and not significantly different from the rest:

- /f/ was marginally different ($p<0.05$) from /ʂ, ʂ̪/ and **not** significantly different from /ɸ, ʂ̪, ʂ̺/;
- /θ/ was marginally different ($p<0.05$) from /ʂ/ and **not** significantly different from /ʂ̪, ç/;
- /s/ was significantly different ($p<0.01$) from all non-English fricative errors;
- /ʃ/ was significantly different ($p<0.01$) from /ç/, marginally different ($p<0.05$) from /ʂ̺/ and **not** significantly different from /ʂ/;
- /h/ was significantly different ($p<0.01$) from /ç/ and **not** significantly different from /x, χ, ħ/.

Third window

In the ANOVA analysis the third window had a significant main effect of phone identity [(1,3112 F= 126.28), $p< 0.01$, $\eta 2 = 0.32$]. Bonferroni post hoc tests indicated that there are

[39] These errors will be discussed further in section 6.2. Language acquisition perspective.

significant contrasts between some phones with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The dental-alveolar (s̪) is significantly different (p<0.01) from /x, χ, h/, marginally different (p<0.05) from /ħ/ and **not** significantly different from /ɸ, f, θ, ç/;
- The laminal (s̺) is significantly different (p<0.01) from /f, x, χ, ħ, h/ and **not** significantly different from /ɸ, θ, ç/;
- The alveolar (s) is significantly different (p<0.01) from all the wider fricatives;
- The postalveolar (ʃ) is significantly different (p<0.01) from /f, x, ħ, h/, marginally different (p<0.05) from /χ/ and **not** significantly different from /ɸ, θ, ç/;
- The retroflex (ʂ) is significantly different (p<0.01) from /x, h/ and **not** significantly different from /ɸ, f, θ, ç, χ, ħ/.

Within the narrower fricatives (s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪) is significantly different (p<0.01) from /s/, marginally different (p<0.05) from /ʂ/ and **not** significantly different from /s̺, ʃ/;
- The laminal (s̺) is significantly different (p<0.01) from /s/, marginally different (p<0.05) from /ʂ/ and **not** significantly different from /s̪, ʃ/;
- The alveolar (s) is significantly different (p<0.01) from all narrower fricatives;
- The postalveolar (ʃ) is significantly different (p<0.01) from /s/ and **not** significantly different from /s̪, s̺, ʂ/;
- The retroflex (ʂ) is significantly different (p<0.05) from /s/, marginally different (p<0.05) from /s̪, s̺/ and **not** significantly different from /ʃ/.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2b):

- The bilabial (ɸ) is **not** significantly different from any of the wider fricatives;

- The labiodental (f) is significantly different ($p<0.01$) from /ç, x, h/, marginally different ($p<0.05$) from /θ/ and **not** significantly different from /ɸ, χ, ħ/;
- The interdental (θ) is significantly different ($p<0.01$) from /x, ħ, h/, marginally different ($p<0.05$) from /f/ and **not** significantly different from /ɸ, ç, χ/;
- The palatal (ç) is significantly different ($p<0.01$) from /f, x, χ, ħ, h/ and **not** significantly different from /ɸ, θ/;
- The velar (x) is significantly different ($p<0.01$) from /f, θ, ç/ and **not** significantly different from /ɸ, χ, ħ, h/;
- The uvular (χ) is significantly different ($p<0.01$) from /ç, h/ and **not** significantly different from /ɸ, f, θ, x, ħ/;
- The pharyngeal (ħ) is significantly different ($p<0.01$) from /θ, ç/ and **not** significantly different from /ɸ, f, x, χ, h/;
- The glottal (h) is significantly different ($p<0.01$) from /f, θ, ç, χ/ and **not** significantly different from /ɸ, x, ħ/.

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 2):

- /f, s, h/ are significantly different ($p<0.01$) from all English target fricatives;
- /θ, ʃ/ are not significantly different from each other, but are significantly different ($p<0.01$) from the rest of the English target fricatives.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[40] and not significantly different from the rest:

- /f/ was significantly different ($p<0.01$) from /s̪/ and **not** significantly different from /ɸ, s̪, ʂ/;
- /θ/ was not significantly different from any non-English fricative errors;

[40] These errors will be discussed further in section 6.2. Language acquisition perspective.

- /s/ was significantly different (p<0.01) from all non-English fricative errors;
- /ʃ/ was not significantly different from any non-English fricative errors;
- /h/ was significantly different (p<0.01) from /ç, χ/ and **not** significantly different from /x, ħ/.

Fourth window

In the ANOVA analysis the fourth window has a marginal main effect of phone identity [(1,3112 F= 16.01, p< 0.01, η^2 = 0.06. Bonferroni post hoc tests indicated that there are significant contrasts between some phones with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The dental-alveolar (s̪) and the laminal (s̺) are marginally different (p<0.05) from /θ/ and **not** significantly different from /ɸ, f, ç, x, χ, ħ, h/;
- The alveolar (s) is significantly different (p<0.01) from /x, h/, marginally different (p<0.05) from /χ/ and **not** significantly different from /ɸ, f, θ, ç, ħ/;
- The postalveolar (ʃ) and the retroflex (ʂ) are significantly different (p<0.01) from /f, θ/ and **not** significantly different from /ɸ, ç, x, χ, ħ, h/;

Within the narrower fricatives (s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪) is **not** significantly different from any of the narrower fricatives.
- The laminal (s̺) is marginally different (p<0.05) from /s/ and **not** significantly different from /s̺, ʃ, ʂ/;
- The alveolar (s) is significantly different (p<0.01) from /ʃ, ʂ/, marginally different (p<0.05) from /s̺/ and **not** significantly different from /s̪/;
- The postalveolar (ʃ) is significantly different (p<0.01) from /s/ and **not** significantly different from /s̪, s̺, ʂ/;
- The retroflex (ʂ) is significantly different (p<0.01) from /s/ and **not** significantly different from /s̪, s̺, ʃ/.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2b):

- the bilabial (ɸ) and the palatal (ç) are **not** significantly different from any of the wider fricatives;
- the labiodental (f) is significantly different ($p<0.01$) from /x, h/ and **not** significantly different from /ɸ, θ, ç, χ, ħ/;
- the interdental (θ) is significantly different ($p<0.01$) from /x, χ, h/, marginally different ($p<0.05$) from /ħ/ and **not** significantly different from /ɸ, f, ç/;
- the velar (x) and the glottal (h) are significantly different ($p<0.01$) from /f, θ/ and **not** significantly different from /ɸ, ç, χ, ħ, h/;
- the uvular (χ) is significantly different ($p<0.01$) from /θ/ and **not** significantly different from /ɸ, f, ç, x, ħ, h/;
- the pharyngeal (ħ) is marginally different ($p<0.05$) from /θ/ and **not** significantly different from /ɸ, f, ç, x, ħ, h/;

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 2):

- /f, θ, s/ are significantly different ($p<0.01$) from /ʃ, h/ but are not significantly different between themselves;
- /ʃ, h/ are only significantly different ($p<0.01$) from /f, θ, s/ but not significantly different between themselves;

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[41] and not significantly different from the rest:

[41] Ibid.

- /f/ was significantly different (p<0.01) from /ʂ/ and **not** significantly different from /ɸ, s̪, ʂ̪/;
- /θ/ was significantly different (p<0.01) from /s̪/ and **not** significantly different from /ʂ, ç/;
- /s/ was significantly different (p<0.01) from all /ʂ/, marginally different (p<0.05) from /ʂ̪/ and **not** significantly different from /s̪, ç/;
- /ʃ/ was not significantly different from any non-English fricative errors;
- /h/ was not significantly different from any non-English fricative errors;

6.1.1.1.4. Kurtosis

Below, there is a graphic with a visual representation of the statistical differences in all fricatives and in all windows and average of kurtosis. These differences are analyzed further in more detail.

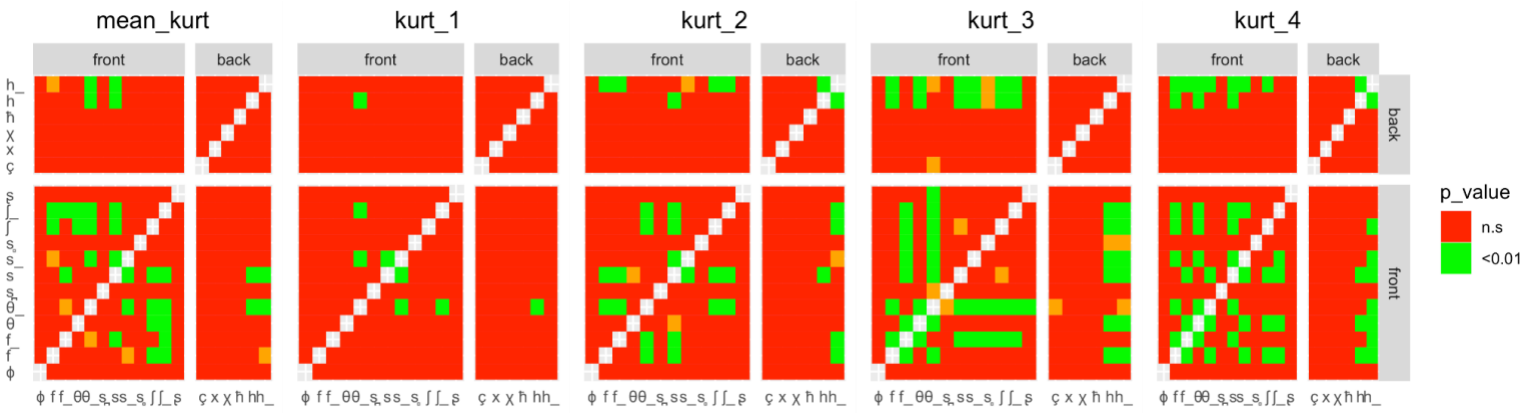

Figure 20 - Visualization of the statistical differences between fricatives in kurtosis

Average

In the ANOVA analysis, the values of the four windows of skewness were averaged and there was a significant main effect of phone identity [(1,8205 F= 8.64), p< 0.01, η2 = 0.018]. Bonferroni post hoc tests indicated that there are significant contrasts between some phones with a more narrow constriction (s̪, ʂ̪, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The dental-alveolar (s̪), the laminal (s̻) and the retroflex (ʂ) are **not** significantly different from any of the wider fricatives;
- The alveolar (s) is significantly different ($p<0.01$) from /h/ and **not** significantly different from the other wider fricatives;
- The postalveolar (ʃ) is significantly different ($p<0.01$) from /f, θ/ and **not** significantly different from /ɸ, ç, x, χ, ħ, h/;

Within the narrower fricatives (s̪, s̻, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪), the laminal (s̻) and the retroflex (ʂ) are **not** significantly different from any of the narrower fricatives;
- The alveolar (s) is significantly different ($p<0.01$) from /ʃ/ and **not** significantly different from /s̪, s̻, ʂ/;
- The postalveolar (ʃ) is significantly different ($p<0.01$) from /s/ and **not** significantly different from /s̪, s̻, ʂ/.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h), the Bonferroni post hoc tests indicated that none of the wider fricatives are significantly different, (hypothesis 2b).

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 2):

- /f/ and /θ/ are significantly different ($p<0.01$) from /ʃ/, however, are not significantly different from the other English target fricatives;
- /s/ is significantly different ($p<0.01$) from /ʃ, h/, however, is not significantly different from the other English target fricatives;
- /ʃ/ is significantly different ($p<0.01$) from /f, θ, s/, however, is not significantly different from /h/;
- /h/ is significantly different ($p<0.01$) from /s/, however, is not significantly different from the other English target fricatives;

The English target fricatives are not significantly different from any of the non-English fricative errors.

Among all fricatives, kurtosis was highest for /x/ (10.91) and lowest for /s̪/ (1.79).

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it is possible to observe:

- /s/ is significant different ($p<0.01$) from the LibriSpeech /s/;
- /f, θ, ʃ, h/ are not significantly different from the LibriSpeech /f, θ, ʃ, h/.

First window

In the ANOVA analysis the first window has a significant main effect of phone identity [(1,3112 F= 3.38), $p<0.01$, $\eta^2 = 0.01$]. The results of the Bonferroni post hoc tests revealed that most of the narrower fricatives (s̪, s̺, s, ʃ, ʂ) are not significantly different from the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h), with the exception of two pairs, (hypothesis 2):

- the alveolar (s) is significantly different ($p<0.01$) from /ʃ, h/.

No other contrasts between fricatives were found.

Second window

In the ANOVA analysis, the second window [(1,3112 F= 10.68), $p<0.01$, $\eta2 = 0.04$] has a significant main effect of phone identity. Bonferroni post hoc tests indicated that there were significant contrasts between some phones with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- the dental (s̪), the laminal (s̺) and the retroflex (ʂ) are **not** significantly different from any of the wider fricatives;
- the alveolar (s) is significantly different ($p<0.01$) from /f, θ, h/ and **not** significantly different from /ɸ, ç, x, χ, ħ/;

- the postalveolar (ʃ) is marginally different (p<0.05) from /h/ and **not** significantly different from /ɸ, f, θ, ç, x, χ, ħ/.

Within the narrower fricatives (s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪), the laminal (s̺) and the retroflex (ʂ) are **not** significantly different from any of the narrower fricatives.
- The alveolar (s) is significantly different (p<0.01) from /s̺, ʂ/, marginally different (p<0.05) from /s̺/ and **not** significantly different from /s̪, ʃ/;
- The postalveolar (ʃ) is significantly different (p<0.01) from /s/ and **not** significantly different from /s̪, s̺, ʂ/;

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated that none of the wider fricatives are significantly different from each other (hypothesis 2b).

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 2):

- /s/ is significantly different (p<0.01) all the English target fricatives;
- /f, θ/ are significantly different (p<0.01) from /s/ but **not** significantly different from /ʃ, h/;
- /ʃ/ is significantly different (p<0.01) from /s/ and marginally different (p<0.05) from /h/, but **not** significantly different from /f, θ/;
- /h/ is significantly different (p<0.01) from /s/ and marginally different (p<0.05) from /ʃ/ but **not** significantly different from /f, θ/.

The English target fricatives are not significantly different from any of the non-English fricative errors, with the exception of /s/ which is significantly different (p<0.01) from /s̺/.

Third window

In the ANOVA analysis, the third window has a significant main effect of phone identity [(1,3112 F= 10.68), p< 0.01, η2 = 0.04]. Bonferroni post hoc tests indicated that there are significant contrasts between some phones with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The dental-alveolar (s̪) is marginally different (p<0.05) from /x, h/ and **not** significantly different from /ɸ, f, θ, ç, χ, ħ/;
- The laminal (s̺) is significantly different (p<0.01) from /h/, marginally different (p<0.05) from /x/ and **not** significantly different from /ɸ, f, θ, ç, χ, ħ/;
- The alveolar (s) and the retroflex (ʂ) are significantly different (p<0.01) from /h/ and **not** significantly different from the rest of the wider fricatives;
- The postalveolar (ʃ) is significantly different (p<0.01) from /f, x, h/ and **not** significantly different from /ɸ, θ, ç, χ, ħ/;

Within the narrower fricatives (s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2a):

- The dental-alveolar (s̪), the laminal (s̺) and the retroflex (ʂ) are **not** significantly different from any of the other narrower fricatives;
- The alveolar (s) is significantly different (p<0.01) from /ʃ/ and **not** significantly different from /s̪, s̺, ʂ/;
- The postalveolar (ʃ) is significantly different (p<0.01) from /s/ and **not** significantly different from /s̪, s̺, ʂ/.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 2b):

- The labiodental (f) and interdental (θ) are significantly different (p<0.01) from /h/ and **not** significantly different from the rest of the wider fricatives;
- The bilabial (ɸ), the palatal (ç), the velar (x), the uvular (χ) and the pharyngeal (ħ) are **not** significantly different from any of the wider fricatives.

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 2):

- /h/ is significantly different ($p<0.01$) from all English target fricatives;
- /f/ and /s/ are significantly different ($p<0.01$) from /ʃ, h/;
- /ʃ/ is significantly different ($p<0.01$) from /f, s, h/;
- /θ/ is significantly different ($p<0.01$) from /h/.

The English target fricatives (f, θ, s, ʃ, h) were not significantly different from any of the non-English fricative errors.

Fourth window

In the ANOVA analysis, the fourth window has a significant main effect of phone identity for only a few phones [$(1,3112\ F= 8.40)$, $p<0.01$, $\eta^2 = 0.03$]. Bonferroni post hoc tests indicated that there are significant contrasts between some phones with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and phones with a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 2):

- The dental (s̪), the laminal (s̺) and the retroflex (ʂ) are not significantly different from any of the wider fricatives;
- The alveolar (s) is significantly different from ($p<0.01$) from /h/ but not from the other wider fricatives;
- The postalveolar (ʃ) is significantly different from ($p<0.01$) from /f, θ/.

Within the narrower fricatives (s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated that only the alveolar and the postalveolar are significantly different ($p<0.01$), (hypothesis 2a).

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated that the only contrast found was between the labiodental (f) and the interdental (θ), which are significantly different ($p<0.01$) from /h/, (hypothesis 2b).

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 2):

- /f, θ, s/ are significantly different (p<0.01) from /ʃ, h/ but not significantly different from the other English target fricatives;
- /ʃ/ is significantly different (p<0.01) from /f, θ, s/ but not significantly different from /h/;
- /h/ is significantly different (p<0.01) from /f, θ, s/ but not significantly different from /ʃ/;

The English target fricatives (f, θ, s, ʃ, h) were not significantly different from any of the non-English fricative errors.

6.1.1.1.5. Peak location

Below, there is a graphic with a visual representation of the statistical differences in all fricatives and in all windows and average of kurtosis. These differences are analyzed further in more detail.

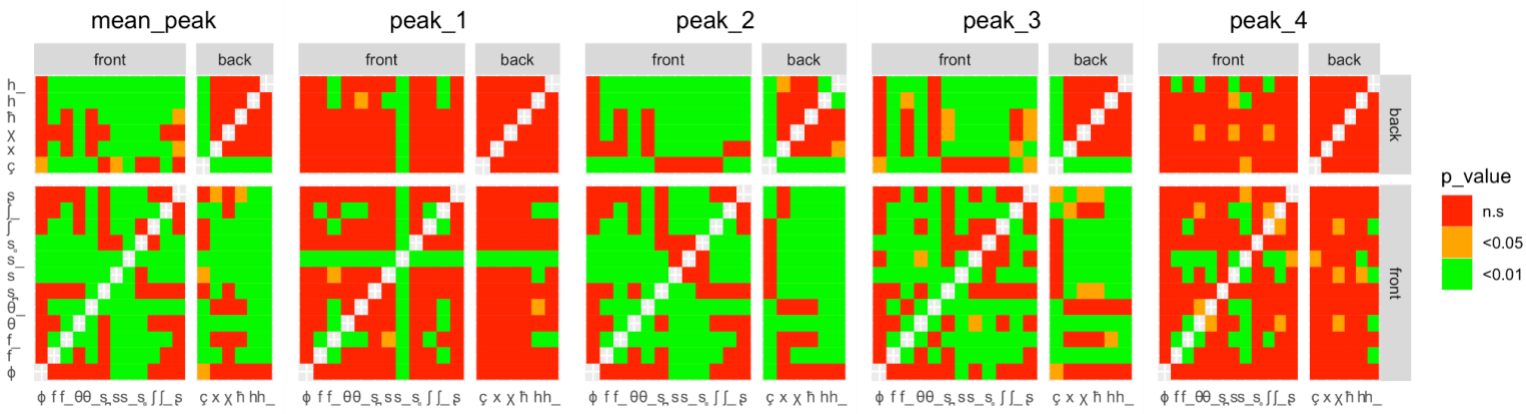

Figure 21 - Visualization of the statistical differences between fricatives in peak location

Average

In the ANOVA analysis, the values of the four windows of the centroid were averaged and there was a significant main effect of phone identity [(1,8205 F= 301.20), p< 0.01, η2 = 0.37]. The results of the Bonferroni post hoc tests revealed the following significant differences

between some front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) and some back fricative (ç, x, χ, ħ, h), (hypothesis 1):

- The bilabial (ɸ) is marginally different ($p<0.05$) from the back palatal fricative (ç), but **not** significantly different from the other back fricatives;
- The labiodental (f) is significantly different ($p<0.01$) from the back fricatives /ç, x, ħ, h/ and **not** significantly different from /χ/;
- The interdental (θ) is significantly different ($p<0.01$) from all the back fricatives;
- The dental-alveolar (s̪) and the laminal (s̺) are significantly different ($p<0.01$) from the back fricatives (ħ, h) and **not** significantly different from /ç, x, χ/;
- The alveolar (s) is significantly different ($p<0.01$) from the back fricatives (x, χ, ħ, h) and **not** significantly different from /ç/;
- The postalveolar (ʃ) is significantly different ($p<0.01$) from the back fricatives (x, χ, ħ, h) and marginally different ($p<0.05$) from /ç/;
- The retroflex (ʂ) is significantly different ($p<0.01$) from /h/, marginally different ($p<0.05$) from the back fricatives (x, ħ) and **not** significantly different from /ç, χ/;

Within the front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1a):

- The bilabial (ɸ) is significantly different ($p<0.01$) from /s̺, s/ and **not** significantly different /f, θ, s̪, ʃ, ʂ/;
- The labiodental (f) is significantly different ($p<0.01$) from /s̺, s, ʃ/ and **not** significantly different from /ɸ, θ, s̪, ʂ/;
- The interdental (θ) is significantly different ($p<0.01$) from /s/ and **not** significantly different from /ɸ, f, s̪, s̺, ʃ, ʂ/;
- The dental-alveolar (s̪) is significantly different ($p<0.01$) from /s/ and **not** significantly different from /ɸ, f, θ, s̺, ʃ, ʂ/;

- The laminal (s̪) is significantly different (p<0.01) from /ɸ, f/ and **not** significantly different from /θ, s̪, s, ʃ, ʂ/;
- The alveolar (s) is significantly different (p<0.01) from /ɸ, f, θ, s̪, ʃ, ʂ/ and **not** significantly different from /s̪/;
- The postalveolar (ʃ) is significantly different (p<0.01) from /f, s/ and **not** significantly different from /ɸ, θ, s̪, s̪, ʂ/;
- The retroflex (ʂ) is significantly different (p<0.01) from /s/ and **not** significantly different from /ɸ, f, θ, s̪, s̪, ʃ/.

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1b):

- The palatal (ç) is significantly different (p<0.01) from all the back fricatives /x, χ, ħ, h/ and no other significant differences were found.

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings, (hypothesis 1):

- /s, h/ are significantly different from all English target fricatives (p<0.01);
- /f/ and /θ/ are not significantly different from each other;
- /θ/ and /ʃ/ are not significantly different from each other;

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[42] and not significantly different from the rest:

- /f/ was significantly different (p<0.01) from /s̪/ and **not** significantly different from /ɸ, s̪, ʂ/;
- /θ/ was significantly different (p<0.01) from /ç/ and **not** significantly different from /s̪, ʂ/;

[42] These errors will be discussed further in section 6.2. Language acquisition perspective.

- /s/ was significantly different ($p<0.01$) from /s̪, ʂ/, marginally different ($p<0.05$) from /ç/ and **not** significantly different from /s̺/;
- /ʃ/ was not significantly different from any of the non-English fricative errors;
- /h/ was significantly different ($p<0.01$) from /ç/ and **not** significantly different from /x, χ, ħ/.

When looking at pairs of fricatives that have adjacent places of articulation (hypothesis 1c), it is possible to observe that:

- the bilabial (ɸ) and the labiodental (f) are **not** significantly different;
- the labiodental (f) and the interdental (θ) are **not** significantly different;
- the interdental (θ) and the dental-alveolar (s̪) are **not** significantly different;
- the dental-alveolar (s̪) and the laminal (s̺) are **not** significantly different;
- the laminal (s̺) and the alveolar (s) are **not** significantly different;
- the alveolar (s) and postalveolar (ʃ) are significantly different ($p<0.01$);
- the retroflex (ʂ) and the palatal (ç) are **not** significantly different;
- the palatal (ç) and the velar (x) are significantly different from each other ($p<0.01$);
- the velar (x) and the uvular (χ) are **not** significantly different;
- the uvular (χ) and the pharyngeal (ħ) are **not** significantly different;
- the pharyngeal (ħ) and the glottal are **not** significantly different;

Among all fricatives, peak location was highest for /s/ (5732Hz) and lowest for /h/ (1336Hz).

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it is possible to observe:

- /f, θ, s, ʃ/ are significant different ($p<0.01$) from the LibriSpeech respective /f, θ, s, ʃ/;
- /h/ is not significantly different from the LibriSpeech /h/.

First window

In the first window of peak location there was a marginal main effect of phone identity [(1,3112 F= 1.96], p< 0.024, η^2 = 0.0073]. The results of the Bonferroni post hoc tests revealed that the front fricative /s/ is significantly different (p<0.01) from the back fricative /h/, (hypothesis 1). No other significant differences were found in this window.

Second window

In the ANOVA analysis, the second window has a significant main effect of phone identity [(1,3112 F= 116.50), p< 0.01, η2 = 0.30]. The results of the Bonferroni post hoc tests revealed the following significant differences between some front fricatives (ɸ, f, θ, s̪, s̪̊, s, ʃ, ʂ) and some back fricative (ç, x, χ, ħ, h), (hypothesis 1):

- The bilabial (ɸ) is marginally different (p<0.05) from the back palatal fricative (ç), but not from the other back fricatives;
- The labiodental (f) is significantly different (p<0.01) from the back fricatives /ç, χ, ħ, h/ and **not** significantly different from /x/;
- The interdental (θ) is significantly different (p<0.01) from all the back fricatives /ç, x, χ, ħ, h/;
- The dental-alveolar (s̪), the laminal (s̪̊), the alveolar (s) and the postalveolar (ʃ) are significantly different (p<0.01) from the back fricatives /x, χ, ħ, h/ and **not** significantly different from /ç/;
- The retroflex (ʂ) is significantly different (p<0.01) from /ħ, h/, marginally different (p<0.05) from the back fricatives /ç, χ/ and **not** significantly different from /x/.

Within the front fricatives (ɸ, f, θ, s̪, s̪̊, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1a):

- The bilabial (ɸ) is significantly different (p<0.01) from /s̪̊, s/ and **not** significantly different from /f, θ, s̪, ʃ, ʂ/;
- The labiodental (f) is significantly different (p<0.01) from /s̪̊, s, ʃ/ and **not** significantly different from /ɸ, θ, s̪, ʂ/;

- The interdental (θ) is significantly different (p<0.01) from /s̪, s̪̊, s/ and **not** significantly different from /ɸ, f, ʃ, ʂ/;
- The dental-alveolar (s̪) is significantly different (p<0.01) from /θ, s, ʃ, ʂ/ and **not** significantly different from /ɸ, f, s̪̊/;
- The laminal (s̪̊) is significantly different (p<0.01) from /ɸ, f, θ, ʃ, ʂ/ and **not** significantly different from /s̪, s/;
- The alveolar (s) is significantly different (p<0.01) from /ɸ, f, θ, s̪, ʃ, ʂ/ and **not** significantly different from /s̪̊/;
- The postalveolar (ʃ) is significantly different (p<0.01) from /f, s̪, s̪̊, s/ and **not** significantly different from /ɸ, θ, ʂ/;
- The retroflex (ʂ) is significantly different (p<0.01) from /s̪, s̪̊, s/ and **not** significantly different from /ɸ, f, θ, ʃ/.

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1b):

- The palatal (ç) is significantly different (p<0.01) from all the back fricatives /x, χ, ħ, h/ and no other significant differences were found.

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings, (hypothesis 1):

- /s, h/ are significantly different from all English target fricatives (p<0.01);
- /f/ and /θ/ are not significantly different from each other;
- /θ/ and /ʃ/ are not significantly different from each other.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[43] and not significantly different from the rest:

[43] Ibid.

- /f/ was significantly different (p<0.01) from /s̪/ and **not** significantly different from /ɸ, s̪, ʂ/;
- /θ/ was significantly different (p<0.01) from /s̪, ç/ and **not** significantly different from /ʂ/;
- /s/ was significantly different (p<0.01) from /s̪, ʂ/ and **not** significantly different from /s̪, ç/;
- /ʃ/ was significantly different (p<0.01) from /s̪/ and **not** significantly different from /ʂ, ç/;
- /h/ was significantly different (p<0.01) from /ç/ and **not** significantly different from /x, χ, ħ/.

When looking at pairs of fricatives that have similar places of articulation (hypothesis 1c), it is possible to observe that:

- the bilabial (ɸ) and the labiodental (f) are **not** significantly different;
- the labiodental (f) and the interdental (θ) are **not** significantly different;
- the interdental (θ) and the dental-alveolar (s̪) are significantly different (p<0.01);
- the dental-alveolar (s̪) and the laminal (s̪) are **not** significantly different;
- the laminal (s̪) and the alveolar (s) are **not** significantly different;
- the alveolar (s) and postalveolar (ʃ) are significantly different (p<0.01);
- the retroflex (ʂ) and the palatal (ç) are significantly different (p<0.05);
- the palatal (ç) and the velar (x) are significantly different from each other (p<0.01);
- the velar (x) and the uvular (χ) are **not** significantly different;
- the uvular (χ) and the pharyngeal (ħ) are **not** significantly different;
- the pharyngeal (ħ) and the glottal are **not** significantly different;

Third window

In the ANOVA analysis, the third window has a significant main effect of phone identity [(1,3112 F= 118.85), p< 0.01, η2 = 0.31]. The results of the Bonferroni post hoc tests revealed

the following significant differences between some front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) and some back fricative (ç, x, χ, ħ, h), (hypothesis 1):

- The bilabial (ɸ) is significantly different ($p<0.05$) from the back palatal fricative (ç), but **not** from the other back fricatives;
- The labiodental (f) is significantly different ($p<0.01$) from the back fricatives /ç, x, h/, marginally different ($p<0.05$) from /χ, ħ/;
- The interdental (θ) is significantly different ($p<0.01$) from the back fricatives /x, χ, ħ, h/ and marginally different ($p<0.05$) from /ç/.
- The dental-alveolar (s̪) is significantly different ($p<0.01$) from the back fricatives/x, χ, h/ and marginally different ($p<0.05$) from the back fricative /ħ/ and **not** significantly different from /ç/;
- The laminal (s̺), the alveolar (s) and the postalveolar (ʃ) are significantly different ($p<0.01$) from the back fricatives /x, χ, ħ, h/ and **not** significantly different from /ç/;
- The retroflex (ʂ) is significantly different ($p<0.01$) from the back fricatives /x, h/ and **not** significantly different from /ç, χ, ħ/;

Within the front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1a):

- The bilabial (ɸ) is significantly different ($p<0.01$) from /s/, marginally different ($p<0.05$) from /s̺/ and **not** significantly different from /f, θ, s̪, ʃ, ʂ/;
- The labiodental (f) is significantly different ($p<0.01$) from /s̺, s, ʃ/ and **not** significantly different from /ɸ, θ, s̪, ʂ/;
- The interdental (θ) is significantly different ($p<0.01$) from /s̪, s̺, s/ and **not** significantly different from /ɸ, f, ʃ, ʂ/;
- The dental-alveolar (s̪) is significantly different ($p<0.01$) from /θ, ʂ/ and **not** significantly different from /ɸ, f, s̺, s, ʃ/;

- The laminal (s̪) is significantly different (p<0.01) from /f/, marginally different (p<0.05) from /ɸ/ and **not** significantly different from /θ, s̪, s, ʃ, ʂ/;
- The alveolar (s) is significantly different (p<0.01) from /ɸ, f, θ, s̪, ʃ, ʂ/ and **not** significantly different from /s̪/;
- The postalveolar (ʃ) is significantly different (p<0.01) from /f, s/ and **not** significantly different from /ɸ, θ, s̪, s̪, ʂ/;
- The retroflex (ʂ) is significantly different (p<0.01) from /s̪, s̪, s/ and **not** significantly different from /ɸ, f, θ, ʃ/;

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts, (hypothesis 1b):

- The palatal (ç) is significantly different (p<0.01) from all the back fricatives /x, χ, ħ, h/ and no other significant differences were found.

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings, (hypothesis 1):

- /s, h/ are significantly different from all English target fricatives (p<0.01);
- /f/ and /θ/ are not significantly different from each other;
- /θ/ and /ʃ/ are not significantly different from each other;

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[44] and not significantly different from the rest:

- /f/ was significantly different (p<0.01) from /s̪/ and **not** significantly different from /ɸ, s̪, ʂ/;
- /θ/ was significantly different (p<0.01) from /s̪/, marginally different (p<0.05) from /ç/ and **not** significantly different from /ʂ/;

[44] These errors will be discussed further in section 6.2. Language acquisition perspective.

- /s/ was significantly different ($p<0.01$) from /s̪, ʂ/ and **not** significantly different from /s̺, ç/;
- /ʃ/ was not significantly different from any of the non-English fricative errors;
- /h/ was significantly different ($p<0.01$) from /ç/ and **not** significantly different from /x, χ, ħ/.

When looking at pairs of fricatives that have similar places of articulation (hypothesis 1c), it is possible to observe that:

- the bilabial (ɸ) and the labiodental (f) are **not** significantly different;
- the labiodental (f) and the interdental (θ) are **not** significantly different;
- the interdental (θ) and the dental-alveolar (s̪) are significantly different ($p<0.01$);
- the dental-alveolar (s̪) and the laminal (s̺) are **not** significantly different;
- the laminal (s̺) and the alveolar (s) are **not** significantly different;
- the alveolar (s) and postalveolar (ʃ) are significantly different ($p<0.01$);
- the retroflex (ʂ) and the palatal (ç) are **not** significantly different;
- the palatal (ç) and the velar (x) are significantly different from each other ($p<0.01$);
- the velar (x) and the uvular (χ) are **not** significantly different;
- the uvular (χ) and the pharyngeal (ħ) are **not** significantly different;
- the pharyngeal (ħ) and the glottal are **not** significantly different;

Fourth window

In the ANOVA analysis, the fourth window has a marginal main effect of phone identity [$(1,3112\ F= 12.74]$, $p< 0.01$, $\eta^2 = 0.05$]. The results of the Bonferroni post hoc tests revealed the following significant differences between some front and some back fricatives, (hypothesis 1):

- the labiodental (f), the interdental (θ), the alveolar (s) and the postalveolar are significantly different ($p<0.01$) from the back fricatives /χ, h/, however, are **not** significantly different from the rest of the back fricatives (ç, x, ħ);

- the retroflex (ʂ) is significantly different (p<0.01) from the back fricative /χ/, marginally different (p<0.05) from /h/ and not significantly different from /ç, x, ħ/;
- the bilabial (ɸ), the dental (s̪) and the laminal (s̻) are significantly different (p<0.01) from the back fricative /χ/, but not significantly different from the rest of the back fricatives.

Within the front fricatives (ɸ, f, θ, s̪, s̻, s, ʃ, ʂ) the Bonferroni post hoc tests indicated that these fricatives are not significantly different, except for the labiodental and the interdental, which are marginally different (p<0.05), (hypothesis 1a).

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts between, (hypothesis 1b):

- the palatal (ç) is significantly different (p<0.01) from the back fricative /χ/, marginally different (p<0.05) from /h/ and **not** significantly different from the rest of the back fricatives;
- the velar (x) is significantly different (p<0.01) from the back fricative /χ/ and **not** significantly different from the rest of the back fricatives;
- the uvular (χ) is significantly different (p<0.01) from /ç, x, h/ and marginally different (p<0.05) from /ħ/;
- the pharyngeal (ħ) is marginally different (p<0.05) from the back fricative /χ/ and **not** significantly different from the rest of the back fricatives;
- the glottal (h) is significantly different (p<0.01) from /χ/, marginally different (p<0.05) from /ç/ and **not** significantly different from the rest of the back fricatives;

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings, (hypothesis 1):

- /h/ is significantly different (p<0.01) from all English target fricatives;
- /f/ and /θ/ are marginally different (p<0.05) from each other but not significantly different from the /s, ʃ/;

- /s, ʃ/ are only significantly different ($p<0.01$) from /h/.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[45] and not significantly different from the rest:

- /f, θ, s, ʃ/ were not significantly different from any of the non-English fricative errors;
- /h/ was significantly different ($p<0.01$) from /χ/, marginally different ($p<0.05$) from /ç/ and **not** significantly different from /x, ħ/.

When looking at pairs of fricatives that have adjacent places of articulation (hypothesis 1c), it is possible to observe that most of these pairs are not significantly different, with the exception of two pairs:

- the labiodental (f) and interdental (θ) are marginally different ($p<0.05$) from each other;
- the velar (x) and the uvular (χ) are significantly different ($p<0.01$) from each other;

Below, there is a table with all the values for the spectral properties.

Major places of articulation	Fricatives	M1	M2	M3	M4	Peak location
Labial	ɸ	2245Hz	1.13	1.310	4.72	2253Hz
	f	3310Hz	1.32	0.637	4.76	2890Hz
Coronal	θ	2976Hz	1.27	0.852	6.04	2559Hz
	s̪	4040Hz	1.35	-0.222	1.79	3763Hz
	s̺	4950Hz	1.12	-0.371	2.14	4868Hz
	s	5619Hz	1.17	-1.560	8.36	5732Hz

[45] These errors will be discussed further in section 6.2. Language acquisition perspective.

Major places of articulation	Fricatives	M1	M2	M3	M4	Peak location
	ʃ	4086Hz	1.11	0.394	2.40	3795Hz
	ʂ	3473Hz	1.10	0.931	3.45	3232Hz
	ç	4803Hz	1.10	-0.415	4.61	4728Hz
Dorsal	x	2003Hz	1.03	1.700	10.91	1769Hz
	χ	2345Hz	1.15	1.243	4.45	1950Hz
pharyngeal	ħ	2192Hz	1.15	1.570	7.50	1644Hz
	h	1732Hz	1.03	1.950	9.33	1336Hz

Table 12 - Spectral properties values

6.1.1.2. Transition information

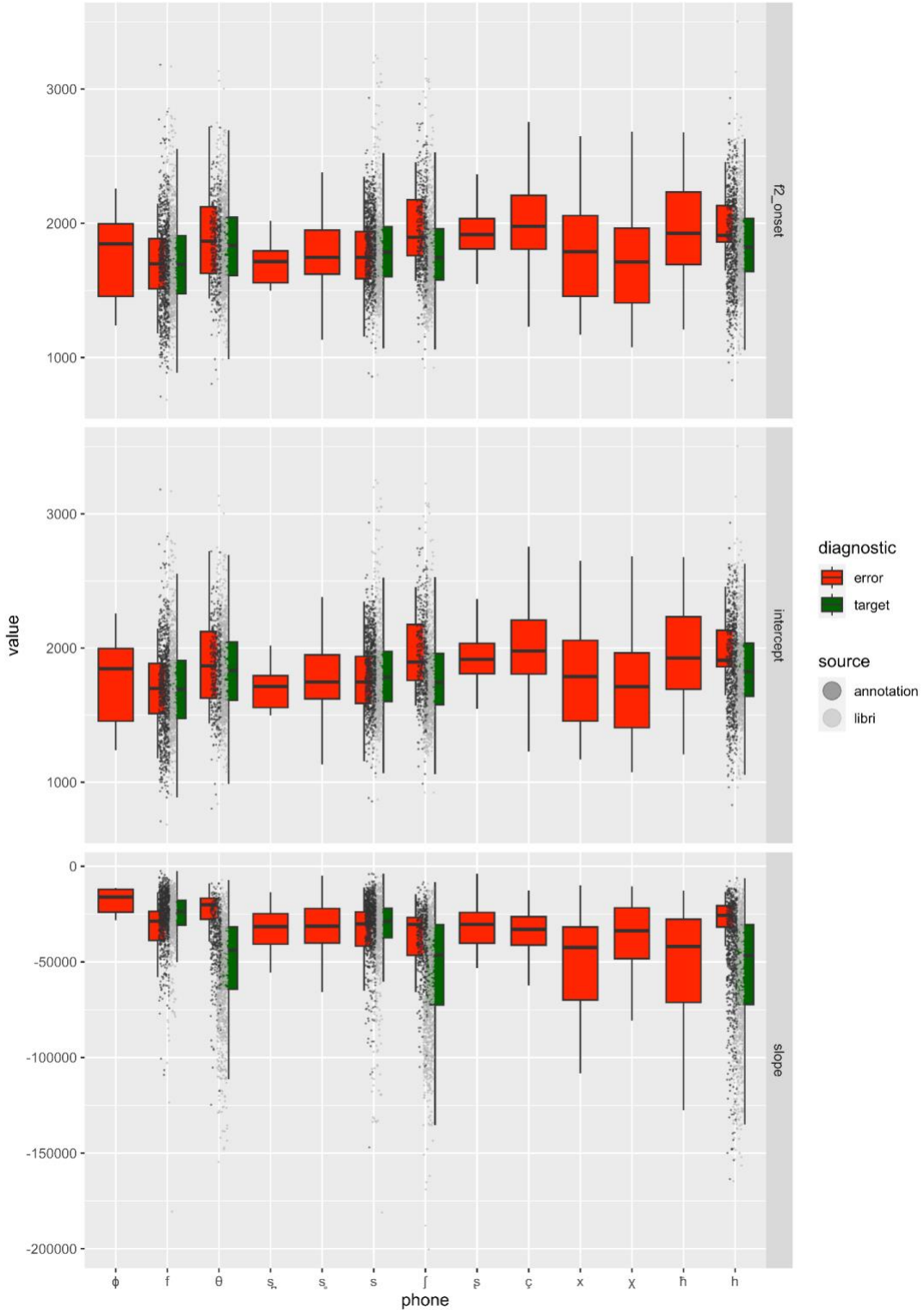

Figure 22 - Transition information measurements

For the measurements F2 onset frequency, intercept and slope, given the chosen vowel /æ/ does not exist in Vietnamese, only good productions of the target vowel will be analyzed.

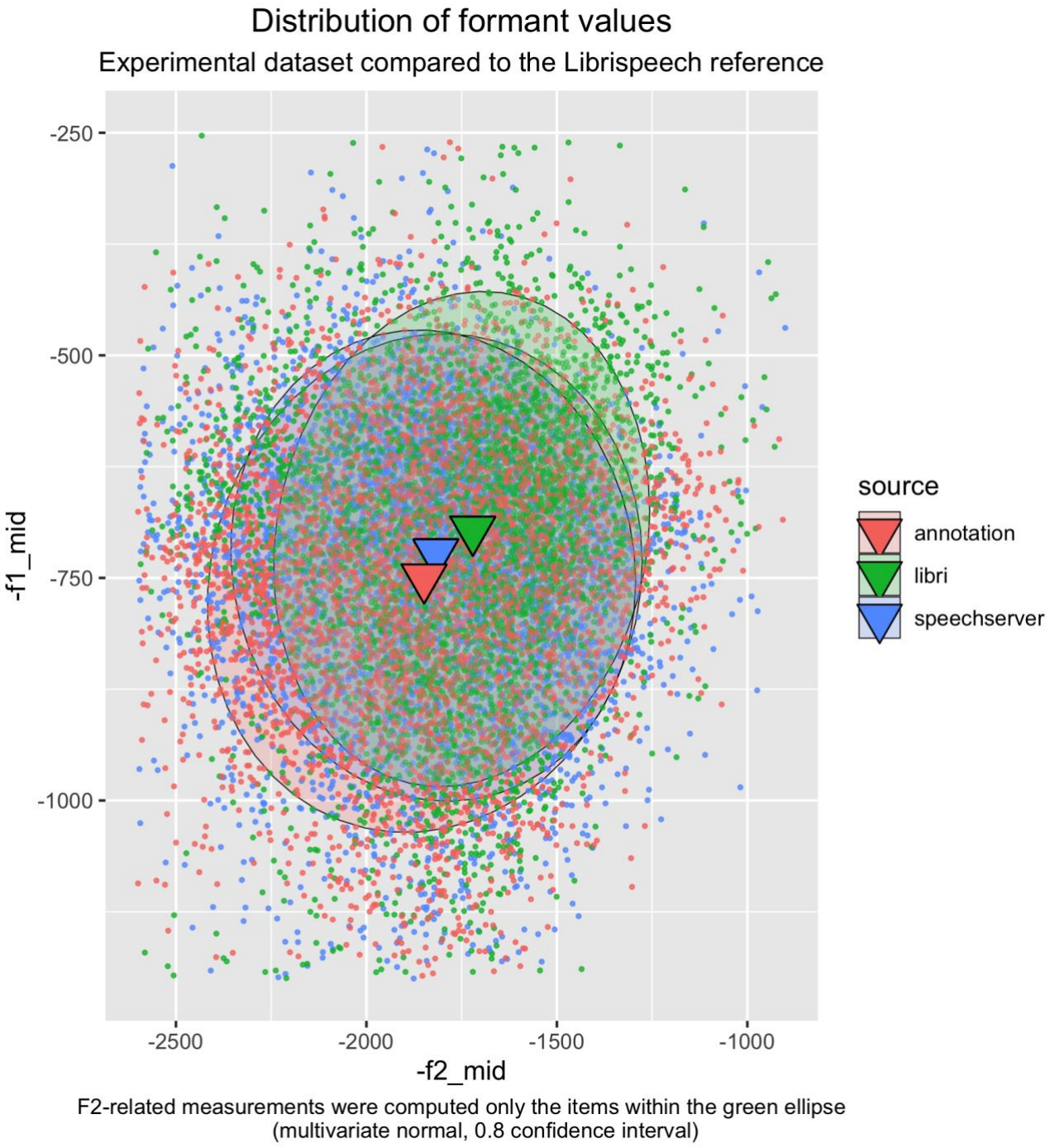

Figure 23 - Distribution of the formant values compared to the LibriSpeech reference database

6.1.1.2.1. F2 onset frequency and intercept

F2 onset frequency and intercept measurements have the same values, therefore will be analyzed simultaneously. Below, there is a graphic with a visual representation of the statistical differences in all fricatives of the F2 onset frequency/intercept and slope measurements. These differences are analyzed further in more detail.

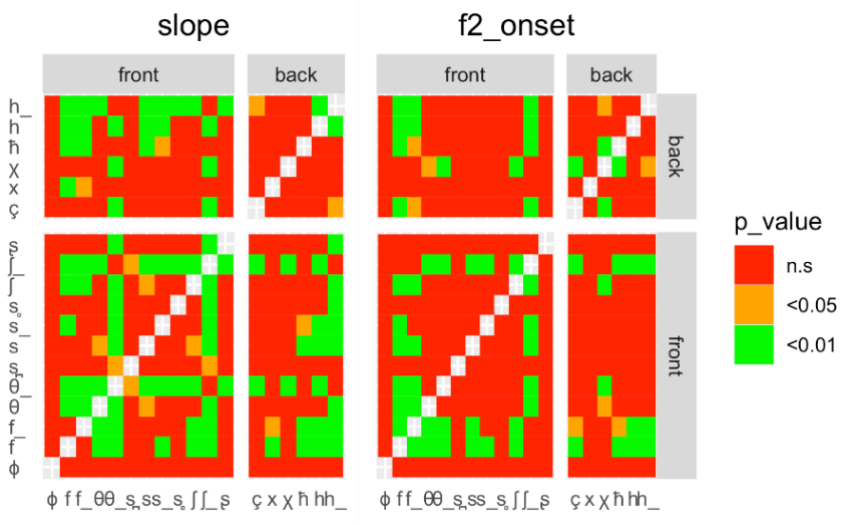

Figure 24 - Visualization of the statistical differences between fricatives in F2 onset frequency/intercept and slope

In the ANOVA analysis, the values of the F2 onset frequency and intercept showed a significant main effect of phone identity [(1,6097 F= 14.43), $p< 0.01$, $\eta^2 = 0.029$]. The results of the Bonferroni post hoc tests revealed the following significant differences between some of the front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) and some of the back fricative (ç, x, χ, ħ, h), (hypothesis 4):

- The bilabial (ɸ), the laminal (s̺) and the postalveolar (ʃ) are **not** significantly different from any of the back fricatives;
- The labiodental (f) is significantly different (p<0.01) from the back fricatives /ç, h/, marginally different (p<0.05) from /ħ/ and **not** significantly different from /x, χ/;
- The interdental (θ) is significantly different (p<0.01) from /ç/ and **not** significantly different from /x, χ, ħ, h/;

- The dental-alveolar (s̪) and the alveolar (s) are marginally different ($p<0.05$) from the back fricatives /ç/ and **not** significantly different from /x, χ, ħ, h/;
- The retroflex (ʂ) is marginally different ($p<0.05$) from /χ/ and **not** significantly different from /ç, x, ħ, h/;

Within the front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts (hypothesis 4a):

- The bilabial (ɸ), the dental (s̪) and the laminal (s̺) are **not** significantly different any of the other front fricatives;
- The labiodental (f) is significantly different ($p<0.01$) from /s, ʃ, ʂ/, marginally different ($p<0.05$) from /θ/ and **not** significantly different from /ɸ, s̪, s̺/;
- The interdental (θ) is significantly different ($p<0.01$) from /ʃ, ʂ/, marginally different ($p<0.05$) from /f/ and **not** significantly different from /ɸ, s̪, s̺, s/;
- The alveolar (s) is significantly different ($p<0.01$) from /f, ʃ, ʂ/ and **not** significantly different from /ɸ, θ, s̪, s̺/;
- The postalveolar (ʃ) is significantly different ($p<0.01$) from /f, θ, s/ and **not** significantly different from /ɸ, s̪, s̺, ʂ/;
- The retroflex (ʂ) is significantly different ($p<0.01$) from /f, θ, s/ and **not** significantly different from /ɸ, s̪, s̺, ʃ/.

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts between (hypothesis 4b):

- The palatal (ç) is marginally different ($p<0.05$) from the uvular (χ) and **not** significantly different from /x, ħ, h/;
- The velar (x), the pharyngeal (ħ) and the glottal (h) are **not** significantly different from any of the back fricatives.

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 4):

- /f/ is significantly different ($p<0.01$) from /s, ʃ, h/ and marginally different ($p<0.05$) from /θ/;
- /θ/ is significantly different ($p<0.01$) from /ʃ/, marginally different ($p<0.05$) from /f/ and not significantly different from /s, h/;
- /s/ is significantly different ($p<0.01$) from /f, ʃ/ and not significantly different from /θ, h/;
- /ʃ/ is significantly different ($p<0.01$) from /f, θ, s/ and not significantly different from /h/;
- /h/ is significantly different ($p<0.01$) from /f/ and not significantly different from any of the other English target fricatives.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[46] and not significantly different from the rest:

- /f/ was significantly different ($p<0.01$) from /ʂ/ and **not** significantly different from /ɸ, s̪, s̺/;
- /θ/ was significantly different ($p<0.01$) from /ʂ, ç/ and **not** significantly different from /s̪/;
- /s/ was significantly different ($p<0.01$) from /ʂ/, marginally different ($p<0.05$) from /ç/ and **not** significantly different from /s̪, s̺/;
- /ʃ/ and /h/ were **not** significantly different from any of the non-English fricative errors.

When looking at pairs of fricatives that have adjacent places of articulation (hypothesis 4c), it is possible to observe that:

- the bilabial (ɸ) and the labiodental (f) are **not** significantly different;
- the labiodental (f) and the interdental (θ) are marginally different ($p<0.05$);
- the interdental (θ) and the dental-alveolar (s̪) are **not** significantly different;
- the dental-alveolar (s̪) and the laminal (s̺) are **not** significantly different;
- the laminal (s̺) and the alveolar (s) are **not** significantly different;

[46] These errors will be discussed further in section 6.2. Language acquisition perspective.

- the alveolar (s) and postalveolar (ʃ) are significantly different ($p<0.01$);
- the retroflex (ʂ) and the palatal (ç) are **not** significantly different;
- the palatal (ç) and the velar (x) are **not** significantly different;
- the velar (x) and the uvular (χ) are **not** significantly different;
- the uvular (χ) and the pharyngeal (ħ) are **not** significantly different;
- the pharyngeal (ħ) and the glottal are **not** significantly different;

Among all fricatives, F2 onset frequency/intercept was highest for /ʂ/ (1979Hz) and lowest for /ɸ/ (1640Hz). If only target fricatives are taken into consideration, then the highest was /ʃ/ (1842Hz) and the lowest was /f/ (1707Hz).

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it is possible to observe:

- /f, θ, s, h/ are not significant different from the LibriSpeech /f, θ, s, h/, respectively;
- /ʃ/ is significantly different ($p<0.01$) from the LibriSpeech /ʃ/.

6.1.1.2.2. Slope

In the ANOVA analysis, the values of the slope showed a significant main effect of phone identity [(1,6097 F= 118.71), $p< 0.01$, $\eta^2 = 0.25$]. The results of the Bonferroni post hoc tests revealed the following significant differences between some of the front fricatives (ɸ, f, θ, ʂ̪, ʂ̺, s, ʃ, ʂ) and some back fricative (ç, x, χ, ħ, h), (hypothesis 4):

- The bilabial (ɸ) is marginally different ($p<0.05$) from /h/ and **not** significantly different from /ç, x, χ, ħ/;
- The labiodental (f) is significantly different ($p<0.01$) from /x, ħ, h/ and **not** significantly different from /ç, χ/;
- The interdental (θ) and the postalveolar (ʃ) are significantly different ($p<0.01$) from /h/ and **not** significantly different from /ç, x, χ, ħ/;

- The dental-alveolar (s̪) and the laminal (s̻) are **not** significantly different from any of the back fricatives;
- The alveolar (s) is significantly different ($p<0.01$) from /x, ħ, h/ and **not** significantly different from /ç, χ/;
- The retroflex (ʂ) is marginally different ($p<0.05$) from /h/ and **not** significantly different from /ç, x, χ, ħ/;

Within the front fricatives (ɸ, f, θ, s̪, s̻, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts (hypothesis 4a):

- The bilabial (ɸ), the dental (s̪), the laminal (s̻) and the retroflex (ʂ) are **not** significantly different any of the other front fricatives;
- The labiodental (f) is significantly different ($p<0.01$) from /θ, ʃ/ and **not** significantly different from /ɸ, s̪, s̻, s, ʂ/;
- The interdental (θ) is significantly different ($p<0.01$) from /f, s/ and **not** significantly different from /ɸ, s̪, s̻, ʃ, ʂ/;
- The alveolar (s) is significantly different ($p<0.01$) from /θ, ʃ/ and **not** significantly different from /ɸ, s̪, s̻, s, ʂ/;
- The postalveolar (ʃ) is significantly different ($p<0.01$) from /f, s/ and **not** significantly different from /ɸ, θ, s̪, s̻, ʂ/;

Within the back fricatives (ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated that there are no significant contrasts (hypothesis 4b).

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 4):

- /f/ is significantly different ($p<0.01$) all English target fricatives, except /s/;
- /θ/ is significantly different ($p<0.01$) from all English target fricatives, except /ʃ/;
- /s/ is significantly different ($p<0.01$) from all English target fricatives, except /f/;

- /ʃ/ is significantly different ($p<0.01$) from all English target fricatives, except /θ/;
- /h/ is significantly different ($p<0.01$) from all English target fricatives.

None of the English target fricative (f, θ, s, ʃ, h) was significantly different from the non-English fricative errors that were found in the annotations.

When looking at pairs of fricatives that have adjacent places of articulation (hypothesis 4c), it is possible to observe that:

- the bilabial (ɸ) and the labiodental (f) are **not** significantly different;
- the labiodental (f) and the interdental (θ) are significantly different ($p<0.01$);
- the interdental (θ) and the dental-alveolar (s̪) are **not** significantly different;
- the dental-alveolar (s̪) and the laminal (s̻) are **not** significantly different;
- the laminal (s̻) and the alveolar (s) are **not** significantly different;
- the alveolar (s) and postalveolar (ʃ) are significantly different ($p<0.01$);
- the retroflex (ʂ) and the palatal (ç) are **not** significantly different;
- the palatal (ç) and the velar (x) are **not** significantly different;
- the velar (x) and the uvular (χ) are **not** significantly different;
- the uvular (χ) and the pharyngeal (ħ) are **not** significantly different;
- the pharyngeal (ħ) and the glottal are **not** significantly different;

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it is possible to observe:

- /θ, ʃ, h/ are significant different ($p<0.01$) from the LibriSpeech /θ, s, ʃ/, respectively;
- /f, s/ are not significantly different from the LibriSpeech /f, s/.

Below, there is a table with all the values for the transition properties from this dataset:

Major places of articulation	Fricatives	F2 onset	Intercept	Slope

Major places of articulation	Fricatives	F2 onset	Intercept	Slope
Labial	ɸ	1640	1640	-18490
	f	1707	1707	-25767
Coronal	θ	1728	1728	-38938
	s̪	1713	1713	-33471
	s̺	1773	1773	-37531
	s	1769	1769	-29744
	ʃ	1842	1842	-33687
	ʂ	1979	1979	-33467
	ç	1916	1916	-34158
Dorsal	x	1836	1836	-47384
	χ	1813	1813	-34903
Pharyngeal	ħ	1823	1823	-52434
	h	1808	1808	-44138

Table 13 - Transition information values

6.1.1.3. Amplitude

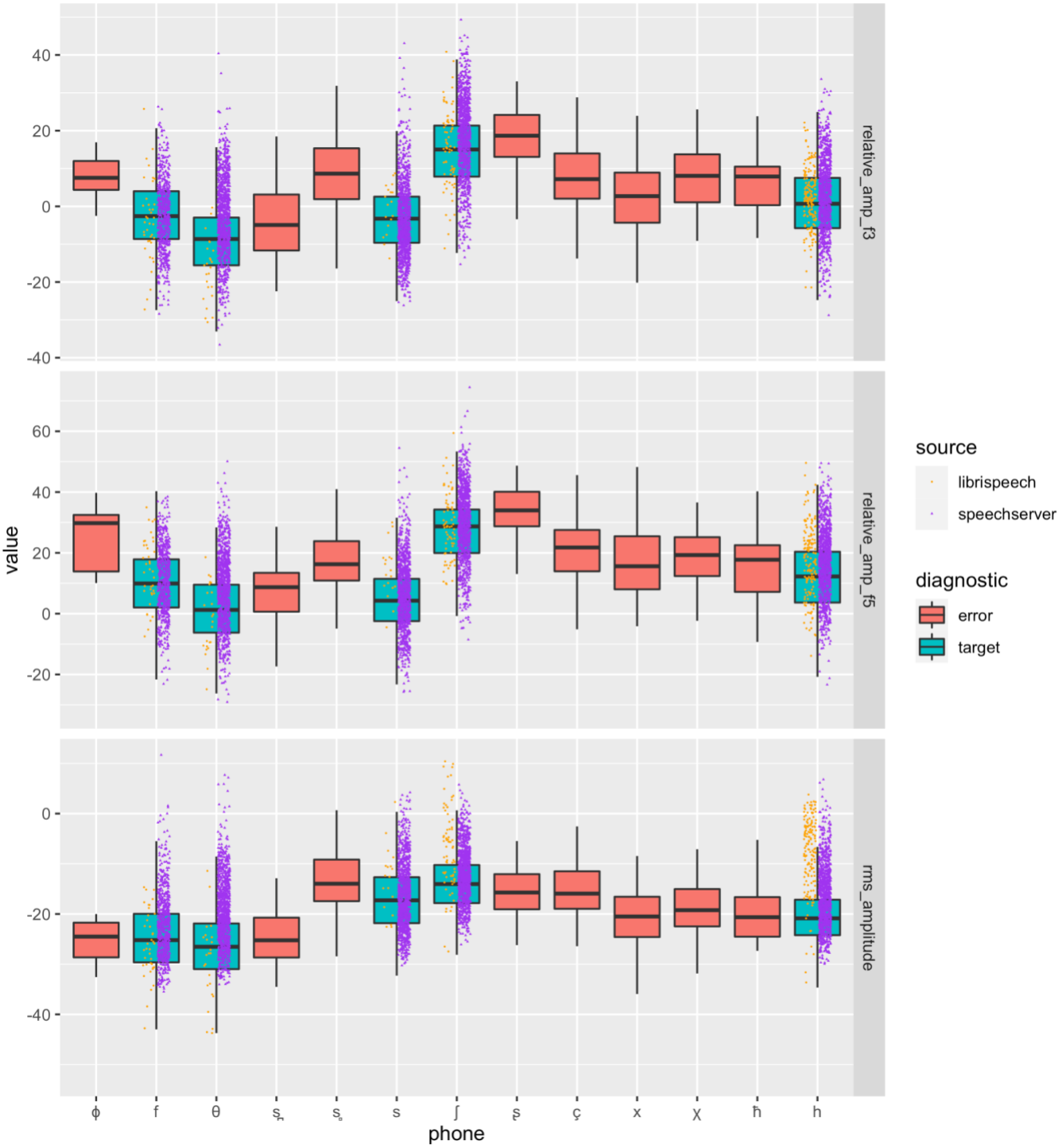

Figure 25 - Amplitude measurements

6.1.1.3.1. Normalized amplitude

Below, there is a graphic with a visual representation of the statistical differences in all fricatives of the normalized amplitude and relative amplitude at F3 and F5, measurements. These differences are analyzed further in more detail.

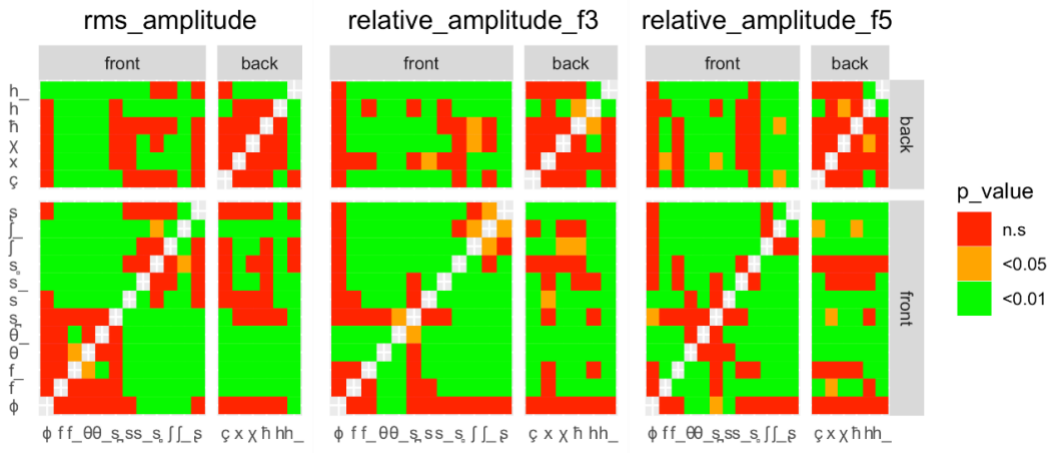

Figure 26 - Visualization of the statistical differences between fricatives in normalized amplitude and relative amplitude at F3 and F5

In the ANOVA analysis, the values of the normalized amplitude showed a significant main effect of phone identity [(1,8205 F= 254.77), p< 0.01, η^2 = 0.34]. Bonferroni post hoc tests indicated that there are significant contrasts between a more narrow constriction (s̻, s̪, s, ʃ, ʂ) and a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 6):

- The dental-alveolar (s̻) was significantly different (p<0.01) from /ç/ and **not** significantly different from any of the other wider fricatives;
- The laminal (s̪) was significantly different (p<0.01) from /ɸ, f, h/ and **not** significantly different from /θ, ç, x, χ, ħ/;
- The alveolar (s) was significantly different (p<0.01) from /f, θ, h/ and **not** significantly different from /ɸ, ç x, χ, ħ/;
- The postalveolar (ʃ) was significantly different (p<0.01) from /ɸ, f, θ, x, χ, h/ and **not** significantly different from /ç, ħ/;
- The retroflex (ʂ) was significantly different (p<0.01) from /f, θ, h/ and **not** significantly different from /ɸ, ç x, χ, ħ/;

Within the narrower fricatives (s̪, s̪̊, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts (hypothesis 6a):

- The dental-alveolar (s̪) was significantly different (p<0.01) from /s, ʃ/ and **not** significantly different from /s̪̊, ʂ/;
- The laminal (s̪̊) was significantly different (p<0.01) from /ʃ/ and **not** significantly different from any of the other narrower fricatives;
- The alveolar (s) was significantly different (p<0.01) from /s̪, ʃ/ and **not** significantly different from /s̪̊, ʂ/;
- The postalveolar (ʃ) was significantly different (p<0.01) from /s̪, s̪̊, s/ and **not** significantly different from /ʂ/;
- The retroflex (ʂ) was not significantly different from any of the narrower fricatives.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts (hypothesis 2b):

- The bilabial (ɸ) was not significantly different from any of the other wider fricatives;
- The labiodental (f) was significantly different (p<0.01) from /ç, x, χ, ħ, h/ and **not** significantly different from /ɸ, θ/;
- The interdental (θ) was significantly different (p<0.01) from /ç, x, χ, ħ, h/ and **not** significantly different from /ɸ, f/;
- The palatal (ç) was significantly different (p<0.01) from /f, θ, h/ and **not** significantly different from /ɸ, x, χ, ħ/;
- The velar (x), the uvular (χ) and the pharyngeal (ħ) were significantly different (p<0.01) from /f, θ/ and **not** significantly different from the rest of the wider fricatives;
- The glottal (h) was significantly different (p<0.01) from /f, θ, ç/ and **not** significantly different from /ɸ, x, χ, ħ/.

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 6):

- /s, ʃ, h/ were significantly different (p<0.01) from all other English target fricatives;
- /f/ and /θ/ were not significantly different from each other.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[47] and not significantly different from the rest:

- /f/ was significantly different (p<0.01) from /s̪, ʂ/ and **not** significantly different from /ɸ, s̪/;
- /θ/ was significantly different (p<0.01) from /ʂ, ç/ and **not** significantly different from /s̪/;
- /s/ was significantly different (p<0.01) from /s̪/ and **not** significantly different from /s̪, ʂ, ç/;
- /ʃ/ was significantly different (p<0.01) from /s̪/ and **not** significantly different from /ʂ, ç/;
- /h/ was significantly different (p<0.01) from /ç/ and **not** significantly different from /x, χ, ħ/.

Among all fricatives, normalized amplitude was highest for /ɸ/ (-25.53) and lowest for /ʃ/ (-12.32).

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it is possible to observe:

- /s, ʃ, h/ are significant different (p<0.01) from the LibriSpeech /s, ʃ, h/, respectively;
- /f, θ/ are not significantly different from the LibriSpeech /f, θ/.

6.1.1.3.2. Relative amplitude

In the ANOVA analysis, the values of the normalized amplitude at F3 showed a significant main effect of phone identity [(1,8179 F= 234.042), p< 0.01, η^2 = 0.33]. Bonferroni

[47] These errors will be discussed further in section 6.2. Language acquisition perspective.

post hoc tests indicated that there are significant contrasts between a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 6):

- The dental-alveolar (s̪) was significantly different ($p<0.01$) from /θ, ç, ħ/ and **not** significantly different from /ɸ, f, x, χ, h/;
- The laminal (s̺) was significantly different ($p<0.01$) from /f, θ, h/ and **not** significantly different from /ɸ, ç, x, χ, ħ/;
- The alveolar (s) was significantly different ($p<0.01$) from /θ, ç, χ, ħ, h/, marginally different ($p<0.05$) from /x/ and **not** significantly different from /ɸ, f/;
- The postalveolar (ʃ) was significantly different ($p<0.01$) from /f, θ, ç, x, h/, marginally different ($p<0.05$) from /χ, ħ/ and **not** significantly different from /ɸ/;
- The retroflex (ʂ) was significantly different ($p<0.01$) from /f, θ, ç, x, χ, ħ, h/ and **not** significantly different from /ɸ/;

Within the narrower fricatives (s̪, s̺, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts (hypothesis 6a):

- The dental-alveolar (s̪) was **not** significantly different any of the narrower fricatives;
- The laminal (s̺) was significantly different ($p<0.01$) from /s/ and **not** significantly different from /s̪, ʃ, ʂ/;
- The alveolar (s) was significantly different ($p<0.01$) from /s̺, ʃ, ʂ/ and **not** significantly different from /s̪/;
- The postalveolar (ʃ) was significantly different ($p<0.01$) from /s/ and **not** significantly different from /s̪, s̺, ʂ/;
- The retroflex (ʂ) was significantly different ($p<0.05$) from /s/ and **not** significantly different from /s̪, s̺, ʃ/.

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts (hypothesis 6b):

- The bilabial (ɸ) was significantly different (p<0.01) from /θ/ and **not** significantly different from any of the other wider fricatives;
- The labiodental (f) was significantly different (p<0.01) from /θ, ç, χ, ħ, h/ and **not** significantly different from /ɸ, x/;
- The interdental (θ) was significantly different (p<0.01) from all the wider fricatives;
- The palatal (ç) was significantly different (p<0.01) from /f, θ, h/ and **not** significantly different from /ɸ, ç, x, χ, ħ/;
- The velar (x) was significantly different (p<0.01) from /θ/ and **not** significantly different from the rest of the wider fricatives;
- The uvular (χ) was significantly different (p<0.01) from /f, θ, h/ and **not** significantly different from /ɸ, ç, x, ħ/;
- The pharyngeal (ħ) were significantly different (p<0.01) from /f, θ/, marginally different (p<0.05) from /h/ and **not** significantly different from /ɸ, ç, x, χ/;
- The glottal (h) were significantly different (p<0.01) from /f, θ, ç, χ/, marginally different (p<0.05) from /ħ/ and **not** significantly different from /ɸ, x/;

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 6):

- /θ, ʃ, h/ were significantly different (p<0.01) from all English target fricatives;
- /f/ and /s/ were not significantly different from each other.

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[48] and not significantly different from the rest:

- /f/ was significantly different (p<0.01) from /s̪, ʂ/ and **not** significantly different from /ɸ, s̪/;
- /θ/ was significantly different (p<0.01) from /s̪, ʂ, ç/;

[48] These errors will be discussed further in section 6.2. Language acquisition perspective.

- /s/ was significantly different (p<0.01) from /s̪, ʂ, ç/ and **not** significantly different from /s̻/;
- /ʃ/ was significantly different (p<0.01) from /ʂ, ç/ and **not** significantly different from /s̪/;
- /h/ was significantly different (p<0.01) from /ç, χ/, marginally different (p<0.05) from /ħ/ and **not** significantly different from /x/.

Among all fricatives, relative amplitude at F3 was highest for /ʂ/ (18.17) and lowest for /θ/ (-4.61).

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it is possible to observe:

- /θ, s, h/ were significant different (p<0.01) from the LibriSpeech /θ, s, h/, respectively;
- /ʃ/ was marginally different (p<0.05) from the LibriSpeech /ʃ/;
- /f/ was not significantly different from the LibriSpeech /f/.

In the ANOVA analysis, the values of the normalized amplitude at F5 showed a significant main effect of phone identity [(1,8179 F= 227.34), p< 0.01, η^2 = 0.32]. Bonferroni post hoc tests indicated that there are significant contrasts between a more narrow constriction (s̻, s̪, s, ʃ, ʂ) and a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h), (hypothesis 6):

- The dental-alveolar (s̻) was significantly different (p<0.01) from /θ, ç, ħ/, marginally different (p<0.05) from /ɸ/ and **not** significantly different from /f, x, χ, h/;
- The laminal (s̪) was significantly different (p<0.01) from /f, θ/ and **not** significantly different from /ɸ, ç, x, χ, ħ, h/;
- The alveolar (s) was significantly different (p<0.01) from /ɸ, f, ç, x, χ, ħ, h/ and **not** significantly different from /θ/;
- The postalveolar (ʃ) and the retroflex (ʂ) were significantly different (p<0.01) from /f, θ, ç, x, χ, ħ, h/ and **not** significantly different from /ɸ/;

Within the narrower fricatives (s̪, s̪, s, ʃ, ʂ) the Bonferroni post hoc tests indicated the following significant contrasts (hypothesis 6a):

- The dental-alveolar (s̪) was marginally different (p<0.05) from /s̪/ and **not** significantly different the other narrower fricatives;
- The laminal (s̪) was significantly different (p<0.01) from /s/, marginally different (p<0.05) from /s̪/ and **not** significantly different from /ʃ, ʂ/;
- The alveolar (s) was significantly different (p<0.01) from /s̪, ʃ, ʂ/ and **not** significantly different from /s̪/;
- The postalveolar (ʃ) and the retroflex (ʂ) were significantly different (p<0.01) from /s/ and **not** significantly different from the other narrower fricatives;

Within the wider fricatives (ɸ, f, θ, ç, x, χ, ħ, h) the Bonferroni post hoc tests indicated the following significant contrasts (hypothesis 6b):

- The bilabial (ɸ) was significantly different (p<0.01) from /θ/ and **not** significantly different from any of the other wider fricatives;
- The labiodental (f) was significantly different (p<0.01) from /θ, ç, χ, ħ/, marginally different (p<0.05) from /x/ and **not** significantly different from /ɸ, h/;
- The interdental (θ) was significantly different (p<0.01) from all the wider fricatives;
- The palatal (ç) was significantly different (p<0.01) from /f, θ, h/ and **not** significantly different from /ɸ, ç, x, χ, ħ/;
- The velar (x) was significantly different (p<0.01) from /θ/, marginally different (p<0.05) from /f/ and **not** significantly different from /ɸ, ç, χ, ħ, h/;
- The uvular (χ) was significantly different (p<0.01) from /f, θ/, marginally different (p<0.05) from /h/ and **not** significantly different from /ɸ, ç, x, ħ/;
- The pharyngeal (ħ) were significantly different (p<0.01) from /f, θ/ and **not** significantly different from /ɸ, ç, x, χ, h/;

- The glottal (h) were significantly different ($p<0.01$) from /θ, ç/, marginally different ($p<0.05$) from /χ/ and **not** significantly different from /ɸ, f, x, ħ/;

As for the English target fricatives (f, θ, s, ʃ, h) there are the following findings (hypothesis 6):

- /f/ and /h/ were significantly different from all English target fricatives, expect from each other;
- /θ/ and /s/ were significantly different from all English target fricatives, expect from each other;
- /ʃ/ was significantly different ($p<0.01$) from all English target fricatives;

Each English target fricative (f, θ, s, ʃ, h) was significantly different from some of the non-English fricative errors that were found in the annotations[49] and not significantly different from the rest:

- /f/ was significantly different ($p<0.01$) from /s̪, ʂ/ and **not** significantly different from /ɸ, s̪/;
- /θ/ was significantly different ($p<0.01$) from /s̪, ʂ, ç/;
- /s/ was significantly different ($p<0.01$) from /s̪, ʂ, ç/ and **not** significantly different from /s̪/;
- /ʃ/ was significantly different ($p<0.01$) from /ç/ and **not** significantly different from /s̪, ʂ/;
- /h/ was significantly different ($p<0.01$) from /ç/, marginally different ($p<0.05$) from /χ/ and **not** significantly different from /x, ħ/.

Among all fricatives, relative amplitude at F5 was highest for /ʂ/ (33.68) and lowest for /s/ (5.73).

[49] These errors will be discussed further in section 6.2. Language acquisition perspective.

When comparing the English target fricatives from this experiment with the English fricatives present in the benchmark dataset, LibriSpeech, it is possible to observe:

- /f, s, h/ were significant different ($p<0.01$) from the LibriSpeech /f, s, h/, respectively;
- /θ, ʃ/ were not significantly different from the LibriSpeech /θ, ʃ/, respectively.

Below, there is a table with all the values for the amplitude measurements from this dataset:

Major places of articulation	Fricatives	Relative amplitude F3	Relative amplitude F5	Normalized amplitude
Labial	ɸ	7.65	25.28	-25.53
	f	-2.58	10.29	-23.56
Coronal	θ	-4.61	6.31	-21.48
	s̪	-4.34	6.74	-24.63
	s̺	8.75	16.67	-13.47
	s	-3.31	5.73	-16.68
	ʃ	17.08	28.96	-12.32
	ʂ	18.17	33.68	-15.71
	ç	8.13	19.78	-15.38
Dorsal	x	1.79	15.91	-20.24
	χ	8.30	18.35	-19.02
Pharyngeal	ħ	6.36	15.79	-19.52
	h	2.18	15.12	-17.12

Table 14 - Amplitude values

6.1.2. Discussion

This section aims to critically evaluate the previously proposed hypotheses and showcase the key findings obtained from the acoustic analysis. The organizational structure of this section is similar to the findings of the acoustic analysis results. Firstly, the discussion focuses on the spectral properties' findings, followed by an exploration of the transition information and amplitude. Lastly, a brief comparison with LibriSpeech data is presented.

6.1.2.1. Spectral properties

In the present study, the more robust spectral properties are the centroid, skewness and peak location. This confirms hypothesis 3 and it agrees with the literature (Forrest *et al.*, 1988; Shadle, & Mair, 1996; Jongman *et al.*, 2000; Nirgianaki, 2014). The kurtosis measurement does not distinguish any fricatives in any of the windows, therefore, it does not seem to be adding information. This is contrary to the findings from Forrest *et al.* (1988), and Shadle & Mair (1996).

Across the different spectral properties, the second and third windows were the most robust windows, with the first and fourth windows not providing much information. However, in the kurtosis and peak location measurements, the fourth window provided relevant information.

The centroid does not make any distinctions between the back fricatives /x, χ, ħ, h/, and peak location, relevant information is only obtained in the fourth window, which is a window with a relatively small effect size.

The third window of the centroid and standard deviation were the only two measurements when all the English target fricatives were distinguishable.

The bilabial fricative failed most distinctions in the spectral properties' measurements. The lack of distinctions with the bilabial fricative could be due to insufficient data. Perhaps a larger number of annotations for the bilabial fricative could allow for a clearer distinction.

Despite the expected difficulties in distinguishing the dental-alveolar and the alveolar fricatives, both fricatives were distinguishable in several windows of several measurements (centroid, standard deviation, skewness, peak).

6.1.2.1.1. Centroid and peak location

The centroid and peak location were not the most informative or relevant measurements when trying to find differences between fricatives produced further back in the oral cavity (back fricatives). Existing literature strongly suggests that measures such as center of gravity and peak location are effective in capturing differences in the length of the front cavity, which aligns with the hypotheses proposed in this study. However, it is important to note that previous studies often focus on a limited number of fricatives, leading to a potential misconception that these measures are universally effective. In reality, their effectiveness may be specific to high-amplitude fricatives. Therefore, a comprehensive examination of multiple fricatives is necessary to obtain a more accurate assessment of the overall performance of these measures.

For most of the fricatives where the hypotheses failed, it can be argued that they have a flatter, low-energy spectrum, which makes the centroid and peak more hazardous.

Both the centroid and peak location should differentiate between some fricatives that are produced in the front of the oral cavity from some fricatives produced in the back of the oral cavity (hypothesis 1). In the present study the two windows (second and third) partially confirm hypothesis 1 for both measurements. In the centroid, out of the eight front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) four were distinct from all the back fricatives (ç, x, χ, ħ, h) and other two were distinct from all back fricatives, expect the palatal. In peak location, out of the eight front fricatives (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) one - the interdental - was distinct from all the back fricatives and at least four other fricatives were distinct from all back fricatives, expect the palatal.

Within back and front fricatives there are some contrasts or lack thereof that are relevant to explain:

1. The bilabial was distinguishable from the palatal in the centroid and peak, even though it was not distinguishable from any of the other back fricatives /x, χ, ħ, h/. This may occur because the back fricatives /x, χ, ħ, h/ have a flatter spectrum and less peaks. The palatal (ç) seems to show a much higher centroid/peak location value than the other back fricatives (see page 120).
2. The dental and the laminal were not distinguishable from the palatal in all windows and average of the centroid and peak location. This may occur due to the palatal being articulated relatively in the same space in the front cavity as the dental and laminal, as the tongue is positioned close to the roof of the mouth in slightly different ways, resulting in a closer proximity.

Within English target fricatives (f, θ, s, ʃ, h), the labiodental and the interdental should be able to distinguish from the glottal (hypothesis 1). Both the centroid and peak location successfully distinguish the two fricatives in the second and third windows. Additionally, in the third window, all English target fricatives are distinguishable. In previous research, the non-sibilants (f, θ) were also distinguishable from the two sibilants (s, ʃ) in average centroid and peak location. This is confirmed for the alveolar (s) and partially confirmed for the postalveolar (ʃ). The postalveolar was distinguishable from the labiodental; however, it was not distinguishable from the interdental. Both fricatives (θ, ʃ) seem to have a relatively similar peak which is unexpected (a flatter spectrum is expected) but the spectrum is wider for the interdental than for the postalveolar. Both /f/ and /θ/ have extremely low values which can be explained by:

- /f/ and /θ/ are somewhat weak acoustically, therefore finding the centroid is difficult and potentially not reliable (Behrens & Blumstein, 1988a). The distribution of these two fricatives is very wide which means some tokens might behave as predicted, however these may be invalidated by other tokens which do not behave as predicted and create extremely low values.

In both measurements, it should be easy to distinguish between the fricatives /ɸ, f/ and the fricatives /ʃ, ʂ/, (hypothesis 1a). This is partially confirmed for the centroid and not so much

for peak location. In the centroid, both the bilabial and the labiodental were distinguishable from the postalveolar. However, they were never distinguishable from the retroflex. In the peak, the labiodental was distinguishable from the postalveolar. No contrasts were found between the bilabial fricative or the retroflex fricative. As for the retroflex, values from Gordon *et al.* (2002) showed a much higher centroid (4535Hz) while in this experiment it is much lower (3473Hz). This fricative shows lower values than expected which suggests this fricative is produced further back than anticipated and has a much flatter spectrum than expected, similarly to the bilabial and labiodental.

Both the centroid as well as peak location should distinguish between /ç, x/ and /h/ (hypothesis 1b). This is confirmed for the palatal and the glottal in the second and third windows of the centroid and peak location. However, the prediction failed for the velar and the glottal, since no window can distinguish the two fricatives in both measurements. Despite the length of the front cavity being different for these two fricatives, they have very flat spectra and therefore it is not possible to find a distinction between the two.

It is expected that fricatives in adjacent places of articulation should be hard to distinguish (hypothesis 1d). The hypothesis is partially confirmed for both the centroid and peak location. In the second window of the centroid six out of the twelve pairs were not distinguishable and in the third window, seven out of the twelve pairs. In the second window of peak location eight out of the twelve pairs were not distinguishable. In the third window nine out of the twelve pairs. The fourth window of peak location does not distinguish ten out of the twelve pairs, however, the two pairs distinguished are important, as explained in the next paragraphs.

The second window of the **centroid** distinguished half of the pairs (six). It distinguished all the pairs from the interdental to the palatal: θ, s̪ / s̪, s̪ / s̪, s / s, ʃ / ʃ, ʂ. The third window of the **centroid** successfully distinguished one pair that none of the other spectral properties was able to distinguish - /ɸ, f/. This is expected to be a very subtle contrast, as even the two annotators disagreed more in the annotations process than most of the other fricatives.

The fourth window of **peak location** also distinguished one pair that none of the other spectral properties was able to distinguish - /x, χ/. Given the small effect size of this window, this distinction may be accidental. Nonetheless, these two fricatives are hard to distinguish in all measurements and a distinction in one of the windows shows that the contrasts that the annotators found had some acoustic truth to them. The same window also showed a distinction between the labiodental and the interdental.

For both the centroid and peak location, the two pairs of adjacent fricatives (/θ, s/, /s, ʃ/) should be distinguishable (hypothesis 1d.i). Both pairs were successfully distinguishable in the second and third windows of the centroid and peak location.

6.1.2.1.2. Standard deviation, skewness and kurtosis

Standard deviation, skewness and kurtosis should differentiate between fricatives with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h). (Hypothesis 2) In the present study two windows (second and third) partially confirm hypothesis 2 for standard deviation and skewness and mostly invalidate the hypothesis for kurtosis.

Standard deviation showed some success in distinguishing between these fricatives, as it successfully differentiated two narrower fricatives from four of the wider fricatives, out of the total eight fricatives considered. Additionally, all other narrower fricatives displayed contrasts with some of the wider fricatives.

Skewness also showed some success, as it successfully differentiated one narrower fricative from all the wider fricatives, and it successfully differentiated two narrower fricatives from five wider fricatives. All other fricatives showed contrasts with some of the wider fricatives.

Kurtosis showed less success. In the third window it only had a few contrasts. In the second and fourth windows, three narrower fricatives /s̪, s̺, ʂ/ didn't show any differences from the wider fricatives.

Between the narrower and the wider fricatives there are some contrasts or lack thereof that are relevant to explain in standard deviation and kurtosis:

1. The bilabial and the uvular fricatives were never distinguishable from the narrower fricatives (s̪, s̪, s, ʃ, ʂ) in standard deviation and kurtosis. These two contrasts have shown to be particularly difficult. Differences can be observed only in the second and third windows of skewness. For the uvular, given the points of articulation are so far back, the constriction could be smaller and have a different size and shape than anticipated which causes unexpected acoustic properties (Denzer-King, 2013). Maybe the size and shape are closer to the narrower fricatives.
2. The pharyngeal was only distinguishable from the dental-alveolar and not from any of the other narrower fricatives, in the second and third windows of standard deviation and kurtosis. Given the point of articulation being so far back, the constriction could be smaller and have a different size and shape than anticipated.
3. The palatal was only distinguishable from the dental-alveolar and not from any of the other narrower fricatives, in the second and third windows of standard deviation. The palatal fricatives are quite variable in terms of size and shape and maybe can be as narrow as some of the narrow fricatives. Perhaps the palatal is better characterized as a narrow fricative. Kurtosis did not make any distinction between the palatal and the narrower fricatives.

In the three measurements (SD, skewness and kurtosis), within the English target fricatives (f, θ, s, ʃ, h), the labiodental, the interdental and the glottal should be distinguishable from the alveolar and postalveolar (hypothesis 2). This is confirmed in the second and third windows of standard deviation. When it comes to skewness, it is partially confirmed, given that it is true for the alveolar but only partially true for the postalveolar. In the kurtosis measurement this is partially confirmed, since only a small number of distinctions is made in the different windows.

This differs from the findings of Jongman *et al.* (2000) regarding standard deviation and skewness. In their study, these fricatives were distinguished in skewness and all windows of standard deviation, except in the second window. These results have some similarities to the kurtosis findings, in which kurtosis failed to distinguish the two sibilants (f, θ) from the two non-sibilants (s, ʃ). Furthermore, Jongman *et al.* (2000) results indicated the two sibilants (s, ʃ) were easily distinguishable in the average of standard deviation, skewness and kurtosis but not in the third window of skewness and kurtosis. In this experiment, the two fricatives are distinguishable in the average skewness and in the third window of skewness and kurtosis. Average standard deviation and kurtosis fail to contribute to the distinction. Jongman *et al.* (2000) results also indicate that the two non-sibilants (f, θ) should be distinguishable in all windows of standard deviation except the second, all windows of skewness except the third and in the first window of kurtosis. In this experiment, the two fricatives are distinguishable only in the third window of standard deviation and skewness, and in the fourth window of standard deviation and kurtosis.

Within narrow fricatives, standard deviation, skewness and kurtosis should make distinctions between the alveolar /s/ and some of the other narrower fricatives (s̪, s̺, ʃ, ʂ), (hypothesis 2a). This is partially confirmed for the second and third windows of standard deviation and kurtosis. In the skewness measurement the hypothesis was confirmed, with the alveolar being distinguishable from all the other narrower fricatives.

Within wider fricatives, standard deviation, skewness and kurtosis should make distinctions between the palatal /ç/ and the other wider fricatives (ɸ, f, θ, x, χ, ħ, h), (hypothesis 2b). Both standard deviation and skewness in the second and third windows were partially successful in distinguishing the palatal and the other wider fricatives. In the two measurements it was possible to distinguish between the palatal and all wider fricatives, except /ɸ/. Kurtosis failed to make any distinction between the palatal and the other wider fricatives. Perhaps it is necessary to gather more data for the bilabial to be more accurately compared.

6.1.2.2. Transition information

In the present study, the more robust transition information measurement is the slope[50]. This invalidates hypothesis 5 and the results from Nirgianaki (2014). The size effect for F2 onset[51] is rather small, confirming Jongman *et al.* (2000) findings.

The bilabial fricative failed all distinctions in the transition information measurements. The lack of distinctions with the bilabial fricative could be due to insufficient data. Perhaps a larger number of annotations for the bilabial fricative could allow for a clearer distinction. Further research needed.

Both the F2 onset frequency and the slope should differentiate between fricatives that are produced in the front of the oral cavity (ɸ, f, θ, s̪, s̺, s, ʃ, ʂ) from fricatives produced in the back of the oral cavity (ç, x, χ, ħ, h). (Hypothesis 4). In the present study this is partially confirmed for F2 onset and for the slope.

In the F2 onset, three of the front fricatives were not distinct from any of the back fricatives. Some of the fricatives showed a few distinctions. One front fricative (f) was distinct from three (60%) of the back fricatives (ç, ħ, h) and four others (θ, s̪, s, ʂ) were distinct from one (20%) of the back fricatives (ç or χ). However, in F2 onset none of the front fricatives were distinguishable from the velar. The coarticulatory effects between the vowel and the fricatives can lead to reduced acoustic differences in F2 onset between front fricatives and the velar fricative. The articulatory movements involved in producing the vowel and transitioning to the fricative can influence the shape and size of the oral cavity. This coarticulatory influence can make it difficult to distinguish front fricatives from the velar fricative based on the F2 onset measurement.

The slope showed some success with these distinctions, by having two front fricatives (f, s) that were distinct from three (60%) of the back fricatives (x, ħ, h). However, two of the front

[50] See methodology section for more information about this measurement and its caveats.
[51] Due to the way the intercept measurement was approached, the intercept values are the same as F2 onset frequency.

fricatives were not distinct from any of the back fricatives. In the slope none of the front fricatives were distinguishable from the palatal and the uvular.

Within English target fricatives (f, θ, s, ʃ, h), the labiodental and the interdental should be easy to distinguish from the glottal (hypothesis 4). This hypothesis is mostly confirmed since these fricatives were always distinguishable, except the interdental from the glottal in F2 onset frequency. Adjacent sounds influence each other through coarticulation. The influence of neighboring sounds can affect the articulatory configuration and length of the front cavity for front fricatives. The interdental and glottal's lack of distinction may occur due to coarticulation between the vowel and the glottal which influences the shape and size of the oral cavity when producing the fricative.

In both measurements, it should be easy to distinguish between the fricatives /ɸ, f/ and the fricatives /ʃ, ʂ/, (hypothesis 4a). This was partially confirmed for F2 onset frequency and not confirmed for the slope. In the F2 onset, the bilabial did not show any distinctions. The labiodental was distinguishable from both fricatives (ʃ, ʂ). In the slope, only the labiodental was distinguishable from the postalveolar.

The fricatives /ç, x/ should be distinguishable from /h/ (hypothesis 4b). This hypothesis was invalidated by the results in both measurements. The small size effect of this measurement could be the reason for this. Furthermore, values tend to descend slightly with dorsals and pharyngeals which suggests the dorsal and pharyngeal fricatives were altered due to the vowel properties. Vowels are produced with specific configurations of the vocal tract in the positioning of the tongue, jaw, and other articulatory structures. These configurations can have an impact on the articulation and shape of the fricative, causing adjustments to the size, constriction, or airflow characteristics of the fricative.

It is expected that fricatives in adjacent places of articulation should be hard to distinguish (hypothesis 4d). This hypothesis is mostly confirmed for both measurements. There were only two distinguishable pairs of adjacent fricatives: f/θ and s/ʃ.

6.1.2.3. Amplitude

In the present study, all amplitude measurements were robust. This validates hypothesis 7 and the results from Jongman *et al.*, (2000)[52] and Wikse Barrow *et al.* (2022).

The bilabial fricative showed the most distinctions in normalized amplitude. Nonetheless, this fricative still showed few distinctions, perhaps due to insufficient data. Perhaps a larger number of annotations for the bilabial fricative could allow for a clearer distinction. Further research needed.

Normalized amplitude and relative amplitude at F3 and F5 should differentiate between fricatives with a more narrow constriction (s̪, s̺, s, ʃ, ʂ) and a wider constriction (ɸ, f, θ, ç, x, χ, ħ, h). (Hypothesis 6) This was partially confirmed for normalized amplitude and relative amplitude at F3, and mostly confirmed for relative amplitude at F5.

In normalized amplitude, the narrow fricative (ʃ) was successfully distinguishable from all wider fricatives, except /ɸ/. Most of the narrow fricatives were distinguishable from three wider fricatives and only one showed a minimal distinction from one wider fricative. The pharyngeal was never distinguishable from the narrower fricatives.

In relative amplitude at F3, two of the narrower fricatives were successfully distinguishable from all wider fricatives, except /ɸ/. All the narrow fricatives were distinguishable from a minimum of three wider fricatives. The bilabial was never distinguishable from the narrower fricatives.

In relative amplitude at F5, three of the narrower fricatives were successfully distinguishable from all wider fricatives, except one (either /ɸ/ or /θ/). All the narrow fricatives were distinguishable from a minimum of two wider fricatives.

[52]Jongman *et al.* (2000) only considers relative amplitude at F3 and relative amplitude at F5 for some of the fricatives. More information in section 5.3. Acoustic measurements.

In the three measurements, within the English target fricatives (f, θ, s, ʃ, h), the labiodental, the interdental and the glottal should be distinguishable from the alveolar and postalveolar (hypothesis 6). This hypothesis is fully confirmed for normalized amplitude and mostly confirmed for relative amplitude at F3 and F5. In relative amplitude at F3 only the labiodental was not distinguishable from the alveolar. In relative amplitude at F5 only the interdental was not distinguishable from the alveolar. This proves that doing these two measurements for all fricatives is useful, unlike Jongman *et al.* (2000), who only uses each measurement for some of the fricatives.

Within narrow fricatives, there should be a distinction between the alveolar /s/ and some of the other narrower fricatives (s̪, s̪, s, ʃ, ʂ), (hypothesis 6a). This is mostly confirmed for relative amplitude at F3 and F5, where /s/ was distinguishable from three of the four narrower fricatives (75%). The alveolar (s) was not distinguishable from the dental (s̪). This was partially confirmed for normalized amplitude, where /s/ was distinguishable from two of the four narrower fricatives (50%).

Within wider fricatives, there should be a distinction between the palatal /ç/ and the other wider fricatives (ɸ, f, θ, x, χ, ħ, h), (hypothesis 6b). This hypothesis was mostly invalidated in all measurements. Out of the seven fricatives, three were distinguishable from the palatal (f, θ, h). Perhaps the more significant amounts of data for these fricatives compared to the other wider fricatives is the reason to only find a distinction for those. It is also possible that these three fricatives are the widest fricatives and therefore are easier to distinguish.

6.1.2.4. LibriSpeech

Each LibriSpeech English target fricative is not expected to be distinct from the exact same fricative in this dataset. Skewness, kurtosis and F2 onset/intercept were the only two measurements that met most of the expectation for the annotated target and the LibriSpeech target in that it should not be different. In skewness the only fricative that was distinguishable was /θ/, in kurtosis it was /s/ and in F2 onset frequency it was /ʃ/. All other measurements

showed differences between the respective fricative in both datasets. The English target fricatives may differ from the LibriSpeech fricatives due to:

- LibriSpeech data only includes fricatives with 16kHz sampled speech. If some fricatives have very large amounts of energy, even if it is diffuse, in the frequency range above 8kHz, then these peaks may not be accurately captured in the acoustic signal by the LibriSpeech data. This could lead to a situation where, for example, the center of gravity is shifted or reduced for these fricatives. As a result, the values for fricatives with high peaks in the frequency range, such as /s/ and /ʃ/ may differ, and this would translate into different numbers for the fricatives in this dataset and the fricatives in the LibriSpeech data.
- A noisier background throughout this dataset than the LibriSpeech data. This may affect the acoustically weaker fricatives more than the rest (θ, h).
- The annotators might have excluded fricatives that were within the range of native production due to very strict criteria.
- LibriSpeech includes a wider range of variants of a given fricative, while our annotations are much narrower.

6.2. Language acquisition perspective

This section will present the results of the second language acquisition analysis on the human annotated data, followed by a discussion regarding these results and considering the hypothesis predicted.

6.1.2. Results

6.1.2.1. The five English fricatives

The data was annotated by two linguists and iterations were done till reaching a certain consensus. Cohen's kappa coefficient was used to measure the agreement and check for reliability of the annotation process.[53]

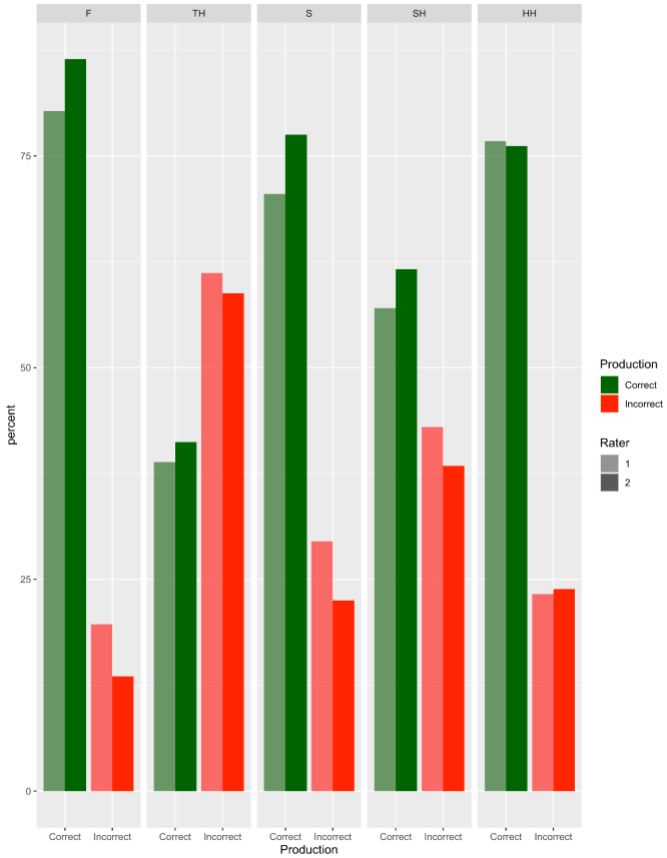

Figure 27 - Correct vs Incorrect productions of the target fricatives for the two raters

[53] There was no cross-reference between correct and incorrect productions for the two annotators, resulting in possible overlap between OKs from one annotator and ERR's from the second annotator in this figure.

As it is possible to observe in figure 27, the target fricatives /f, s, h/ have higher percentages of correct productions and /θ, ʃ/ have a lower percentage of correct productions. This is true for both annotators' results, even though one of the annotators seems to flag more errors than the other. However, the five fricatives had a relatively high percentage of incorrect productions. Within the incorrect productions there were substitutions, deletions and insertions. All the errors made by Vietnamese speakers present in the annotations can be found in figure 28.

Each fricative had the following amount of correct productions averaged across the two speakers:

- /f/ had 83% correct productions.
- /θ/ had 38% correct productions.
- /s/ had 74% correct productions.
- /ʃ/ had 43% correct productions.
- /h/ had 75% correct productions.

During the annotations of human annotated data at least 10% of all audios were excluded, given their lack of quality. This presented an issue, since it created the necessity to discard many possible mispronunciations due to this. However, it is not surprising, considering this is data from users who can record in any conditions. Poor quality audios are sometimes related not only with poor conditions of recording and bad audio quality, but also because of the volume of air made by the users which goes directly to the microphone.

Cohen's kappa coefficient was calculated to assess inter-rater agreement for *IN* audios and *OUT* audios. For the classification of /f/, the kappa coefficient was found to be k = 0.53 (95% CI: 0.45-0.62), indicating moderate agreement between the two raters. For the classification of /θ/, the kappa coefficient was found to be k = 0.98 (95% CI: 0.96-1), indicating almost perfect agreement between the two raters. For the classification of /s/, the kappa coefficient was found to be k = 0.44 (95% CI: 0.37-0.51), indicating moderate agreement

between the two raters. For the classification of /ʃ/, the kappa coefficient was found to be k = 0.51 (95% CI: 0.42-0.60), indicating moderate agreement between the two raters. For the classification of /h/, the kappa coefficient was found to be k = 0.43 (95% CI: 0.36-0.49), indicating moderate agreement between the two raters.

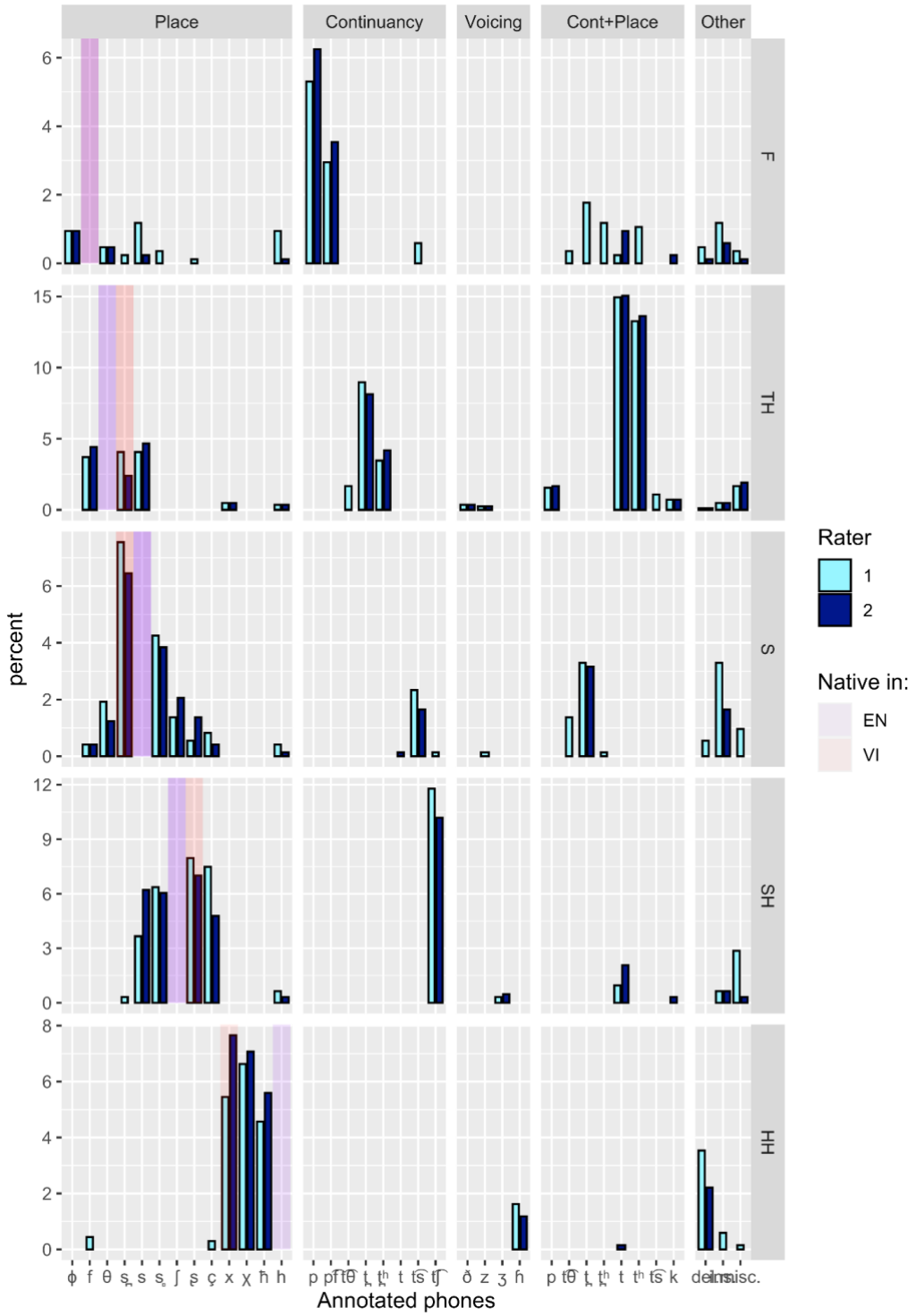

Figure 28 - Annotated phonemes in all five English fricatives by the two annotators

Within the errors for the labiodental fricative /f/, the most common errors were the substitution for the voiceless bilabial stop /p/ and the voiceless labiodental affricate /p͡f/. Both sounds are non-continuant (-continuant), meaning that they create an obstruction and partially block the airflow at a specific point of articulation, unlike /f/. Cohen's kappa coefficient was calculated to assess inter-rater agreement for the classification of /f/. The kappa coefficient was found to be k = 0.47 (95% CI: 0.39-0.55), indicating moderate agreement between the two raters.

Within the errors for the interdental /θ/ fricative, the most common errors were the substitution for the voiceless unaspirated and aspirated dental stops /t̪, t̪ʰ/, the unaspirated and aspirated alveolar stops /t, tʰ/, the voiceless labiodental fricative /f/ and the voiceless alveolar fricative /s/. The dental stops and the alveolar stops are non-continuant (-continuant) unlike /θ/ but share the same place of articulation (+anterior). Cohen's kappa coefficient was calculated to assess inter-rater agreement for the classification of /θ/. The kappa coefficient was found to be k = 0.92 (95% CI: 0.90-0.95), indicating almost perfect agreement between the two raters.

Within the errors for the alveolar /s/ fricative, the most common errors were the substitution for the voiceless dentalized fricative /s̪/ and the voiceless laminal alveolar fricative /s̻/. All these sounds share their place of articulation (coronal). Cohen's kappa coefficient was calculated to assess inter-rater agreement for the classification of /s/. The kappa coefficient was found to be k = 0.87 (95% CI: 0.82-0.93), indicating almost perfect agreement between the two raters.

Within the errors for the postalveolar /ʃ/ fricative, the most common errors were the substitution for the voiceless postalveolar affricate /t͡ʃ/, the voiceless retroflex /ʂ/, the voiceless palatal /ç/, the voiceless alveolar fricative /s/ and the voiceless dentalized /s̪/. Cohen's kappa coefficient was calculated to assess inter-rater agreement for the classification of /ʃ/. The kappa coefficient was found to be k = 0.73 (95% CI: 0.66-0.80), indicating substantial agreement between the two raters.

Within the errors for the glottal /h/ fricative, the most common errors were the substitution for the voiceless velar fricative /x/, the voiceless uvular fricative /χ/ and the voiceless pharyngeal fricative /ħ/. Both the uvular and pharyngeal were found to be distinguishable from the glottal in the acoustics section, although in few measurements. The velar was never distinguishable from the glottal. Cohen's kappa coefficient was calculated to assess inter-rater agreement for the classification of /h/. The kappa coefficient was found to be k = 0.68 (95% CI: 0.56-0.80), indicating substantial agreement between the two raters.

6.2.1.1. Discussion

The phones that exist in the inventory of the first language and second language simultaneously are not expected to be hard to produce for learners. (Lado, 1957; Flege, 1987) For Vietnamese native speakers it is the labiodental /f/ and glottal /h/ fricatives. (Hypothesis LA_1[54]) In this dataset, the two fricatives had high numbers of misproductions, 17% and 25%, respectively. This means the hypothesis failed.

The labiodental fricative may have caused problems due to its spelling, since all the words present in the data use the letter "f" which does not exist in the Vietnamese alphabet. The same sound is represented by "ph" in the Vietnamese language. This probably confused the speakers. Therefore, LA_H1a is verified. Moreover, the two top errors were /p, p͡f̪/ which shows an attempt to produce the phone by increasing the degree of stricture (fortition). (Nguyen, 2007) Learners tend to overemphasize the phonetic differences between L1 and L2 sounds by exaggerating some phonetic features. This tendency has been observed in L2 acquisition of other languages as well. (Odisho, 2020).

As for the glottal fricative, it is unexpected to observe such a high frequency of errors. Perhaps this is related to the following vowel. Since the vowel /æ/ does not exist in the Vietnamese language it might have influenced the productions of the speakers. The closest Vietnamese low vowel is /a/, thereby speakers may be trying to do some fronting for the vowel,

[54] All language acquisition hypotheses will have an index of LA (language acquisition) to avoid confusion with the detection hypothesis.

but this results in hyper-articulation and modification of /h/. In future studies, more vowel contexts should be used for this type of experiment. It is also important to note that one of the most common misproductions was for the velar fricative which also exists in Vietnamese. However, it is surprising to observe that one of the main misproductions was of the uvular fricative, which does not exist in Vietnamese, but it is also a dorsal sound (produced by lifting the back of the tongue) like the velar. This fricative has a close place of articulation and once again, maybe in trying to produce the vowel or over pronouncing the target phone, the uvular fricative was produced. It is possible that the vowel /æ/ influenced the way that the fricative was produced, leading to the production of the uvular and pharyngeal instead. This could be due to the fact that these sounds are produced further back in the mouth. Additionally, it is possible that the Vietnamese speakers were overemphasizing /h/, leading to the overproduction of these sounds.

The phones which exist in the inventory of the native language but are phonetically different, both acoustically and articulatory, from the second language phone are expected to be hard (Flege, 1987). For Vietnamese native speakers it is the alveolar fricative /s/. (Hypothesis LA_2b). In this dataset the alveolar fricative had 26% misproductions, confirming hypothesis LA_2b. The most common misproduction was the voiceless dentalized fricative /s̪/. This is expected because the Vietnamese language has the voiceless dental fricative /s̪/, which is articulatory and acoustically (as shown in section 6.1. Acoustic analysis) close to the English /s/, as it differs mainly in its dentalization. Due to positive transfer[55] from the native language, Vietnamese speakers may produce the English /s/ as /s̪/ by applying the same articulation pattern as in Vietnamese. Another common misproduction was the voiceless laminal alveolar fricative /s̻/ which is acoustically similar to the postalveolar, as it is possible to observe in section 6.1. Acoustic analysis. This misproduction may occur because some Vietnamese speakers may not have a clear distinction between the English alveolar /s/ and the postalveolar /ʃ/ sounds. Therefore, they may produce the English /s/ with a postalveolar articulation, resulting in a laminal alveolar fricative. Finally, there were also some misproductions of the

[55] Concept is present in section 3.4. The L2 acquisition of fricatives.

voiceless retroflex /ʂ/. This misproduction may occur because Vietnamese also has a retroflex fricative /ʂ/ and this sound is articulatorily similar to the alveolar. Vietnamese speakers may have a tendency to overgeneralize the production of /ʂ/ to the English /s/ sound, resulting in a retroflex rather than alveolar fricative. Overall, the common misproductions of the English /s/ sound by Vietnamese speakers are due to the influence of both articulatory and acoustically similar sounds in their native language and the lack of clear distinction between certain English sounds.

The phones which exist in the phonemic inventory of the second language but do not exist in the phonemic inventory of the native language are expected to be hard (Flege, 1987). (Hypothesis LA_3) These phones are the interdental fricative (θ) and the postalveolar fricative (ʃ). These were the two fricatives with more misproductions confirming hypothesis LA_3. Both phones had 62% and 57% misproductions, respectively.

When attempting to produce the interdental fricative (θ), Vietnamese speakers produced mainly the voiceless dental stop /t̪/ and voiceless aspirated dental stop /t̪ʰ/, the voiceless alveolar stop /t/ and the voiceless aspirated alveolar stop /tʰ/. These substitutions are expected as Vietnamese uses the Latin alphabet and has "th" in the spelling of the language (Tam, 2005; Bui, 2016). This may have caused confusion in the speakers. This confirms hypothesis LA_3.a.ii.. There were also misproductions of the labiodental (f) and the dental /s̪/. This can be explained by the fact that these two fricatives exist in the Vietnamese inventory and have a close place of articulation to the interdental fricative. (hypothesis LA_3.a.i..)

When attempting to produce the postalveolar fricative (ʃ) the production of the voiceless retroflex /ʂ/ and the voiceless dentalized /s̪/ fricatives are expected since they are the fricatives with the closest place of articulation existent in the Vietnamese inventory. (Hypothesis 3c) In here, the non-native sounds were perceived as the closest articulatory phone in their native language. (Best & Tyler, 2007) The production of the voiceless postalveolar affricate /t͡ʃ/ and the voiceless palatal /ç/ are unexpected.

7. Conclusion and future work

The opportunity to access and analyze data from thousands of speakers in real life conditions is a luxury that most language acquisition studies do not have. It made possible the study of an unusual amount of data and an unusual number of speakers, which provided insights into the production of voiceless fricatives by Vietnamese speakers.

The different measurements reported in the literature, which are normally used to distinguish only a handful of fricatives, proved to be reliable to distinguish a much wider range of fricatives. This includes fricatives which are very similar.

Most studies are made in a laboratory or minimally controlled setting and have just a few speakers: Jongman *et al.* (2000), Nirgianaki (2014) and Wikse Barrow *et al.* (2022) all analyzed data from 20 speakers and Jones & Nolan (2007) from 5 speakers. In this study, the data is from outside of a controlled lab setting and it contains hundreds of speakers. Despite these challenging conditions, the measurements were proved to be reliable.

From the spectral properties, the centroid and peak location were the most successful in distinguishing major places of articulation and distinguishing minor places of articulation, similarly to what was reported by Jones & Nolan (2007) and by Nirgianaki (2014). On the other hand, it was demonstrated that some spectral properties, particularly standard deviation, did not distinguish major or minor places of articulation well. This might be due to the large number of fricatives being compared. This measurement demonstrated to be more successful when examining the same subset of fricatives that other research have evaluated, failing only in the distinction of two fricatives when comparing to Jongman *et al.* (2000) and providing more differences than Nirgianaki (2014).

From the transition information, the slope was the most robust measurement, despite it not being measured in the typical manner. F2 onset and the intercept were not as robust, which was similar to the results of Jongman *et al.* (2000). As it was not feasible to have numerous tokens per speaker, the slope measurement could not be replicated in the same way

as other research (Sussman & Shore, 1996; Jongman *et al.*, 2000). However, this measurement has been successful in the past. In order to differentiate more fricatives in future experiments, it would be desirable to have more tokens per speaker.

From the amplitude measurements, all measurements were successful in distinguishing major and minor places of articulation. Similarly to Jongman *et al.* (2000) and Nirgianaki (2014). Moreover, it was shown that both relative amplitude at F3 and relative amplitude at F5 are useful measurements to distinguish several fricatives and not only a subset, like in previous studies.

Furthermore, the average of the window values in the spectral properties should be calculated considering only the two most robust windows, second and third.

As anticipated in the language acquisition perspective of this thesis, the interdental and postalveolar fricatives presented the most significant difficulties for these speakers.

The impact of the Vietnamese alphabet is particularly evident in the articulation of fricatives such as /f/ and /θ/. The labiodental can be spelled with an "f" or with "ph" in English. However, in this dataset there were only words with "f". The spelling for the labiodental is always "ph" in Vietnamese. This is a factor that could be considered for next studies. The interdental is always spelled with "th" in English. The same spelling in Vietnamese is most likely interpreted as a combination of two different sounds /t+h/, resulting in different productions.

Future research will be made to assess the extent to which the acoustic metrics presented here could classify the fricatives in terms of place of articulation, in line with Jongman *et al.* (2000). Additionally, it might be helpful to collect additional samples of some particular fricatives in order to create a model that will be statistically meaningful. Another option would be to exclude some of these fricatives from the analysis, as they could be accidental or an artifact of the recordings.

Equally important would be to select a greater number of vowels that follow the fricative, as the learner may perform differently depending on vowel context. The front vowel

/æ/ does not exist in the Vietnamese phonetic inventory and this could have an impact in the production. If both existing and nonexistent vowels of the phonetic inventory of the native language of the speaker are included, this can be controlled.

Future studies need to look into how much the characteristics described here influence how fricatives are perceived in their place of articulation, as well as how this information can be applied to better teaching strategies and student feedback.

8. Bibliography

Artstein, R. (2017). *Inter-annotator Agreement*. In: Ide, N., Pustejovsky, J. (eds) Handbook of Linguistic Annotation. Springer, Dordrecht. https://doi.org/10.1007/978-94-024-0881-2_11

Baddeley, A. D. & Hitch, G. J. (1974). *Working memory*. In: G. Bower (Ed.), Recent advances in learning and motivation, Vol. 8. (pp. 47-90), New York: Academic Press.

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

Best, C. T. (1995). *A direct realist view of cross-language speech perception*. In W. Strange (Ed.) Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research. Baltimore: York Press, Timonium, MD.

Best, C. T., Tyler, M.D., Bohn, O., & Munro, M.J. (2007). *Nonnative and second-language speech perception: Commonalities and complementarities*. Second language speech learning: The role of language experience in speech perception and production. Amsterdam: John Benjamins, pp. 13-34.

Brants, Thorsten. (2000). *Inter-Annotator Agreement for a German Newspaper Corpus*. In the Second International Conference on Language Resources and Evaluation (LREC-2000).

Bui, S.T. (2016). *Pronunciation of Consonants /ð/ and /θ/ by Adult Vietnamese EFL Learners*. Indonesian Journal of Applied Linguistics, Vol. 6 No. 1, July 2016, pp. 125-134.

Cheon, Sang Yee. (2005). *Production and Perception of Phonological Contrasts in Second Language Acquisition: Korean and English Fricatives*. PhD diss., University of Kawai'i.

Demirezen, M. (2016). *Perceptual Identification and Perception of Sibilants of English Language by Turkish English Majors*. Procedia - Social and Behavioral Sciences, 232, 750-758.

Denzer-King, R. (2013). *The acoustics of uvulars in Tlingit*. M.A. thesis. The State University of New Jersey

Emerich, G. H. (2012). *The Vietnamese Vowel System*, PhD thesis. University of Pennsylvania Publicly Accessible Penn Dissertations. 632.

Flege, J. E. (1987). *The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification*. Journal of Phonetics, 15, 47-65.

Flege, J. E., MacKay, I.R.A., and Meador, D. (1999). *Native Italian Speakers' perception and production of English vowels*. The Journal of Acoustical Society of America, 106(5), 2973-2987.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). *Statistical analysis of word-initial voiceless obstruents: Preliminary data*. J. Acoust. Soc. Am. 84, 115–124.

Fox, R.A., Jacewicz, E., Eckman, F.R., Iverson, G.K., & Lee, S. (2009). *Perception versus Production in Korean L2 Acquisition of English Sibilant Fricatives*. In M.-G. Pak's (Ed.), Current issues in unity and diversity of languages (pp. 2661-2680). Seoul: The Linguistic Society of Korea.

Fu, H.L., Rodman, R.D., McAllister, D.F., Bitzer, D.L., & Xu, B. (1999). *Classification of Voiceless Fricatives through Spectral Moments*.

Gordon, M., Barthmaier, P., & Sands, K. (2002). *A cross-linguistic acoustic study of voiceless fricatives*. Journal of the International Phonetic Association, 32, 141 - 174.

Goto. H. (1971). *Auditory perception by normal Japanese adults of the sounds 'L' and 'R'*. Neuropsychologia 9: 317-323.

Hedrick, M. S., and Ohde, R. N. (1993). *Effect of relative amplitude of frication on perception of place of articulation*. J. Acoust. Soc. Am. 94, 2005–2027.

Hoang, T. Q. H. (1965). *A phonological contrastive study of Vietnamese and English*. M.A. thesis Lubbock, Texas: Texas Technological College, USA.

Isbell, Dan. (2016). *The Perception-Production Link in L2 Phonology.* MSU Working Papers in Second Language Studies. 7. 57-67.

Jones, M. J., & Nolan, F. J. (2007). *An acoustic study of North Welsh voiceless fricatives*. In Proceedings of the XVIth International Congress of Phonetic Sciences (pp. 873-876).

Jongman, A., Wayland, R., & Wong, S. (2000). *Acoustic characteristics of English fricatives*. The Journal of the Acoustical Society of America, 108, *3 Pt 1,* 1252-63.

Kirby, J. P. (2011). *Vietnamese (Hanoi Vietnamese)*, Journal of the International Phonetic Association, 41 (3): 381–392.

Kitikanan, P. (2016). *L2 English fricative production by Thai learners*. Doctoral dissertation, University of Newcastle upon Tyne, United Kingdom.

Ladefoged, Peter. (1967). Three Areas of Experimental phonetics. London: Oxford University Press.

Ladefoged, P., & Johnson, K. (2015). A Course in Phonetics (Seventh Edition). Boston, MA: Cengage Learning.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell. ISBN 0-631-19814-8.

Lado, R. (1957). *Linguistics Across Cultures*. Ann Arbor: University of Michigan Press.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge Studies in Speech Science and Communication. Cambridge University Press, Cambridge. ISBN 0-521-26536-3

Major, R. (2001). F*oreign Accent: the Ontogeny and Phylogeny of Second Language Acquisition*. Mahwah, NJ: Lawrence Earlbaum Associates, Inc.

McFarland, D. H., Baum, S. R., and Chabot, C. (1996). *Speech compensation to structural modifications of the oral cavity*, J. Acoust. Soc. Am. 100, 1093–1104.

Mella, O., Fohr, D., & Bonneau, A. (2015). *Inter-annotator agreement for a speech corpus pronounced by French and German language learners*. In Workshop on Speech and Language Technology in Education, ISCA Special Interest Group (SIG) on Speech and Language Technology in Education, Sep 2015, Leipzig, Germany.

Nirgianaki E. (2014). *Acoustic characteristics of Greek fricatives*. *The Journal of the Acoustical Society of America*, *135*(5), 2964–2976. https://doi.org/10.1121/1.4870487

Nguyen, D. D., Chacon, A., Payten, C., Black, R., Sheth, M., McCabe, P., Novakovic, D., & Madill, C. (2022). *Acoustic characteristics of fricatives, amplitude of formants and clarity of speech produced without and with a medical mask*. International Journal of Language & Communication Disorders, 57(2), 366-380.

Nguyen, T.D. (2014). *Some common pronunciation problems facing Vietnamese learners of English*. Faculty of Foreign Languages.

Nguyen, T. T. (2007). *Difficulties for Vietnamese when pronouncing English: Final Consonants*. Dalarna University, School of Languages and Media Studies.

Odisho, E.Y. (2020). *Different Degrees of Accent in Foreign/Second Language Learning: a Case of Overcompensation*. Linguística: Revista de Estudos Linguísticos da Universidade do Porto, 171-186.

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). *Librispeech: An ASR corpus based on public domain audio books*. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5206-5210.

Polivanov, E. (1931). *La perception des sons d'une langue étrangère*. In Travaux du Cercle Linguistique de Prague 4; in Le Cercle de Prague (change, 3) Paris, 1969, pp.111-14.

Roach, P. (1984). *English phonetics and phonology: a practical course*. Cambridge: Cambridge University Press, 1983. Pp. x, 212. *RELC Journal*. 15, 117-118. doi:10.1177/003368828401500113

Shadle, C. H. (1990). *Articulatory-Acoustic Relationships in Fricative Consonants*. In speech production and speech modelling, NATO ASI Series D-Vol. 55 (W. J. Hardcastle & A. Marchal, editors), pp .187-209. Dordrecht: Kluwer

Shadle, C.H., & Mair, S.J. (1996). *Quantifying spectral characteristics of fricatives*. Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, 3, 1521-1524 vol.3.

Stevens, K. N. (1999). *Acoustic phonetics*, Cambridge, MA: MIT Press.

Sussman, H. M., and Shore, J. (1996). *Locus equations as phonetic descriptors of consonantal place of articulation*, Percept. Psychophys. 58, 936 – 946.

Tam, H.C. (2005). *Common pronunciation problems of Vietnamese learners of English*. Journal of Science Foreign Language.

Tomiak, G. R. (1990). *An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents*. Doctoral dissertation, SUNY Buffalo.

Tuan, L.T. (2011). *Vietnamese EFL learners' Difficulties with English consonants*. Studies in Literature and Language, 3, 56-67.

Wikse Barrow, C., Włodarczak, M., Thörn, L., & Heldner, M. (2022). *Static and dynamic spectral characteristics of Swedish voiceless fricatives*. The Journal of the Acoustical Society of America, 152(5), 2588-2600.

Tools:

"IPA Chart, http://www.internationalphoneticassociation.org/content/ipa-chart, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2018 International Phonetic Association."

L1-L2 Map, Jacques Koreman, Olaf Husby and Preben Wik, available in https://l1-l2map.hf.ntnu.no/, Copyright © 2009 Craig Thompson

Acoustic analysis scripts:

Dicanio's Vowel acoustic script:

https://www.acsu.buffalo.edu/~cdicanio/scripts/Vowel_Acoustics_for_corpus_data_2.praat

Dicanio's Spectral Tilt Script for Praat:

https://www.acsu.buffalo.edu/~cdicanio/scripts/Get_Spectral_Tilt_2.praat

Dicanio's Spectral Envelope Script for Praat:

https://www.acsu.buffalo.edu/~cdicanio/scripts/Get_Spectral_Envelope.praat

Dicanio's Spectral Moments Script for Praat:

https://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives_4.0.praat

Reetz's spectrum script: https://github.com/HenningReetz/Praat-scripts/blob/main/Spectrum/Spectrum_5_0_0.praat

Kawahara's intensity script:

http://user.keio.ac.jp/~kawahara/scripts/get_intensity_minmax.praat