UiT The Arctic University of Norway

Faculty of Humanities, Social Sciences and Education

The effect of perception training on production and perception of English back vowels by German learners

Fabian Neumann

Master’s Thesis in Theoretical Linguistics LIN-3990 May 2024

Table of Contents

Foreword

The present study investigates whether perception training with feedback can improve the production and perception of German native speakers, when they speak English. An online experiment with three training rounds was created, which all participants had to complete. Perception and production performance were tested before and after the training. The quantitative analysis showed no significant improvements between the tests. No correlation between perception and production could be found. A qualitative analysis of the data provided insights into the perception and production of English back vowels in an L2 context. Speech patterns were analysed with regard to deviations from English native speakers. The data were interpreted in the context of current models of second language speech learning. The obtained results contribute to the sparse amount of available literature in this field of study.

Forord

Den foreliggende studien undersøker om persepsjonstrening med tilbakemelding kan forbedre produksjonen og persepsjonen til tyske morsmålstalere når de snakker engelsk. Et nettbasert eksperiment med tre treningsrunder ble opprettet, som alle deltakerne måtte gjennomføre. Persepsjons- og produksjonsprestasjoner ble testet før og etter treningen. Den kvantitative analysen viste ingen signifikante forbedringer mellom testene. Ingen korrelasjon mellom persepsjon og produksjon ble funnet. En kvalitativ analyse av dataene ga innsikt i persepsjonen og produksjonen av engelske bakre vokaler i en L2-kontekst. Talemønstre ble analysert med hensyn til avvik fra engelske morsmålstalere. Dataene ble tolket i sammenheng med gjeldende modeller for andrespråksinnlæring. De oppnådde resultatene suplerer den sparsomme mengden tilgjengelig litteratur om dette emnet.

How effective is perception training for the improvement of an L2 speaker's production and perception performance?

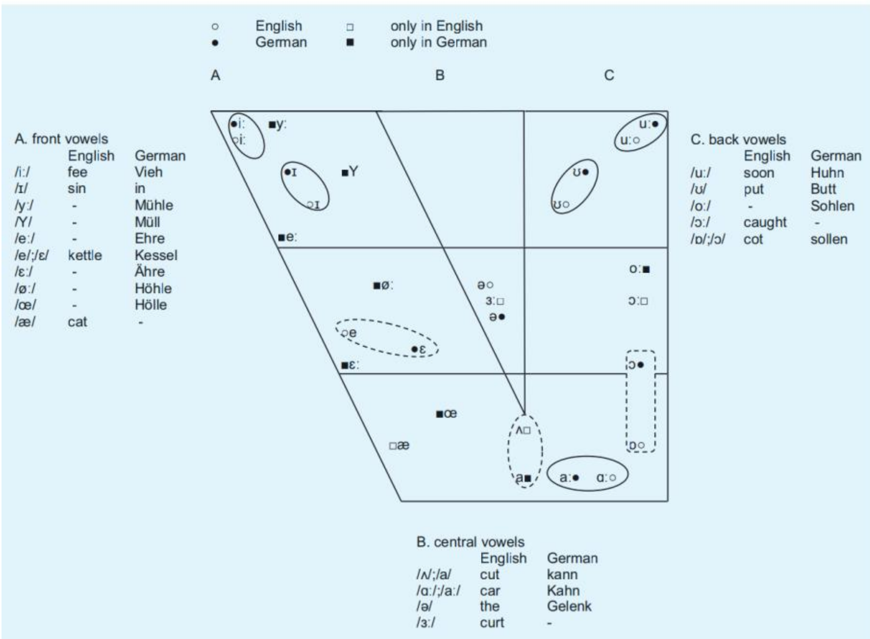
1 Introduction

The motivation that led to writing this work is of an almost traumatic nature. A few years ago, I found myself in a transcription course, faced with the task of transcribing an English text according to IPA guidelines. The initial attempts to phonemically transcribe the text were a disaster. I particularly remember the uncertainty in choosing between the English back vowels, as their distribution is different in my German native tongue. With this work, I have the chance to look back and reappraise my shortcomings from the past.

British English and German have sound systems that, in some instances, overlap and in others have meaningful differences. These differences can affect the perception performance of German learners of English and also lead to inconsistencies or errors in speech production. L2 phonology and second language (L2) speech acquisition models are concerned with phenomena that occur whenever adults learn a new language. In terms of phonology, these theories speak of the categorisation of speech sounds and aim to describe the process that a learner of a foreign language has to go through in order to obtain full command of the new phonemic system and its phonetic equivalents. A very influential model is James Emil Flege's Speech Learning Model (SLM) from 1995. His theory suggests that a learner can utilise the same learning mechanisms present in first language acquisition, although their perceptual sensitivity undergoes alterations due to age, L1 exposure, and other external factors. The SLM, along with other theories of cross-linguistic influence, shall form the backbone of this thesis. It contributed significantly to the formulation of my hypotheses.

For this work, the focus lies on the vowel realisations and perceptions of German-speaking L2 learners of English. The thesis aims at presenting a clearer picture of the contrasts between German and British English back vowels. It must be emphasised that the phenomena under investigation, in the form they appear here, are unique to Received Pronunciation (RP) variants of British English (as the most prominent example), and a comparison with speakers of other variants, such as General American (GA), may result in different observations or the absence of the described processes. The thesis seeks to gather answers from within the L2 phonology framework on the one hand and from an empirical data collection on the other. Many surveys have been conducted looking at English front vowels as potentially disturbing factors in L2 acquisition of English (see for example Bohn & Flege, 1992; Kautsch, 2010; Sönning, 2020) but less light has been shed upon the bottom right corner of the vowel space (see Kautzsch, 2012). A quick comparison of English and German phonemic systems reveals that the two languages exhibit phonologically different sounds in the lower right corner of the vowel space. However, these sounds are generally very similar, phonetically speaking. The opposition between phonetics and phonology will reappear over the course of this work, according to ongoing debate in the corresponding literature. Figure 1 provides a very useful overview that nicely visualises the topic of interest. The main focus of this work lies on the British English phonemes /ɒ/, /ʌ/, /ɑː/ and /ɔː/ which appear here in the lower right corner of an English/German vowel chart.

Figure 1: *English and German phoneme systems contrastively (Kortmann, 2020: 137)*



In that corner, we find differences in quality for vowel phonemes, which occur in both languages. These differences in vowel quality will be of interest in this thesis. From an educational standpoint, it is interesting to see whether these potential errors can be unlearned. The good news is that, judging by theories like the SLM, such errors can be unlearned, so that native-like command of a language is possible even in learners of higher age.

Regarding the structure of this work, first, an appropriate background will be drawn, against which findings from my experiment will be presented later on. The theoretical backdrop is mainly concerned with L2 phonology and acquisition theory, such as Flege‘s SLM and Best‘s Perceptual Assimilation Model (PAM) but an excursus into Universal Grammar (UG) will also be featured. The theoretical part seeks to establish the context between my empirical work and the current state of research within the field of language experience in second language speech learning.

The empirical part consists of an online experiment where German L2 speakers (learners) of English had to perform a number of tasks: a perception task as well as a production task, both before and after perception training. The findings provide interesting insights into the phonetic nature of German realisations of English back vowel phonemes. The perception tests revealed difficulties in discrimination and identification with regard to certain phonetic contrasts within the English language. My findings are based on data from a diverse group of women and men across a variety of age groups. A prerequisite for participation was that German was the participants' native language, which could, however, include candidates with other native languages. The data was made accessible and analysed using the programs R-Studio, Excel and most importantly, Praat for the phonetic analysis.

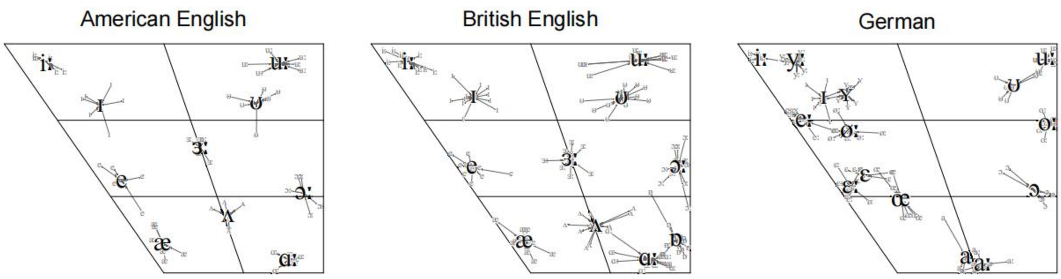
The broader field of speech learning research is often dismissed in colloquial remarks such as *That's just the way it is* or *Just accept that you sound German*. Indeed, a superficial examination of the subject material does not make the situation at hand appear particularly complex. When looking at the vowel charts of English and German, respectively, the matter seems straightforward, with German and English exhibiting different sets of phonemes that simply have to be acquired by a learner. But the phonetic reality of this part of the vowel space is in fact more complex than contrastive observation on a phonemic level may suggest, and so the question remains of how exactly these differences can be learned (and perceived). The study gives insights into how Germans perceive and produce English back vowels. Another concern of the study is to test the effectiveness of online training and to see to what extent the participants improve their performances in both listening and speaking abilities. An online approach has many advantages over a stationary (lab) approach as it allows for more flexible participation. The downside of such a procedure is, of course, the lack of control and the uncertainty about each participant‘s intention. The concept of online surveys for the collection of (authentic) speech samples is put to the test here and presents an ambivalent outcome. During the course of the experiment, I want to test if simple production and perception tasks can enhance competence in perception and production. Rooted in the SLM framework, the overall research question is a fairly simple one.

Can perception training with feedback significantly enhance the speaker's performance and perception competence?

In order to establish a solid groundwork upon which this question can be answered, I want to take a closer look at the distribution of English back vowels and the way in which native speakers of German realise them when acquiring English as an L2. Open-back vowels pose a problem for Germans, as some of them receive a (slightly) different phonetic realisation than their German counterparts and may also appear in differing phonemic contexts. German has a close, mid-rounded back vowel /o/ and a more open /ɔ/ which it shares with its Anglo-Saxon brother. Furthermore, it has the open front vowel /a/ and long /aː/ respectively (Wiese, 2006; Kortmann, 2020 see table 1). British English makes use of an additional three vowels that occupy a similar position within the vowel space, with varying German vowels corresponding to them on a phonemic level, namely /ɒ/, /ʌ/ and /ɑː/ (Roach, 2004). The purpose of this study is to test how these sounds are realised by German speakers and whether potential improvement in production can be achieved through perception training. Furthermore, the work aims to fill a gap in the scientific discourse by addressing a phenomenon that has received little research attention thus far. Taking into consideration the popularity of German and English as linguistic guinea pigs, this seems odd and deserves further academic commitment.

Chapter 3 provides a detailed description of the demographic composition of experiment participants. Here is just a brief overview to provide a rough idea of the testing conducted. All participants had an overall good command of the English language. However, there was no testing involved to determine each participants proficiency. Four questions were asked to gain some insight into each participant's speaking habits and overall experience with the use of the English language (see chapter 3). The conditions for participation were kept very general in order to achieve the highest possible participation rate. The justification for this approach is briefly outlined along the following lines: Overall, Germans' L2 English proficiency is quite high. For my purposes, it is important to know that the English language has an increasing influence on the life of an average German-speaking individual. Raymond Hickey's collective volume *English in the German speaking World* (2019) introduces aspects of an Anglo-German interplay. He describes a long history of English in the German-speaking area. Other authors point to socio-economic conditions and demographic developments in a globalised German society. From my own experience, I can say that English is taught at an early age in primary school and is omnipresent in people’s daily and professional lives (see also Hüllen, 2005). However, certain characteristics remain, which can easily be detected in the form of a typical German accent. Hickey (2019) refers to this as „common features across large numbers of speakers“ and hypothesises that „reasons for this [...] lie [...] in the structure of German, both phonological and morphosyntactic, and [...] in the manner in which the language was, and still is, taught in schools with typical non-native-like features being passed on from generation to generation“ (5). Another issue of relevance is the question of which variety of English is desirable for a learner of English or which variation should form the exonormative base for an intermediate learner (Hilgendorf, 2005). In my data collection, I also encounter this question because hardly any participant consistently adheres to a particular variation of English. Over the course of this paper, I will demonstrate how this instance presents a significant difficulty to most learners. In any case, it is interesting to have a systematic look at the matter because for people who already have good command of an L2, fine-tuning their speaking habits is the next step. This can be quite challenging, both on the phonetic and phonological level, due to the multifaceted diversity of English varieties. Interest in English L2 phonology stems from the simple fact that not all dialects use the exact same sounds in their respective inventories. To correctly master these differing sounds, is essential, to make oneself understood. To foreign accents, Flege (1988a) attests great potential of having a range of negative outcomes for the L2 user, including (negative) social evaluation and problems of intelligibility. In the long run, this is of interest for language educators who profit from concrete findings in L2 speech perception. Bent, Bradlow and Smith (2007) find that inaccurate realisations of vowels have greater consequences for a speaker's intelligibility than, e.g., errors in consonant production. A thorough examination of German and English back vowels thereby receives legitimacy because it has the potential to significantly further the participant‘s performance and provide insight into how (in this case auditive) training can accelerate the acquisition of new sounds. The study will point out the problems that arise when German learners of English are confronted with English back vowels in a context-free setting. The survey aims to demonstrate the pattern used by Germans to identify and produce the English vowels in question. The following vowel chart shows the phoneme inventories of GA, RP and standard German, respectively. This points to a factor that has caused considerable concern for my participants in both the production and perception parts, as GA and RP differ with regard to the vowel phonemes in question.

Figure 2*: Vowel quality in American English, British English, and German: Survey of auditory descriptions of Monophthongs (Sönning, 2020: 54)*



The minimal quadruple c*ot /ɒ/, caught /ɔ:/, cut /ʌ/, cart /a:/* illustrates the problem Germans face when acquiring English. Judging by their L1, Germans are expected to realise these pairs with four different sounds, just like a (British) English native speaker would, but unlike GA speakers. This is due to an overlap between the German and English phoneme systems, respectively. The following assumptions are based on and motivated by various literature by Flege, Best, Kautzsch and Sönning among others and my own experience. I expect my participants to perform as follows under production testing:

1. English /ɒ/ will most likely be realised as German /ɔ/.

2. English /ɔ:/ will most likely be realised as German/o:/.

3. /ʌ/ and /ɑ:/ will most likely be realised as German /a/ and /a:/ respectively, making a qualitative distinction between the *cut* and *cart* vowels impossible.

4. I hence expect difficulties when facing the /ʌ/ and /ɒ/ contrast (both in perception and production) because these vowels are not part of the standard German vowel inventory and are prone to English vowel mergers (see chapter 2.2 ff.).

5. It is expected that Germans compensate for the lack of qualitative distinction by a quantitative change in the length of the corresponding vowel and the application of L1 vowel qualities, so-called L1 to L2 transfer. I expect the following pattern:

English /ɒ/ → German /ɔ/

English /ɔ:/ → German /o:/

English /ɑ:/ → German /a:/

English /ʌ/ → German /a/

(English /ɒ/ --> German /a/ depending on context)

In order to cross-check these assumptions, German speech samples of German native speakers were also collected. The conducted experiment seeks to find out whether these assumptions are in fact true. Listening to target-oriented input from native English speakers may help the participants become more confident in using the appropriate phoneme in the corresponding situation and hence speak more accurately (if you know you are doing it right you automatically speak in a more confident way). This corresponds to findings by Birdsong (2007) who suggests that, in addition to linguistic awareness in the form of phonetic training, high (intrinsic) motivation favours successful learning of a new language. According to the following descriptions, a certain motivation for participation in the experiment can be inferred. This assumption is supported by the rather high time investment that was required of the participants.

The foregoing lines serve to introduce and contextualise the present work. They outline the key aspects to be addressed throughout the work, including the composition of the experiment participants, the theoretical background of the study, and the work's motivation and objectives. Additionally, it highlights the importance of a heightened awareness of the phonetic reality of speech utterances. Thus, the foregoing preamble provides an overview of the fundamental concepts and themes that will be further explored in the subsequent sections of this work. It thereby leads to the following two hypotheses:

H0 : Perception training has no significant effect on L2 learner's active performances.

H1 : Perception training has a significant effect on L2 learner's active performances.

H0: Good performance in perception does not have positive influence on performance in production

H1: Good results in perception yield more native like performance in production.

The listed hypotheses are based on findings by Flege et al. (see e.g.: Flege, 1988; Flege et al., 1995; Flege & Bohn, 2021). Flege's work underscores the importance of perception training in L2 acquisition, demonstrating that with appropriate training, learners can significantly improve their phonetic skills, leading to better overall language proficiency. In order to test my hypotheses, a survey is conducted, data is collected and the received feedback is put into context with pre-existing findings in order to contribute to the linguistic discourse. The realisation of English back vowels by German L2 speakers of English is a phenomenon with a short history of academic coverage. Contrastive studies on English and German are countless and readily available. However, a contrastive look at back vowels seems to be a niche where little research has been conducted so far. Works such as Kortmann (2020), Sönning (2020), Kautzsch (2010 and 2012) and Iverson and Evans (2009) should be mentioned in this particular field. Comparative data on vowel quality and general *academic inspiration* largely come from these authors and their works. Especially the various visualisations in these works have partly influenced my own elaboration. Kautzsch (2012) deserves mentioning in particular because in his work he portrays the phenomenon in question in much more detail than any other author in the field.

Another important aspect of my work is the distinction between purely theoretically motivated differences in phonology and the physical reality of the observed phenomena in phonetic terms (see Briere, 1966; Strange, 2007). To achieve this, common misconceptions need to be addressed, and phenomena that frequently arise outside the scientific discourse must be explained. I refer here to keywords such as vowel mergers of different kinds or long versus short vowels, which are likely familiar to many native speakers of English, and I attempt to embed these into the scientific discourse. Part of my theoretical background involves the rethinking of traditional linguistic assumptions on a phonological level and raising the readers' awareness of the phonetic reality of utterances, which is of great importance in the context of this work and in the field of second language speech learning in general.

2 Literature review and background information

2.1 Phonology vs. Phonetics

In order to draw conclusions about the phoneme systems of two languages, the first step is to consult a vowel chart of the corresponding language. These charts can give us an idea of the phoneme inventory of the language in question. They show all the entities that can make a difference in meaning when moved around between words. A good example of this are minimal pairs. When looking at the words *cart* and *cut,* we see that the vowel quality and quantity change, and with it, the meaning of the word. A phonemic transcription of these English words displays /kɑːt/ and /kʌt/ respectively, and is therefore a representation of the opposition of these two English sounds. Another example of a minimal pair in British English is the *caught* and *cot* vowels; /ɔː/ vs. /ɒ/. Quality and length differ, resulting in different meanings. German learners of English are hence facing a separation of the vowel space in four different areas, which are crucial to conveying meaning in British English. However, this area of the vowel space displays a different phonetic ground plan when it comes to German. To understand why a quick look at a table that displays such phonemic contrasts can be misleading, a little history of phonological research must be included.

IPA (international phonetic association) symbols are conventions that have historically developed over time and are surprisingly resilient when confronted with findings that indicate their defectiveness. The problem that arises is that in the literature, no uniform standards for phonemic transcription can be found. This circumstance becomes particularly apparent when considering the phonemes /a/ and /ʌ/. According to the IPA definition, /ʌ/ corresponds to an open-mid back unrounded vowel, thus serving as the unrounded counterpart to /ɔ/. However, upon examining the phonetic realisation of this phoneme by British native speakers, it becomes evident that this sound, by virtue of its physical properties, is located much lower on the vowel chart. Indeed, there is almost a phonetic overlap between the British phoneme /ʌ/ and the German phoneme /a/ (as my own findings suggest, see chapter 4). This phenomenon is illustrated in the following pages. In English language literature, almost exclusively /ʌ/ is used for the phonemic transcription of a sound that, according to its phonetic properties, would be more accurately described as an open central unrounded vowel (and should therefore be transcribed in a different way). In the practical part of this work, this circumstance becomes very clear, especially in contrast to German pronunciation. Since the English conventions in phonemic transcription prescribe the notation /ʌ/ for an open, central, unrounded vowel, I will adhere to that. However, it must be mentioned that an examination of a vowel chart according to IPA standards presents a distorted image of phonetic reality. In any event, this is the case when two or more languages are contrastively compared.

In order to understand this circumstance, a brief historical excursion into the beginnings of phonological research is necessary. The IPA maps vowels based on tongue position and corresponding lip rounding. This goes back to Daniel Jones and his descriptions of cardinal vowels from 1917! Although ground-breaking at the time, the notation shows considerable weaknesses in a context where every country and every school of thought establishes their own interpretations of the matter. A contrasting examination of the English and German phoneme systems initially yields an impression that does not correspond to the phonetic facts because of oversimplification of the vowel space in general. Where German and English are relatively close to each other (linguistically speaking) we get a different picture in a cross-linguistic context that includes more distant languages, which are utterly at variance with any Germanic sound system. There is a need to look at phonetics and the physical reality of sounds to grasp the matter in all its facets. This insight was already gained by Eugene Briére in 1966 and published in his book *Quantity before Quality in Second Language Composition.* Strange (2007) remarks that „the comparison of L1 and L2 phonological categories cannot be performed at the level of abstract phonological classes or distinctive features“ (37). She suggests that research should focus on „actual phonetic realisations“ (ibid.) of phonological categories in the way they are perceived by the listener. For learners of a foreign language, the issue always arises about what input they are provided with during their acquisition process. Clearly problematic is that the theoretical approach to foreign language instruction, generally cannot lead to working out phonetic differences in a foreign language on the basis of input, that is questionable at best (after all, we listen with our ears and not with our eyes). Flege, Best and other authors argue in a similar way and their ideas reappear over the course of this thesis. Hence, this thesis aims to bridge the gap between phonemic description and phonetic reality, at least with regard to a niche phenomenon that has not received much academic attention so far. The considerations that led to writing this paper will be briefly explained over the next few paragraphs.

2.2 The thing with mergers

As much as the British love talking about the weather, maybe their true passion lies in the field of linguistics and talking languages. Admittedly, they have a good reason for that, as English is spread over the entire globe and few other languages have such diverse manifestations in accents and idioms. To better contextualise the results of this study, some general terms need to be clarified. This includes so-called vowel mergers that occur in various variations of English. A vowel merger, in a diachronic perspective, combines two different (though phonetically similar) sounds into a single, new sound. At the outset, I already mentioned that my experiment primarily focuses on sounds associated with the RP variant of British English. This is important because other varieties may behave differently within the same linguistic context. The back vowels of English are a very volatile terrain. This means that many variants of English use a different pronunciation for these vowels (see Roach, 2004; Labov, 2006; Kortmann, 2020). Even native speakers who want or need to improve their pronunciation, struggle with these sounds. With regard to the lower right corner of the vowel chart, one might hear terms like *Hell's Corner* from YouTube coaches who try to teach English pronunciation to aspiring actors and L2 learners (@accenthelp is a YouTube channel dedicated to teaching English pronunciation). I will describe some of these mergers to later address these concepts in the analysis of my data.

Two vowel mergers are of interest here. I take the names from Labov et al. (2006). The first is the so-called *father-bother* merger*.* Varieties that employ this merger create homophones out of word pairs that, in the context of my study, are treated as minimal pairs. Examples are *bodge* and *barge* or *sob* and *Saab,* among others (a full list can be found in the appendices). According to Labov et al., this merger is dominant in GA, Irish and Scottish English and Indian and other Asian varieties of English. If uttered by a speaker who employs such a merger, the immediate consequence is that the assumed difference in vowel quality and quantity ceases to exist. For a learner, this is quite a significant obstacle on the path to a more native-like pronunciation. To avoid mistakes in this area, the learner must be aware that different variants of English have altered phoneme systems with different phonetic realisations. This is something that, in theory, is mentioned in foreign language instruction but is hardly ever actively taught (I remember vividly how teachers would urge me to pick one standard and stick to it but this happened primarily with regards to orthography rather than any serious feedback on the way I spoke). Besides myself, Kautzsch (2010, 2012) and Sönning (2020) also observe that only a few learners are consistently able to apply one standard variety of English to their speech, be it RP or GA, which are the standard varieties one chooses to adhere to in high school and university (see Hickey, 2019). Looking at the occurrence of this merger, it quickly becomes clear what great difficulty the L2 learner faces. GA is the common *Hollywood* variation of English, and native input in this variation is omnipresent. If one does not actively attempt to distinguish it from other variants, the use of the *father-bother* merger may seem almost arbitrary in context with other varieties of English. Again, we come across Birdsong's (2007) remarks on intrinsic motivation and how important it is for a more native-like performance in a target language. Also, Guion and Pederson (2007) assert that a participant's attention can be channelled towards a specific phenomenon, thereby improving phonetic distinction. However, phenomena like mergers are seldom questioned by the average learner of an L2. In the rare cases where phenomena like vowel mergers surface during the course of foreign language acquisition, it most probably happens due to very high intrinsic motivation or interest in the matter (like making it the main topic of one‘s master thesis). Since the participants in my study come from different backgrounds, this is an aspect that ought to be kept in mind when browsing through the data at hand. I do not assume that participants were in any way familiar with such phonological circumstances. However, the effects these mergers had on my participants‘ performances were undeniable.

But that is not all. A second merger, the so-called *cot-caught* merger, leads to complete confusion for the learner, especially against the backdrop of German as an L1. Varieties that exhibit this merger are again GA, Irish and Scottish, among others (see Labov et al., 2006). In these varieties, the minimal pair of *cot* vs *caught* becomes a homophone. For the same reasons as just explained, this variant is also very widespread, for example, in film and television. According to Labov (and as a logical consequence), the spread of these two mergers can result in a situation where parts of Northern America nowadays do not distinguish between the /ɔ/, /ɒ/ and /ɑ/ sounds, leading to homonomy of the words *stalk*, *stock* and *stark*. (For the sake of completeness, the concept of rhoticity must be mentioned here, which typically ensures that vowels followed by an *r* are also pronounced with an *r*. This is not the case in non-rhotic variants like RP, where *r* is pronounced in onset position alone (see Roach, 2004). Here, the *calf-cough* opposition serves as a better example). My female native speaker exhibited some of these mergers in her speech, at least partly. This demonstrates how even native speakers apply the mergers in some cases. This detail had profound and unforeseen consequences for my study's final data set.

To show that the devil is in the details here, Labov et al. provide an example where they found that approximately 15% of GA speakers exhibit a specific merger between the vowels /ɑ/ and /ɔ/ before the consonant /n/, resulting in homophony between words like *Don* and d*awn* while maintaining a distinction between words like *cot* and *caught.* This phenomenon, termed the *Don–dawn* merger*,* represents a distinct vowel shift that occurs alongside the more widespread merger between *father* and *bother* observed in all speakers who have undergone that merger (by sheer coincidence, this pair is also featured in one of my perception rounds and variation in performance could be detected). As one would expect from the esteemed Master Labov, there is a plethora of research and publications on all conceivable mergers and phonetic environments that favour the use of one over the other. Going into detailed descriptions of all the parameters at play would be a thesis in and of itself. My aim is simply to highlight the existence, significance and complexity of these phenomena, as they can be relevant for analysing the data I have gathered. For someone interested in aligning their spoken English with a universal standard like RP or GA, being informed about these phenomena is highly beneficial. At the same time, the average learner is (almost) never provided with such information during the learning process and hence variation in L2 production must be expected. I take these mergers to be important factors in my later interpretation of the received data.

A quick summary looks as follows: especially Northern American variations exhibit the *cot-caught* and *father-bother* merger. As a result, the minimal pairs that exist in RP become homophones in GA. We then find a threefold distinction in GA back vowel qualities, or, if both mergers apply, /ɔ/, /ɒ/ and /ɑ/ can even fall into one phoneme (see Labov's remarks on Canadian English). A remaining opposition always remains in the shape of the phoneme /ʌ/. The widespread occurrence of these mergers makes it very difficult for learners to recognise, let alone apply, the phonetic differences that RP still maintains in the appropriate context. It cannot generally be assumed that learners are aware of this fact without explicit instruction (Birdsong, 2007). (However, taking RP as a base eliminates the question of whether or not to apply the merger to a given word. In theory, the German phoneme system favours the use of an RP-oriented pronunciation). Given the varied nature of my findings, the role of these mergers must therefore be considered in my data analysis. Ultimately, it must be taken into account during the data analysis that some participants may not be familiar with the differences between the vowels due to the variety of English they have chosen (consciously or not). Regarding my own survey, I discovered some variation in the pronunciation of certain words, especially by my female native speaker. Later on, I will discuss the consequences of this for my study in general.

2.3 Developments in L2 speech learning

The field of study for second language speech learning is steadily growing. It is a complex field, including a wide range of disciplines such as linguistics, psychology and pedagogic approaches, among others. From a linguistic point of view, primarily phonology and phonetics are of interest when it comes to contrastive analysis of L1 and L2 variation. These contrasts can be conditioned by a myriad of external factors, which, by naming them all, would burst the limits of my thesis. The following lines contain some factors on which the relevant literature has largely reached a consensus: The perception (and production) of a language other than one's native tongue can be conditioned by age, developmental factors, quantity and quality of exposure to the L2, the general use over time, variance in L2 instruction or teaching (German English teachers tend to have strong German accents when they speak) and of course, the corresponding L1. Also, differences like motivation and aptitude of a learner play a role here (Bohn & Flege, 1992; Hickey, 2019). This is just to name some potential influences in the field of linguistics. There is no such study or experiment that takes into account all the factors that are at work here at the same time. When dealing with phenomena in this field, the focus often lays on specific attitudes, taking into account some of the afore-mentioned aspects but never taking an extensive account of all the variables at play. Another facet, which shall not be of interest here but nonetheless must be mentioned, is that the effects of non-native patterns in perception and production can have heavy effects on learner‘s participation in the target linguistic community (Brown & Kachru, 1995). Therefore, it is critical to address these issues in a globalised world where many people are multilingual or at least strive for proficiency in a language other than their mother tongue. That alone provides a justification for conducting research in this area.

When looking at the developmental stages that the field went through, we see that up until the late 1950s, not much energy was channelled towards questions of second language speech perception. A short chronological overview shall act as the basis upon which the outcomes of this article rest. In early L2 research, the focus was on the phonemic level rather than its physical counterpart in phonetics. Bloomfield (1933) stated that adult L2 learners can „ ignore the rest of the gross acoustic mass that reaches [their] ears“ (in Hockett, 1958: 24) because features of their L1 pre filter all linguistic input. In 1939 Trubetzkoy took a similar stand by saying that the sound system of an L1 prevents the acquisition of native-like L2 speech as an adult. Where the structuralists of the Prague school did not favour exhaustive analysis of the phonetic nature of the sounds in question, scientific discourse took off with Robert Lado‘s 1957 book *Linguistics across Cultures,* where he formulated his Contrastive Analysis Hypothesis (CAH), a publication that addressed L2 pedagogy from a behaviourist perspective. The CAH represented the initial formal explanation of cross-linguistic influence in L2 acquisition. According to the CAH, language learning entails the development of a set of habits that are transferred to the L2. Lado refers to successful habit shifts as positive transfers, and he attributes errors to negative transfers (or interference). He suggests that, through a contrastive analysis, it should be possible to identify challenging structures. He claims that similar structures should be straightforward, while different structures should present difficulties to the learner. The basic elements of contemporary speech models are already discernible in Lado‘s line of thinking. The CAH, while ground-breaking in its time, is not without its weaknesses. Some of these weaknesses include oversimplifying the process of language learning by primarily focusing on structural differences between languages and neglecting the role of other factors such as cultural differences, motivation, and individual learning strategies (Birdsong, 2007). While the CAH acknowledges the concept of transfer, it oversimplifies the ways in which learners transfer linguistic elements from their native language to the L2, thereby failing to address the complexities of transfer phenomena such as overgeneralisation or interference (Munro & Bohn, 2007). Moreover, the CAH primarily focuses on surface-level linguistic structures and does not consider the underlying cognitive processes involved in language acquisition, neglecting factors such as pragmatics, discourse, and sociolinguistic aspects of language use. This static perspective may not fully capture the complexity of language acquisition and use in real-life contexts, as the CAH tends to view language systems as static entities, failing to recognise the dynamic nature of language change and variation (Flege & Port, 1981). Also, CAH approaches often place much emphasis on potential negative transfer from the native language to the target language, overlooking the potential benefits of positive transfer or language universals that could facilitate learning (Grosjean, 1998). Some critics have raised concerns about the validity of the CAH (see Kohler, 1981), questioning its ability to predict language learning difficulties accurately, as successful language learning depends on various factors that go beyond structural similarities and differences. Moreover, predictions from the CAH may not always be applicable across different learner populations, language pairs, or learning contexts, overlooking the variability among language learners and treating them as homogeneous groups rather than considering their individual differences. Addressing these weaknesses has led to the development of more comprehensive theories and approaches to language learning and teaching. An important episode followed in the shape of Briére‘s *Quantity before Quality in Second Language Composition.* In 1966 it paved the way for a more modern approach by including physical representations of speech samples (phonetics) into the equation. This is an important remark when conducting cross-linguistic investigations. Briére argues that quantity within the input leads to fluency and automaticity and emphasises the role of practice and repetition in consolidating linguistic structures. These topics are still frequently revisited, even to this day. In 1967 Lenneberg and Chomsky hypothesised about a critical period (CP) within which native-like acquisition of a language is possible. According to them, this period lasts until puberty sets in and is conditioned by the higher neurological plasticity of the brain. CP states that native-like performance cannot be obtained by adults by means of passive exposure to linguistic input alone. Some of these strong claims have later been refuted by empirical findings; see, e.g., Flege and other theorists, who can indeed provide empirical evidence of adults acquiring native-like competence.

However, by the end of the decade, all this was not enough to deliver satisfactory answers to the hows and whys of underlying cognitive mechanisms and processes. The discourse shifted towards underlying perceptual phenomena that distinguish L1s and L2s during the 1970s. Miyawaki et al. (1975) explore the l-r distinction Japanese learners of English have to acquire in order to make themselves understood (other publications see e.g. Goto, 1971 in auditory perception by Japanese adults of the l-r contrast or Strange & Dittmann, 1984 on effects of discrimination training on the perception of r vs. l). Another focus was put on the analysis of voice onset times (VOT) of Spanish learners of English by Elman, Diehl & Buchwald (1977). This research is noteworthy because it deals with concrete phonetic phenomena rather than descriptive data on a phonemic level. Empirical work from the 1970s gave rise to the establishment of productive research programs on L2 perception that shape the linguistic discourse to this day.

During the 1980s, research increasingly shifted towards questions concerning the perception of phonetic differences in a foreign language. The focus was on understanding how adults could recognise and learn these elements that differentiate meaning. Prominent figures in this field include Winifred Strange, Catherine Best, and James Flege. On one hand, studies were conducted to examine the learning of entirely unfamiliar sounds (Strange & Dittmann, 1984; Best et al., 1988 -> PAM), while on the other hand, models were developed in an attempt to simulate the learning curve of a foreign language (Flege & Hillenbrandt, 1984 -> SLM). Models that have not lost their relevance to this day include James Emil Flege‘s Speech Learning Model (SLM) and Cathrine Best‘s Perceptual Assimilation Model (PAM). Both models account for the listener's difficulties when perceiving an L2 language. Best‘s PAM focuses on foreign or *naive* sounds, whereas Flege‘s SLM seeks to model the learning procedure of a language. It must be explicitly stated that Emil Flege‘s influence over the past 30 or so years cannot be overestimated. The following pages canvass the nature of the SLM and put it into perspective with regard to other approaches in the field. Current research is all concerned with „exploring new ways of testing theoretical predictions, of evaluating phonetic similarity and of training listeners to perceive and produce non-native sounds“ (Munro & Bohn, 2007: 5). For aspiring linguists, this is exciting news, as opportunities for empirical studies are seemingly endless. The explosive nature of Flege‘s and Best's publications ensures that the state of research and the further development of the mentioned models do not stagnate. Both the SLM and the PAM have undergone updates in the past and remain significant for ongoing scientific discourse (see SLM-r and PAM-L2). With a brief summary of the historical development of the research field, I hope to have provided a context into which my own findings can be integrated.

2.4 What is the role of transfer? The concept of similarity in phonetics

The term *transfer* is a central concept in the two major speech learning models, SLM and PAM. It refers to the application of knowledge or skills from one context or language to another. In the context of language learning, transfer can occur positively where previously acquired knowledge or skills make learning of a new language easier. However, negative transfer is also possible, where interference from the native language impedes the acquisition of the target language. Understanding transfer is crucial for comprehending how learners adapt their existing linguistic knowledge to new language contexts. This abstract yet vital concept will be further elucidated before delving into the details of the aforementioned models.

When talking about the acquisition of phonemes and subsequently whole utterances, one is concerned with concepts of segmental learning. „The task of segmental learning is to divide up the *perceptual space* in a way that corresponds to language specific categories" (Zsiga, 2013: 455). These categories represent the phonetic nature of a sound, which differs from one language to another. For my purposes, this vowel space is defined by the range of the first (F1) and second (F2) formants, which represent the oscillation of our voice in the unit Hertz (Hz). These formants can be used to plot speech samples in a coordinate system. Such a system then illustrates the auditory space of a language learner, both in terms of what can be produced and what can be perceived. What the SLM and other models focus on is how the learner learns about phonetic differences and how implementation at the phonemic level can be promoted. The concept of phonetic similarity is central to both the SLM and the PAM. To make these phonetic differences tangible, empirically grounded models are needed, as well as concrete physical proof such as the visualisation of sounds in a coordinate system, as described earlier. As previously stated, two aspects are of interest here. These are the degree of phonological difference (something that can easily be explored by means of empiric research e.g. by tracking formant values and comparing them) and the learner's perception of such differences. Perception is a dubious area, as it has not yet been conclusively determined how these differences are perceived or processed by the brain (see Zsiga, 2013: 455 ff). What is largely agreed upon in the literature is that L1 exerts the greatest influence on the process of second language acquisition and transfer from L1 to L2 frequently takes place. Here, I adopt Zsiga's metaphor of a *mental* or *cognitive* valley. The sounds learned in the native language are mentally (in some way or another) firmly anchored. New sounds either inevitably slide into an existing valley (we refer to this as L1 transfer to L2) or they remain somewhere in between. This can result in sounds being realised inconsistently or, in the long run, finding their place in the *wrong mental valley*. (Think of it as recognising shapes, e.g., the Apple logo. Just because I recognise the logo does not mean I am able to draw it on a piece of paper. The more complex the shape, the longer it takes to establish a mental representation of it). Speech learning models try to identify the concrete processes that are at work here. It seems plausible that learning to distinguish between newly learned sounds entails a lot of learner-specific factors like age, exposure to L2 or personal aptitude. When conducting research in the field, all studies have to draw a line as to how much individual information will be obtained and taken into account during data interpretation. To many learner-specific factors can render any survey into a never-ending story with little explanatory power. During the 1980s, a reassessment of the significance of transfer, considering the contrast between similarity and dissimilarity, caused an explosion of research that led to the development of numerous perceptual models of second language acquisition. The most prominent models will be briefly introduced.

2.5 Theories in L2 speech learning

2.5.1 The PAM-(L2)

Traditionally, within the field of L2 speech learning, a focus was put on perceptual difficulties that learners may experience when starting to learn a new language. To achieve this, the starting points for a contrastive examination must be clearly defined. Models such as SLM and PAM examine the degree of equality between non-native contrasts. Non-native contrasts vary in difficulty. A study by Best et al. (1988) suggested that native English speakers can easily distinguish between Zulu click consonants and other (pulmonic) consonants. Best characterises this as the quality of a sound to assimilate to the native L1 sound space. As clicks are not speech sounds by English standards, they can be easily detected and assimilation does not take place so that a separate phonetic category (or *valley*) is established. These findings led to the formulation of the PAM in 1993 and 1995 respectively. The PAM puts a focus on non-native speech perception, often using the phrasing of *naive perception* for scenarios where participants are unfamiliar with the language in question. In essence, the PAM puts a focus on monolinguals with no *a priori* contact with another language. It states among other things that a „high degree of similarity between two foreign categories may be associated with patterns of assimilation that lead to poor discrimination“ (Bohn & Munro, 2007: 6). This means that if two sounds are phonetically similar, assimilation to the next best L1 sound can be expected. The PAM is the product of efforts aimed at establishing phonetic differences within a foreign language. This encompasses the qualitative contrasts of two (or more) sounds within a foreign language. However, research in naive sound perception does not nourish promising outcomes when trying to model L2 acquisition. In 2007 Best and Tylor refined the PAM with regard to L2 learning. The PAM-L2 specifically applies its principles to L2 acquisition, focusing on how learners come to perceive and produce L2 sounds based on their L1 phonetic and phonological systems. An important change in PAM-L2 is the focus on phonetic opposition within the target language as a result of L1 and L2 interference. „While the SLM restricts its focus to a single [L2] category, a PAM-L2-informed approach deals with pairs of [target/L2] sounds“ (Sönning, 2012: 10).

Table 1: *Assimilation Patterns according to PAM-L2 (Sönning, 2012:10)*

Pattern	Assimilation	Discrimination
Two Category	Assimilated to different L1 categories	Excellent
Single Category	To same L1 category, equal goodness-of-exemplar ratings	Poor
Category Goodness	To same L1 category, different goodness-of-exemplar ratings	Intermediate
Uncategorized-Categorized	Only one phone is perceived as a speech sound and assimilated	Very good
Uncategorized-Uncategorized	Neither phone is perceived as a speech sound and assimilated	Poor to excellent

The table shows how PAM-L2 categorises assimilation patterns and to what extent perceptional learning may occur. Before going over the table in more detail, here are some basic assumptions that form the foundations of the PAM concept.

„PAM‘s assimilation types describe the possible outcomes of first contact with an unfamiliar phonological system“ (Best & Tyler, 2007: 14). With regard to L2 learning, something that is referred to as intra-language by some (see Sönning, 2012 or Zsiga, 2013) evolves on the way to mastering the sound system of the new language. Very soon after the learning process of a new language commences, it is no longer possible to speak of naive perception. PAM-L2 seeks to outline the changes within this intra-language as it develops towards L2 proficiency. It categorises potential cognitive scenarios into concrete cases of exposure to the L2 and seeks to draw general consequences that allow for universally valid conclusions. With regards to the table in Sönning the following can be stated:

A two-category assimilation poses little difficulty to the learner as categories X and Y correspond to two different categories in L1. (e.g. the opposition of *barn* vs. *born* caused no difficulty for German native speakers as the same contrast is present in German *bahne* vs. *Bohne*). The single-category pattern behaves differently. Two L2 sounds are assimilated to the same L1 sound with no perceivable difference between the two sounds (an example is the *cot-caught* merger which, if applied, can eliminate the opposition between the two vowels). An uncategorised sound is not perceived as a speech sound. A prime example are click sounds, as featured in Best et al. (1988). If one or two sounds are not perceived as forming part of a language, discrimination is fairly easy, as Best et al. suggest in their study (see also Frieda & Nozawa, 2007 where inexperienced listeners were more successful in discriminating naive sounds). PAM-L2 identifies six perceptual assimilation patterns that characterise how learners categorise non-native sounds: identity, category goodness, similarity, contrast, difference, and no response (Best & Tyler, 2007). By analysing these patterns, PAM-L2 provides insights into the perceptual challenges faced by second language learners and the factors that influence their ability to discriminate and acquire new phonetic contrasts. Overall, the model offers insights into the perceptual processes underlying second language acquisition, contributing to a deeper understanding of how learners navigate the complex terrain of phonetic learning.

2.5.2 The SLM

In 1995, Flege formulated the SLM, which seeks to categorise a learner‘s acquisition process of L2 sounds. The SLM‘s main principle is that the „mechanisms and processes used in learning the L1 sound system remain intact over the life span“ (Flege, 1995: 239). Logically, the converse argument to CP then states that native-like performance in an L2 can be achieved by late or adult learners. That is a bold claim that certainly contradicts other prominent assumptions in linguistics to which we will come later (see CPH, Lenneberg & Chomsky, 1967). The SLM further posits that learners undergo a sequence of stages as they acquire new phonetic categories. Initially, learners rely on their native language phonetic categories to perceive and produce sounds in the L2 (Flege, 1995). With increased exposure and practice, learners gradually shift towards forming new, language-specific phonetic categories. The model highlights the influence of age on the ability to attain native-like pronunciation, suggesting that younger learners have a greater potential for achieving native-like proficiency due to a more flexible perceptual system. Also, the SLM emphasises the role of individual variability and the importance of factors such as motivation, aptitude, and language input in shaping a learners' speech development (see Birdsong, 2007).

The SLM tries to model phonetic acquisition over time. It states that phonetic similarities between an L1 and an L2 sound can pose major difficulties for learners because the phonetic space of an L1 phoneme occupies a potential L2 sound category and hence makes it harder to establish new L2 sound categories in a phonetically similar environment. SLM and PAM think of phonetic similarity as the key issue in determining where difficulties for L2 learners actually lie. However, the SLM looks at L2 acquisition over time, whereas the PAM is mainly concerned with unfamiliar perceptual distinctions in an unknown language. When looking at speech development over time, factors like age and exposure to native input are important considerations that have to be included in the equation. Sönning states that „perceptual learning as a result of [L2] experience yields the modification of established L1 categories and the creation of new [L2] phonological representations. L1 and [L2] representations coexist in the same phonological space“ (Sönning, 2012: 8). Once again, this is referred to as a form of intra-language (Zsiga, 2013) and within this space, new categories must be established. The ability to do so depends on successful sound discrimination in the target language. All of this relates to the perceptual abilities of the learner and therefore, Flege coined the term *perceived phonetic dissimilarity.* From this, two central points of his model can be derived. Firstly, the assignment of a sound in the native language to a new sound in the L2, and secondly, the knowledge about the degree of deviation between the L1 sound and the L2 sound. Of course, this knowledge can only exist if noticeable differences between the sounds can be detected. Flege calls it *equivalence classification* if a similar sound blocks the formation of a new sound category. This then means that no new categories can be formed if the difference between the two sounds cannot be perceived. This is of central importance for this work because it theoretically models exactly what has been puzzling the participants of my study. Theoretical knowledge about a phonetic difference between two sounds alone can hardly have an effect without the phonetic-physical parameters being recognised and perceived. (That's exactly why reading and understanding a vowel chart is only half the battle). However, awareness of this fact can be used to identify the difference that was initially imperceptible and once this step is taken, nothing stands in the way of the formation of new sound categories. Flege assumes that perception improvements ultimately lead to production improvements. Neither the SLM nor the PAM, model developments in production, though the SLM and PAM attribute a central role to performance in perception . Hence, researchers have been concerned with the question of whether or not perceptual training can in fact improve the production of L2 speech sounds. At this point, scholars are at odds and present the reader with a variety of results from studies that support either one or the other. A link between perception training and an eventual rectification of production patterns is central to Flege‘s SLM. It states that changes in perception should ultimately lead to changes in production (Bohn & Munro, 2007). These valuable insights fuelled the motivation for the present survey.

2.5.3 The SLM revised

In Flege's revised Speech Learning Model (SLM-r) from 2021, there is a continued exploration of the processes involved in L2 speech acquisition, building upon the concepts established in the original SLM. The SLM-r proposes advancements in research on age-related changes in speech perception and production, as well as the role of individual differences and variability among learners. While the core principles of the original SLM remain, the SLM-r offers a more versatile understanding of the dynamic interaction between cognitive, social, and environmental factors in shaping second language speech development (Flege & Bohn, 2021). Additionally, the SLM-r emphasises the importance of considering a learners' prior linguistic experience and their ability to adapt and reorganise phonetic representations over time. Overall, the SLM-r represents a refinement and extension of Flege's original model, providing a more comprehensive framework for studying the complexities of L2 speech acquisition in different contexts.

2.6 How does age interact with L2 acquisition ?

Now that the latest version of the SLM also mentions the factor of age repeatedly, the question arises as to how age affects the learning process and either promotes or inhibits it. In the literature, there is a significant consensus that late acquisition of a foreign language limits proficiency in that language. Early and late learners typically differ in how they perceive and produce the L2. Perhaps the most noticeable of these differences is a foreign accent that many learners exhibit, but other domains of language can also be affected, such as syntax, morphology or semantics. Flege et al. (1995b, 1999) investigate the relationship between age of onset of learning (AOL) and accent and find that early language acquisition also leads to better active and passive performance. While there seems to be agreement in the literature on this matter, there are also studies that report instances where L2 learners perform on par with native speakers (Bohn & Flege, 1992; Birdsong, 2007).

Such surveys contradict the proposition of the Critical Period Hypothesis (CPH) (Lenneberg & Chomsky, 1967). The CPH posits that there is a specific window of time during which language acquisition is most effective and efficient. This hypothesis suggests that there is an optimal age range, typically in childhood, during which individuals are more adept at acquiring language, especially their first language and potentially subsequent languages. Beyond this critical period, language acquisition becomes more challenging, and individuals may struggle to attain native-like proficiency. The exact duration and timing of the critical period remain topics of debate among researchers, but the general idea is that early exposure to language is crucial for achieving full native-like proficiency in a language. The offsets for learning linguistic structures vary widely, ranging from a few months (Hyltenstam & Abrahamsen, 2003) to around the onset of puberty (Lenneberg, 1967). Scovel, e.g. specifically refers to the completion of the 12th year of life (Scovel, 1988). Diller (1981) considers the critical period to be surpassed by around 6-8 years of age. Biological factors are always a welcome excuse for CPH proponents that lead them to the assumption that „a post pubertal L2 learner will inevitably have a foreign accent“ (Scovel, 1988: 185). Flege in 1988, put up some resistance and spoke of an almost pessimistic approach within the CPH line of thinking calling for a „need for explaining the partial approximation that is observed in L2 speech production“ (Flege, 1988: 363). His assertion is supported by diverse research by countless authors, which has again and again demonstrated that individuals can be successful in learning an L2 with native-like performance in listening and speaking. Armed with this understanding, the question arises as to whether successful acquisition of a foreign language is even possible or what extralinguistic factors enable accent-free language learning in adulthood. The CPH is critical in scholarly discourse because it serves as a counterpoint to the assumptions of authors such as Best and Flege.

2.7 Universal Grammar in L2 phonological theory

In literature, the CPH is often associated with nativist approaches that posit the innate ability of every human to acquire language. This so-called Language Acquisition Device (LAD) came into being through the generative ideas of Noam Chomsky and his postulated Universal Grammar (UG). In the context of L2 phonological theory, Chomsky's UG plays a role in understanding how learners acquire the phonological systems of second languages. UG, which emerged during the late 1950s and early 1960s (see Chomsky, 1957 and 1965) posits that humans are born with an innate linguistic knowledge that shapes the way they acquire and use language. This innate knowledge provides a set of principles and parameters that guide language acquisition across different languages, as stated in the Government and Binding Theory (Chomsky, 1981). Principles are true for all languages, whereas parameters vary cross-linguistically. With regards to L2 learning, questions arise about whether or not adults still have access to UG when learning a new language and whether the speech system of the adult learner makes use of the same UG principles that children use (see Mitchell et al., 2012). Referring back to Zsiga's *cognitive valleys* and the question of how linguistic input is anchored in the brain, UG serves as a very convenient means of interpretation because it assumes an innate ability to process linguistic input from the day we were born. In such a nativist approach to language acquisition we assume that children „based on this universal grammar, [...] construct their L1 grammar by inferring the relevant settings from the input" (Sönning, 2012: 19).

When applied to L2 phonological theory, UG suggests that learners bring their innate linguistic knowledge, including phonological principles, to the process of acquiring a second language's phonological system. While the phonological systems of different languages may vary in their specific phonetic inventories, there are underlying universal principles that govern the organisation and structure of phonological systems across languages (Archibald, 1994). Here, one of UG's main ideas resurfaces. Ungrammaticality can be detected due to its lack in the input. Archibald refers to this as *inductive learning,* which eventually relies on repetition but also takes a CP into account. To gather understanding of these underlying principles, can help educators design effective instructional strategies that align with a learners' innate linguistic predispositions (or in my case design an experiment that appeals to potential participants). One intended purpose of UG may be to explain why certain phonological processes or constraints are common across languages. These processes include syllable structure, stress patterns, phonotactic constraints (permissible sound sequences), and phonemic contrasts (Singleton, 2005). Learners rely on their innate linguistic knowledge, as guided by UG, to analyse and internalise these phonological patterns of the L2. However, UG also allows for flexibility and variation in how these universal principles are instantiated in specific languages. This flexibility accounts for the fact that while languages share common underlying principles, they also exhibit variation in their phonological systems. Therefore, learners must adjust their L1 phonological knowledge to accommodate the specific phonetic and phonological features of the L2. The concept of UG is relevant in the context of L2 acquisition because it suggests that there are underlying linguistic principles shared by all human languages. According to UG theory, humans are born with an innate linguistic knowledge that guides language acquisition. This means that when individuals learn a second language, they draw upon this innate knowledge to internalise the grammar and structure of the new language. However, the question remains whether this LAD is linked to a CP. Furthermore, understanding UG can help explain why certain aspects of language learning may be easier or more difficult for L2 learners. It provides insights into the cognitive processes involved in language acquisition and has the potential to influence language teaching methodologies (Martohardjono & Flynn, 1995). By recognising the universality of certain linguistic principles, educators can tailor their approaches to better facilitate L2 learning and address common challenges faced by learners.

Theories exist that try to model this process of L2 acquisition. The term that must be mentioned in this context is *biological endowment*. Advocates of nativism have quite varied views on how Chomsky's ideas can be implemented in the L2 acquisition context. For instance, Martohardjono and Flynn (1995) argue that non-innate aspects of language learning can indeed be subject to the CPH. However, even after the CP, innate language features remain accessible to the learner. Some nativists, on the other hand (see Bley-Vroman, 1989; Schachter, 1988) take the stance that language learning after puberty can no longer occur with or through UG. Ageing thus denies access to the UG components, and learning new languages is then limited to „general problem solving mechanisms, which are seen as not constrained by maturation“ (Singleton, 2005: 274). The question remains as to which features are innately preconditioned and which ones are not.

In summary, UG provides a theoretical framework for understanding how learners acquire the phonological system of a second language. It suggests that learners rely on innate linguistic principles to analyse and internalise the phonological patterns of the L2, while also accommodating the specific phonetic and phonological features of the target language. If we assume an innate LAD that enables language learning in the first place, the question arises as to why this genetic disposition should disappear with age. Therefore, it is quite sensible to consider generative thoughts from the nativist school in the context of L2 acquisition. The assertion made by Flege that native-like foreign language acquisition is still possible even at an advanced age can be supported by the assumption of an LAD.

3 Methodology

This chapter provides details about methods and data. I provide a summary of the materials, procedures, and participants utilised to collect the data. The final paragraph gives an outline of key aspects of the statistical analysis that was used for the data in question. I considered both the collection of production and perception pre-test data as vital. Little research that examines the phenomenon of English open-back vowels in a setting with German speakers in detail has been conducted. Therefore, I wanted to find out if the Germans in fact performed as I expected them to. As other data (regarding these specific vowels) was scarcely available, I relied on collecting my own pre-test data, which then formed the starting point and acted as a reference for the following post-tests. Subsequently, I wanted to test if exposure to the phenomena could have an impact on their spoken performance after all. Here I refer back to my hypotheses, which I mentioned during my introduction. Creating an online study to collect such language-related data must strike a balance between feasibility and effectiveness to keep the participants happy over the course of the survey. I will further elaborate on what this means in the discussion section.

3.1 Participants

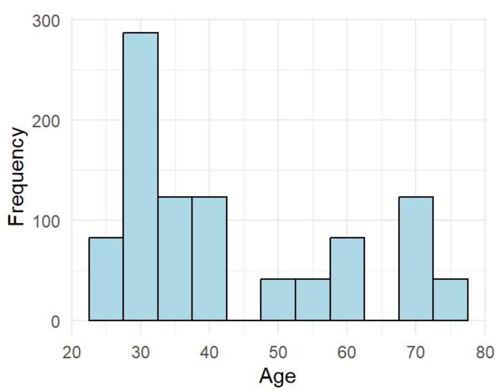
Data collection for this thesis took place via an online survey. No payments or other disbursements were involved. Candidates participated on voluntary terms which for one thing was a necessity but for another gave reason to believe that participation might have been considered out of a personal interest in the matter. Hence, the group of participants was made up of friends and family members. The experiment was further distributed amongst students of English at the University of Cologne via the local student association. A remote collection of data implied, of course, that participants had to be trusted to perform their tasks to the best of their knowledge and belief. Instructions were given as detailed as possible to avoid unnecessary confusion. Also, the study needed a simple structure with minimal allowance for potential misinterpretation. Therefore, participants were solely asked to read out specific words during production testing and choose between two possible words during the perception task. A simple set-up guaranteed smooth operation with as many usable datasets as possible. Judging by the received data, it was eventually decided that a total of 23 datasets, with 9 male participants and 14 female participants, were in fact usable. Participants did not have the possibility of asking questions during the experiment. Because the incentives for participation were pretty low (to say the least) each part of the experiment did not take longer than 10 minutes to complete, to make it more appealing to the audience. It took participants approximately 60 minutes to complete the experiment in all its parts. Participants accessed the study via a computer or mobile device, like a smart phone or tablet. This, of course, meant varying qualities of the produced sound files. Also, background noise posed a considerable difficulty, which, unfortunately, rendered some of the datasets unusable. Also, time seemed to have been a factor, as some participants decided to leave the survey halfway through, despite the fact that they were producing usable data during the first rounds of the study. 43 participants started the first round of the survey, but only 23 people finished the entire experiment.

A three-step approach over three days for the different parts was aimed for. By and large, participation mainly took place over the weekend, so the given time frame of one week was more or less satisfied. Participants were asked for the following background information:

- age
- sex
- whether or not they speak a dialect
- how many additional languages they speak
- how often they hear English of native speakers
- how often they use English in general

This information was obtained to get an idea of the demographics of the participants and to further deduce influences on their performance during the testing. The age range extends from 25-73 years, which is due to the fact that participants were recruited from close family, friends and students at the University of Cologne, respectively. My participants were 43 years old on average. The median was 37 years, with an SD of 15,4 years. The frequency represents the number of speech tokens.

Table 2: *Age Distribution*



Contrary to most research in the field, the group of participants in this survey has a rather wide age range and is not limited to a specific age or professional group. The sample nicely represents a random sample of the population. 9 participants were men and 14 participants were women. 5 participants said they were able to speak a (German) dialect (4 women, 1 man). 50% of female participants spoke no additional languages. One spoke 4 additional languages. One spoke 3 additional languages and three spoke one additional language. 4 male participants stated they speak no additional languages. Two participants spoke two additional languages and three male participants spoke one additional language. Participants were also asked in which contexts they spoke English and in what situations they listened to English native speakers talk. When asked about their speaking performance, 4 answers were possible: never, once or twice a week, every day and on holidays. When asked about confrontation with native English, they could answer that they hear it rarely, once or twice a week, every day, rarely but reading English texts happens on a regular basis. The following table shows the answers given.

Table 2: *Results questionnaire*

Sex	speaking				listening			
	never	On holidays	1-2 per week	daily	rarely	Rarely, but reading texts regularly	1-2 per week	daily
female	1	7	1	5	4	4	2	4
male	1	2	3	3	4	0	2	3

3.2 Native speakers

In order to make comparisons between participants and native English speech samples, two native speakers of a British variety that can be considered close to RP were asked to read out the minimal pairs from the afore-mentioned lists. One was male and one was female. Both had a background in academia, with a masters and a PhD, respectively. Both grew up in south east England and for the entirety of their life's have never lived outside the UK. The native speakers were 33 and 47 years old, respectively. However, upon request, the female native speaker stated that she was in frequent contact with speakers of GA varieties of English through her work. Unfortunately, this fact is mirrored in some instances, where she produced the vowels in question with a bit of an American twang. This was a situation that was not immediately noticeable and was only identified later on in the evaluation process (discussed further in the Discussion section). The native speaker's data was recorded by means of their smartphones. They chose a quiet environment with no interfering sounds. Audio samples were saved as .wav files to guarantee a good enough quality for a later Praat analysis. The native speakers read every entry from the word list once (see Appendix). As a result, I had an extensive list of audio files from which I could choose during the design process of my experiment. Not every speech sample provided by the native speakers was used in the survey.

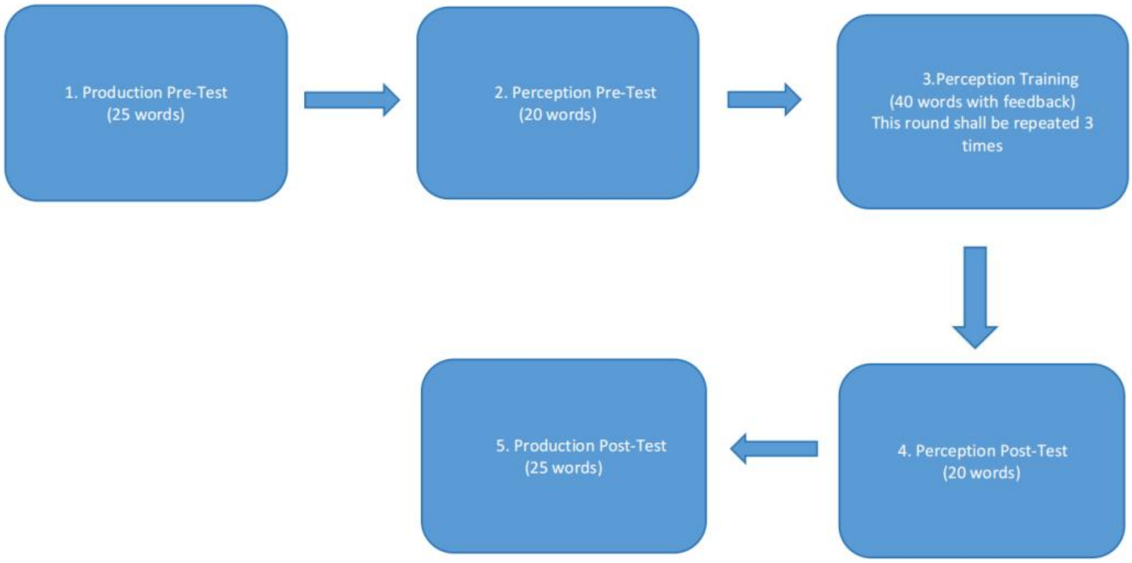
3.3 Experiment design

Now, in order to investigate the perception and production of British English back vowels, I set up an experiment to test those. As the biggest problem for German English learners is the apparent miscategorisation of certain English back vowels in a minimally contrastive context, the first step was to come up with as many minimal pairs that contain the vowels in question as possible. These minimal pairs had the following pattern:

cot; containing the /ɒ/ sound , caught; containing /ɔ:/, cut; containing the /ʌ/ sound; and cart; containing the /a:/ sound

The appendix provides a list of all the minimal pairs that were featured throughout the experiment. The overall idea was to present these minimal pairs to participants in an appropriate manner through production and perception tasks. Accordingly, the study consisted of 5 different parts: a production pre-test, a perception pre-test, three perception training rounds, a perception post-test and a production post-test. Participants were asked to perform the different parts over the course of three days.

***Figure 3**: This set-up gives an overview of the experimental composition*



I recorded participants' answers at pre-/post-tests and during training, however, only the pre- and post-test data was effectively analysed for the current work. Overall, I expected to see improvements in post-tests in perception and production compared to the pre-tests.

Participants were contacted by giving them the URL to my experiment. It started with the afore-mentioned background questionnaire. The experiment itself began with a production pre-test, during which participants read out given words once. To get a nearly untainted impression of my participants, the experiment began with the production task. This formed the basis for a later, empiric analysis. The words appeared in a randomised order. On each trial, participants said the word displayed in the middle of the screen, in black ink, and had to read it. The recording was initiated as soon as the word appeared on the screen and the time-out was unlimited. By hitting the space button they continued to the next word. Participants were not able to repeat their recordings. However, the time they were given to produce an utterance was unlimited, allowing participants to repeat the word in question if desired. In such a case, I chose the last token to be part of the analysis.

3.4 Stimuli

The task's purpose was to confront the reader and listener with English back vowels spoken by native British English speakers. The objective of the word list was then to have a basis from which to choose. The words in the experiment were chosen because of their simplicity. Ideally, monosyllabic words in the pattern of C_V_C (consonant_vowel_consonant) were included. I used Becker et al. *Green Line Oberstufe* as a basis for choosing familiar vocabulary. Becker 2021 is a school book that is nowadays frequently used in German classrooms. Some rather exotic words found their way into the survey as completions to the arrays of the previously mentioned minimal pairs (e.g. pot, part and, rather unusual putt).

Round 1 started with 25 words; 5 words contained the phoneme /a:/, and five contained the phoneme /ɔ:/. Another 8 words contained the /ɒ/ vowel and the remaining 7 words had /ʌ/. The same distributions applied to the final production test. During the two testing rounds, biases were attributed to the phonemes /ɒ/ and /ʌ/ as these do not form part of the German standard vowel inventory and therefore were of particular interest in the context of this survey.

Round 2 was a perception pre-test during which participants heard a word and had to decide between two given answers, which were minimal pairs. This round consisted of 20 words. The sound was presented to the participants once after an onset of three seconds on each slide, without a time limit. The pre-test data was collected to determine whether or not Germans (in general) were in fact immune to the subtle differences between the English and German vowels in question. Each round included 4 samples of the /a:/ and /ɔ:/ sounds, respectively and 6 samples of the /ɒ/ and 6 of the /ʌ/ sounds, respectively, as problems were expected with regard to this contrast.

Round 2 was followed by a perception training round during which participants heard a word and had to decide between two given answers. The answers were again minimal pairs. On the next slide, they saw the correct answer in the form of written feedback. Every training session consisted of 40 words. Here, the participants were confronted with their potential wrongdoing. In doing so, I hoped to raise awareness for the different vowel qualities of English and the phonemic character of the vowels in question. As the participants were presented with minimal pairs, their awareness of the vowel qualities in question was maximised. Participants went through three sessions of training with feedback before proceeding to the final post-tests (without feedback). The order of stimuli in training was random. The training rounds were not subjected to a later analysis and merely served as a means of training to familiarise my participants with the phenomena in question.

To conclude the study, a perception and production test were implemented. The participants heard a word and had to decide between two given answers (minimal pairs) which word they heard, similar to the perception pre-test. The perception task consisted of 20 words and the following production part consisted of 25 words, respectively, with the same distribution of vowel qualities as initially stated. Here, participants were asked one more time to read out the given words, which were then recorded. Expectations were that after completing the perception training, participants would be more accurate in choosing the right answer and producing more appropriate sounds when reading out the presented material. All featured words formed at least one minimal pair (some came in groups of three or four). Rounds 1 and 2 should contain rather simple vocabulary. Rounds 3 through 5 further contained also lesser-known words to check if a learning process was in fact taking place.

3.5 German data collection

To draw a clearer picture of the phonetic discrepancies between the two languages, a miniature data collection of German speech samples was also conducted. 3 male and 3 female German native speakers were asked to produce the words *Sohle, Soll, Kahn* and *kann* and record their performances. These words contain the German phonemes /oː/, /ɔ/, /aː/ and /a/. Using Praat, the vowel qualities of these samples were extracted and used as a means of comparison. This was an important step in order to fully grasp the phonetic background features of the vowels in question. The collected data constituted a decent point of reference as it mirrored the phonetic reality of a German native speaker, which has an impact on speech performance in English. (see SLM and other models). The speech samples came from speakers of a standard variety of German which is generally referred to as *Hochdeutsch* or High German. The speakers were between 29 and 68 years old. As a native speaker of German, I considered differences in regional variation among those speakers as marginal and therefore the sample was representative. The recordings were done using a mobile device and were sent to me via email. The analysed files had the .wav format. The following graphs show the distribution of the German phonemes /oː/, /ɔ/, /aː/ and /a/ and their British counterparts within their corresponding vowel spaces. The graphs were later used to contrast the German and English phoneme systems on a phonetic or empirical basis. (The German native speakers who provided the sound files for these graphs also participated in my experiment). Note the inverted x and y axes according to linguistic convention and the separation by gender.

Figure 4: *Contrastive vowel chart English - German male*

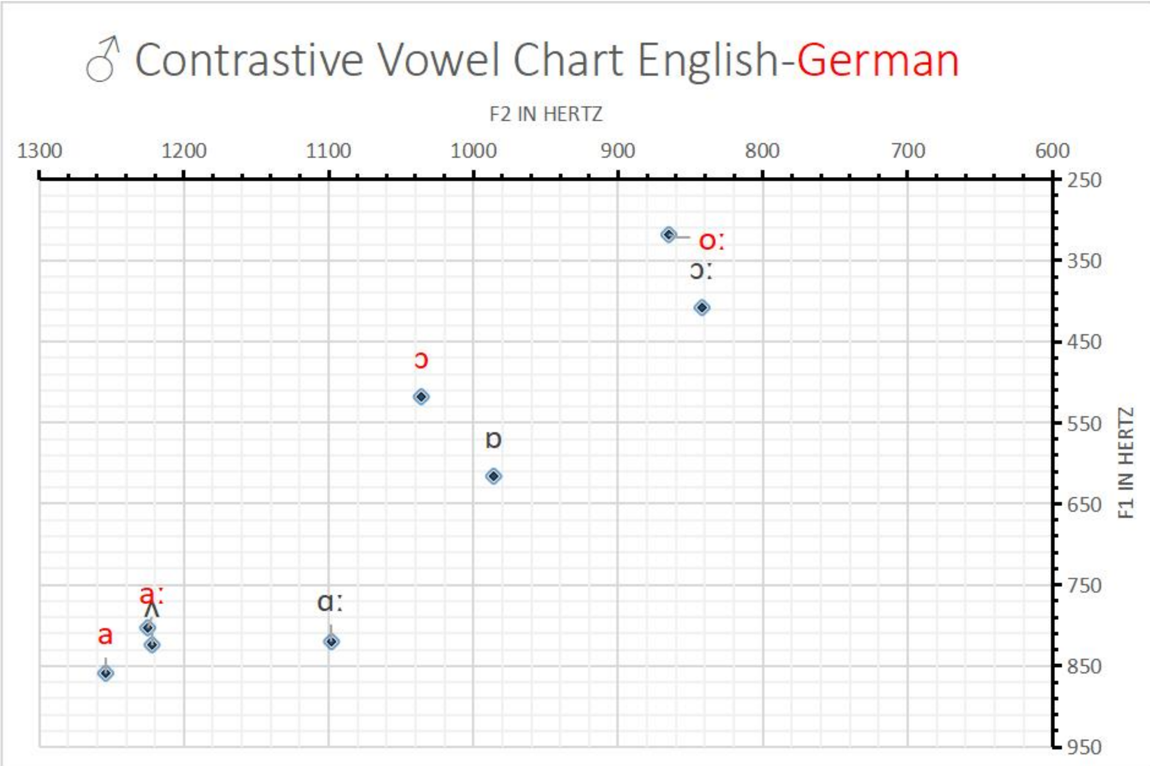


Figure 5: *Contrastive vowel chart English - German female*

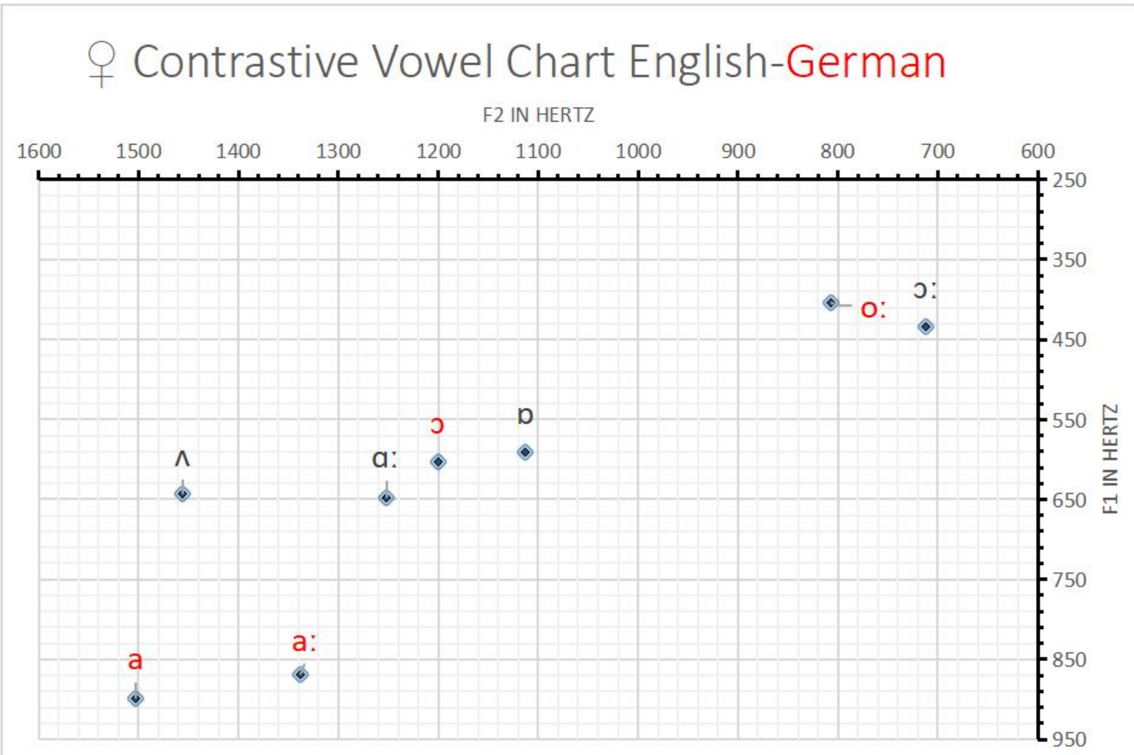
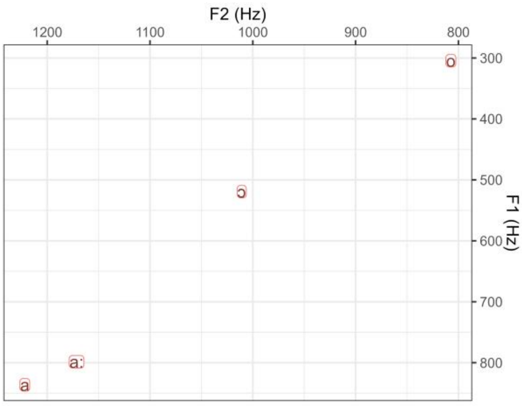
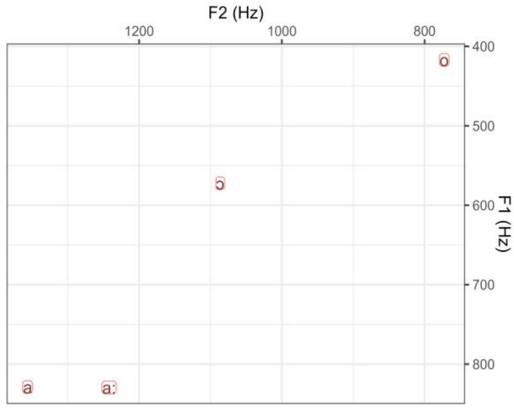


Figure 6: *German formant values*



1. German Male NS



2. German Female NS

3.6 Data collection

The experiment data was collected using E-Babylab. E-Babylab is an open-source browser-based tool for unmoderated online developmental studies. I was kindly granted access to this tool by the staff of UiO (University of Oslo). As the main concern was the collection of speech samples and the presentation of audio input, most of the very elaborate features of E-Babylab were not made use of. E-Babylab provides a user-friendly graphical interface for creating and managing studies, users, participant data, and stimulus material, all without the need for programming expertise. It supports a wide range of audiovisual media for stimulus presentation, along with further measures such as webcam recordings, audio captures, key presses, mouse-click/touch positions, and reaction times (Lo et al., 2023). For this experiment, audio stimuli needed to be presented and key presses as well as mouse positions needed to be tracked. During all the tests, reaction times were also recorded, but they were not analysed in the current work. As E-Babylab is a browser-based tool, the designed study could be shared directly by distributing the URL among the participants. In E-Babylab, each round of the experiment was designed separately. The first part contained slides with general information on the experiment and collected some personal data, as explained above. Participants were given the possibility to exit and/or resume the experiment as they pleased.

In order to distribute the URLs, a Jimdo web page was created, explaining the purpose of the study to participants and neatly presenting the URLs to them. The website can be accessed under the following domain: https://fabian-neumann-1.jimdosite.com

The written German text will also be featured under the appendices. The survey was conducted in German. All information and instructions were given in German. Participants consented to audio files and personal data being used in this study by starting the survey.

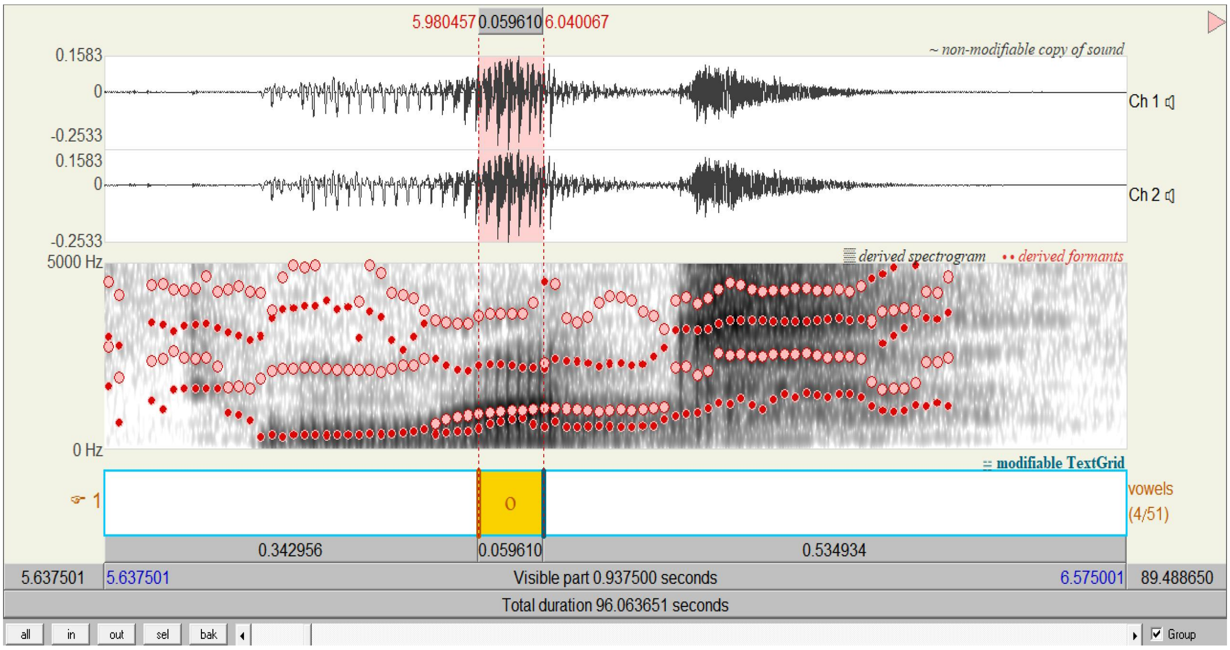
As each part of the experiment constituted an independent entity under a separate URL, a system of codes needed to be created in order to allocate the received data later on. Each participant was asked to identify with a code of the following kind: The first two letters of their first name followed by the first two letters of their place of birth, followed by their year of birth. This code system turned the data management into a much more pleasant experience and simultaneously ensured that the survey remained anonymous (at least for the people I did not personally know).

Before the distribution of the experiment commenced, it had been approved by the Norwegian Agency for Shared Services in Education and Research (sikt.no) (meldeskjema 643739).

3.7 Data processing

The most important programs that were used in data processing were R-Studio, Excel and Praat. The data obtained from E-Babylab required a thorough refactoring. Excel sheets needed to be reformated and audio files needed to be converted into a format accepted by Praat. Babylab provided audio files in a .webm file format. With the help of the VLC Player, the sound files were converted into a .wav file for further processing. Excel sheets were handled by an R-script that filtered relevant data and restructured them so they would be easier to work with. The data of each participant was stored under the code initials+place+year. In order to effortlessly run the data through the corresponding R scripts, a fixed system of storage logistics was essential. The preprocessed data was analysed using Praat. First, an annotation of the spoken data was prepared. To achieve this, all sound files were manually processed. This means that every vowel was extracted by hand to ensure an accurate representation of the sound in question. On and off sets were identified using a spectrogram to find the most stable part of the word. Simultaneously, this allowed for the checking of the formants of each vowel (they should look smooth and even). A text grid was created and the vowels were annotated using *oo* for /ɔː/, *o* for /ɒ/, *aa* for /aː/ and *a* for /ʌ/. Later on, an R-script assigned IPA symbols to this code.

Figure 7: Segmentation of a vowel using Praat



The next step was to create a formant object in Praat based on the sounds at hand. In order to do this, a script was written that scrolled through the different intervals and by coming across an interval that had a label (o,oo,a,aa) cross-checked the same time points in the format object and reported them back in a table. The script needed to know what sound to use and how many time points to extract before going to create a table with some column names about the time, the index and the first three formats (only data for F1 and F2 were used). The script then worked through the Praat text grid, through each interval and for each interval that did not have an empty label found the time point for the beginning and end of that vowel. By the end, the script terminated itself in a loop. By this means, a table of formants was generated using Praat's Burg algorithm, which analysed the spectrum of the afore-marked vowel segments. This algorithm was able to analyse the manually input vowel segments and divide them into 10 time intervals. (For drawing the vowel charts and data analysis, eventually only the 5th time interval was used.) To enable a more accurate analysis, time intervals at the beginning or end of the vowel were selectively omitted. All vowels should have the same articulation at the beginning and at the end to draw a clearer picture of each vowel's phonetic nature without any interfering sounds before or after. The created table was later sent into a Praat script that generated a .csv file for further processing in R-Studio and Excel. The algorithm extracted three formants from the speech data (but only F1 and F2 were of interest during the course of my study). It also provided the absolute time every participant needed to read out the presented words and the time for each individual word (divided into ten time intervals). Unfortunately, the corresponding words for each vowel had to be filled in manually for every instance. A line of the finished dataset looked like this:

Table 3: *Visualisation of the data set*

vowel	word	time_ index	v_time	time_abs	F1	F2	MD	Pre_Post _Test	Participant
o	long	5	0,027	3,139	495	859	2.1	pre	XXYYZZZZ

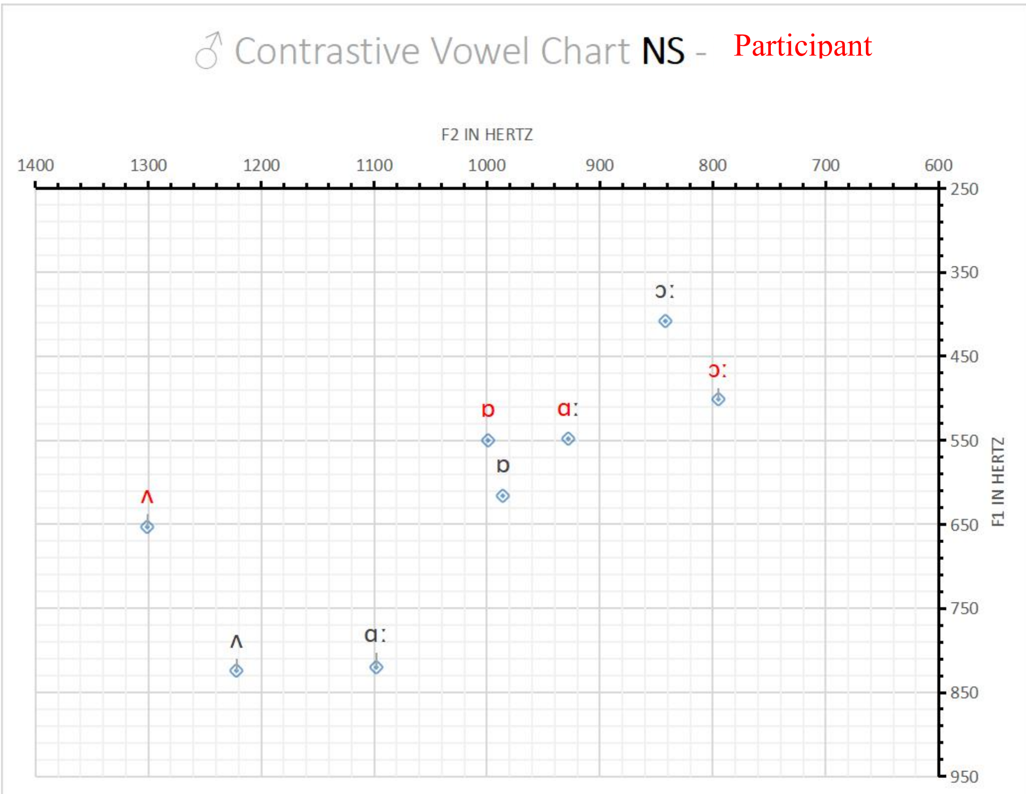
Sex	Age	Dialect	additional languages	frequency spoken	frequency listen to NS
male	68	nein	2	auf Reisen	selten

Annotations were made by manually checking each word for an ideal range of the corresponding vowel (see figure 7). The formant settings needed to be kept in mind. A formant ceiling of 5000 Hz for men and 5500 Hz for women was adapted (Schweinberger, 2016). The program displayed 5 formants with a dynamic range of 30 dB and a window length of 0.04 seconds. The same settings were also applied to the Burg algorithm for formant extraction later on in the process. By checking each sound file manually, I could make sure that the vowels were featured as purely as possible. All Praat scripts and corresponding formant tables were stored and available upon request. An article by Martin Schweinberger (2016) served as a blueprint for my approach to creating the vowel charts. I followed his model for analysing my data and creating the vowel charts.

R-studio was used to generate a personal vowel space of each participant by plotting the first formant against the second (x= F1, y=F2). The overall idea was that „relative distances between distributions of vowel tokens [...] can then be characterised within and across languages“ (Strange, 2007: 39). In my case, it was German and English. I chose to stick to the Hertz voice frequencies. Other units may be Barks, Log Hz or Mels (see Strange, 2007). I compared female participants to a female native speaker, and male participants to a male native speaker of British English. This I did to counteract something that is referred to as the *speaker normalisation problem* in the literature (see e.g. Bradlow, 1995 and Strange, 2007). The problem lay in the simple fact that spectral structures can deviate due to the different nature of each individual's vocal tract. In an environment where sounds lie phonetically close to each other, this could have had problematic consequences.

The original Praat script was able to track formant values and draw vowels on a graph. I averaged over them and the result can be seen as a static point within the corresponding vowel space. This ensured that preceding and following sounds interfered less with the vowels in focus. The script, therefore, averaged over the vowel values and gave an average visualisation of the vowel sound in question. The same was done with the native speakers for both English and German. I kept female and male audio data apart at all times (to at least counteract the *speaker normalisation problem* in broad strokes). For my purposes, I decided to narrow the formant values down to the 5th time point of the segmented section for the quantitative testing that followed. The final product was a vowel space of the following kind, produced in Excel:

Figure 8*: Example of a vowel chart in Excel*



The red colour points to the English tokens my German participant produced and the black colour stems from my native speakers, either male or female. This was done for each participant individually, also because quite a substantial number of participants asked for individual feedback (and after going through my rather time consuming study without any form of compensation I was happy to give back a little graph in return).

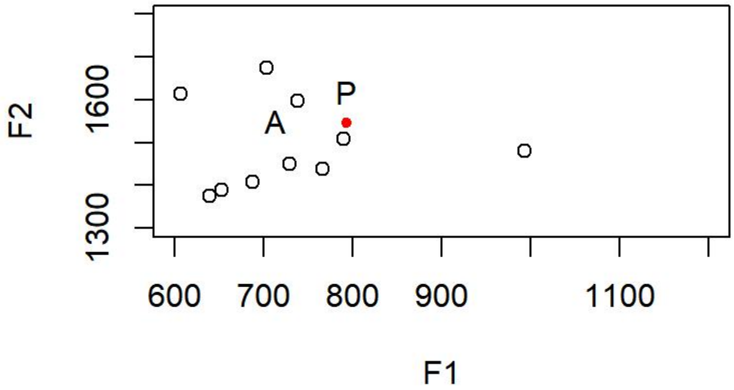
This approach made for a nice case-by-case visualisation of native data against learner samples. However, as Strange remarks „statistical comparisons of mean differences will establish the reliability of differences in central tendencies within and across languages, but will not tell us much about the similarities and differences in the distributions of vowel categories across languages“ (Strange, 2007: 45). The individual tendencies of every participant were portrayed here, but to make statistically relevant statements, a different approach should be employed. In the following section, I will outline the statistical methods chosen to analyse the data and explain why I chose them. My study aims not only to describe the obtained data but also to evaluate it quantitatively and make it accessible, because only then can general conclusions be drawn about German learners of English. A simple comparison of average values is not appropriate at this point. I will attempt to explain why this is the case in the following section.

3.8 Quantitative methods

3.8.1 Production

The data relevant to the investigation ultimately focused on the two formants (F1 and F2) as my dependent variables, as well as the number of correct and incorrect answers during the perception tests. All additional information served to support the results or was added to the dataset through statistical applications in the later stages of the analysis (e.g. MD). My main research question was whether or not online perception training could improve my participants' performance in both reception and expression. As a result, I conducted a production and perception pre- and post-test. In order to receive usable insight into the participants‘ production of the target vowels, I measured the Mahalanobis Distances (MDs) between each uttered vowel token and the distribution of the same vowel uttered by a male or a female native speaker. The MD is a „unitless, scale-invariant measure of the distance between a point and a distribution that takes into account the distance, measured in standard deviations, along the principal component axes of the distribution“ (Kartushina et al., 2015: 820). The following graph illustrates this issue. Point P represents a sound produced by a participant (let us take the word *cut* with the short /ʌ/ as a concrete example). Distribution A now shows all instances in which a native speaker produced the short *a* sound. Male and female data is treated differently, to the extent that the two sexes are never mixed during any calculations with the *speaker normalisation problem* in mind. The following chart draws on data from a female participant.

Figure 9: *Visualisation of a vowel token against the NS distribution of that same vowel*



By calculating the MD for pre- and post-test data, I was able to deduct changes in the pre- and post-distances by taking into account both F1 and F2. A reduced distance in the post-test compared to the pre-test data indicated improved production performance. A larger distance in the post-test data suggested an unsuccessful completion of the training rounds and hence no positive learning outcome for the participant. This approach harbours the advantage that it „allows the natural variability of native speech production to be taken into consideration when assessing the accuracy of production“ (Kartushina et al., 2015: 820). My hypotheses are tested by means of a linear mixed-effects model. The use of such a model allows for the consideration of both fixed effects (such as session, vowel, and sex) and random effects (such as individual differences between participants) in the analysis, providing a more comprehensive understanding of all the factors influencing speech production performance. Additionally, the Type II Wald chi-square tests used in the Anova function from the R- *car* package were chosen for assessing the significance of fixed effects in mixed-effects models, especially in a case like this where the sample size was quite big.

3.8.2 Perception

With regard to the perception tasks, I looked at the vowel contrasts, which seemed to pose the biggest struggle for the participants. During the course of the study, participants were presented with minimal pairs and they had to choose which word they heard. This resulted in a simple correct/false opposition, in which every token received either a correct or a wrong answer. From a descriptive point of view, the average error rate can provide insights into the participants‘ performances. A lower error rate in post-perception indicates an improved perception process after the completed training. However, there are doubts about whether the perception exercises were conducted seriously by one participant. One exceptionally high error rate suggested that the remotely conducted study may not have been approached with the necessary seriousness. This is an issue that will be discussed in more detail in the discussion section.

For the perception data, a generalised linear mixed-effects model (glmer) is fitted using the *glmer* function from the *lme4* package in R. This model analyses the effect of session, vowel, sex and their interactions on the binary response variable *Answer* (correct/incorrect), with random intercepts for subject. Similar to the speech production analysis, a mixed-effects model is chosen to account for the structure of the data, where multiple measurements are taken from the same participants across different vowels and sessions. This allows for the consideration of both fixed effects and random effects. The significance tests (Type II Wald chi-square tests) provide information about the significance of each fixed effect in the model (both in perception and production).

I further wanted to test if my assumptions with regard to perception's influence on production are significant. A linear mixed-effects model was also used to analyse the relationship between perception and production. Hence, the last model includes fixed effects for perception, type of vowel and a random effect for *Subject* to account for individual variability among participants. The dependent variable is the average MD, representing production performance per participant and vowel. *Subject* in the context of my analysis refers to the random effect included in the linear mixed-effects model to account for individual differences between participants. The linear mixed-effects model allows for the estimation of fixed effects, which indicate the relationship between the predictor variables (perception in this case) and the outcome variable (production). A linear mixed-effects model is a sensible choice for the analysis because it appropriately models the nested structure of the data, allows for the inclusion of fixed and random effects, and provides estimates of the relationships of interest while accounting for individual variability and dependencies within the data. This approach ensures that the statistical inferences are valid and interpretable within the context of my study design. With these explanations in mind, I want to provide the relevant data and interpret my results in the next section.

4 Results

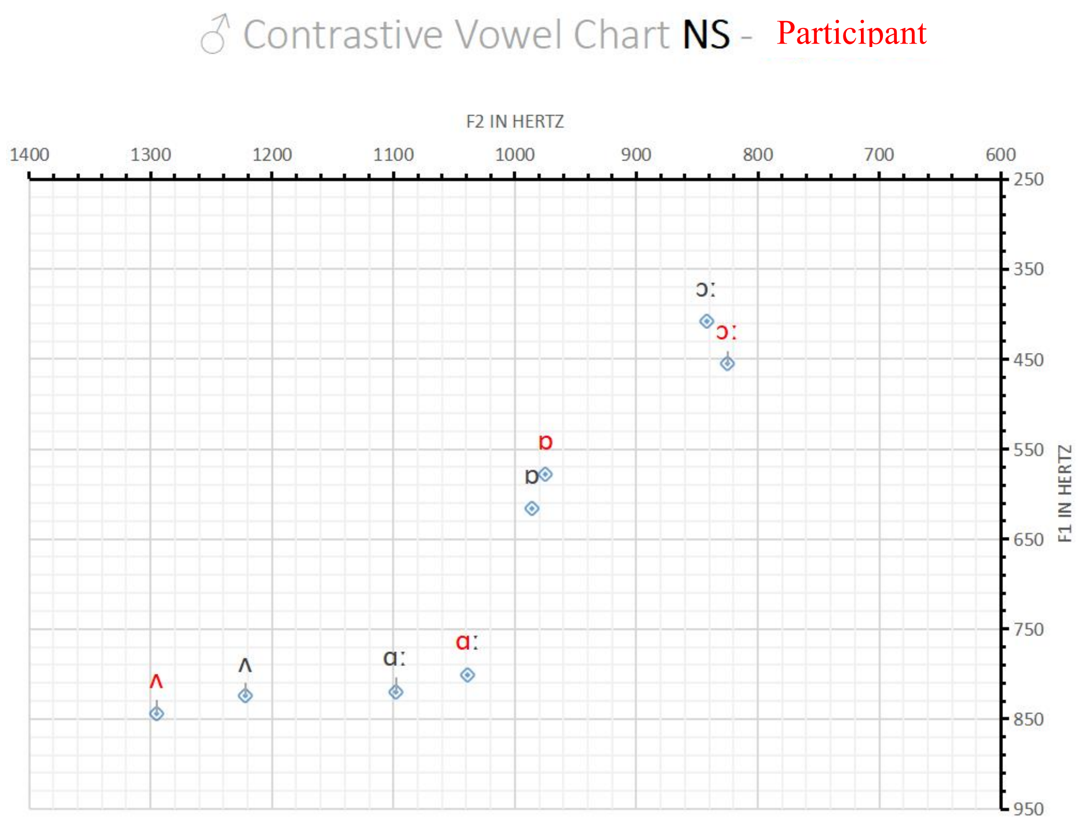
Describing the results of my study occurs in two steps. Firstly, the descriptive results of the performance and perception tests need to be presented. Secondly, I include the statistical analysis of both datasets and check for a significant outcome of my training. I attempt to elucidate emerging patterns by exemplifying them with data specific to some of my participants. Additionally, factors such as proficiency in a (German) dialect or the frequency of exposure to spoken and heard English are taken into account. The goal is to paint as comprehensive a picture as possible, depicting my German native speakers' realisations on one side and my native English speakers' reference measures on the other. The individual vowel charts depicted here represent average values. These are sufficient for an initial visualisation of the corresponding results. I kept male and female data separate at all times.

With regard to the production data, four different visual patterns arose when plotting the results in a vowel chart. I will provide examples for every pattern I identified in the form of individual vowel charts from participants that produced the pattern in question particularly well. The data I chose for visualisation purposes here are from the post-production rounds and therefore represent active performances after the survey had been completed by the participants. As a reference point for comparing the two languages, my male and female native speakers served as benchmarks. The data was plotted against each other so that visual distinctions could be made using the F1 and F2.

4.1 Native-like performance

The first scenario is pretty straightforward, as it describes an instance in which a participant performed on par with the respective native speaker. Please keep in mind that differences in voice height may cause deviations. The distance between the corresponding vowels and the realisation of four very distinct vowel tokens for every sound are important aspects. Out of 23 participants, I consider 6 to have a distinct four-way distribution of vowel tokens. The assessment of these findings is based on the presented vowel charts, so it is done in a visual and purely qualitative manner. The post-graph of a male participant shows an equal distribution of the four phonemes in question according to a native British English pattern.

Figure 10 : Vowel chart showing participants data against NS realisations

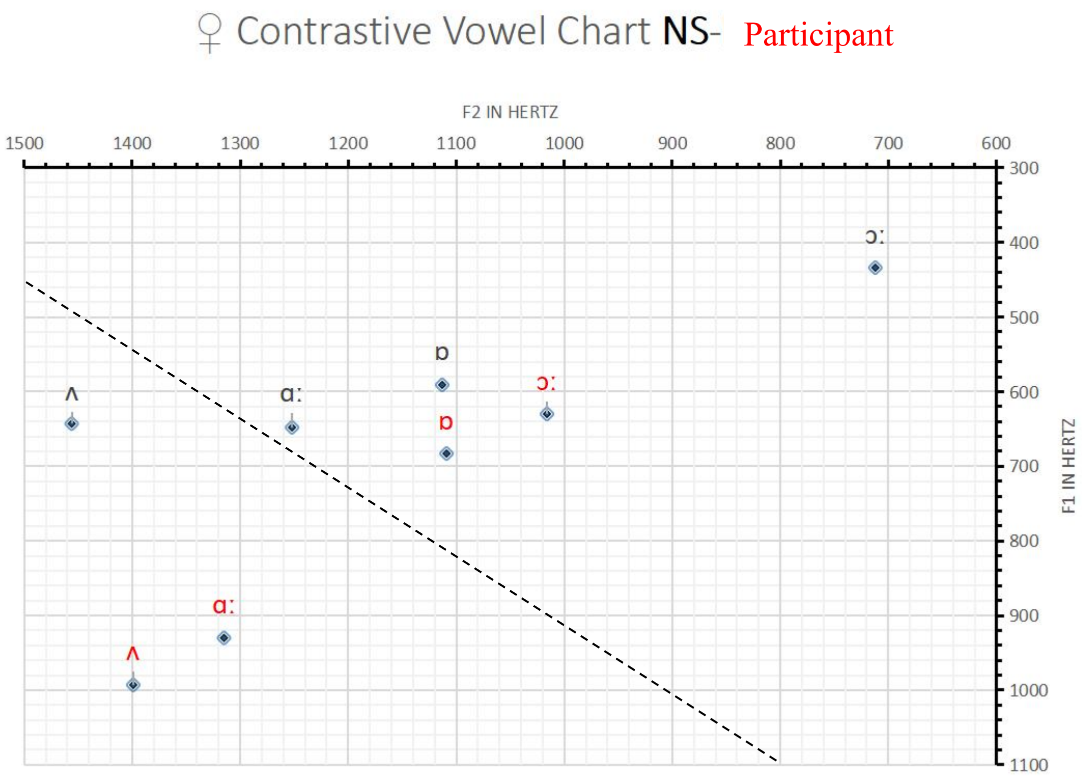


As the following examples suggest, this four-way distinction between the corresponding English phonemes is not self-evident. What is described in the first chapter of this thesis resurfaces in my participant data as well. Several vowel mergers reappear over the course of the experiment. However, I have no means of telling whether this is because of a deliberate choice to not sound British in the first place or because of transfer mechanisms from L1 into L2. Referring back to my thoughts from the introduction section, the data delivers ambiguous outcomes since the variation in production among my participants is quite high (as the following graphs suggest).

4.2 The cot-caught merger

First of all, it must be clearly stated that no participant showed signs of a completed merger in his or her speech. The phenomena I refer to are tendencies, which originate from the fact that I used the average values of the first two formants to create my graphs. The following chart shows the clustering of /ɒ/ and /ɔː/ halfway between the native realisations of the same sounds. Here we see a visualisation of the aforementioned *cot-caught* merger. Simultaneously, we observe a clustering of /ʌ/ and /ɑː/ somewhere between the corresponding native sounds, which supports my initial assumption that the integration of the English phoneme /ʌ/ may pose difficulties for German native speakers. I noticed such cluster formation in 6 out of 23 participants. What we find in this and similar examples is the contrast in German between phonemes /a/ and /a:/, which are phonetically very similar and primarily differ by quantitative features, as well as the *long and short o* of German /ɔ/ and /o:/, which can be quickly confused due to differing quantitative features (German only has a short /ɔ/, but has a long /o:/ instead). Given the background just described, the assumption that L1 transfer plays a role in segmenting the vowel space in this case seems like a reasonable proposition.

Figure 11 : *Vowel chart showing participants data against NS realisations*

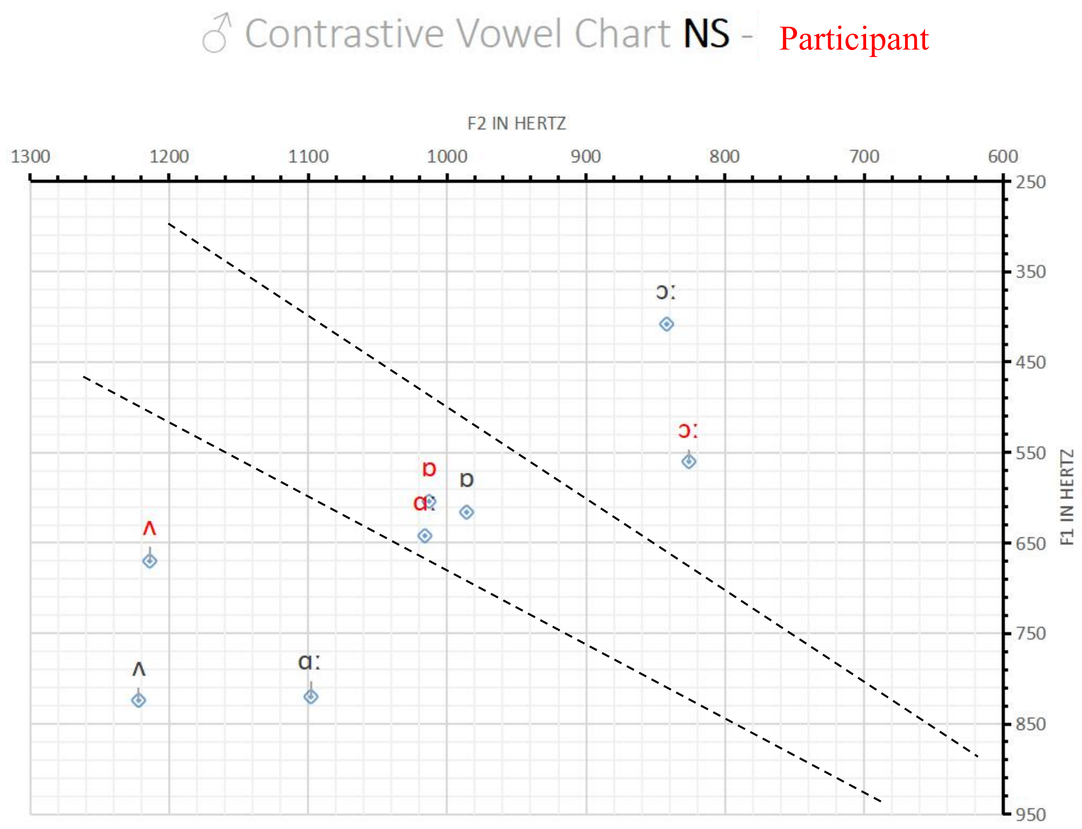


The graph can be neatly divided into two halves, with one half containing the long and short *o* in the upper right corner (thus showing the *cot-caught* merger), and the *a*'s in the lower left corner, respectively. I would like to preemptively note that this contrast caused quite a bit of confusion during the perception tests as well.

4.3 The father-bother merger

In the next example, we come across the *father-bother* merger. This phenomenon of the English language not only causes headaches for L2 learners. As previously mentioned, this merger is a feature of American English as well. For the occurrence of this pattern in the context of my study, I would like to propose two hypotheses once again. The first assumption relates to the potential active use of a GA-oriented pronunciation. In this case, one might even assume an active use of the merger. Scenario number two may represent the product of two factors, which I will outline below: The transfer from L1 becomes evident and struggles with the phonemes /ɑ:/ and /ɒ/ are bound to occur. If we consider a classic vowel chart, /ɑ/ and /ɒ/ form a pair distinguished by the rounding of the lips, where the rounding of the lips is absent entirely or partially in a variant of GA. This circumstance would become clearer if the third formant were also included in the investigation, in addition to F1 and F2. With the existing two formants, distinguishing between the sounds is only possible to a limited extent. In the vowel chart, we see how /ɑ:/ and /ɒ/ meet halfway between their native counterparts. The overlap of the two vowels was observed in 8 out of 23 cases. During perception rounds, the *bodge-barge* contrast caused difficulties for almost all my participants. In some instances, the two vowels were even interchanged and ended up on opposite sides of their native counterpart.

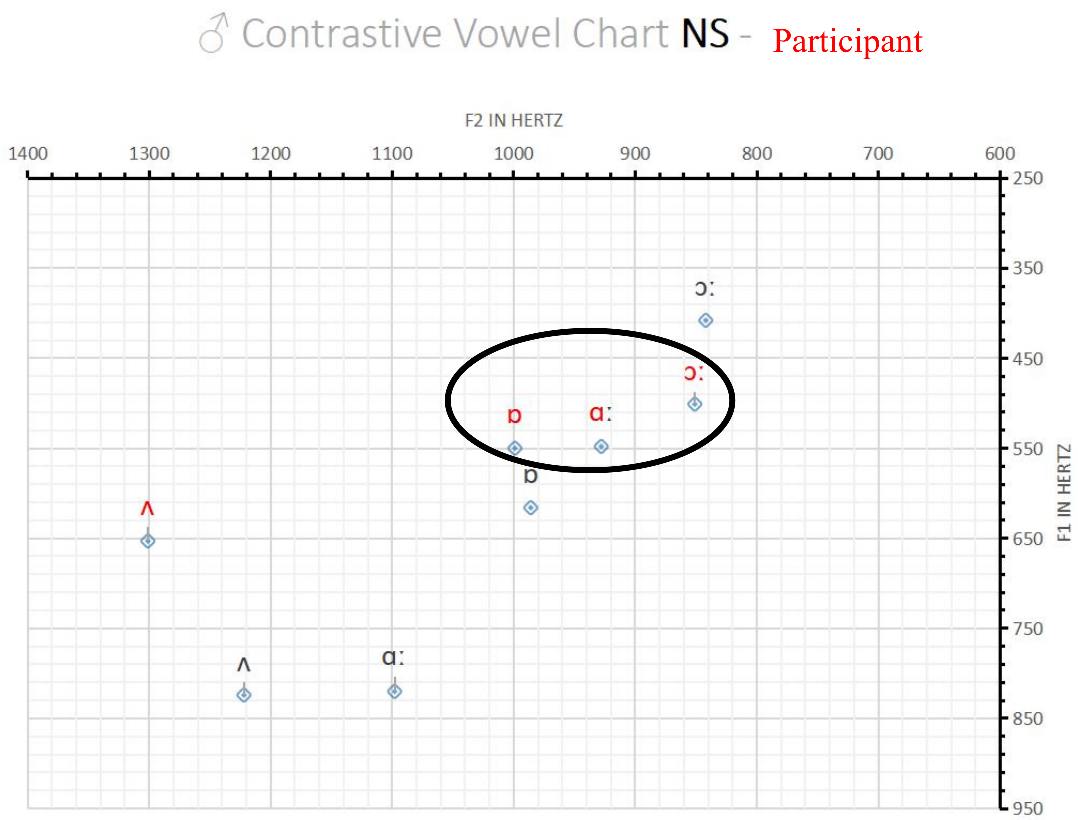
Figure 12 : *Vowel chart showing participants data against NS realisations*



4.4 Both mergers in one speaker

The last pattern to describe appeared prominently in only 4 participants. It involves the occurrence of the *cot-caught* merger in conjunction with the *father-bother* merger, ultimately resulting in three sounds collapsing into one, despite originally representing a fourfold vowel distinction (depending on context). Once again, it is not apparent whether these realisations are intentional in order to adhere to a GA reference standard or if L1 transfer from German to English is at play. According to Labov et al. (2006) the vowels in question are typically pronounced as a back, open-mid vowel /ɔ/ in instances where the merger applies (see e.g. Canadian English). That aligns with the findings of my study (see graph below). Examples also appeared where the vowels were further apart, yet the distance to /ʌ/ was still the greatest. This may indicate an irregular realisation of vowels /ɑ:/, /ɒ/ and /ɔ:/ possibly attributable to uncertainty on the part of the speaker. In addition, other patterns also emerged that could not be explained based on the considerations made so far and therefore should not be further considered. (A complete collection of all the produced vowel spaces can be found in the Appendix. I included pre- and post-test data from every participant and contrasted the post-test data with the NS data by means of a graph).

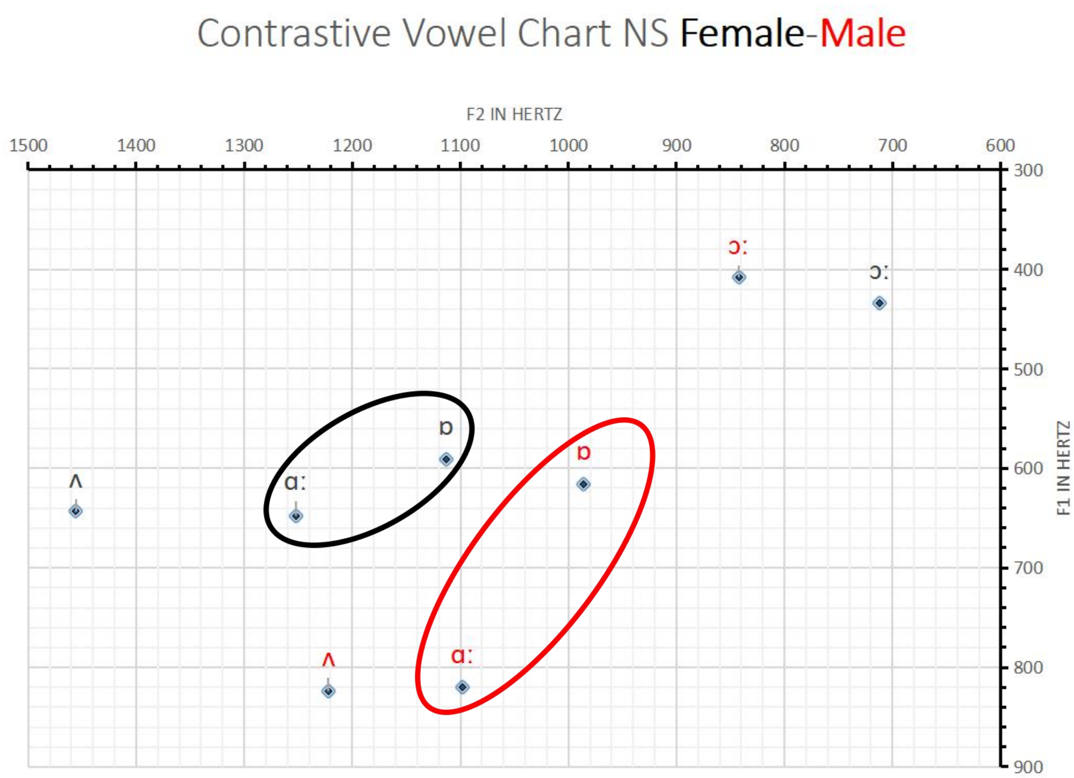
Figure 13: *Vowel chart showing participants data against NS realisations*



4.5 Issues in NS speech data

I must note that my male reference data in general is closer to a prototypical RP variant of English than my female NS data. /ʌ/ and /ɑ:/ in particular are higher up in the chart than their male counterparts. However, this may be due to a higher voice pitch. We find an approximation of the two phonemes /ɑ:/ and /ɒ/, which hints towards elements of the *father-bother* merger and modifies the data in favour of a GA-like accent. In the context of research into RP, this was unforeseen and quite a significant shortcoming of my work. Later in the discussion chapter, the potential implementations of this are explored. To visualise the problem at hand, I would like to provide a comparison of both my male and female NS speech data.

Figure 14: *Contrastive view of male vs. Female NS data*



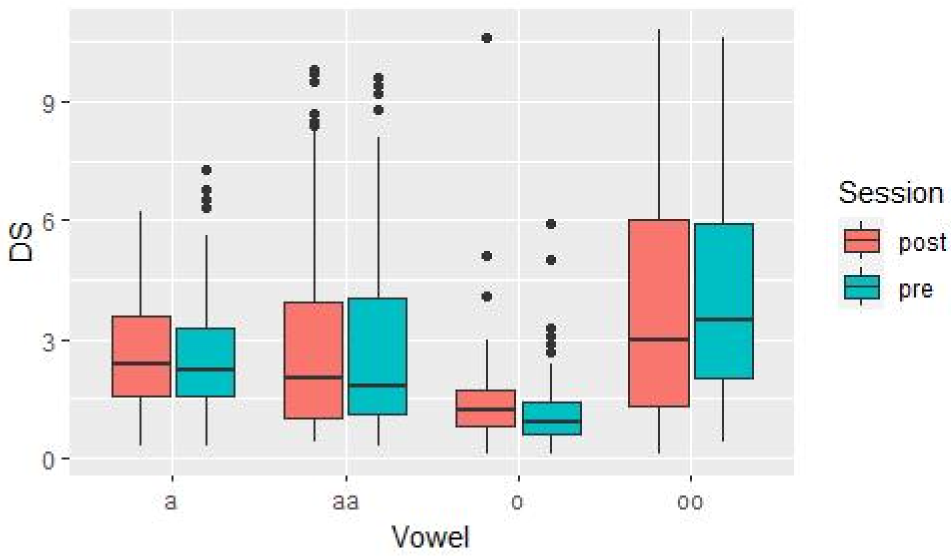
This comparison demonstrates that intra-linguistic variability is high, and the phenomenon of back vowels in English can be quite complex. A more careful selection of native speakers would have been absolutely necessary. On the other hand, the data demonstrates how difficult it is for a learner to internalise the systematics of these sounds, given the fact that even natives tend to struggle when confronted with various varieties of a language.

4.6 Quantitative analysis

4.6.1 Pre- vs. post-testing in production

In order to draw quantitative conclusions about the effectiveness of my survey, MDs were calculated for pre-test as well as post-test scenarios. The MDs are a reference for whether or not learning effects took place during the course of the training rounds. After analysing all the data obtained, it can be concluded that a limited amount of perception training does not significantly enhance the speakers performances in production and does not lead to an improvement in their respective production patterns. The following graph gives an indication of this instance by plotting the MDs (DS) against the corresponding vowels.

Figure 15: *pre/post view of MDs against corresponding vowels*



A longer MD post is indicated by a protruding red bar in three out of four cases. Only the MD of the long *aa* shortens from the pre-test to the post-test. If the preceding training had been successful, this would have been evidenced by a shortened MD across the board. However, the trend in this dataset is unfortunately the opposite, indicating that the offered training was not effective.

My hypotheses are tested using a linear mixed effect model and an ANOVA to analyse the differences among group means in the sample. Before running the analyses, the distribution of the MDs was inspected via the *qqnorm* function in R, and MDs below 11 were excluded from the analyses. The analyses were run on the remaining 1128 data points. The dependent variable was log-transformed MD; the fixed factors were session (pre vs. post), vowel (with the tested vowels), session-by-vowel interaction (to test for potential differences in the effects of training across vowels), and sex (to test for potential differences between male and female participants). The random structure included a session/participant slope to model potential differences in the effect of training across participants. An Anova function from the *car* package was used to assess the significance of the main effects in the model. The latter revealed a significant vowel-by-session interaction and a significant vowel effect but the other two factors were not significant (see data below).

Table 4*: Analysis of Deviance Table (Type II Wald chisquare tests)*

	Chisq	DF	Pr(>Chisq)
Session	0.5135	1	0.47362
Vowel	432.4200	3	< 2e-16 ***
Sex	0.4893	1	0.48424
Session:Vowel	9.0665	3	0.02842 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

I would like to add some explaining words to this data: The main effect of session (pre vs post) is not significant (p = 0.47362), meaning there is no overall improvement in English speech production performance from pre-test to post-test sessions. However, this does not necessarily imply that perception training has no effect at all. As my descriptive data analysis revealed, there were, in fact, some participants who performed better during the post-test. The quantitative data suggests that the results were not uniform across all participants or vowels. According to this quantitative data, I must disconfirm my initial hypothesis. The vowel-by-session interaction is significant (p = 0.02842), indicating that the effect of training on English speech production varied across different vowels. Perception training has a different impact depending on the vowel produced. I will try to address that in the discussion section and relate it to L2 learning models. Lastly, the main effect of sex is not significant (p = 0.48424), suggesting that there are no significant differences in English speech production between male and female participants in this study.

A follow-up *emmeans* function revealed that training was effective in increasing the MD for the *o* vowel only (p= 0.0210) but for the other vowels, the effect of training was not significant.

Table 5: *Emmeans function results per vowel*

Vowel = a:

contrast	estimate	SE	df	t.ratio	p.value
post-pre	0.026095	0.0988	193.7	0.264	0.7919

Vowel = aa:

contrast	estimate	SE	df	t.ratio	p.value
post-pre	0.000423	0.1097	275.4	0.004	0.9969

Vowel = o:

contrast	estimate	SE	df	t.ratio	p.value
post-pre	0.187092	0.0796	89.2	2.350	0.0210

Vowel = oo:

contrast	estimate	SE	df	t.ratio	p.value
post-pre	-0.155597	0.0978	186.1	-1.591	0.1133

While perception training was effective in increasing (slightly worse in post as compared to pre) the MD for the *o* vowel, it did not lead to significant improvements for the other vowels tested. This reinforces the earlier findings of a significant interaction between vowel and session, indicating that the effectiveness of perception training varies depending on the specific vowel being produced.

In summary, the significant vowel effect and vowel-by-session interaction support the idea that perception training can influence English speech production, but the lack of a significant session effect suggests that improvements may not be uniform across all participants or vowels. Further investigation into why certain vowels show greater improvement and how individual differences might play a role could provide deeper insights into the efficacy of perception training for English speech production in German participants. The question now arises whether the observed production results are reflected in any way by the results of perception, and whether correlations between listening comprehension and one's own language production can be identified.

4.6.2 Perception results

Perception data came in the form of the presented words, the auditive input, the vowel that was presented, the vowel opposition that was presented, the response time for each vowel, the candidate's answer and the participant's identification and answers to the initial questionnaire on English performance. The answer column received two possible outcomes; correct and false. If the wrong answer was chosen, the label false appeared. For the false answers, I manually added the corresponding vowel oppositions in an attempt to correlate the wrong outcomes of the perception round with the patterns displayed within the production tasks. Excluding one participant, the error rate for both pre- and post-tests is 2,4 false identifications of vowel oppositions per participant. By including the participant, we get an error rate of 2,6 in pre and 2,9 in post. It was not clear whether the participant scored so low out of carelessness or inability. In any case, there is no significant improvement from the pre-test to the post-test rounds.

However, the described merger occurrences during production resurface within the perception data. The highest number of errors were made when asked about the vowel opposition of /ɒ/ versus /ʌ/ with 25 false identifications in pre and 30 wrong instances in post, closely followed by the *cot-caught* merger with 17 false identifications in pre and 25 wrong instances in post perception. The /ɑ:/ versus /ʌ/ opposition is only twice miscategorised in pre-testing but five times in post-testing. /ɒ/ and /ɑ:/ get a total of 12 erroneous answers in pre and only 4 in post. (I consider my choice of words unlucky in this instance. In almost all examples, the wrongly categorised minimal pair was *bodge-barge,* two words that may not have been part of my participants active vocabulary and were further uttered by my female NS with a significant *father-bother* merger).

In order to tackle the statistical analysis of my perception data, I opted for the following approach: An additional three columns were added to the dataset, which contained the proportion of correct answers for the perception pre- and post-test as well as a total score for both tests. The columns provided information on how many percent of the given responses were correct. Participants' accuracy on each trial (false/accurate) was then incorporated into a generalised linear mixed-effects model with Type II Wald chi-square tests to check for statistical significance of effects (Session, Vowel Session by Vowel interaction and Sex).

Table 6: *Analysis of Deviance Table (Type II Wald chisquare tests) for perception data*

	Chisq	DF	Pr(>Chisq)
Session	0.5440	1	0.46077
Vowel	42.7958	3	2.719e-09 ***
Sex	0.2424	1	0.62247
Session:Vowel	8.2449	3	0.04121 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The session (pre vs. post) did not have a significant overall effect on perception scores (because $p > 0.05$). This suggests that there was no significant improvement or change in perception scores from before to after the training sessions, taking all vowels into account. The type of vowel significantly affects perception scores ($p < 0.001$). This means that some vowels may be perceived more accurately or differently than others, regardless of the training session. There was no significant difference between male and female participants in terms of perception scores ($p = 0.62247$). Sex did not influence how participants perceived the vowels in English. There is a significant interaction between session and vowel type. This interaction suggests that the effect of the session (training) on perception scores varies depending on the vowel being tested. This promotes the assumption that the training effect may differ across different vowel sounds. I now want to take a look at the vowel data on a case-by-case basis:

__Table 7__: Perception scores for the different vowels

vowel = a:

contrast	estimate	SE	df	z.ratio	p.value
post-pre	-2.124	0.765	Inf	-2.775	0.0055

vowel = aa:

contrast	estimate	SE	df	z.ratio	p.value
post-pre	0.643	1.241	Inf	0.518	0.6044

vowel = o:

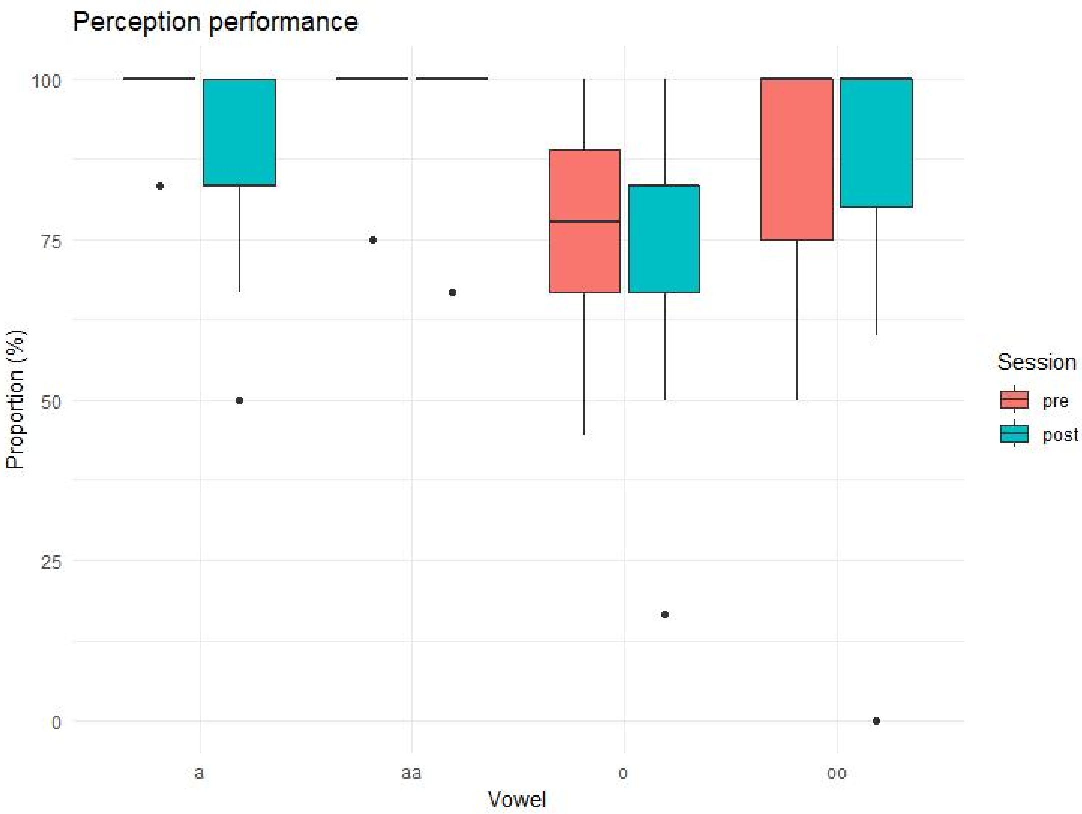
contrast	estimate	SE	df	z.ratio	p.value
post-pre	-0.119	0.276	Inf	-0.430	0.6673

vowel = oo:

contrast	estimate	SE	df	z.ratio	p.value
post-pre	0.394	0.551	Inf	0.716	0.4739

The perception scores for the *a* sound significantly decreased after training. This indicates that participants' perception of /ʌ/ worsened after training. After training, the perception scores for the remaining vowels did not change significantly. Participants' perception of /ɒ/, /ɑ:/ and /ɔ:/ remained similar before and after training. While there was no significant overall effect of the session (pre vs. post) on perception scores either, the interaction effect suggests that the training may have had different impacts on the perception of different vowels. In the discussion section, I would like to delve deeper into these issues against the background of SLM and PAM, as well as the phonetic peculiarities of different English variants. In summary it can be said that the training had the most significant impact on the perception of the vowel *o*, where there was a substantial increase in the median proportion of correct answers. The perception of vowels *a* and *oo* showed little to no improvement, indicating that participants were already fairly accurate in perceiving these vowels. Vowel *aa* showed some improvement, both in median performance and consistency among participants (see graph below).

Figure 16: *Perception accuracy for tested vowels.*



4.6.3 Is perception related to production?

In a final step, the initial hypothesis stating that listening comprehension and production are related and may positively influence each other will now be examined. To achieve this, average values for perception (Answer_prop) and production (DS_mean), for each speaker and vowel were determined, which were then integrated into a mixed-effects model. The mixed effect model then examined the relationship between perception and production data across the different vowels by using MD as the dependent variable and perception score, vowel and their interaction as fixed factors. Participants were included in the model as random intercepts. I deemed this an appropriate approach as observations were not independent (in my case, multiple responses per subject). An Anova (target model) output provided a Type II Wald chi-square test for each term in the model, assessing significance. Based on the analysis, there was no significant relationship between perception and production of vowels at either the pre-test or post-test sessions (see data below).

Table 8: *Pre-test; Correlating production and performance data*

	Estimate	Std.Error	df	tvalue	Pr(>|t|)
(Intercept)	3.70641	1.25261	87.0000	2.959	0.00398**
Answer_prop	-0.01155	0.01225	87.0000	-0.942	0.34855
Vowel aa	0.17642	0.47183	87.0000	0.374	0.70939
Vowel o	-1.71565	0.53885	87.0000	-3.184	0.00202**
Vowel oo	1.56504	0.49267	87.0000	3.177	0.00206**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Pre-test; Fixed effects: DS.mean ~ Answer_prop + Vowel + (1 | Subject)

Table 9*: Post-test; Correlating production and performance data*

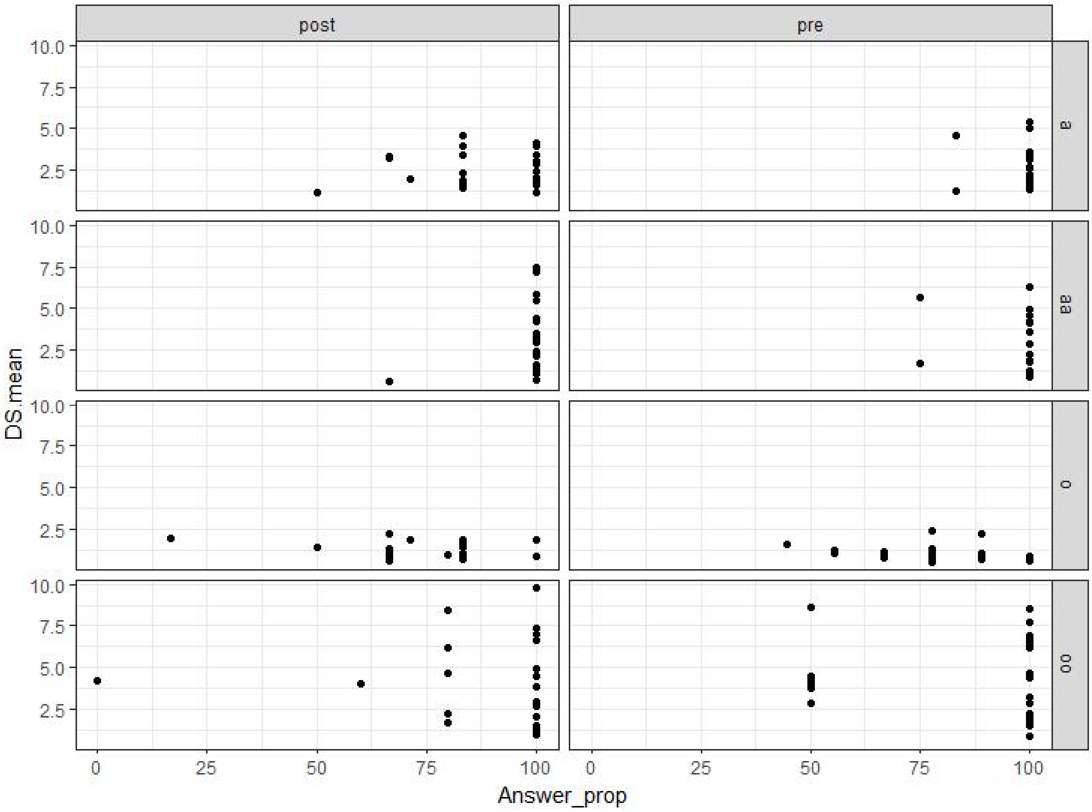
	Estimate	Std.Error	df	t-value	Pr(>|t|)
(Intercept)	2.6472498	1.0526893	87.0000	2.515	0.01375 *
Answer_prop	-0.0009828	0.0112542	87.0000	-0.087	0.93061
Vowel aa	0.3638854	0.5234710	87.0000	0.695	0.48882
Vowel o	-1.2376995	0.5330475	87.0000	-2.322	0.02257 *
Vowel oo	1.4512445	0.5098809	87.0000	2.846	0.00552 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Post-test; Fixed effects: DS.mean ~ Answer_prop + Vowel + (1 | Subject)

The coefficient for Answer_prop is -0.01155, but it is not statistically significant (p = 0.34855). This means that the perception score does not significantly predict the production score at the pre-test session. The same is true for the post-test data. I provide a graph to visualise my findings. It illustrates the relationship between participants' perception performance (measured as the proportion of correct answers in percentage) and their mean discrimination scores, represented by MDs, for different English back vowels. The data is divided into sessions conducted before (pre) and after (post) the training. I assume that the lack of a relationship seems to be due to the lack of variance in the perception task.

Figure 17: *Proportion of correct answers in perception against mean discrimination scores (MDs) in production*



4.6.4 Addendum to the questionnaire data

After visually inspecting the data for additional languages, German dialect and competence, a quantitative approach to significance testing is not pursued. The average error quotient in listening comprehension was 14% with an SD of 0,081. The average mean deviation in production was 2,677 (MDs) with an SD of 3,432. Therefore, I do not expect significant differences in general language abilities such as proficiency in a dialect and/or other foreign languages, as well as active and passive English skills, among my participants. In the discussion, I will examine two examples to illustrate why further investigation is not helpful here.

5 Discussion

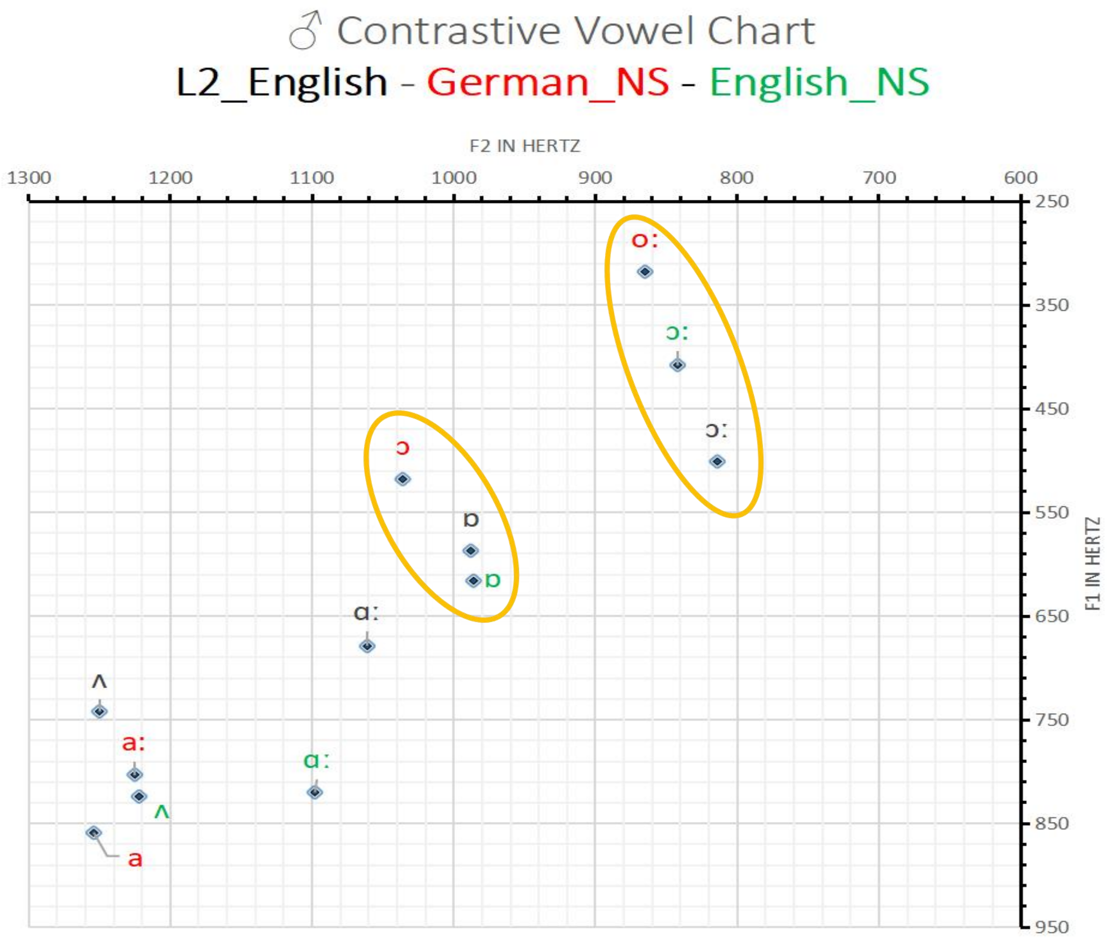
In this last chapter, I would like to attempt to put together the various puzzle pieces of my investigation and, based on my qualitative data analysis, find explanations for my participants' production data. With reference to Kautzsch (2012) I ask the same question as to why Germans do not apply L1 transfer in a more prominent way and exhibit hints of vowel mergers in their spoken performances (at least on a case by case basis). With regard to speech models like the SLM and PAM, I would like to interpret the performances of my participants. Regardless of the quantitative observations that the experiment brought along, the survey does indeed possess a certain descriptive power as it delivers concrete data on German productions (and perceptions) of English back vowels (a phenomenon with a rather short history in research). Even though no significant changes were observed through my intervention, the experiment still provides data that can be used to interpret L2 acquisition models. In this regard, I would like to address the data that suggests some kind of L1-L2 transfer as well as provide explanations based on insights from established speech models. Furthermore, in this context, intra-linguistic factors such as the various vowel mergers must not be overlooked. The question now arises whether my initial assumptions regarding my participants' performances could be confirmed by the data. Backed by findings in Kautzsch (2012) and Sönning (2020) I am asking the question of the lack of consequent L1 transfer in my participants' production data. The miniature German dataset suggests that direct L1 transfer would favour a more native-like (British) pronunciation, so why is it that some participants lose themselves in the imitation of phenomena of the English language like the mentioned vowel mergers?

5.1 Target language vs. native language vs. L2 intra-language

Let us take a look at the following vowel space, which contrastively shows the vowels that my English NS produced over the course of the data collection, as well as realisations of the German words *Sohle* representing /o:/, *Soll* representing /ɔ/, *Kahn* representing /a:/ and *kann* representing /a:/. German data is red, and English NS data is green. I further added the mean production values of all my German male participants (as differences in pre- and post-test were not significant I consider this a reasonable thing to do). The L2 speech data appears in black colour. This threefold presentation of data neatly shows similarities and deviations between English RP, (standard) German, and the intra-language of my participants on their way to forming an L2 vowel space. It is noticeable that there seem to be counterparts in the German phoneme system for the English vowels /ɔ:/ and /ɒ/

(orange circles) and the same quantitative patterns can be found with regard to the low central back vowels (both English and German have a pair of short and long low central back vowels). So, why don't German native speakers simply resort to phonetic transfer and thus spare themselves the acoustic excursions that I found in my data? (For the sake of completeness, it must be mentioned that the same patterns are to be observed in the female data as well. I chose the male data in this instance because of the more appropriate realisations of my male NS in the speech samples).

Figure 18: *The English (green) and German (red) phoneme systems as opposed to participants' L2 intra-language space (black)*



Let us walk through the vowels one by one: L2 /ɔ:/ is closer to native /ɔ:/ than it is to German /o:/. This suggests that, in addition to vowel quantity, a difference in quality is also recognised and can be well implemented on average. For L1 transfer, a more closed realisation of the sound would be expected. The anticipated transfer gives way to an approximation towards the English original. We also find an almost native-like realisation of the /ɒ/ sound, which is contrary to what I assumed earlier. Again, the distance between L2 realisation and native sound is smaller than between L2 and the German equivalent. With regard to short and long mid-back vowels, this may suggest that my participants, perhaps subconsciously, indeed follow the model of British pronunciation, favoured by the similarity of British and German phonemes in this area of the vowel space. The lower left corner of the graph shows an approximation of L2 long low central vowels to the British original samples. Since German also distinguishes its corresponding vowels quantitatively, this agreement is not surprising. What is surprising in this instance is the fact that L2 data is again further away from the German reference in terms of qualitative features. Although the L2 and native English data are not identical, there is no doubt that segmentation according to the British model occurs among German learners. However, the individual examination of each result from my survey makes it clear that there is considerable variation across all participants. Therefore, the above graph should only be considered an average of all my male participants.

5.2 L1 to L2 transfer - put into perspective

So let us focus on the different pronunciation patterns that I was able to observe in my experiment. I have shown that some of my participants do apply mergers that are also known from GA speeches. This is not intended to be a criticism of my participants' performance. Instead, it is interesting to find explanations for this situation. When we look at the comparison of German and English phonetic spaces, there should actually be no reason for Germans to apply a merger, as RP provides a convenient alternative (for those who have active knowledge about this). Unfortunately, the reality looks different, creating a situation where the well-established phonetic repertoire of the Germans encounters irregularities in the English language, which indeed occur very frequently. Let's look at the interplay between the *cot-caught* and *father-bother* mergers which were clearly evident in some participants and also caused problems in the listening comprehension exercises. In a case where the first merger applies, /ɔː/ and /ɒ/ become one phoneme. Which variant should the learner adapt to? Likely, a new sound must be learned (see Kautzsch 2012) which creates uncertainties on the learners end. The result is a scenario where the same phoneme is pronounced differently in changing contexts by L2 learners. We may also assume a scenario where /ɔː/ and /ɒ/ become /ɔ(ː)/. Now let's add the *father-bother* merger to the equation: /ɑː/ and /ɒ/ usually become /a(ː)/ in this instance (Cruttenden & Gimson, 2008). How is an L2 learner supposed to know whether a word that in RP has /ɒ/ is uttered with /ɑ/ or /ɔ/? Why fall back on a merger that would not occur in the L1 or a British variant in the first place? Answers to these questions can be provided by the aforementioned works by

Birdsong, Hickey, Flege, or Best. In this context, I would like to spotlight the following three keywords: motivation, input (or exposure), and transfer.

The input available to learners likely contains a lot of American English, which is naturally omnipresent due to its prevalence in the media. As Collins et al. (2019) point out, the use of mergers is not always regular or transparent unless there is extensive contact with speakers of the given variant. Since foreign language instruction also does not further address these phenomena, the occurrence of such mergers remains a somewhat arbitrary phenomenon for the learner. Kautzsch (2012) examines in his study how German learners implement a British or American accent. In contrast to my own study, there is clarity regarding the desired standard of each participant. This cannot necessarily be assumed for speakers without a university background. However, setting that aside, my study yields similar results (different candidates pursue different standard varieties at least subconsciously). I agree with Kautzsch that learners do not always replicate English varieties one-to-one. Therefore, a distinction between GA and RP does not make sense in the context of my data, since I must assume that participants do not attribute the shift in back vowels to a specific GA or RP accent. A more meaningful statement could have been made if participants had been asked at the beginning of the study which variant of English they aimed to achieve. Since this did not happen, I cannot assume a guided approach was taken in the realisations of the produced tokens. However, the acoustical analyses presented here reveal that in many cases, learners use sounds that differ from both their L1 and L2, which can be viewed as an empirical manifestation of inter-language (see Moreno-Torres et al., 2016 and figure 18). Therefore, the likely cause of the learners' inconsistencies cannot be solely attributed to transfer, but also to a lack of awareness of the highly diverse nature of the input surrounding them (Bohn & Flege, 1992). Judging by ideas put forward in Kautzsch (2012) „it seems that the acquisition of the [GA] system is more inconsistency prone than the acquisition of [an RP] system“ (230). The data I have collected also supports this claim (a fourfold distinction was obtained on average. No participant showed a complete merger in his or her speech.) The data from my study demonstrate how learners, who may not necessarily be aware of these differences, are influenced by both standards. In line with assertions from Hickey (2019), I attribute this to the absence of these phenomena in foreign language instruction, among other factors. This also ties back to L2 speech models. If phonetic differences are to be recognised and produced, it is important that learners are aware of the patterns they need to consider (see Birdsong (2007) on motivation and phonetic training of the ear and Krashen (1982) who describes *affective motivational filters* which tend to allow structured learning of such phenomena only after puberty). For instance, a simple question at the beginning of the study could have been: Are you aware of concrete differences between GA and RP? Quite possibly, the answers with regard to the vowels in focus here would have been negative. My observations align with those in Sönning (2020) and Kautzsch (2012), who have presented one of the few contrastive comparisons of English and German back vowels to this date. While the primary aim of the study, to create greater awareness of the back vowel phenomena in English L2 speech through targeted listening comprehension tasks, was not successful, there is still a small achievement in the form of a qualitative data collection. A previously overlooked issue was examined in greater detail, which has contributed to enriching the sparse data on this subject. The work thus aligns itself with a series of investigations into L2 acquisition of English by German learners, which have been taking place since the structuralist times of the Prague School (see e.g. Jørgensen, (1969), Bohn & Flege, (1992), Iverson & Evans, (2009), Mitterer et al. (2020) and Schlechtweg et al., (2023) to name a few additional papers with a focus on the German sound system).

5.3 Integration of results in L2 models

What is still missing is an integration of these phenomena into the context of the L2 acquisition models mentioned in Chapter 2. This requires a systematic approach, which I have attempted to depict in the paragraph below:

Flege (1995) remarks that with increased exposure, a gradual shift towards the L2‘s standards may occur in a learner. In the context of my study, I do not have any means of telling how many years of exposure my participants had prior to my survey (simply because I did not ask for this information). For my argument, I use the average values from figure 18 and my own general findings as a baseline. On average, figure 18 shows a fourfold distinction between the produced vowels. The question now is, what happens if this fourfold system gets interrupted by a vowel merger? Let us assume that the British /ɒ/ is represented in a German‘s intra-language in one way or another. It occupies a spot that, in GA, can be substituted by /ɑ:/ or /ɔ/. How can new categories be established if the input is so variable? Clearly, L1 transfer of some kind must account for the fact that my German participants, on average, did not exhibit a consequent vowel merger in their speech. However, looking at perception, we can see some confusion with regard to these mergers. In absolute figures, the /ɒ/ vs. /ɔ:/ contrast caused the most erroneous answers during pre- and post-testing (as pre- and post-testing were not significantly different I consider both rounds in this analysis).

This I attribute partly to the quantitative features of the vowels, ergo shifts in vowel length once a merger is applied. The same is true for the /ɒ/ - /ʌ/ opposition, which also caused a significant amount of wrong perception (though slightly less then /ɒ/ vs. /ɔ/). This means ultimately that my participants, on average, produced better than they perceived, which is at odds with the thoughts put forward in Flege (1988, 1995).

The changes with regard to vowel quantity are rather interesting. When applying the two described mergers, the quantitative opposition between the vowels is typically eliminated, something that the German phoneme system heavily relies on (see Lindsey & Wells, 2019; Cruttenden, 2008). If quantitative cues that could contribute to the identification of the vowel are also missing, the *perceived phonetic dissimilarity* described by Flege can no longer be applied. The affected speaker thus does not know which phoneme (within the intra-language) the sound should be assigned to, leading to mis-categorisation. In this context, it would be interesting to investigate British native speakers' performance in the same situation. Would they struggle to perceive merged phonemes in a context free environment? The data suggest that even they would have difficulties categorising a merged sound in such a scenario. (see *phonetic recalibration* in the following paragraph). This brings me to the last phenomenon I want to tackle within the framework of an L2 model. The data from my listening comprehension exercises indicate that after the *o-oo* contrast, the *o-a* opposition was fraught with a similarly high error rate. How can this be? In the English literature, however, there is no mention of a merger following the (made up) *cop-cup* pattern. So why do we find these uncertainties in perception?

Firstly, the quantitative opposition between the two sounds is missing. Both have a short pronunciation. The ability to distinguish these pairs is remarkably poor. This is described in the PAM-L2 as a *single category pattern* and leads to the absence of differentiation in the presented opposition. This may be due to the fact that, as figure 6 shows, the German low central vowels are very dependent on the quantitative difference between them. If the *father-bother* merger is present in one form or another, then discrimination between /ɒ/ and /ʌ/ becomes challenging because Germans do not add qualitative changes to their low central vowels. A clear pick may be prevented due to uncertainties with regards to vowel quality because of equal *goodness of exemplar* ratings (Best & Tyler, 2007). Here we also encounter the criticism mentioned earlier regarding the transcription conventions of English linguistics (English /ʌ/ = German /a/); thus, two phonemes but (almost) one phonetic realisation. Since production apparently does not undergo as much influence from these factors as passive input does, this naturally raises the question to what extent my assumptions contradict those of Flege and other authors who claim that well-developed perception enables accurate production. Given that my dataset is quite small and my intervention did not succeed, I would like to point out that my assertion should not carry much weight in light of such limited data. It is also important to note in the argumentation that the visualised average values, mask the individually occurring patterns in my participants' speech, and thus, conclusions drawn from the average can differ from those drawn from individual examinations of each participant. (Take a candidate who applied the *cot-caught* merger and put it next to someone exhibiting the *father-bother* merger: the two datasets would counterbalance each other, yielding a dataset that does not depict the occurrence of any of the two mergers). In this context, making final statements about the described issue is difficult without first questioning the participants about the variety they are aiming for.

5.4 Remarks on English orthography

I also want to touch on English orthography, as it could indeed contribute to a fourfold distribution in the phoneme system. Lindsey and Wells (2019) state that the short *o* is mostly written with an *O*, which should promote its realisation as /ɒ/ instead of /ɑ/, at least for a German learner. I am referring to a phenomenon that has gained notoriety under the name McGurk illusion. In 1976, McGurk and MacDonald attempted to manipulate the linguistic perception of their subjects with erroneous transcriptions, which showed words like *da* when in reality the subject heard *ga* (McGurk & MacDonald, 1976). They therefore investigated the multisensory nature of speech, which was further researched in the following years under the term phonetic recalibration (it must be mentioned though that the McGurk illusion is typically referred to as a mismatch between visual (dynamic) cues and audio cues. However, McGurk and McDonald can been seen as the pioneers of this specific field and therefore deserve mentioning here). It is conceivable that such a dissonance could arise for a German learner of English when the word *cop* is presented to them with the phonetic realisation of /kɑp/. Both the learner's intra-language and English orthography would favour the realisation of /kɒp/ instead. To stretch this comparison, in the McGurk illusion, there is a conflict between the auditory system that hears the syllable /kɑp/ and the visual system that sees the written *o* in cop (Vroomen & Bart, 2012). In *phonetic recalibration of speech* Keetels et al. showed that listeners „adjust their phonetic boundaries in accordance with disambiguating orthographic information and that these adjustments show a rapid build-up“ (Keetels et al., 2016: 943). I take this line of research to explain the worsened production results at post-test as compared to pre-test results. Taking into account the complexities that arise with regard to English back vowels, this could be an interesting topic for further research in the field of phonetic recalibration.

5.5 UG in L2 speech learning

I would also like to share a last thought on the implementation of a UG supported approach as well: The inclusion of UG in theories of L2 speech learning is outnumbered and less common compared to models like Flege's and Best's. However, the application of nativist ideas to the issue is indeed interesting. The question arises whether occasional instances of vowel mergers in non-native speakers can be attributed to reasons rooted in UG. When faced with the phenomenon of vowel mergers, the learners encounter varied input from the L2. They receive positive feedback from many different directions, so the process of fully acquiring one variant or another cannot be completed subconsciously. Archibald (1994) describes this type of acquisition as *inductive learning*, which particularly emerges before the end of the CP. Assuming this, one must infer that UG mechanisms are no longer available to an adult learner. To manage the chaos of sounds, one therefore resorts to the *general problem solving mechanisms* described by Singleton (2005), which are available to every adult learner after crossing the CP (should be available). However, this logical and rational step must be consciously undertaken and proves to be difficult without specific reference to the corresponding phenomenon. During this argument, the question of whether UG is available after the CP would also be answered. Krashen (1982) answers this question with *no* by remarking that *affective filters*, in the form of motivation and logical thinking, get „strengthened at puberty thanks to the onset of formal operations“ (216). Such a process cannot be termed innate since it does not occur automatically. Testing this hypothesis requires a different type of data collection, though. Therefore, this excursus should only be considered as a side note. From a rational point of view, the assumption of this hypothesis is nevertheless comprehensible because few learners seem to be able (consciously or unconsciously) to separate different variants of the English language from each other in their own performance and by doing so achieve native-like performance in an L2 (see van Leussen & Escudero, 2015; Levis & Zhou, 2018; Mora et al., 2022).

5.6 General remarks on the experiments‘ implementation

The preparation of an empirical work requires a sound understanding of the state of research, a specific phenomenon to investigate, careful planning and implementation of the experimental idea and last but not least, good knowledge of the applicable statistical methods for data analysis. The phenomenon under investigation stemmed from my own concern, seeking clarity about a linguistic phenomenon that had long been a source of confusion. The theoretical knowledge that developed from this curiosity has greatly influenced this work. I would like to share some thoughts on the empirical work and not hesitate to criticise my own approach below.

First, a few critical words about the online experiment conducted here, using an online tool for empirical data collection. The E-Babylab, in my case, largely served its intended purpose of collecting speech data. However, I would like to highlight a few points that lead me to critically view the use of such a tool for future surveys. The most obvious point is certainly that participants must be entrusted with a great deal of confidence that they will conduct the survey to the best of their abilities. There is no means of control available to the researcher once the experiment is posted online; the researcher can only wait to see if the collected data meets expectations. Intervention during the experiment is not possible. I tried to prevent this by providing precise formulations and instructions. Nevertheless, I found myself in a situation where participants misinterpreted some of my instructions, and the experiment was not completed by a total of 25 people. In some cases, where I knew the candidates personally, I could point out ambiguities, thus ensuring that the affected person ultimately completed the experiment. In stationary data collection, such problems are much easier to address. Since the data collection involved noise-sensitive data, it was extremely important to ensure a quiet environment during recording. This is a given in laboratory studies, but it becomes a significant issue in an online experiment. When collecting remote data, problems such as ticking clocks or traffic noise may arise during testing (one participant clearly took part in my study during a bus ride). Therefore, one must accept that some datasets can become unsuitable for analysis due to external factors that would never have arisen in a stationary scenario. Apart from these factors, on-site data collection is also desirable in terms of the data's uniformity. All of my collected data ended up differing significantly from each other, reflecting the audio quality of the received samples. This is not surprising, as different devices were used for the recordings. While software can address this to some extent, good sound quality in the first place must always be the preferred option.

In this context, the set-up of such an online study must also be mentioned: remote data collection must strike a balance between utility and efficiency. This means, in simple terms, that the experimental design must not be too complicated in order to encourage broad participation while ensuring at the same time that the collected data are valid and usable. I believe that the implementation of my experiment has been largely successful (at least as indicated by the feedback from my audience). I would like to offer less criticism of the collected data itself and instead focus on the material provided by my female NS. The noticeable traces of a somewhat non-standard RP accent in hindsight are a major weakness of my study. On several occasions, the variation introduced by my female native speaker into the original material caused many of my participants, including myself, to hesitate. The incidental use of the *father-bother* merger resulted in some minimal pairs being pronounced unclear, almost entirely eliminating the intended contrast. In these cases, participants were literally asked to note differences that were not present. The NS also exhibited some instances of a post-vocalic *r*, which forced me to exclude some of her recordings from the experiment design. Because of this instance, the study contains more male NS tokens than female ones. In my opinion, this also explains the lack of success of my intervention. A uniform quality of the NS data would have potentially provided a different result, at least in the perception part. If this work had not been bound by a deadline, the female native speaker could have been replaced by a more suitable candidate. While these discrepancies are frustrating, they also highlight the extent of variation within the English language among native speakers of English and the significant challenges faced by language learners. Since I repeatedly refer to the concept of mergers and the differences between RP and GA throughout my work, this circumstance also provides the opportunity to delve deeper into these phenomena. As I did not initially consider the distinction between these two variants at the beginning of the study, I was able to obtain imprecise but nevertheless insightful information through the varied data of my female NS. Without the divergent data from my female NS, the American variants, as well as the patterns observed in my participants, would not have been discussed as accurately as they ultimately were. Since I did not initially consider distinguishing my participants into different dialects or affiliation groups at the beginning of the study, the sporadic merger phenomena observed in my female NS favoured the approach I ultimately chose.

Here, I also want to briefly address the composition of my participants, as the broad cross-section of the population present in my study differs from the typical test groups found in the literature. This has both advantages and disadvantages: a wider age range and the diverse backgrounds of the respondents could have provided the study with greater significance if it had been successful. To achieve this, of course, even more participants would have been desirable. With regard to common research practices, which usually limit themselves to recruiting participants from the academic field, I find my approach at least equally appropriate and would recommend its replication in future studies. However, one of the disadvantages of my survey is, of course, that my group showed considerable heterogeneity, and generalisations could only be based on gender. Categorising respondents into various competence levels would require significantly more effort, and it would be questionable whether such an approach could be implemented through the online method chosen here. (Keep in mind that the data that I did receive with regard to dialect and general language competence, were not diagnostically conclusive). With the recruitment method I chose, the participants' English proficiency is as varied as their ages. Particularly with minimal pairs that have somewhat unusual additions, this can lead to words not being recognised and therefore being mispronounced. As the hurdles to participating in my study were not very high, such outcomes had to be expected. With more time and greater incentives for potential participants, this obstacle could be circumnavigated in the future (e.g. gift cards etc.). Since participation in my study was entirely voluntary, I assume that my participants had at least a superficial interest in the research. Here, I refer again to Birdsong (2007) and his discussions on intrinsic motivation in language learning (see also Bongaerts, 2000). According to Birdsong, high motivation is necessary but, in most cases, not sufficient for achieving native-like pronunciation. I would agree with this assertion in light of the discussed L2 speech models. For quantitative testing of such effects, empirical data must be obtained that can provide information about competence and, if applicable, motivation. My attempt to do exactly that unfortunately did not succeed, also due to the insignificance of the testing itself.

6 Conclusion

To appropriately conclude this work, the insights gained will be summarised on the final pages of my thesis. The complexity of this topic ultimately stems from the variability of the English language, which presents itself with some irregularities in its phoneme system(s). Accordingly, an investigation that initially appears almost trivial, quickly becomes a puzzle with some mismatched pieces. Here, I am of course referring to the occurrence of vowel mergers in many English varieties, which make native-like learning of any variant of English pretty challenging. The focus of my work lay on a contrastive reflection on English and German back vowels which exhibit such mergers. The purpose of this work was hence twofold: to gain clarity on these phenomena and to conduct a concrete experiment to find out how the back vowels of the English language are handled in an L2 context with German native speakers and to train them in perception and production tasks. The idea behind the experiment was to train participants by having them engage directly with the relevant phonetic phenomena. This targeted engagement aimed to foster improvements in both their perceptual abilities and their spoken language production. While the conducted experiment did not yield quantitatively valid results, the gain in knowledge, regarding the patterns of English back vowels was substantial.

For someone specifically interested in the various realisations of English back vowels, many questions arise initially, as the literature on this topic is not as extensive as it is for other, more prominent examples in phonology. Due to the *loose* participation criteria, my group of candidates became very heterogeneous, which means that the qualitative data obtained can provide interesting insights into the linguistic habits of my compatriots. This alone is interesting in this context because to achieve targeted improvements in pronunciation (and listening comprehension), there must be clarity about the general starting conditions. My study shows interesting variances within the test group and thus offers a starting point to further explore the phenomenon of English back vowels. At this point, it is necessary to address the number of participants: While 23 participants in such a linguistic experiment constitute a reasonably good sample, the fact that the collected data was highly diverse, suggests that a larger number of participants would be desirable in future research. Additionally, surveying the participants about whether they identify with a particular variant of English, would be useful to better handle the different speech patterns. I believe one of the greatest challenges during the course of this work was dealing with the different, sometimes mixed, variants of the English language. Existing literature in the field usually distinguishes between a British and an American variant. Such a distinction would have made it easier to classify my results from the beginning. This includes the selection of my native speakers as well as the use of complex words derived from the collected minimal pairs. (For example, the deliberate use of an American native speaker could have been a valid option to specifically highlight the differences between the variants). To avoid such problems in the future, it is advisable to gather information about the participants' linguistic competence in the L2 and thoroughly check the performance of the NS as well. Unfortunately, the questionnaire I created (for exactly that purpose) was somewhat imprecise in this regard and provided information that did not significantly aid in addressing my hypotheses (not to mention that the results did not allow for meaningful statistical tests).

Overall, the general lack of significant results after my training sessions is, of course, unfortunate. However, statistical analyses also suggest that isolated improvements did occur and that a more effective experimental design could have amplified this effect. I am mainly referring to the use of speech tokens from native speakers that were not always suitable, which unfortunately happened in my case. Additionally, the selection of presented stimuli could have been better tailored to the language level of my participants, as some minimal pairs were too complex. Another point is the limited time frame provided to my participants. Longer training sessions would certainly have had a positive effect. When placing the experiment in its appropriate context and considering the shortcomings of the work, speech learning models such as PAM or the SLM can, nonetheless, be applied to the obtained data. These models can provide insights into how the participants performed and why the results occurred in the shape they did. As previously mentioned, it is important to acknowledge all limiting factors that I attempted to highlight in the discussion section. Due to the lack of statistical significance of the intervention, the data analysis is largely confined to evaluating the qualitative data and patterns of the participants. The lack of a positive effect from my training may also be seen as evidence that English back vowels are indeed a complex phenomenon that is difficult for Germans to learn.

On this note, I am pleased to announce that the qualitative data I have gathered largely align with those of other authors who have worked in the field of (English and German) back vowels. However, when compared with other works, it becomes clear that the implementation of my experiment did not proceed optimally, and several errors that were my own fault ultimately prevented my laboriously constructed experiment from yielding meaningful results. However, it is important to consider that this work was conducted under a strict time line and that the learning curve for such a master's project is indeed very steep. Entering the field of L2 speech learning may seem complex and has little to do with the preconceptions described in the introduction that make engaging with this area appear easy. As the phenomenon of English and German back vowels has received little attention in linguistic literature to date, I can only encourage further research in this area. They say that one learns from one's mistakes, and so, upon completing this work, I have identified many points of connection that would justify the initiation of further studies. It can only be an advantage that the number of topics from which future work can be drawn is seemingly limitless. With the tools I have acquired in the course of this work, I find further exploration of this area both exciting and beneficial for my own, daily language use.

Works cited

Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research*, *10*(3), 215–240. https://doi.org/10.1177/026765839401000303

Becker, D. (2021). *Green Line—Oberstufe. Schülerbuch] / von Dr. Daniel Becker [und 15 weiteren* (Ausgabe für Baden-Württemberg, [ab 2021], 1. Auflage). Ernst Klett Verlag.

Bent, T., Bradlow, A. R., & Smith, B. (2007). Phonemic errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In O.-S. Bohn & M. J. Munro (Hrsg.), *Language Learning & Language Teaching* (Bd. 17, S. 331–347). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.28ben

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360. https://doi.org/10.1037/0096-1523.14.3.345

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Hrsg.), *Language Learning & Language Teaching* (Bd. 17, S. 13–34). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.07bes

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn & M. J. Munro (Hrsg.), *Language Learning & Language Teaching* (Bd. 17, S. 99–116). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.12bir

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Hrsg.), *Linguistic Perspectives on Second Language Acquisition* (1. Aufl., S. 41–68). Cambridge University Press. https://doi.org/10.1017/CBO9781139524544.005

Bloomfield, L. (1933). *Language*. University of Chicago Press.

Bohn, O.-S., & Flege, J. E. (1992). The Production of New and Similar Vowels by Adult German Learners of English. *Studies in Second Language Acquisition*, *14*(2), 131–158. https://doi.org/10.1017/S0272263100010792

Bohn, O.-S., Flege, J. E., & Munro, M. J. (Hrsg.). (2007). *Language experience in second language speech learning: In honor of James Emil Flege*. Benjamins.

Bongaerts, T., Mennen, S., & Slik, F. V. D. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica*, *54*(2), 298–308. https://doi.org/10.1111/1467-9582.00069

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, *97*(3), 1916–1924. https://doi.org/10.1121/1.412064

Brière, E. J. (1966). QUANTITY BEFORE QUALITY IN SECOND LANGUAGE COMPOSITION. *Language Learning*, *16*(3–4), 141–151. https://doi.org/10.1111/j.1467-1770.1966.tb00816.x

Brown, K., & Kachru, B. B. (1995). The Other Tongue: English across Cultures. *TESOL Quarterly*, *29*(1), 204. https://doi.org/10.2307/3587821

Catford, J. C. (1988). *A practical introduction to phonetics*. Clarendon Press ; Oxford University Press.

Chomsky, N. (1957). *Syntactic Structures*. De Gruyter. https://doi.org/10.1515/9783112316009

Chomsky, N. (1965). Persistent Topics in Linguistic Theory. *Diogenes*, *13*(51), 13–20. https://doi.org/10.1177/039219216501305102

Chomsky, N., Leuninger, H., & Chomsky, N. (1981). *Regeln und Repräsentationen* (1. Aufl). Suhrkamp.

Collins, B., Mees, I. M., & Carley, P. (2019). *Practical English phonetics and phonology: A resource book for students* (Fourth edition). Routledge, Taylor & Francis Group.

Cruttenden, A., & Gimson, A. C. (2008). *Gimson's pronunciation of English* (7th ed). Hodder Education.

Diehl, R. L., Elman, J. L., & Buchwald, S. (1977). Phonetic context effects on the identification of stop consonants. *The Journal of the Acoustical Society of America*, *62*(S1), S78–S79. https://doi.org/10.1121/1.2016382

Diller, K. (1981). SYNTAX, SEMANTICS, AND PRAGMATICS—SECOND LANGUAGE: GENERAL DISCUSSION. *Annals of the New York Academy of Sciences*, *379*(1), 291–295. https://doi.org/10.1111/j.1749-6632.1981.tb42016.x

Flege, J. E., Aoyama, K., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r) Applied. In R. Wayland (Hrsg.), *Second Language Speech Learning* (1. Aufl., S. 84–118). Cambridge University Press. https://doi.org/10.1017/9781108886901.003

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Hrsg.), *Second Language Speech Learning* (1. Aufl., S. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 101–131). Lawrence Erlbaum Associates Publishers.

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, *16*(1), 1–26. https://doi.org/10.1016/0167-6393(94)00044-B

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, *84*(1), 70–79. https://doi.org/10.1121/1.396876

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America*, *76*(3), 708–721. https://doi.org/10.1121/1.391257

Flege, J. E., & Port, R. (1981). Cross-Language Phonetic Interference: Arabic to English. *Language and Speech*, *24*(2), 125–146. https://doi.org/10.1177/002383098102400202

Frieda, E. M., & Nozawa, T. (2007). You are what you eat phonetically: The effect of linguistic experience on the perception of foreign vowels. In O.-S. Bohn & M. J. Munro (Hrsg.), *Language Learning & Language Teaching* (Bd. 17, S. 79–96). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.10fri

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*, *9*(3), 317–323. https://doi.org/10.1016/0028-3932(71)90027-3

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149. https://doi.org/10.1017/S136672899800025X

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Hrsg.), *Language Learning & Language Teaching* (Bd. 17, S. 57–77). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.09gui

Hickey, R. (Hrsg.). (2019). *English in the German-Speaking World* (1. Aufl.). Cambridge University Press. https://doi.org/10.1017/9781108768924

Hilgendorf, S. K. (2005). „Brain Gain statt[instead of]Brain Drain“: The role of English in German education. *World Englishes*, *24*(1), 53–67. https://doi.org/10.1111/j.0883-2919.2005.00387.x

Hockett, C. F. (1958). *A course of modern linguistics*. New York, NY: Macmillan.

Hüllen, W. (2005). *Kleine Geschichte des Fremdsprachenlernens*. Schmidt.

Hyltenstam, Kenneth & Abrahamsson, Niclas. (2003). Age of onset and ultimate attainment in near-native speakers of Swedish as a second language. in: K. Hyltenstam & K. Fraurud (eds.), Multilingualism in Global and Local Perspectives. Stockholm: Stockholm University & City of Stockholm (pp. 319-340).

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877. https://doi.org/10.1121/1.3148196

Johnson, Jim (Regisseur). (2024, May 22). *Hell’s Corner IPA: [ɑ] [ɒ] [ɔ] Vowels* [YouTube Video]. https://www.youtube.com/watch?v=sALAE6iqLWk

Jørgensen, H. P. (1969). Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantenfrequenzen. *Phonetica*, *19*(4), 217–245. https://doi.org/10.1159/000258629

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832. https://doi.org/10.1121/1.4926561

Kautzsch, A. (2010). Exploring L1 Transfer in German Learners of English: High Front Vowels, High Back Vowels and the BED/BAD Distinction. *Research in Language*, *8*, 63–84. https://doi.org/10.2478/v10015-010-0003-8

Kautzsch, A. (2012). Transfer, Similarity or Lack of Awareness? Inconsistencies of German Learners in the Pronunciation of LOT, THOUGHT, STRUT, PALM and BATH. *Research in Language*, *10*(2), 225–241. https://doi.org/10.2478/v10015-011-0029-6

Keetels, M., Schakel, L., Bonte, M., & Vroomen, J. (2016). Phonetic recalibration of speech by text. *Attention, Perception, & Psychophysics*, *78*(3), 938–945. https://doi.org/10.3758/s13414-015-1034-y

Kohler, K. J. (1981). Contrastive Phonology and the Acquisition of Phonetic Skills. *Phonetica*, *38*(4), 213–226. https://doi.org/10.1159/000260025

Kohler, K. J. (1995). *Einführung in die Phonetik des Deutschen* (2., neubearb. Aufl). E. Schmidt.

Kortmann, B. (2020). *English linguistics: Essentials* (2nd revised, updated and enlarged edition). J.B. Metzler Verlag.

Krashen, S. (1982). Accounting for child-adult differences in second language rate and attainment. InChild-Adult Differences in Second Language Acquisition, Stephen D. Krashen, Robin C. Scar-cella and Michael H. Long (eds.), 202–226. Rowley, MA: Newbury House.

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Mouton de Gruyter. https://doi.org/10.1515/9783110167467

Lado, R. (1957). *Linguistics across Cultures: Applied Linguistics and Language Teachers*. University of Michigan Press, Ann Arbor.

Lenneberg, E. H., & Chomsky, N. (1967). *Biological foundations of language*. Wiley.

Levis, J. M., & Zhou, Z. (2018). Accent. In J. I. Liontas (Hrsg.), *The TESOL Encyclopedia of English Language Teaching* (1. Aufl., S. 1–5). Wiley. https://doi.org/10.1002/9781118784235.eelt0002

Lindsey, G., & Wells, J. C. (2019). *English after RP: Standard British pronunciation today*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-04357-5

Lo, C. H., Hermes, J., Kartushina, N., Mayor, J., & Mani, N. (2023). e-Babylab: An open-source browser-based tool for unmoderated online developmental studies. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02200-7

Martohardjono, G., & Flynn, S. (1995). 6 Is There an Age-Factor for Universal Grammar? In D. Singleton & Z. Lengyel (Hrsg.), *The Age Factor in Second Language Acquisition* (S. 135–153). Multilingual Matters. https://doi.org/10.21832/9781800418240-008

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories* (Third edition). Routledge.

Mitterer, H., Eger, N. A., & Reinisch, E. (2020). My English sounds better than yours: Second-language learners perceive their own accent as better than that of their peers. *PLOS ONE*, *15*(2), e0227643. https://doi.org/10.1371/journal.pone.0227643

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331–340. https://doi.org/10.3758/BF03211209

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different conditions: The use of non-lexical materials and masking noise. *Phonetica*, *79*(1), 1–43. https://doi.org/10.1515/phon-2022-2018

Moreno-Torres, I., Mariën, P., Dávila, G., & Berthier, M. L. (2016). Editorial: Language beyond Words: The Neuroscience of Accent. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00639

Munro, M. J., & Bohn, O.-S. (2007). The study of second language speech learning: A brief overview. In O.-S. Bohn & M. J. Munro (Hrsg.), *Language Learning & Language Teaching* (Bd. 17, S. 3–11). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.06mun

Roach, P. (2004). British English: Received Pronunciation. *Journal of the International Phonetic Association*, *34*(2), 239–245. https://doi.org/10.1017/S0025100304001768

Schachter, J. (1988). Second Language Acquisition and Its Relationship to Universal Grammar. *Applied Linguistics*, *9*(3), 219–235. https://doi.org/10.1093/applin/9.3.219

Schlechtweg, M., Peters, J., & Frank, M. (2023). L1 variation and L2 acquisition: L1 German /eː/-/ɛː/ overlap and its effect on the acquisition of L2 English /ɛ/-/æ/. *Frontiers in Psychology*, *14*, 1133859. https://doi.org/10.3389/fpsyg.2023.1133859

Schweinberger, Martin. (2024, Mai 14). *Schweinberger, Martin. 2016. Creating a personalized Vowel Chart with PRAAT and R.* http://www.martinschweinberger.de/blog/persvowelchart/, date.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech* (1. print). Newbury House Publ.

Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours. *IRAL - International Review of Applied Linguistics in Language Teaching*, *43*(4), 269–285. https://doi.org/10.1515/iral.2005.43.4.269

Sönning, L. (2020). *Phonological variation in German Learner English* [Otto-Friedrich-Universität]. https://doi.org/10.20378/irb-49135

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131–145. https://doi.org/10.3758/BF03202673

Trubetzkoy, N. (1939). Grundzüge der Phonologie. *Travaux du cercle linguistique de Prague 7.Christiane Baltaxe, trans. Principles of Phonology.* Berkeley: University *of California* Press. 1969.

Van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*. Weinreich, U. (2010). *Languages in contact: Findings and problems*. De Gruyter Mouton.

Vroomen, J., & Baart, M. (2011). Phonetic Recalibration in Audiovisual Speech. In M. Murray & M. Wallace, *The Neural Bases of Multisensory Processes* (Bd. 20111651, S. 363–380). CRC Press. https://doi.org/10.1201/9781439812174-24

Wiese, R. (2006). *The phonology of German* (Reprinted). Oxford Univ. Press.

Wiese, R. (2010). *Phonetik und Phonologie* (1. Aufl.). utb GmbH. https://doi.org/10.36198/9783838533544

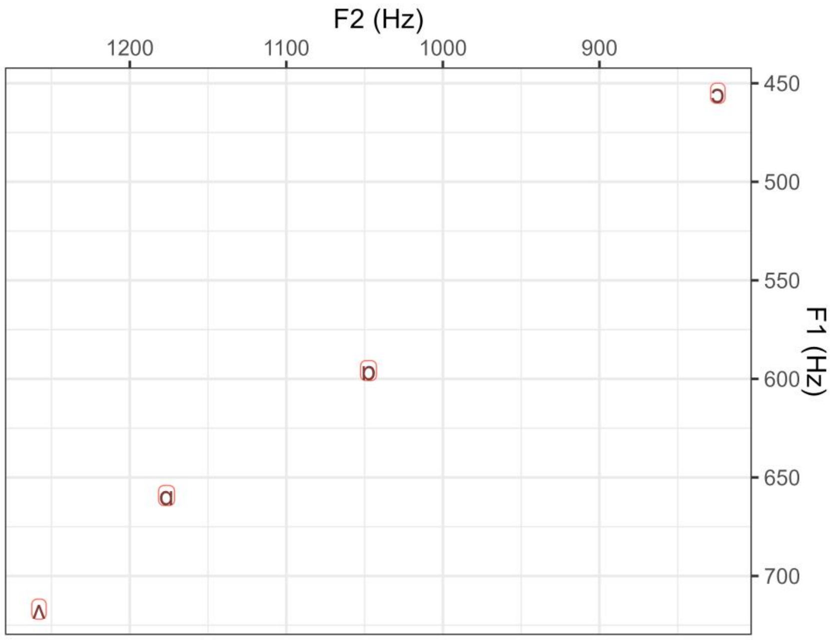
Zsiga, E. C. (2013). *The sounds of language: An introduction to phonetics and phonology*. Wiley-Blackwell.

Appendix 1

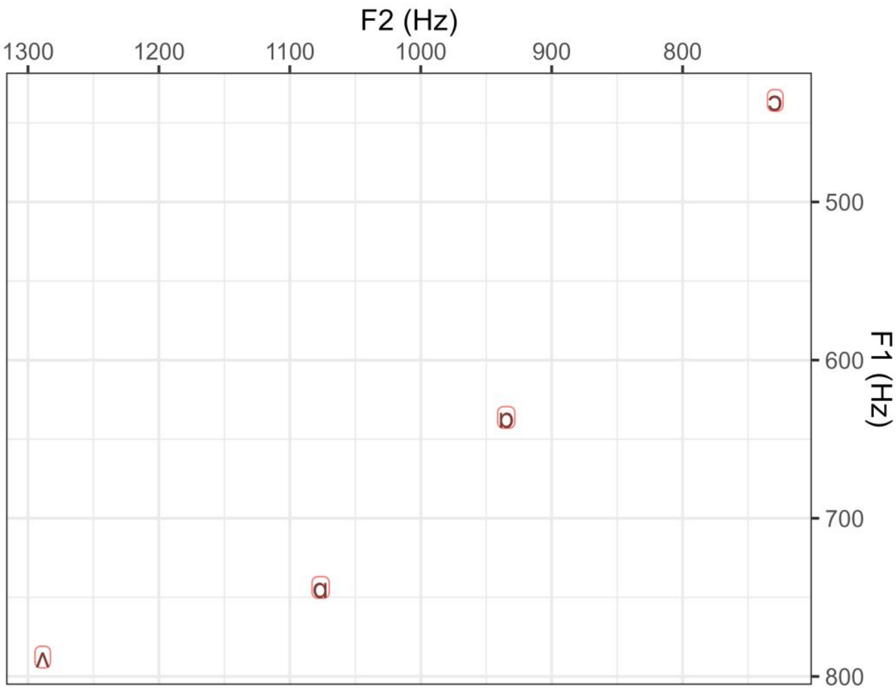
Individual formant values, pre and post in an *alphabetical* order (this allows the author to retrace the individual graphs)

Female participant 1

PreTest

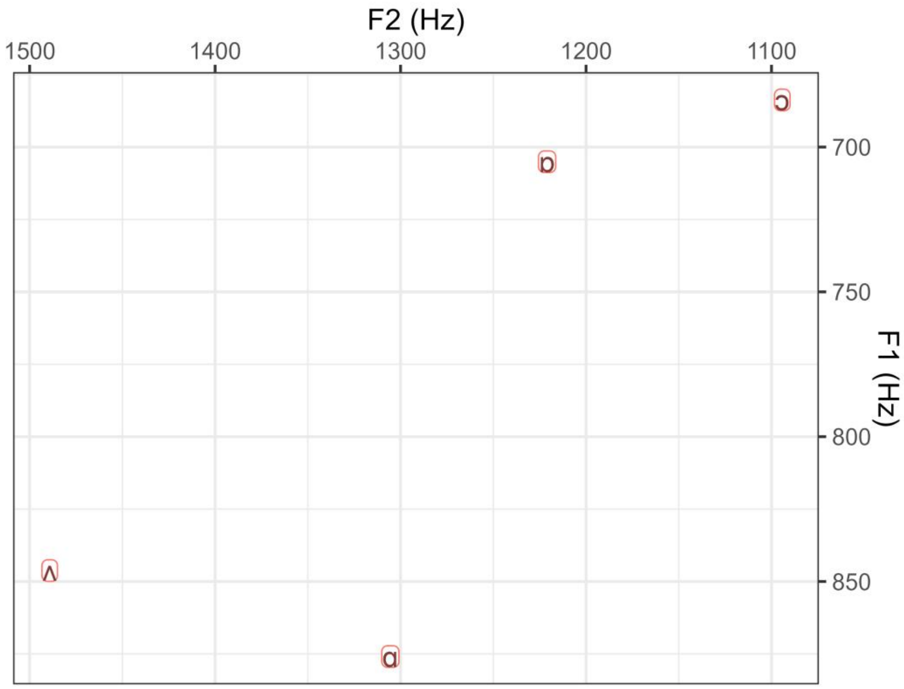


PostTest:

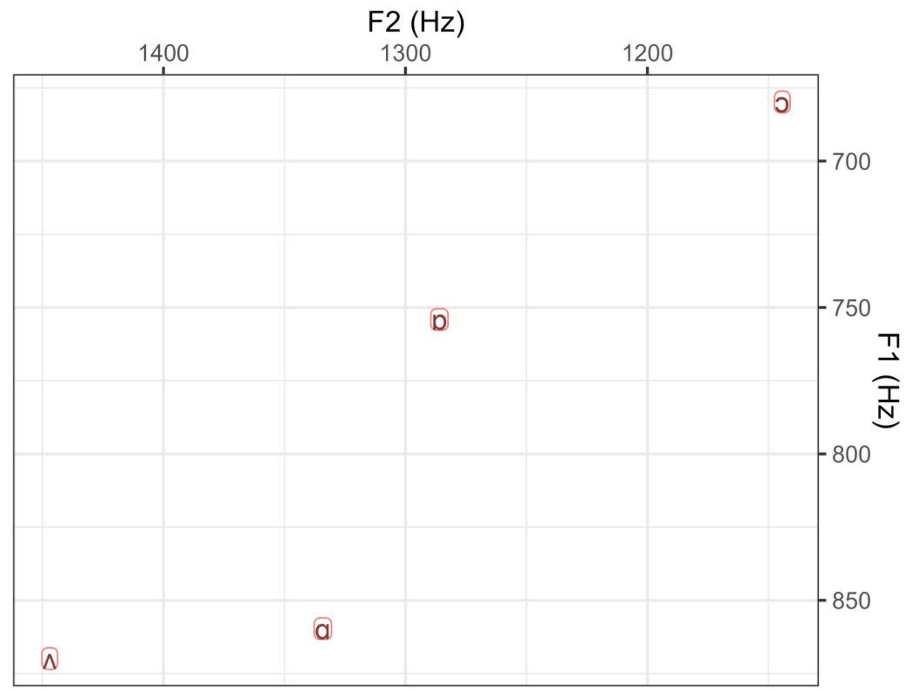


Female participant 2

PreTest:

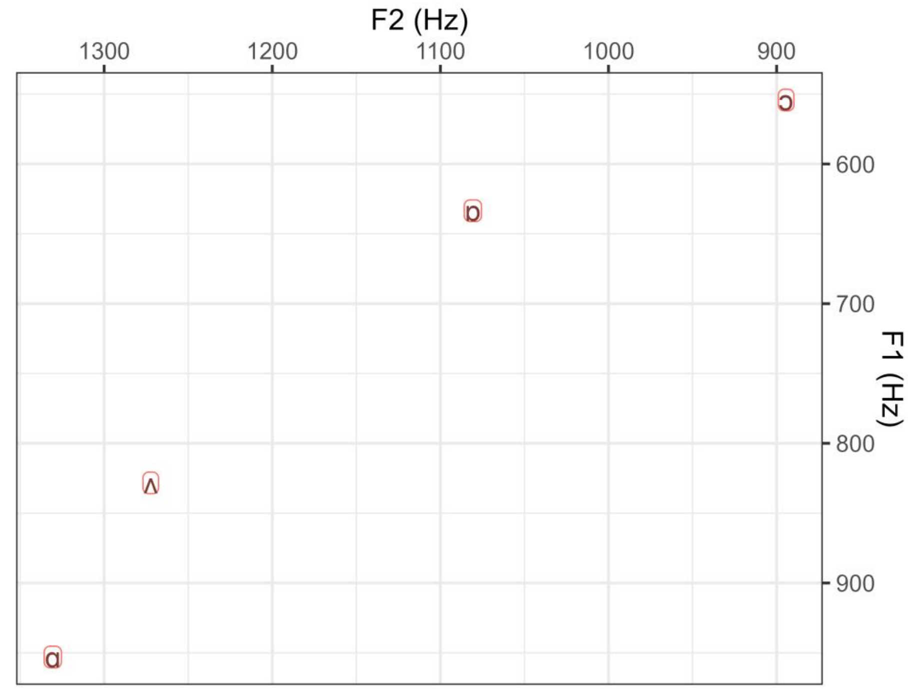


PostTest:

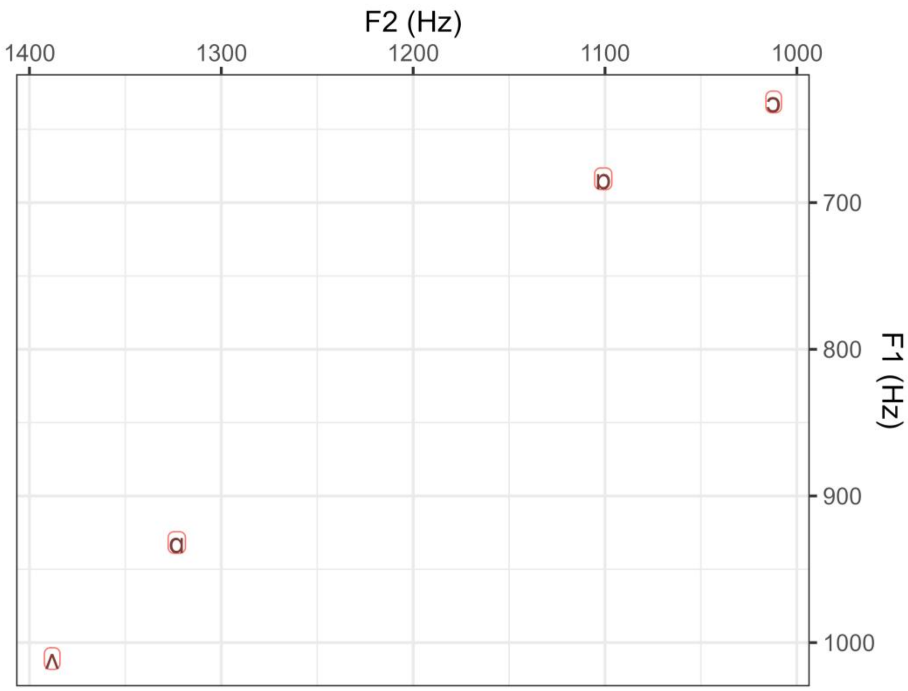


Female participant 3

PreTest:

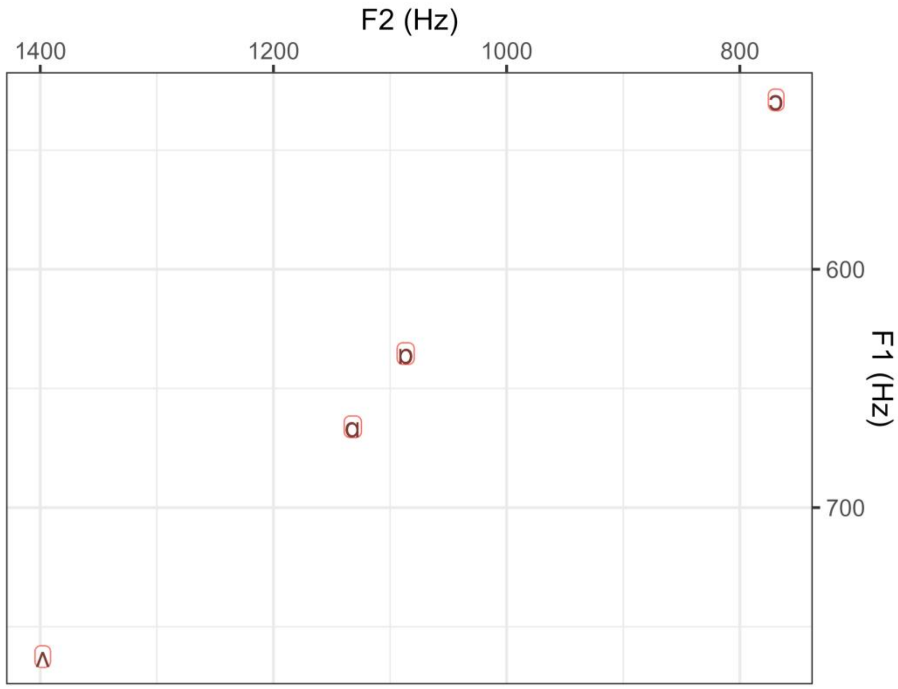


PostTest:

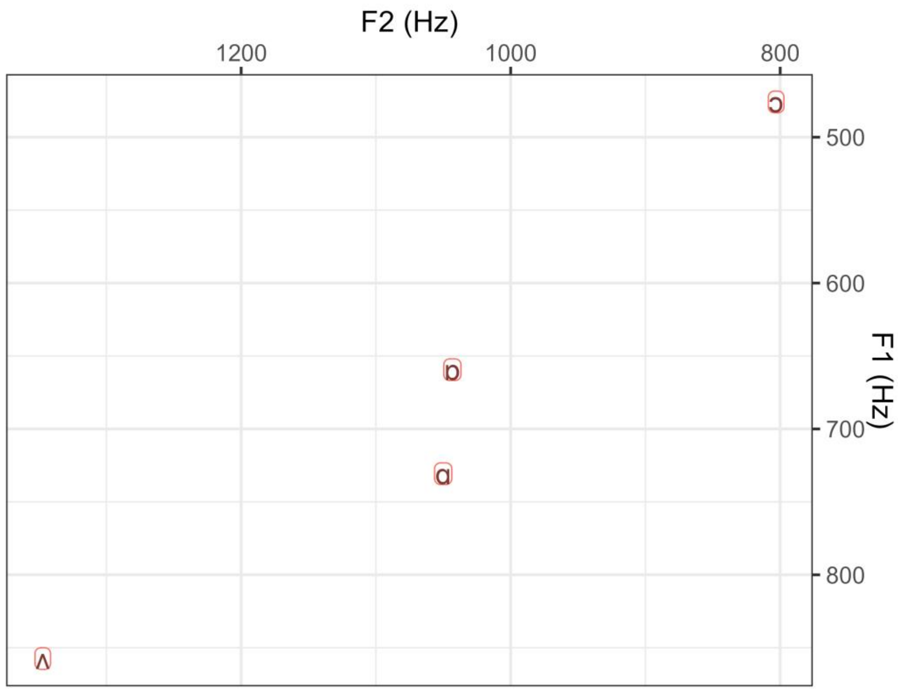


Male participant 1

PreTest:

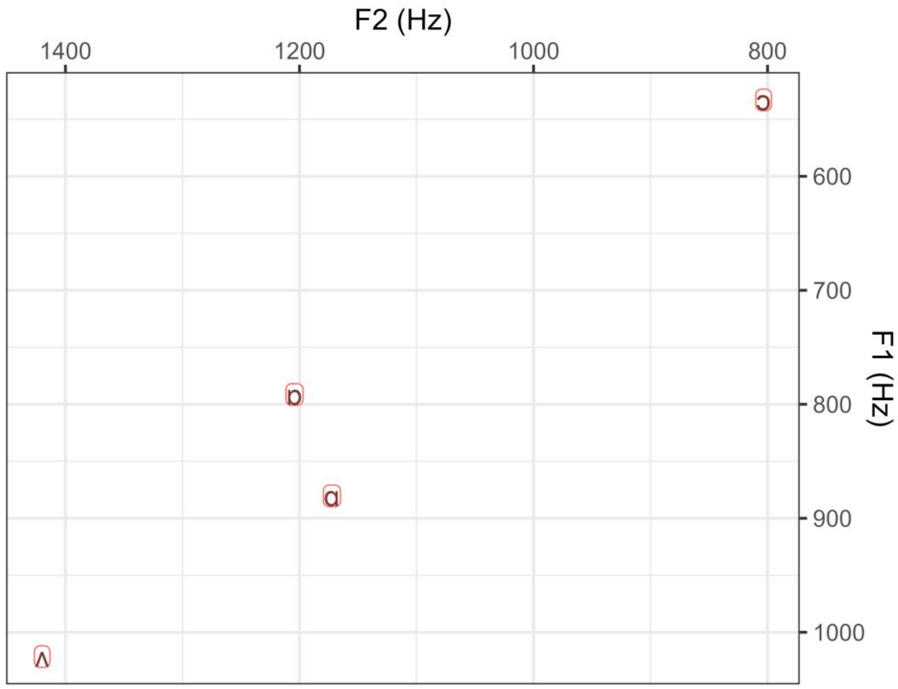


PostTest:

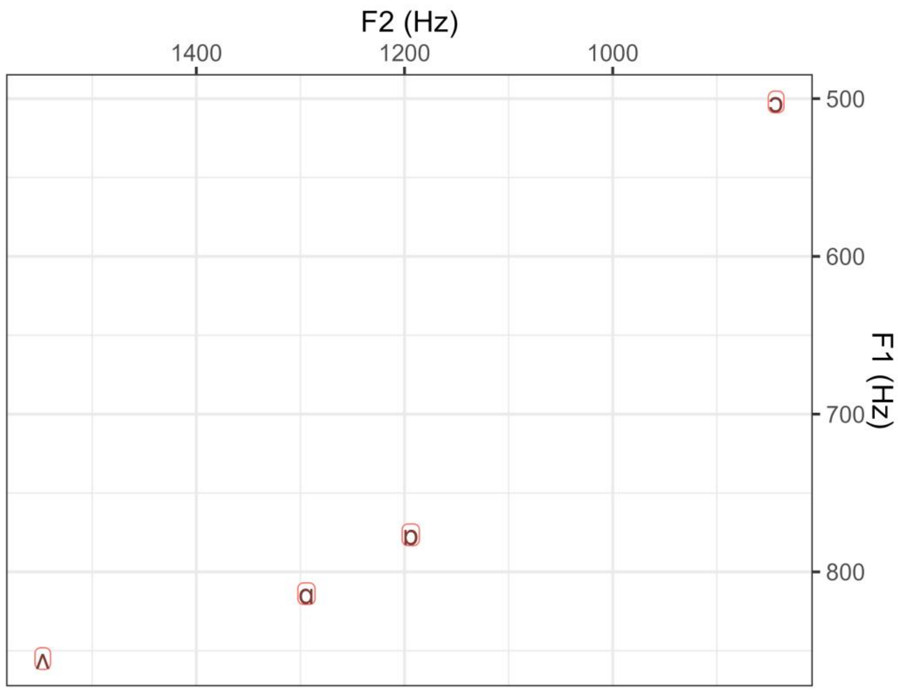


Female participant 4

PreTest:

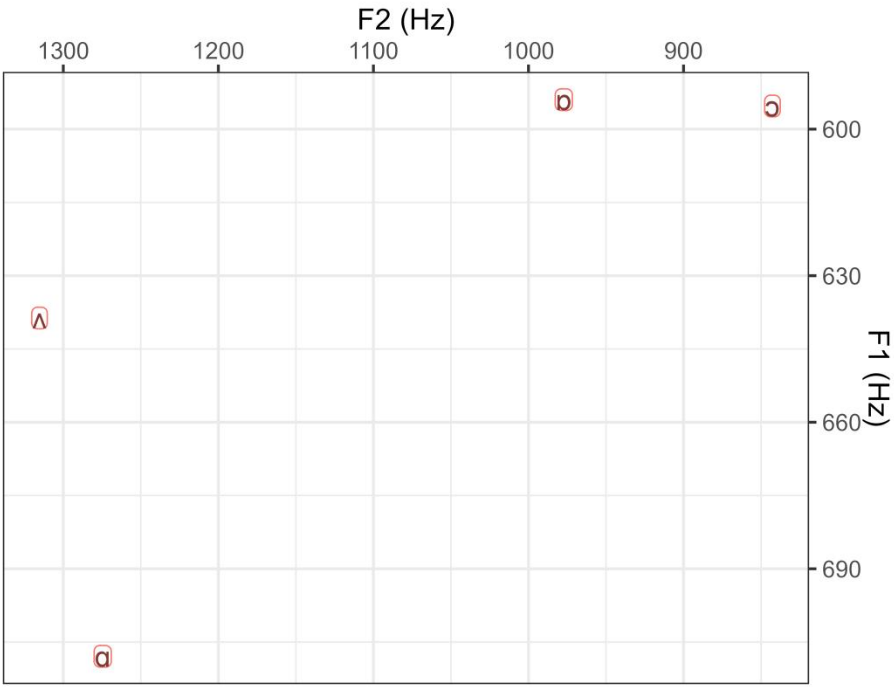


PostTest:

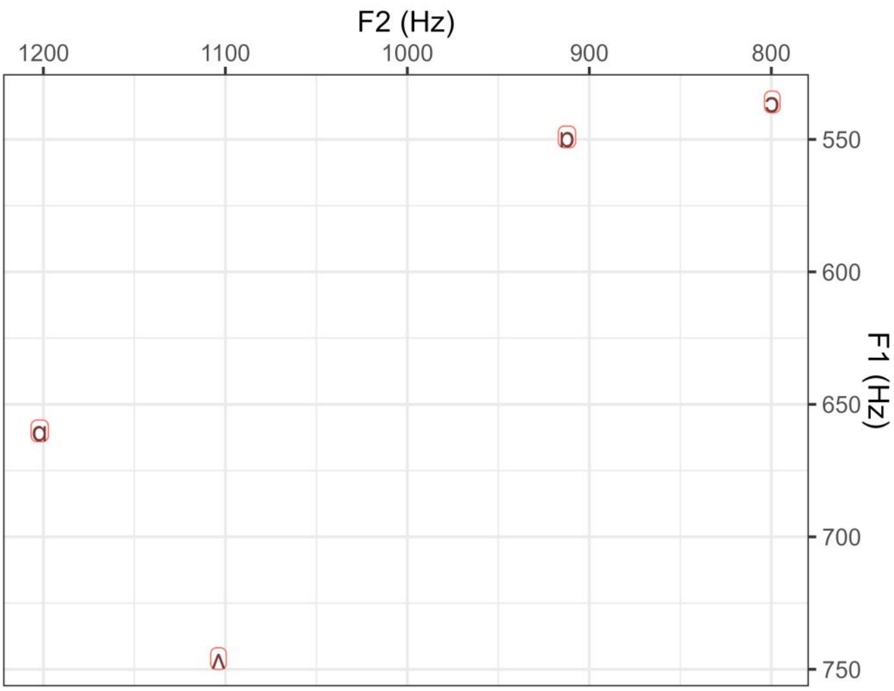


Male participant 2

PreTest:

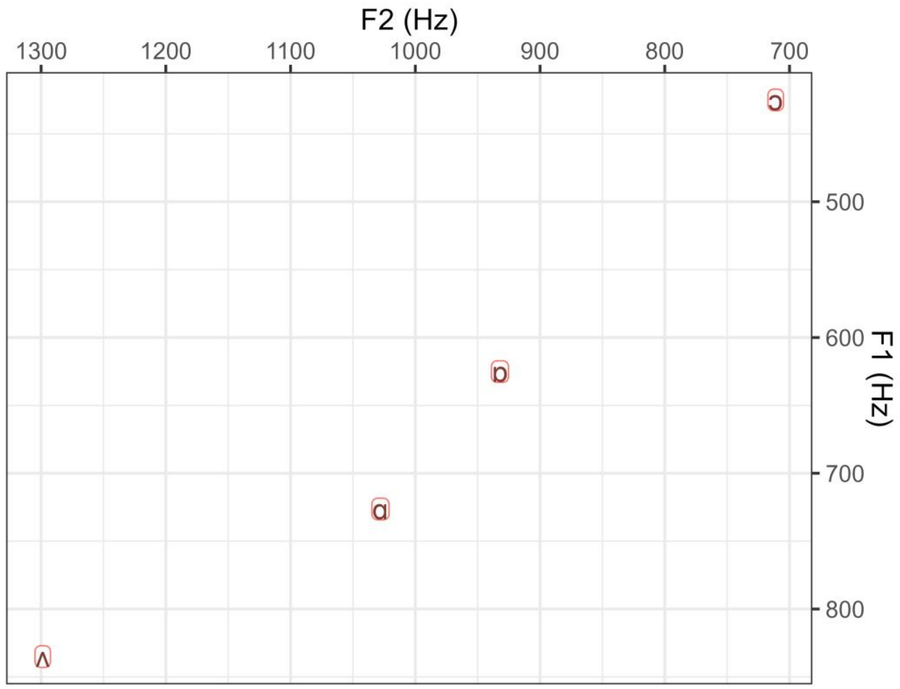


PostTest:

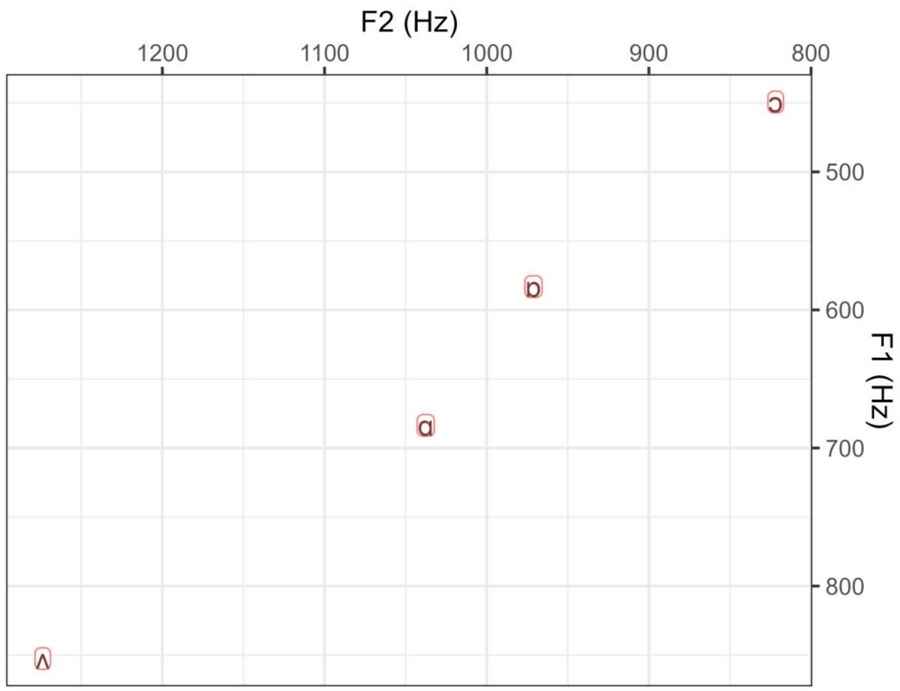


Male participant 3

Pretest:

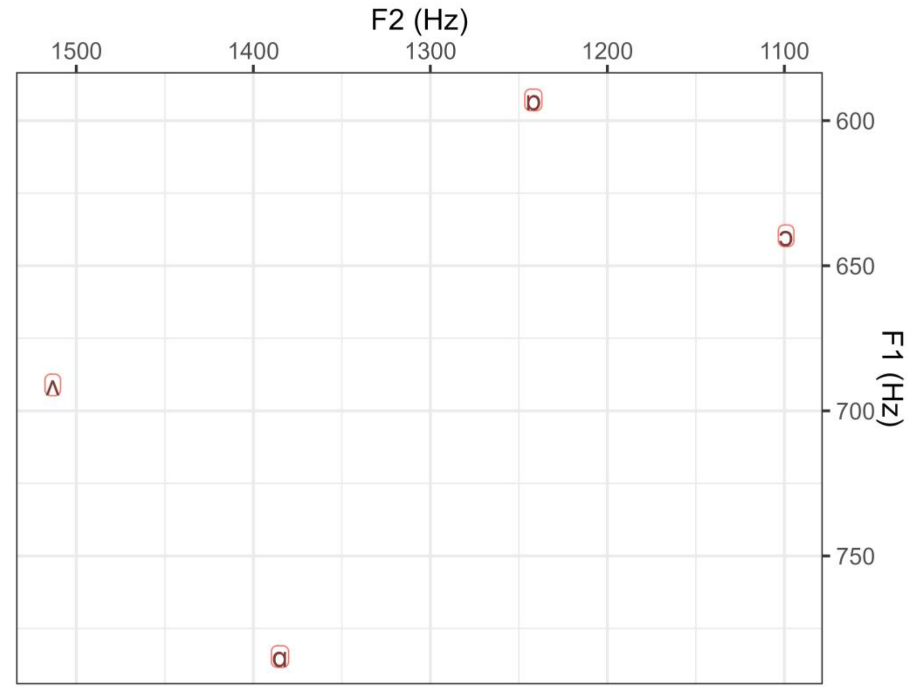


PostTest:

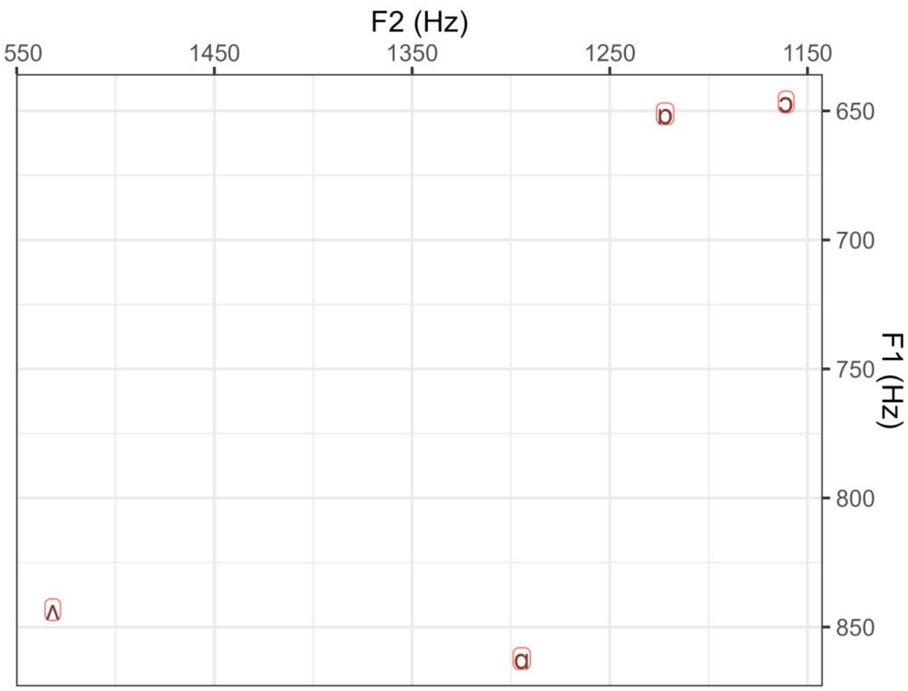


Female participant 5

PreTest:

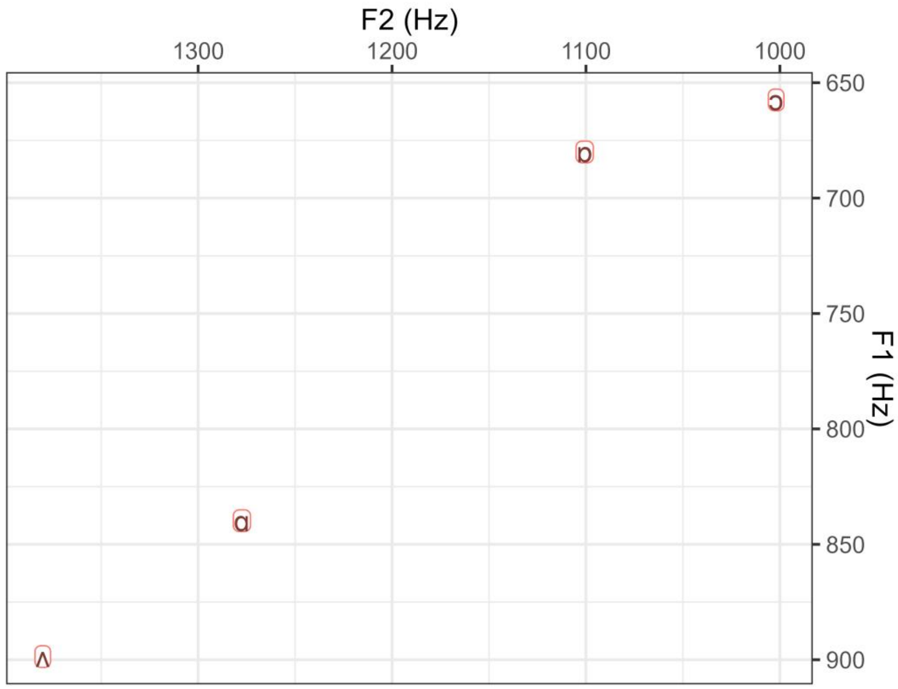


PostTest:

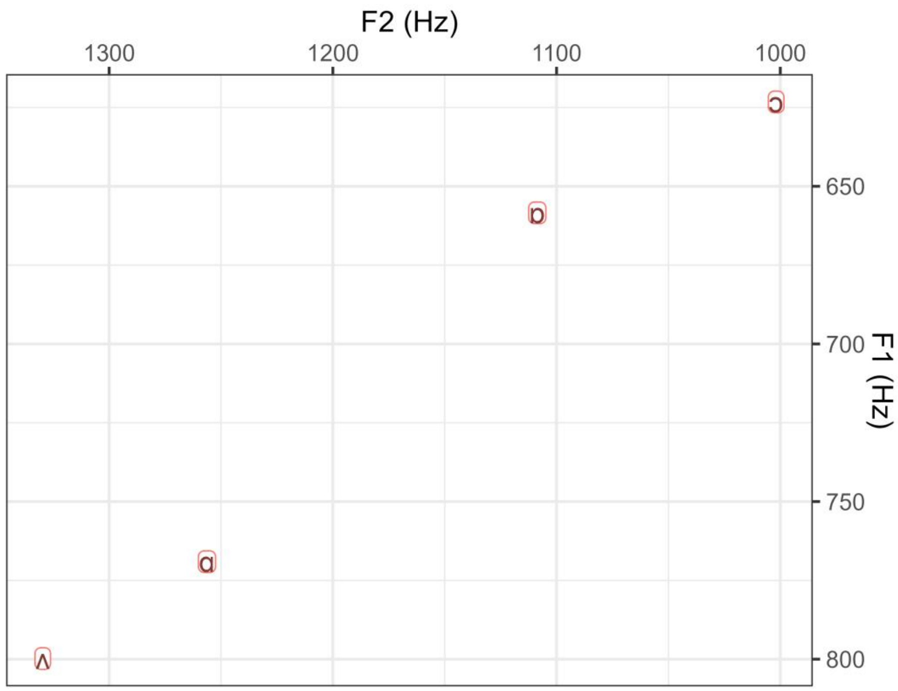


Female participant 6

PreTest

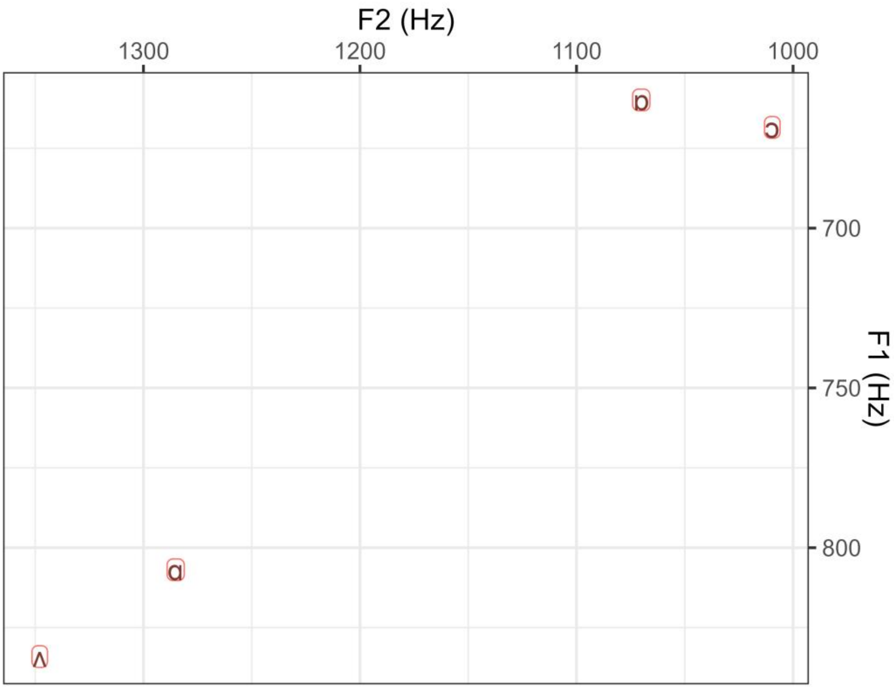


PostTest

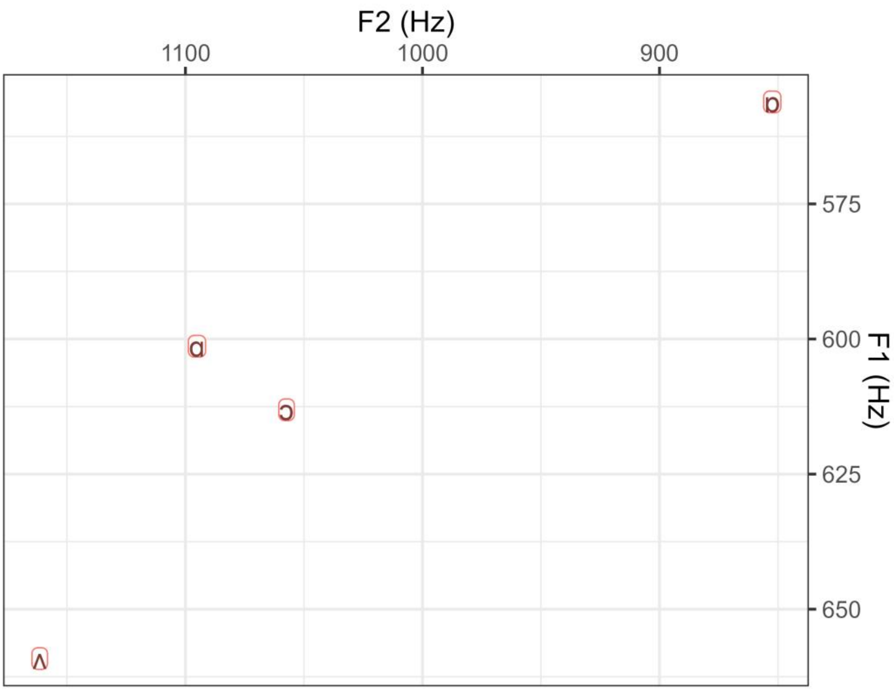


Female participant 7

PreTest

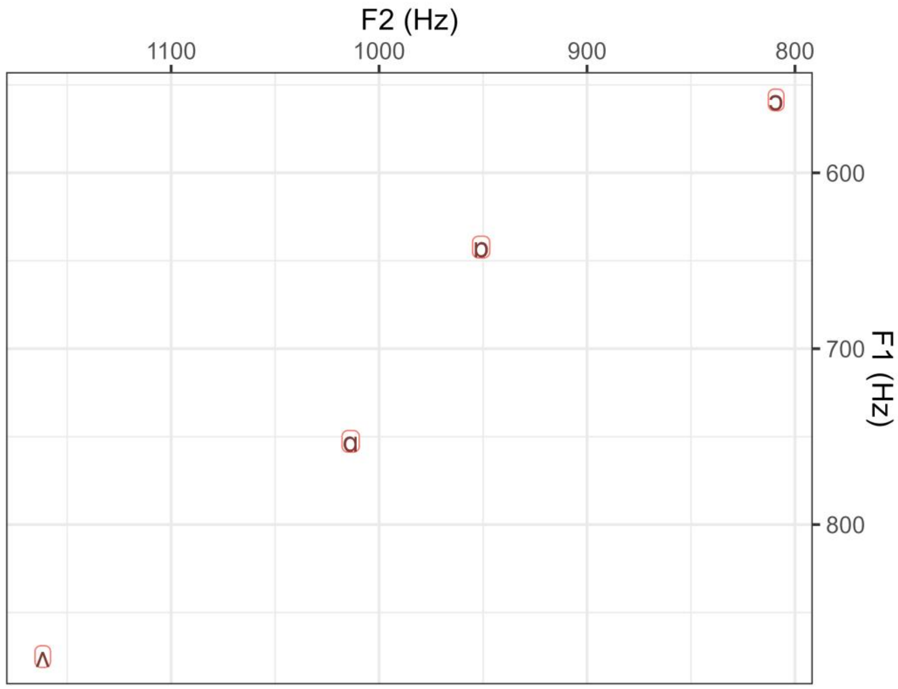


PostTest

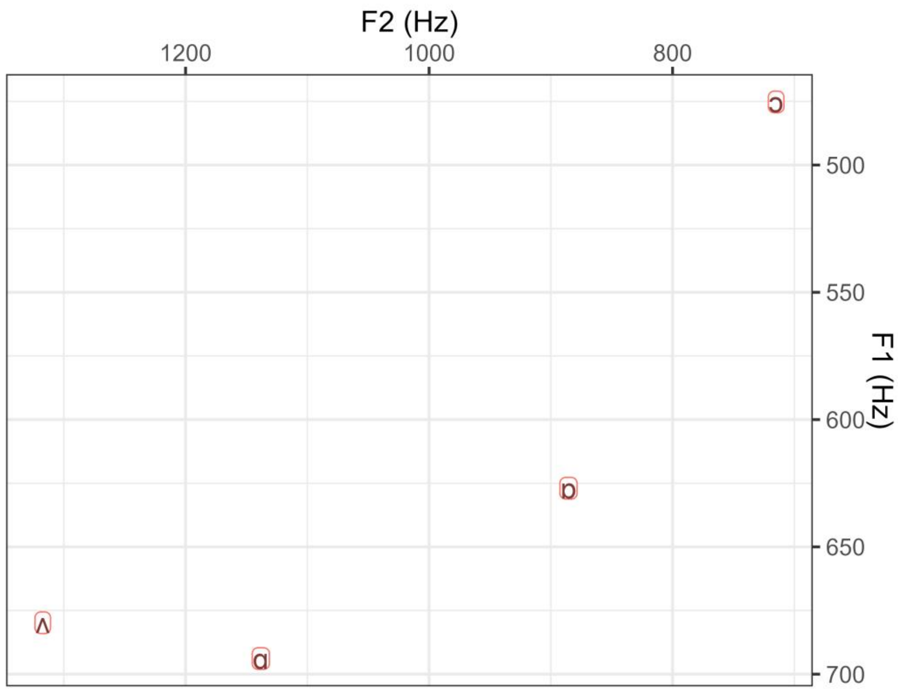


Male participant 4

PreTest

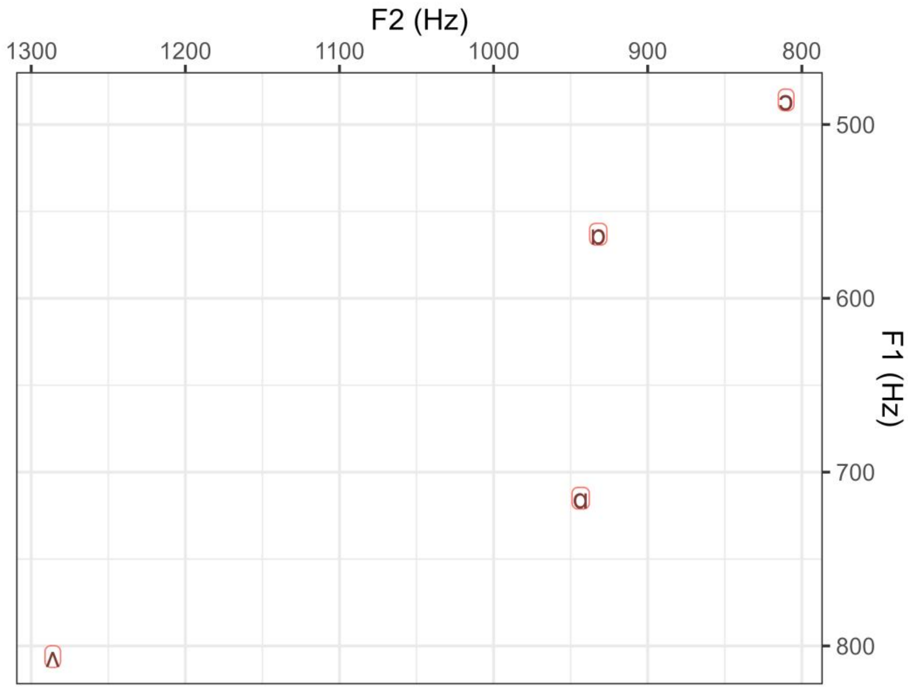


PostTest

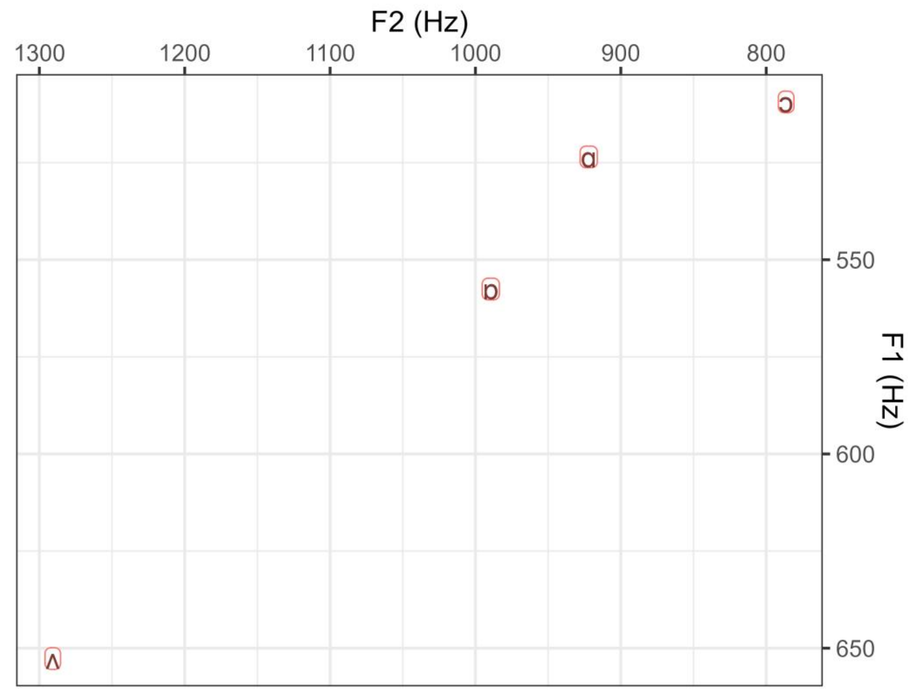


Male participant 5

PreTest

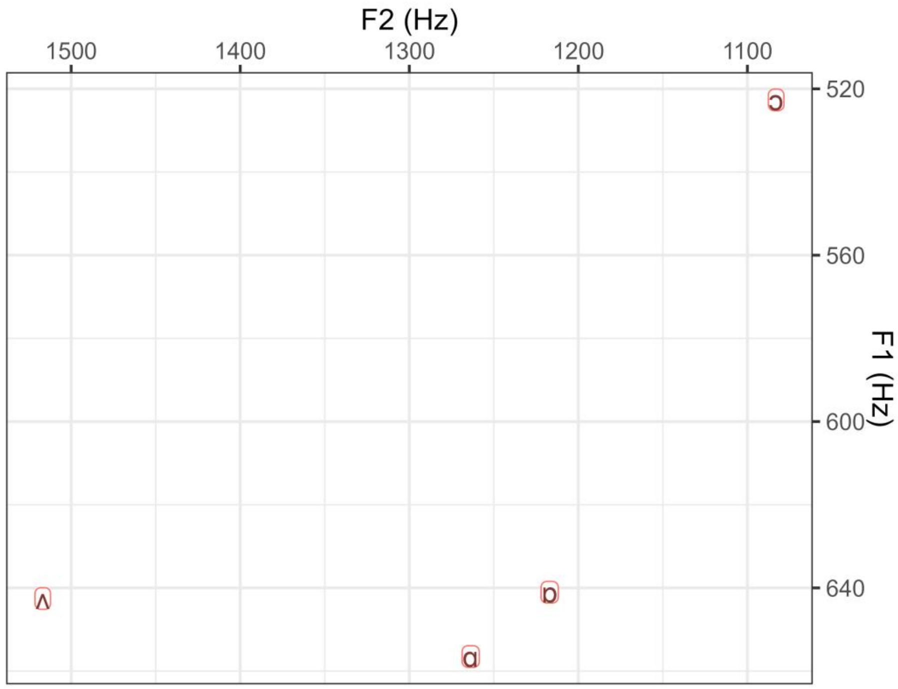


PostTest

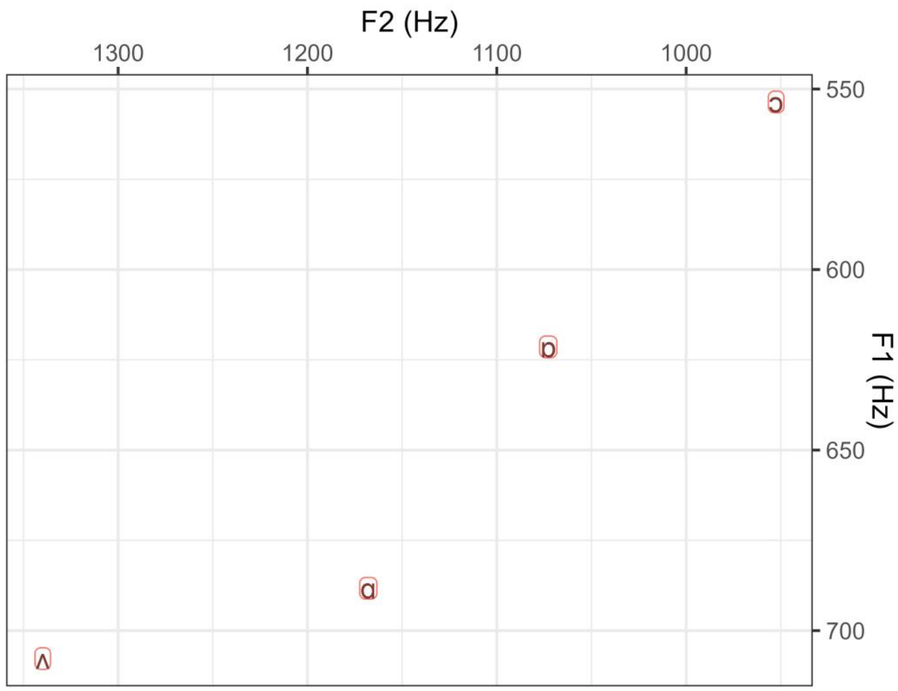


Female participant 9

Pretest

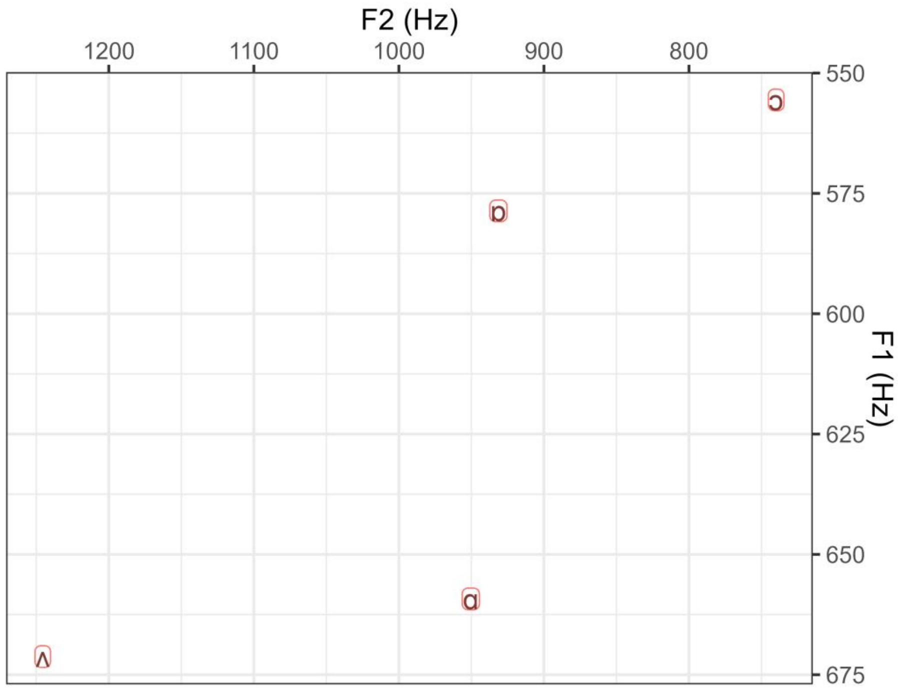


PostTest

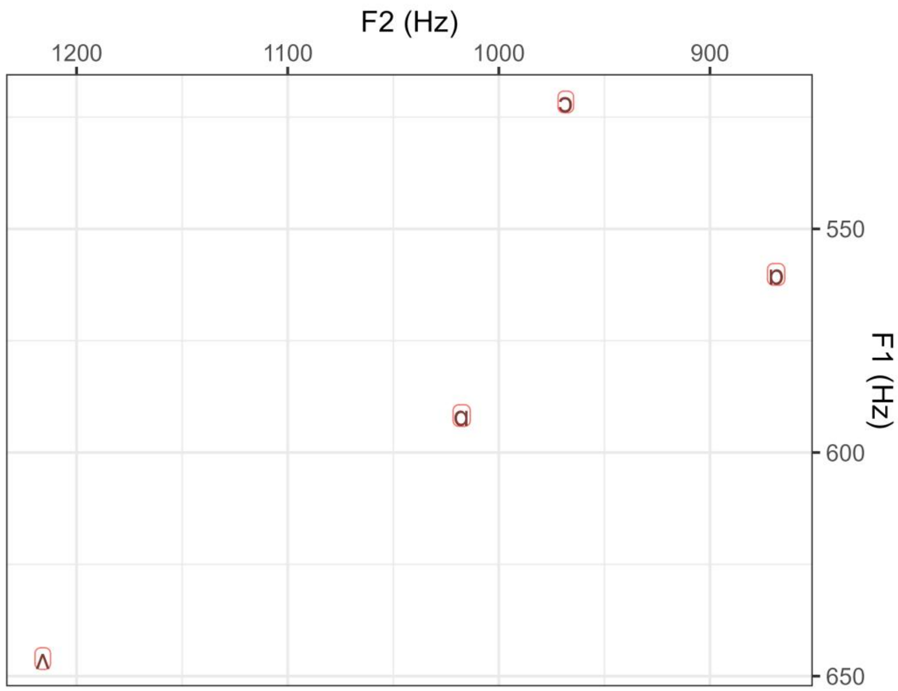


Male participant 6

Pretest

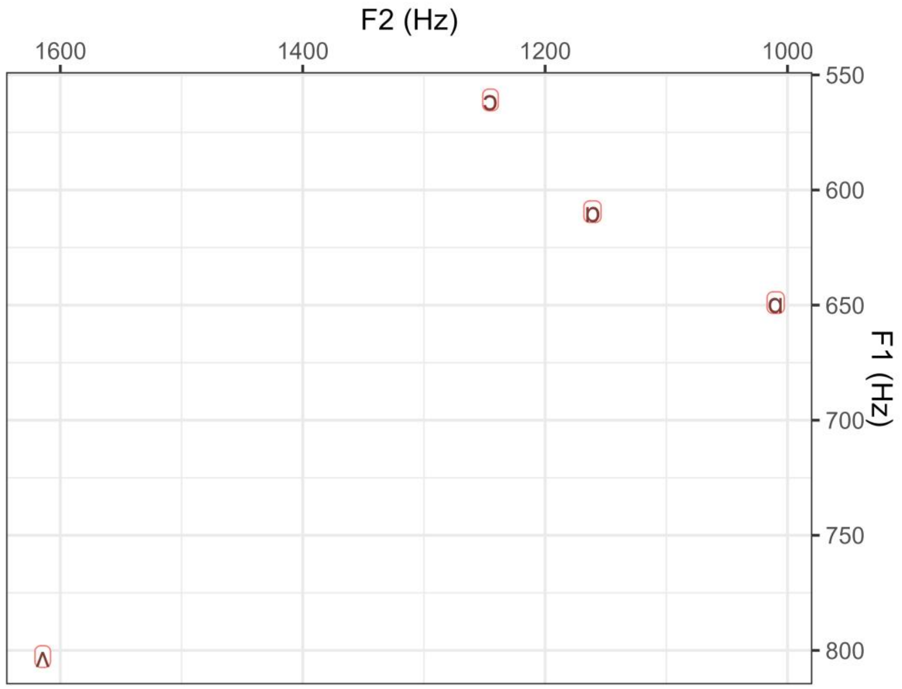


Postest

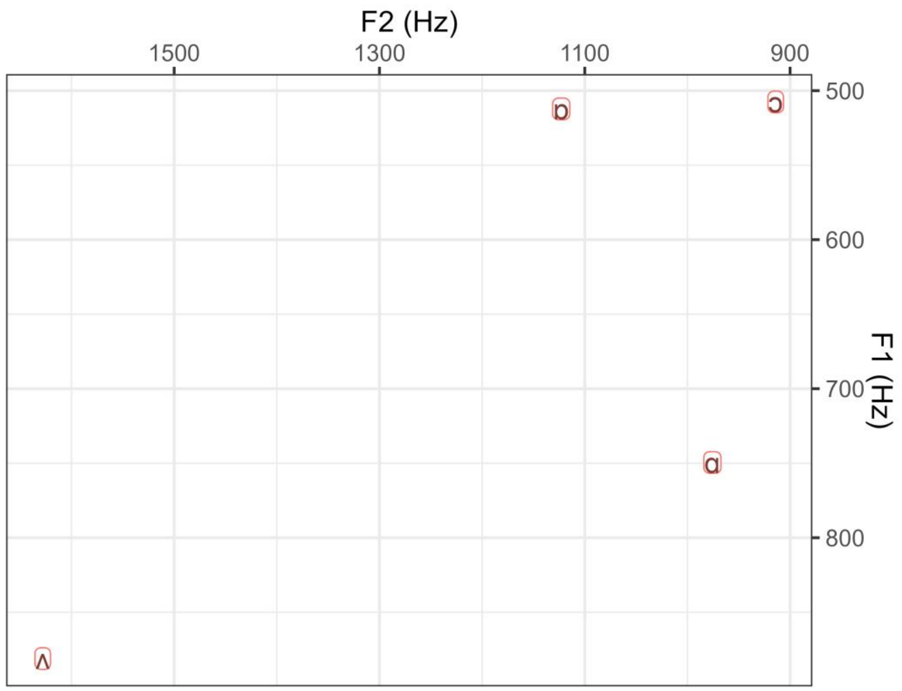


Female participant 10

Pretest

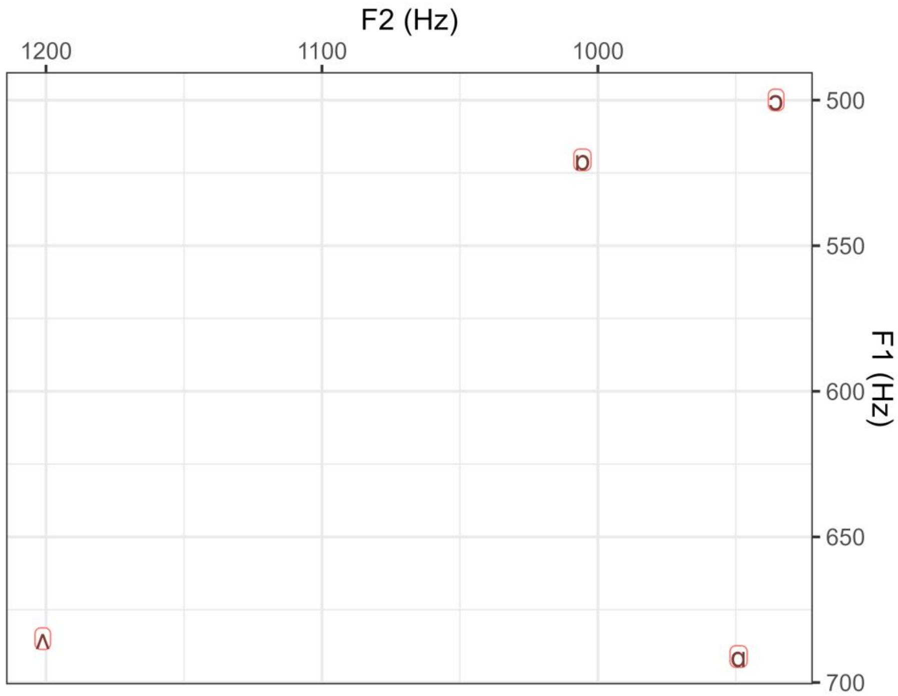


Posttest

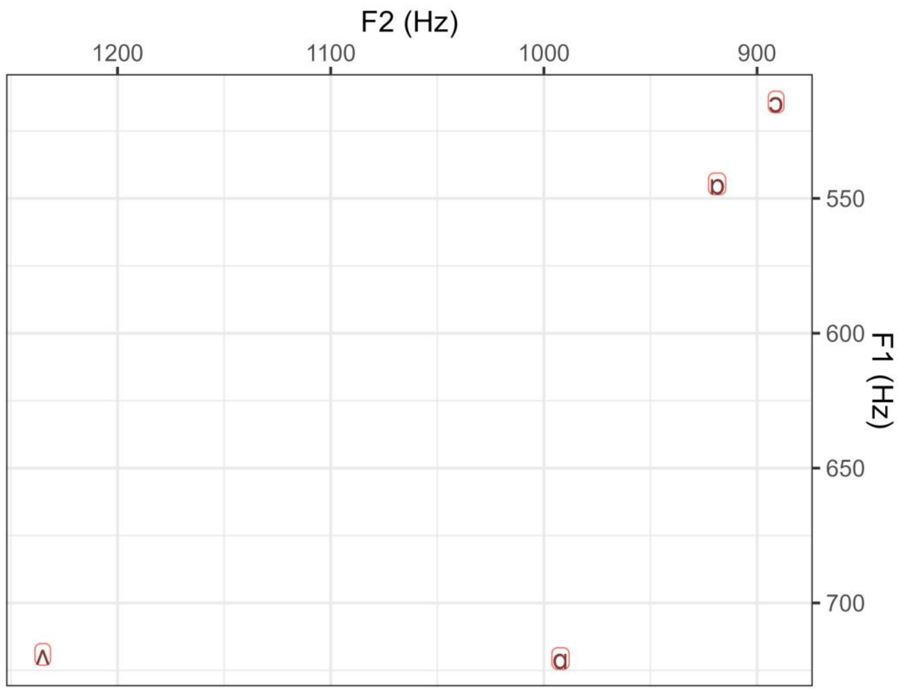


Male participant 7

Pretest

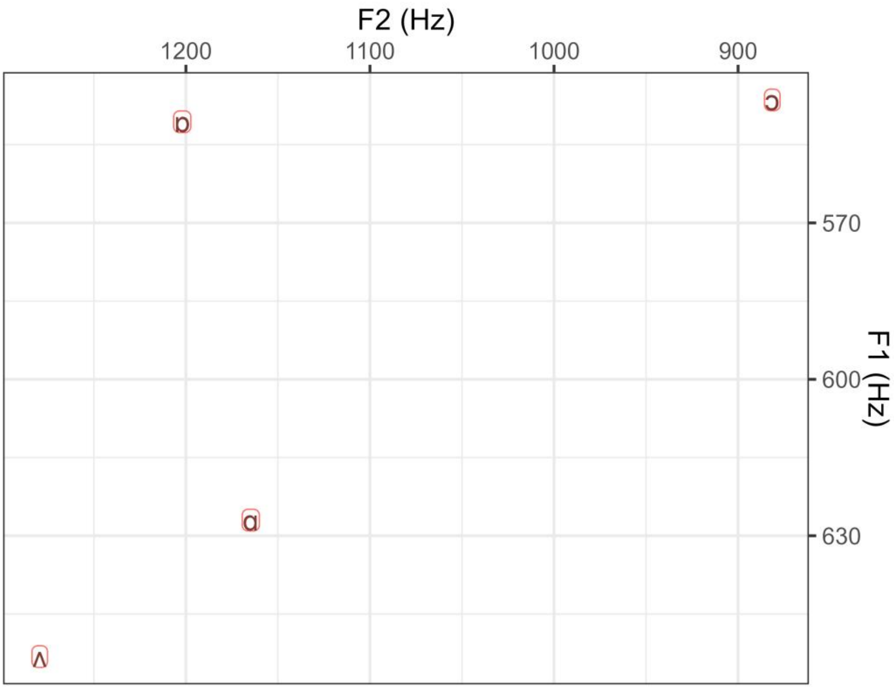


Posttest

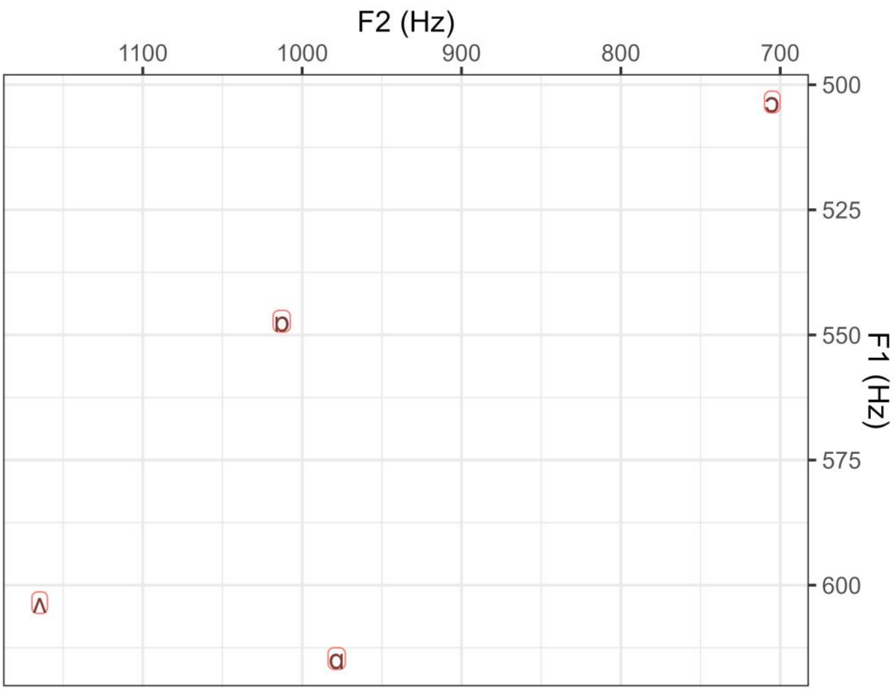


Male participant 8

Pretest

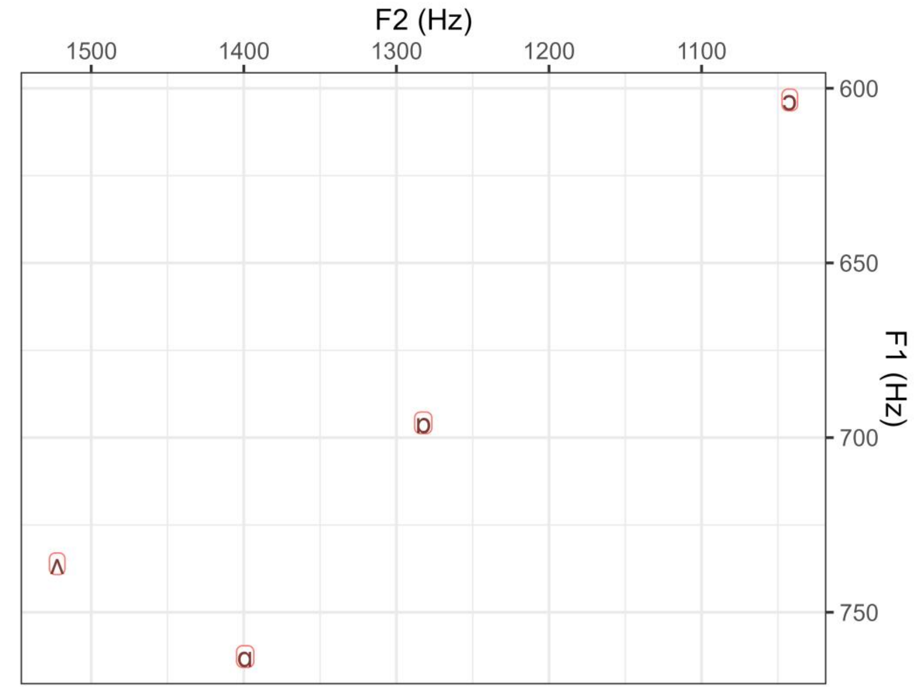


Posttest

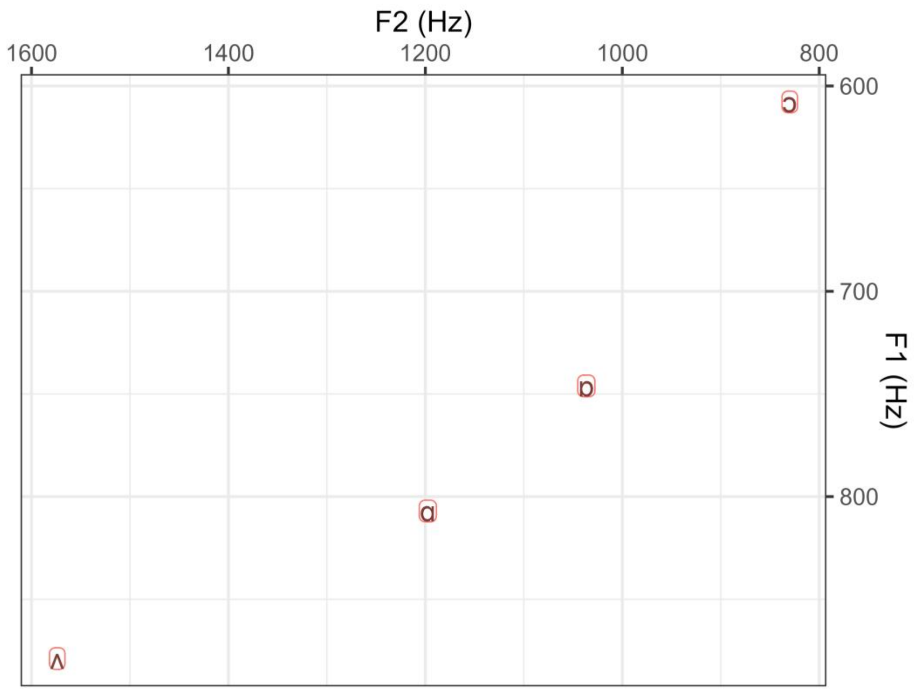


Female participant 11

Pretest

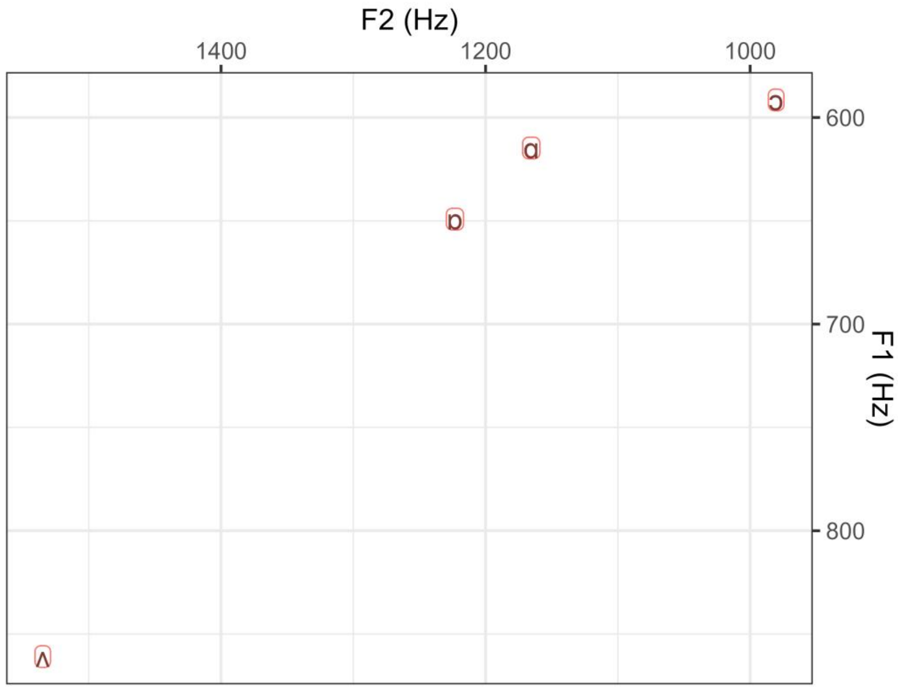


Posttest

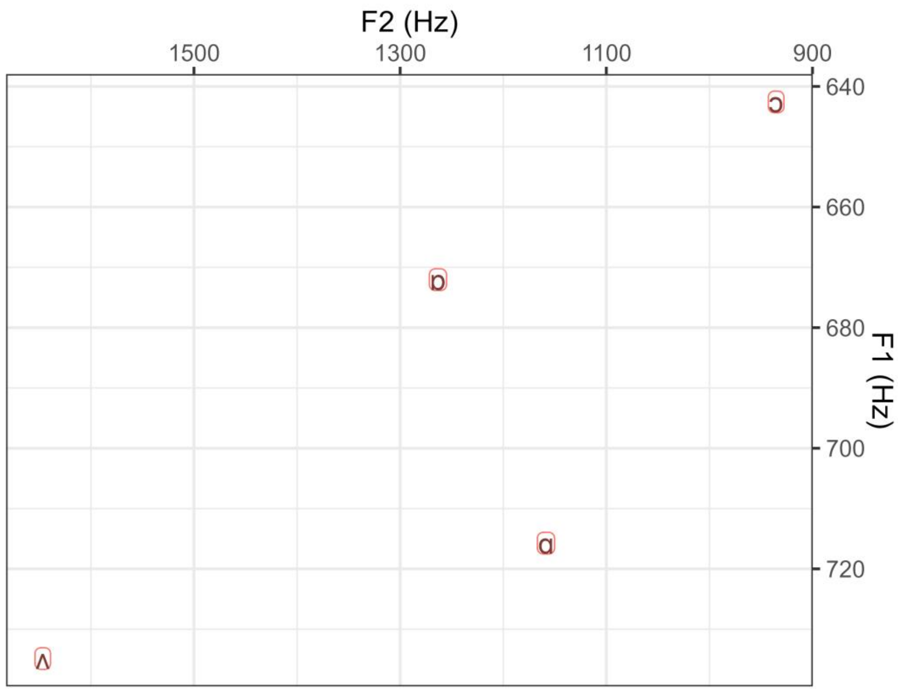


Female participant 12

Pretest

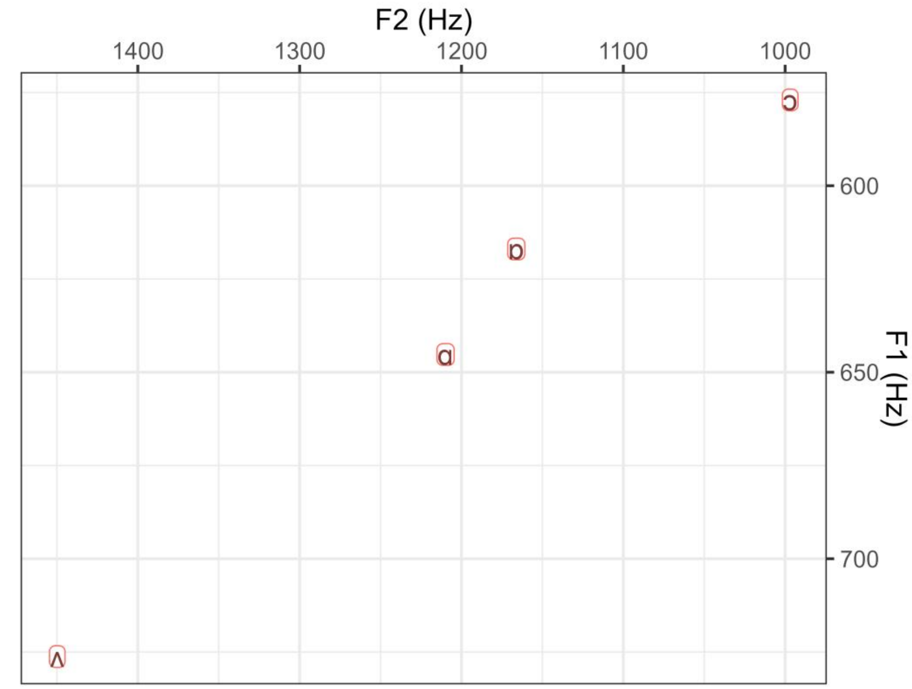


Posttest

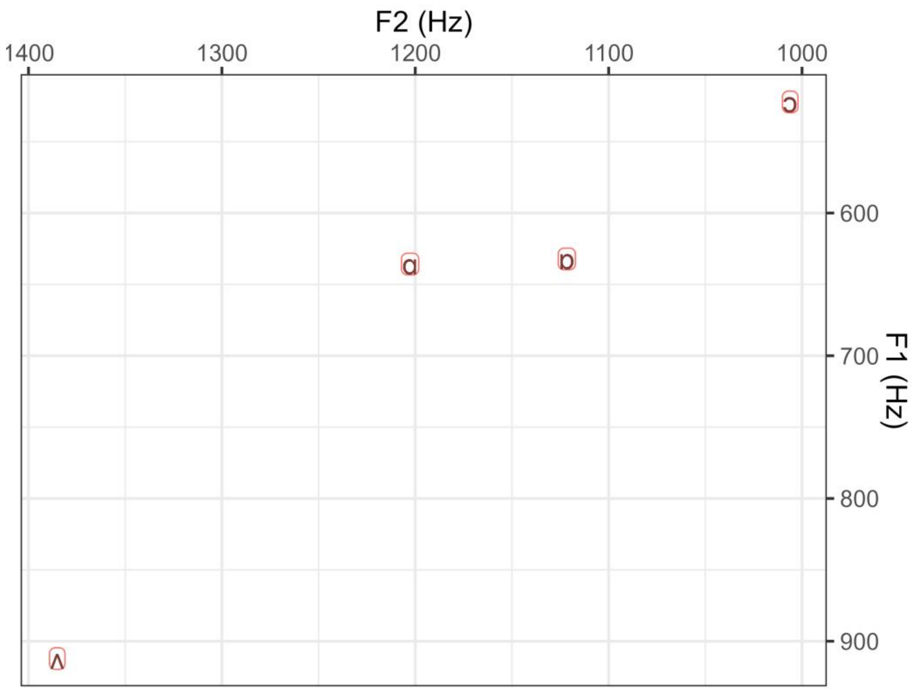


Female participant 13

Pretest

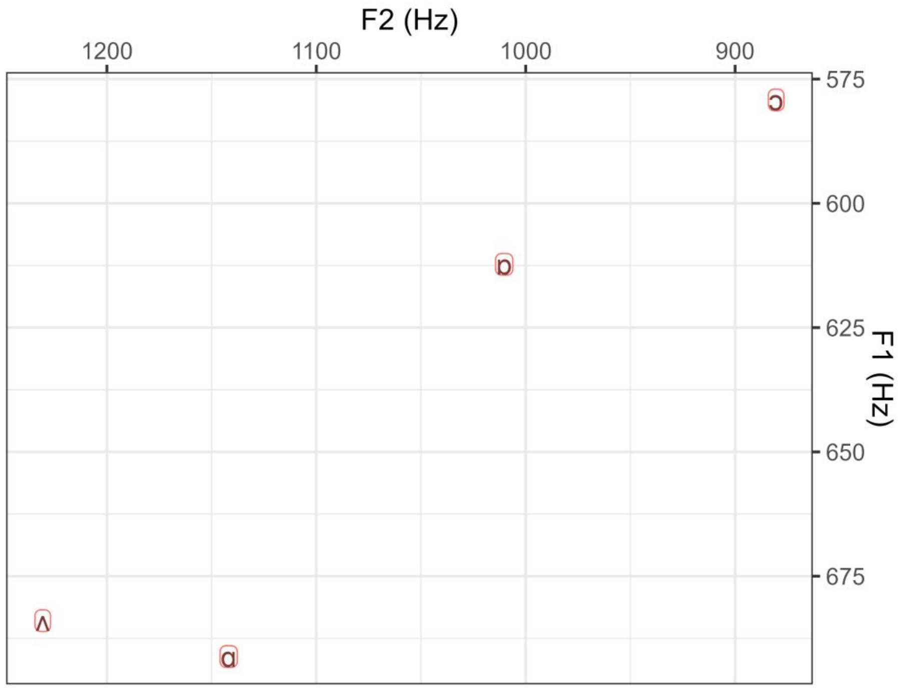


Posttest

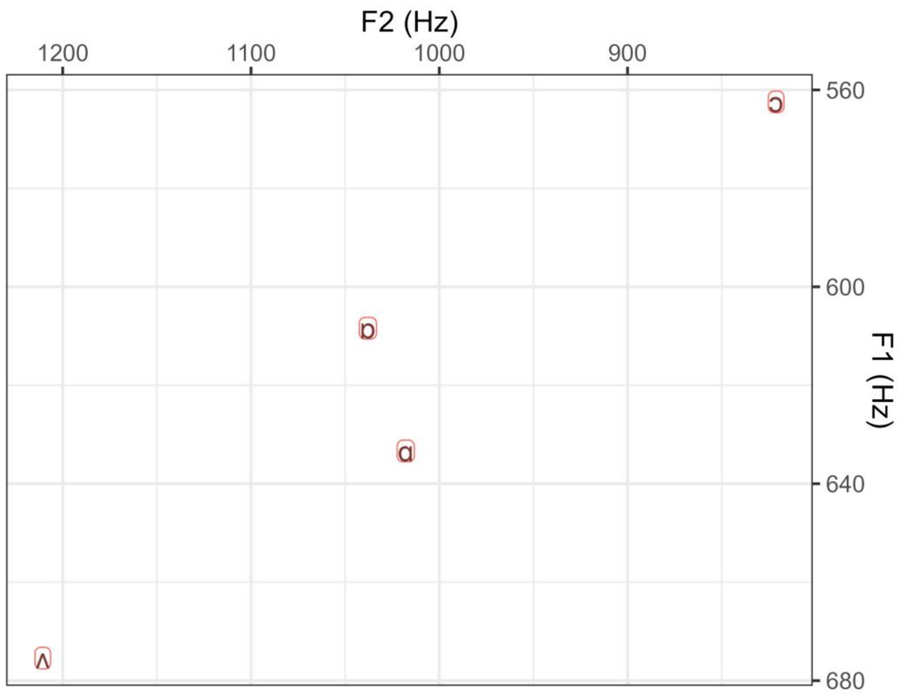


Male participant 9

Pretest

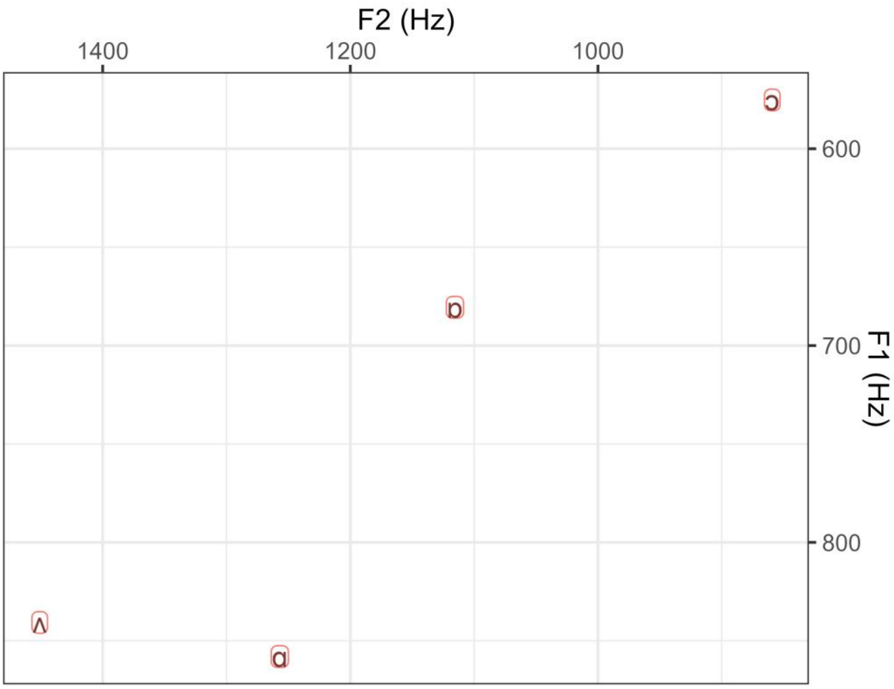


Posttest

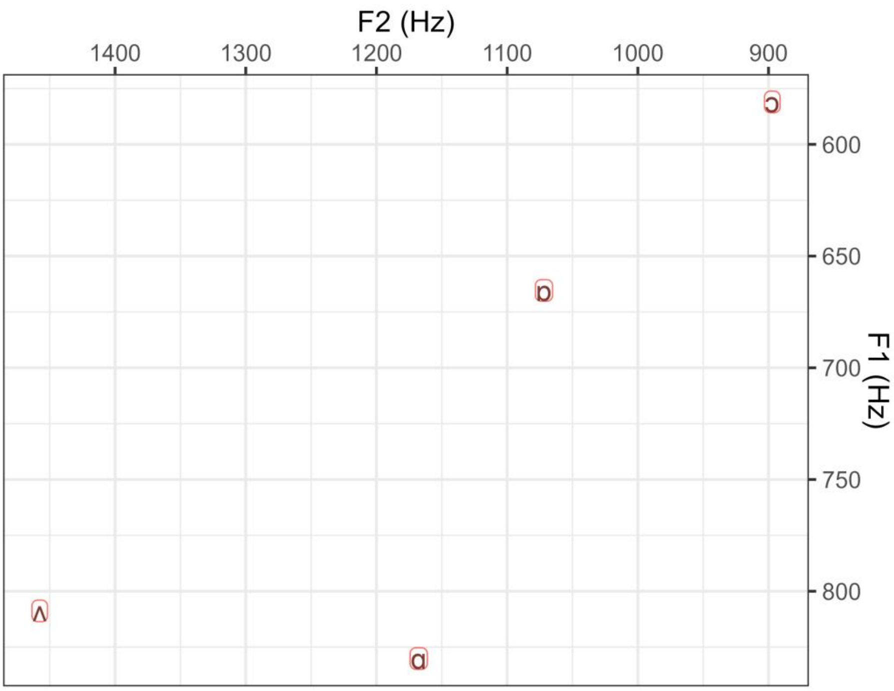


Female participant 14

Pretest

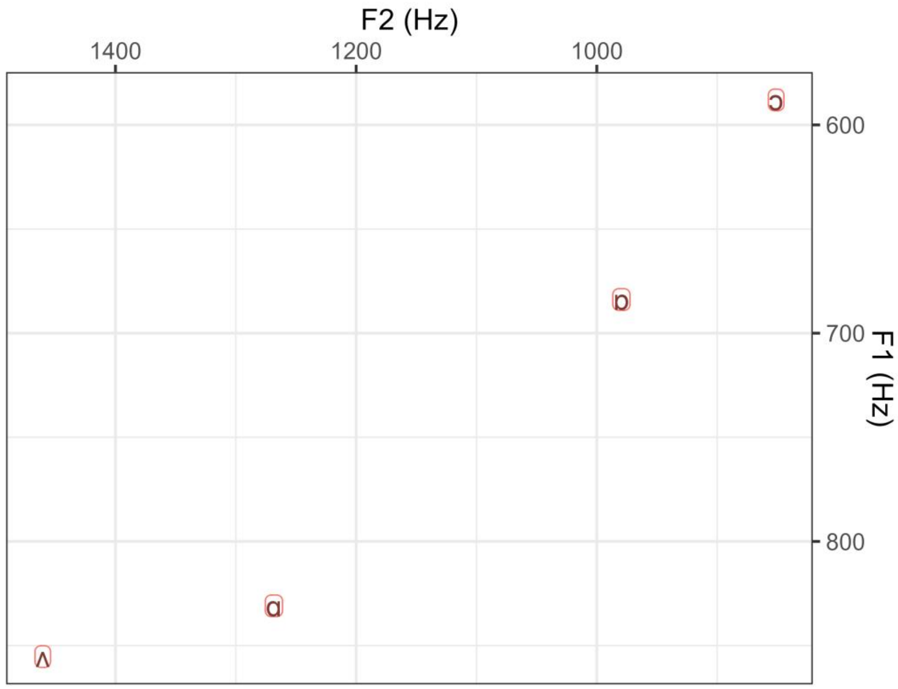


Posttest

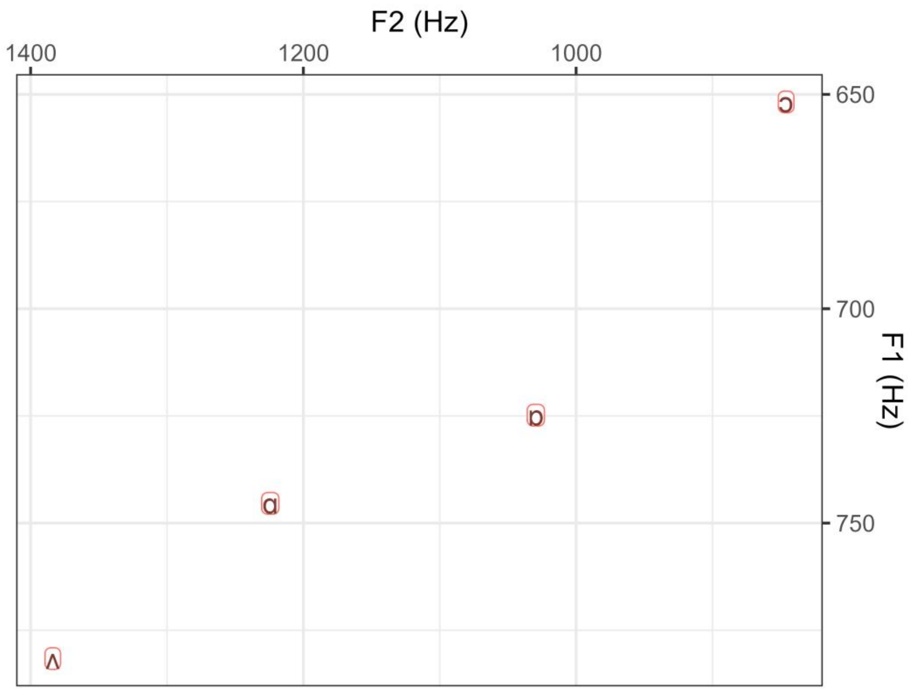


Female participant 8

Pretest

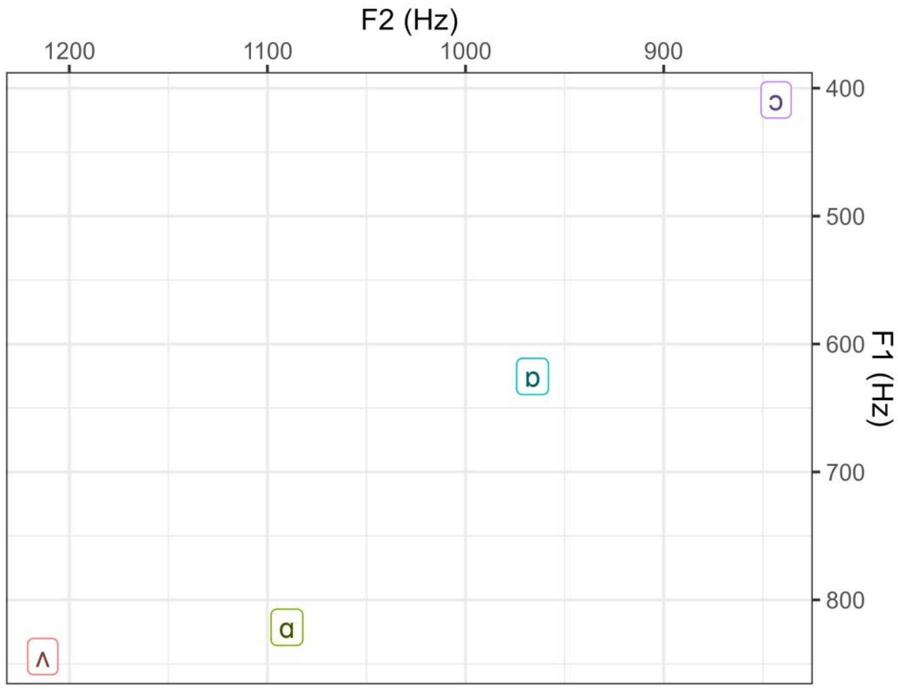


Posttest

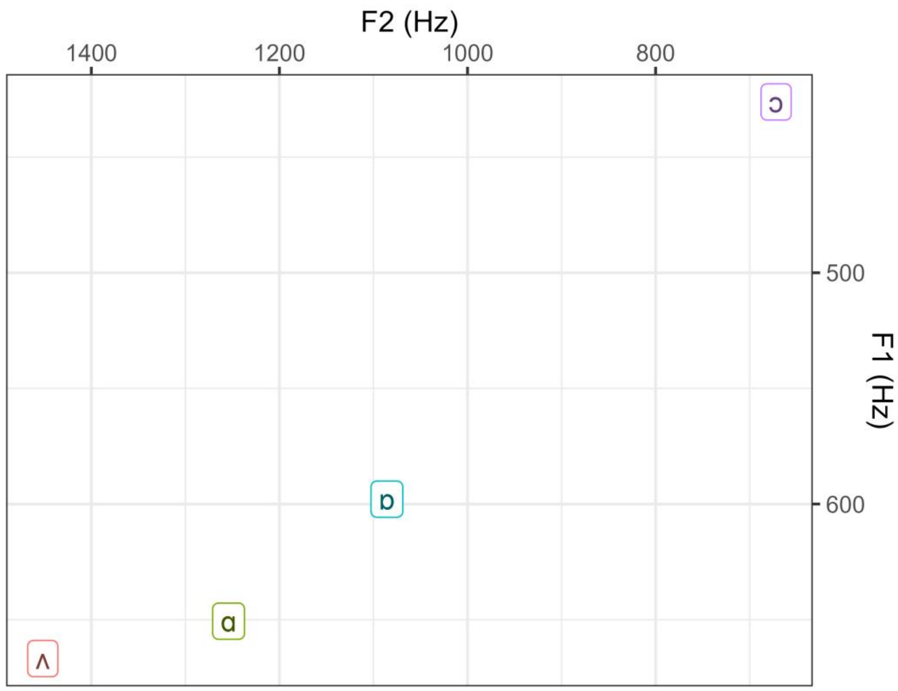


Realisation of /aː/, /ʌ/, /ɔː/ and /ɒ/ in British English (RP)

Male:

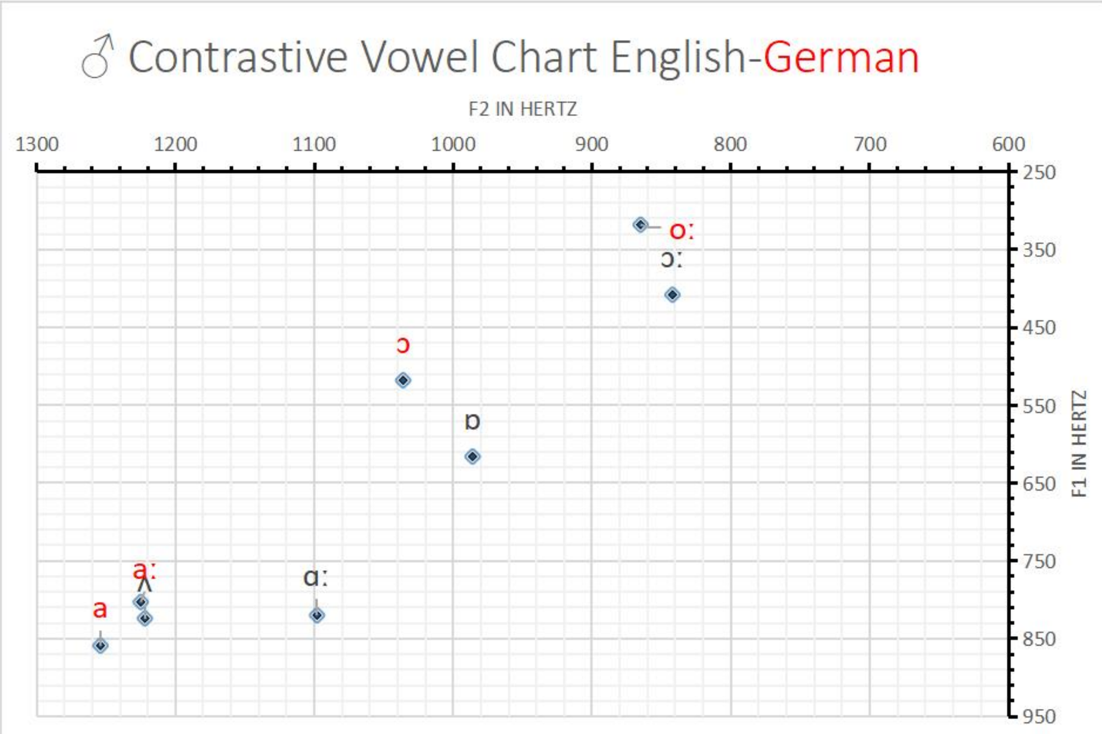


Female:

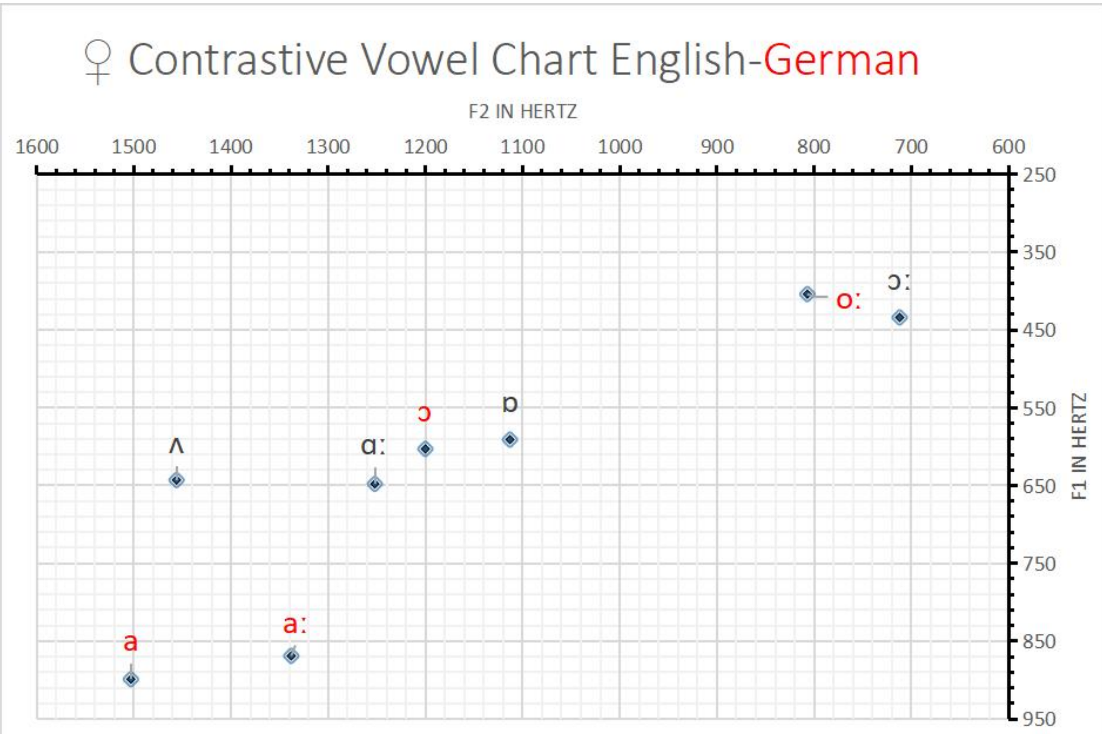


Comparison English /ɑ:/, /ʌ/, /ɔ:/ and /ɒ/ against German /a:/, /a/, /ɔ/ and /o:/

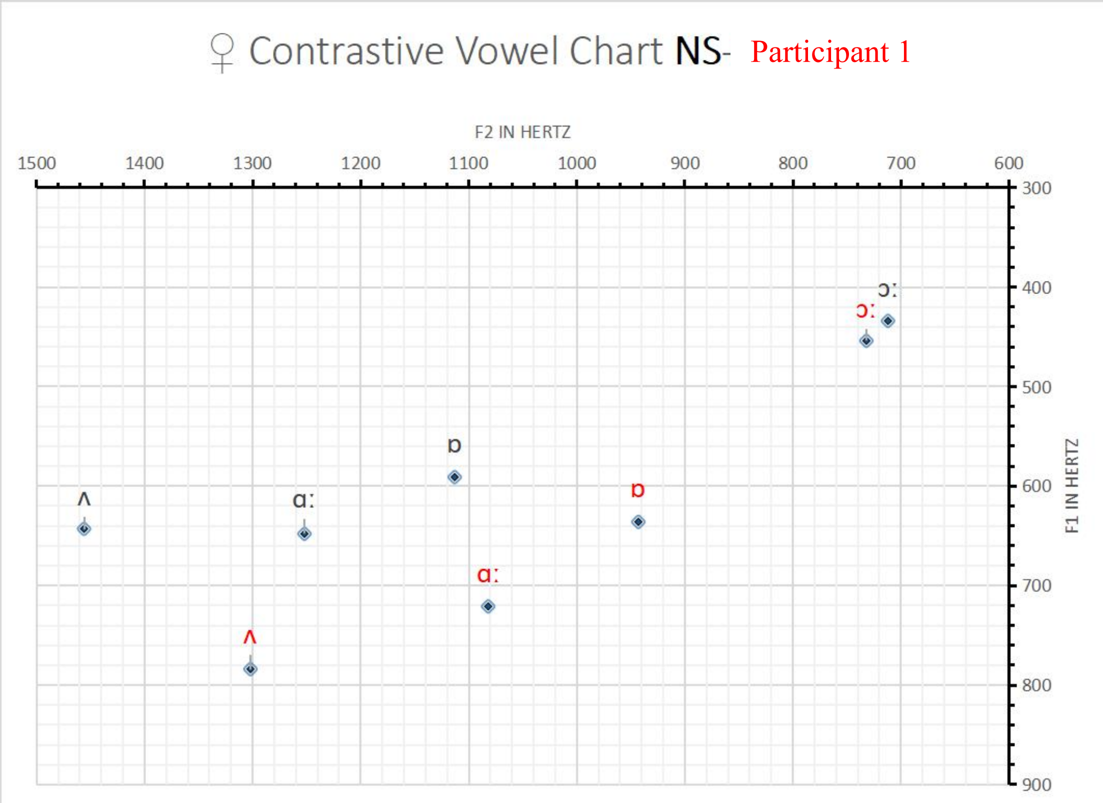
Male:

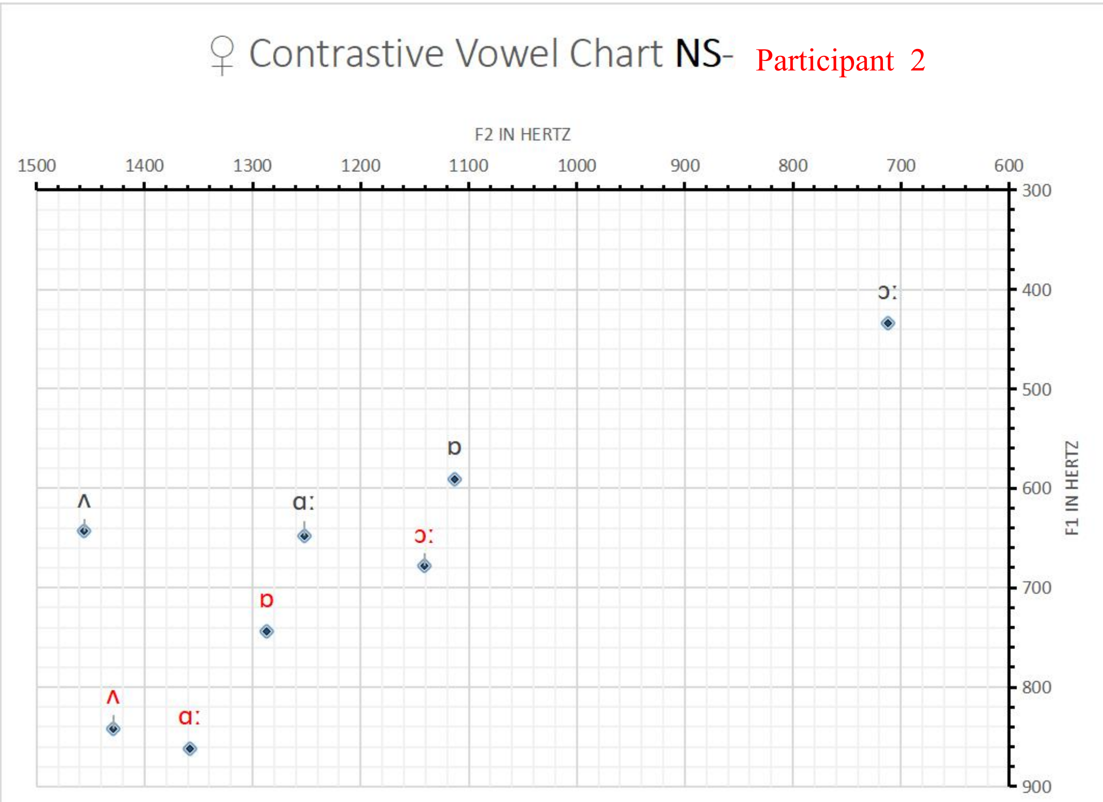


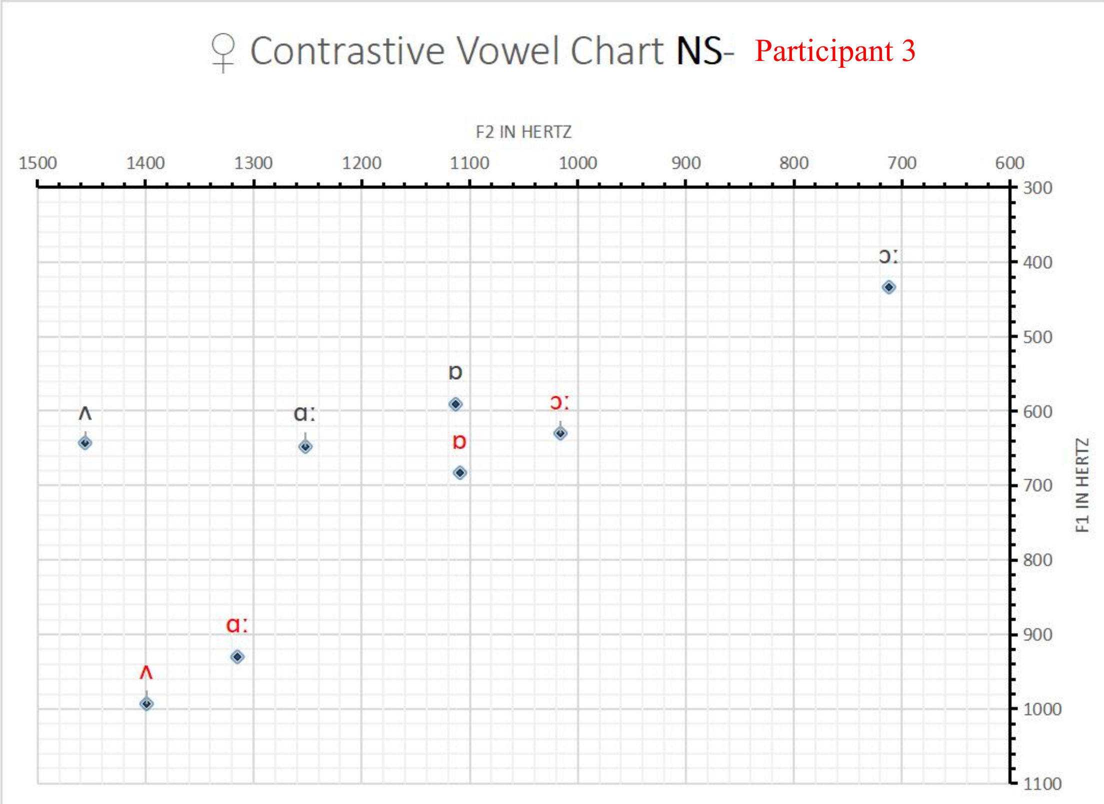
Female:

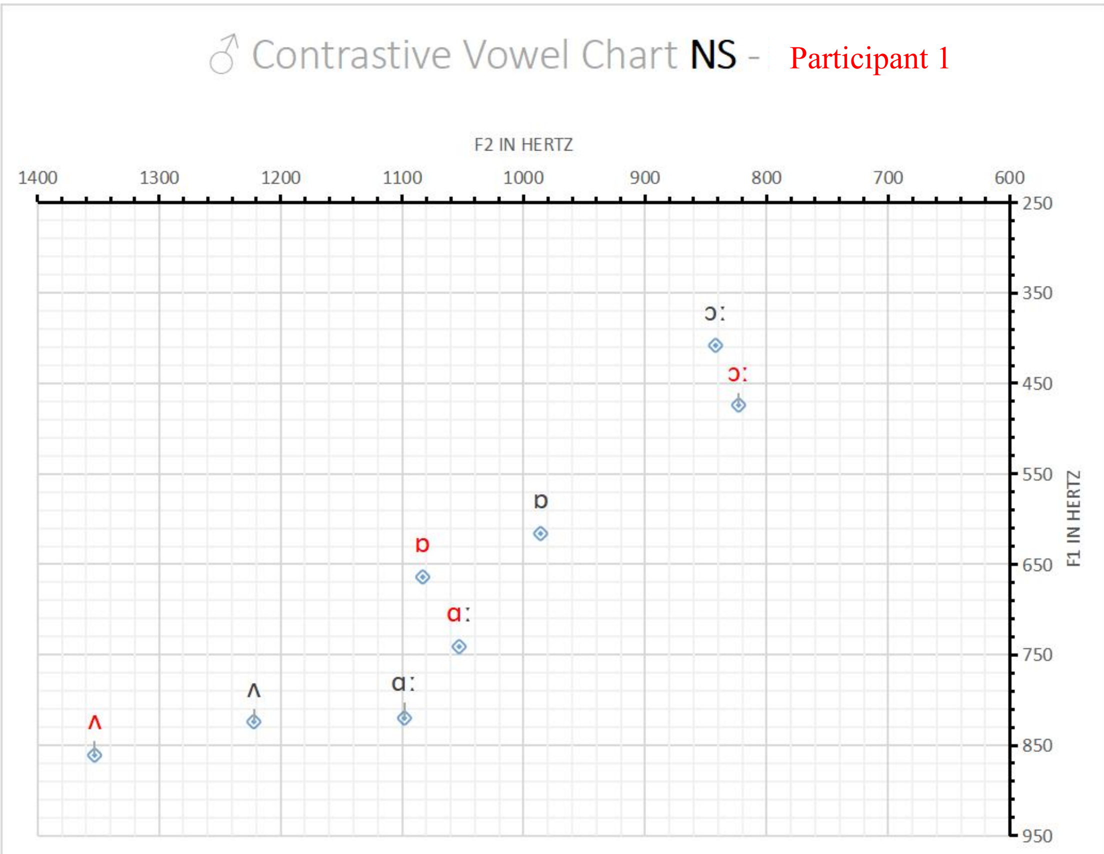


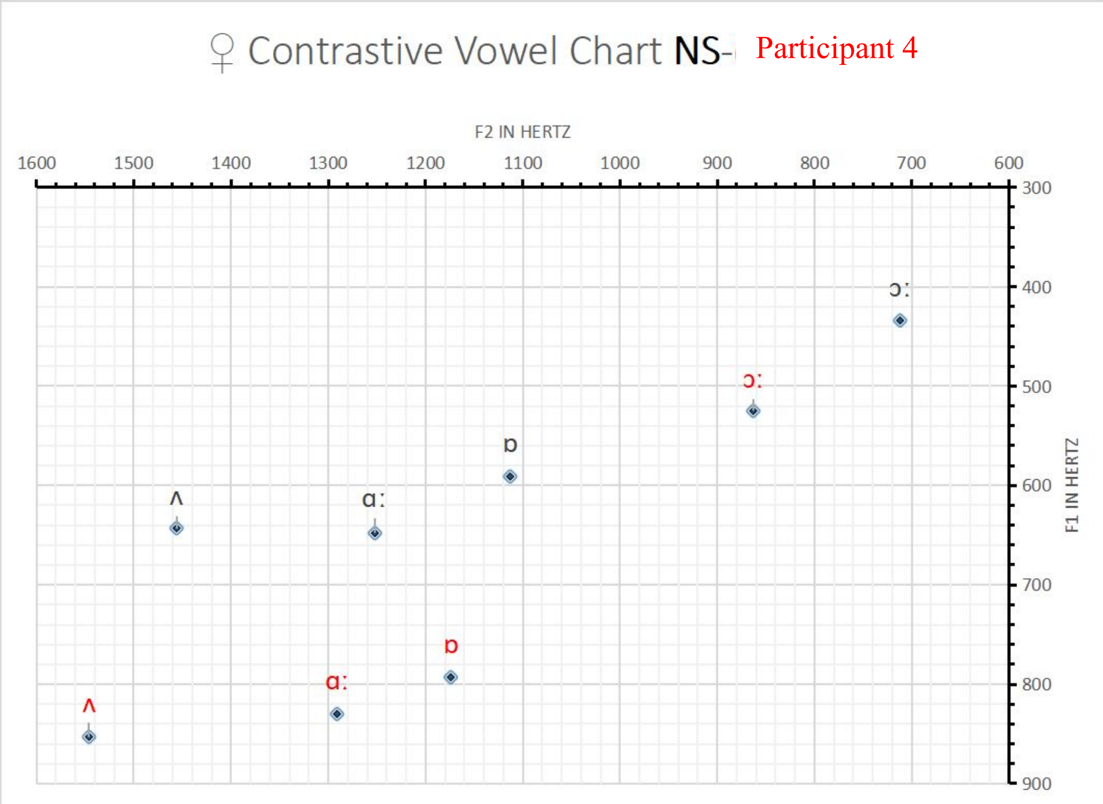
POST-test data of my participants against native speakers (NS) (alphabetical)

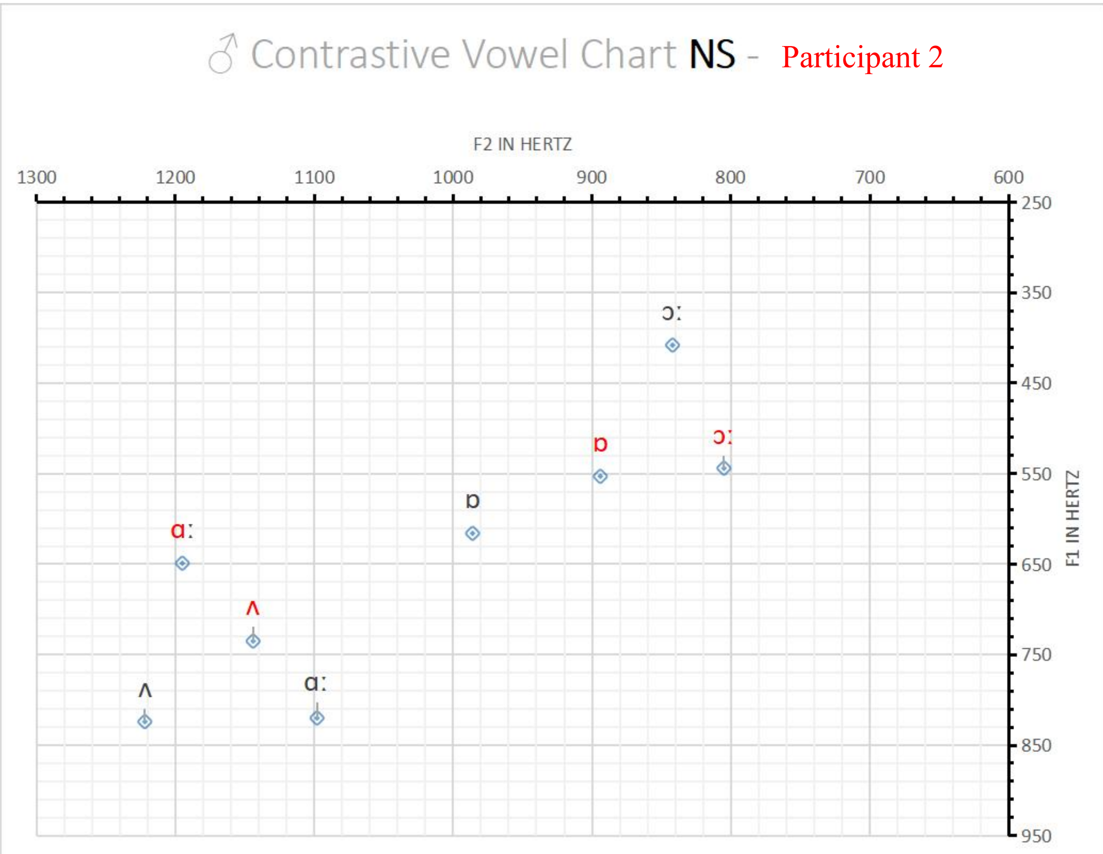


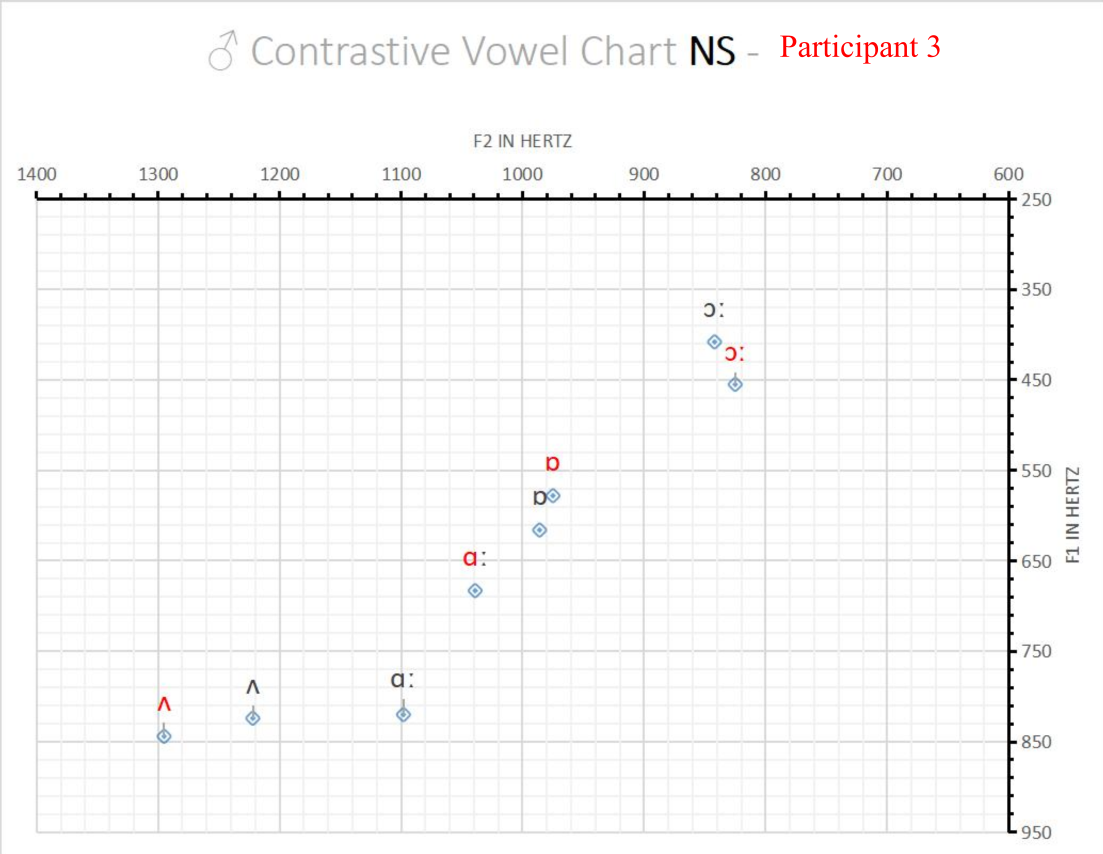


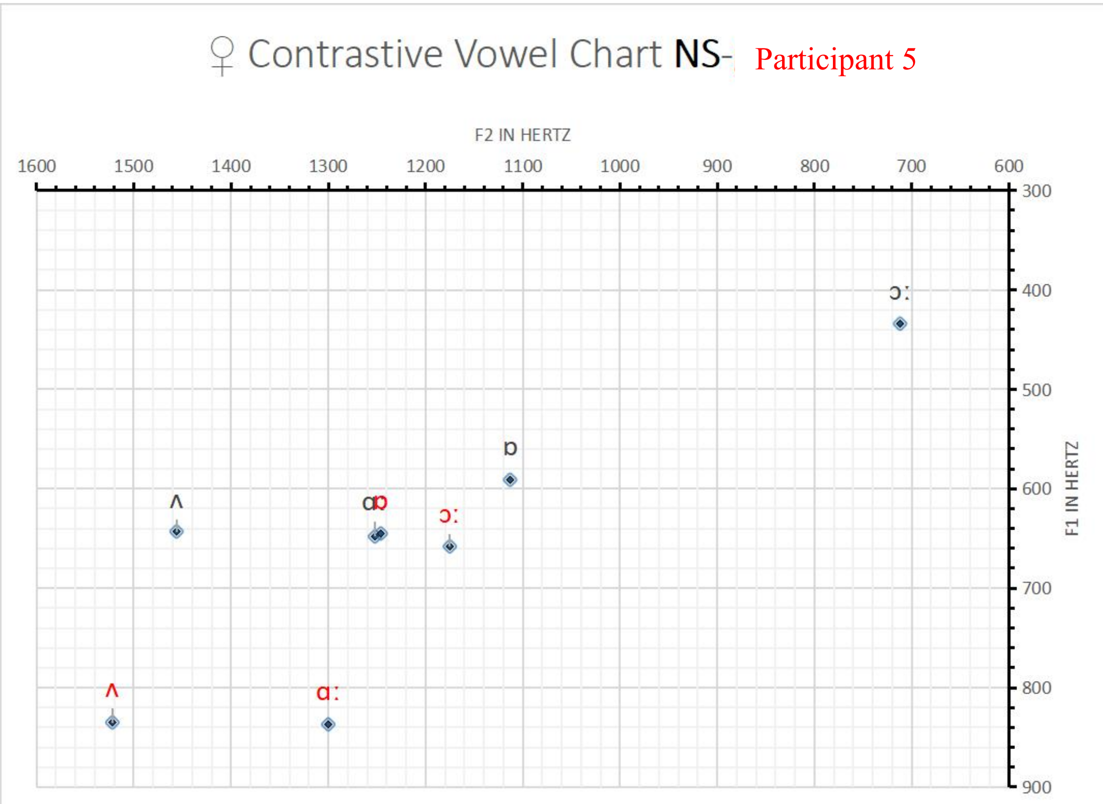

♀ Contrastive Vowel Chart NS- Participant 3
F2 IN HERTZ
1500
1400
1300
1200
1100
1000
900
800
700
600
300
400
500
600
700
800
900
1000
1100
F1 IN HERTZ
ɔː
ɒ
ʌ
ɑː
ɔː
ɒ
ɑː
ʌ

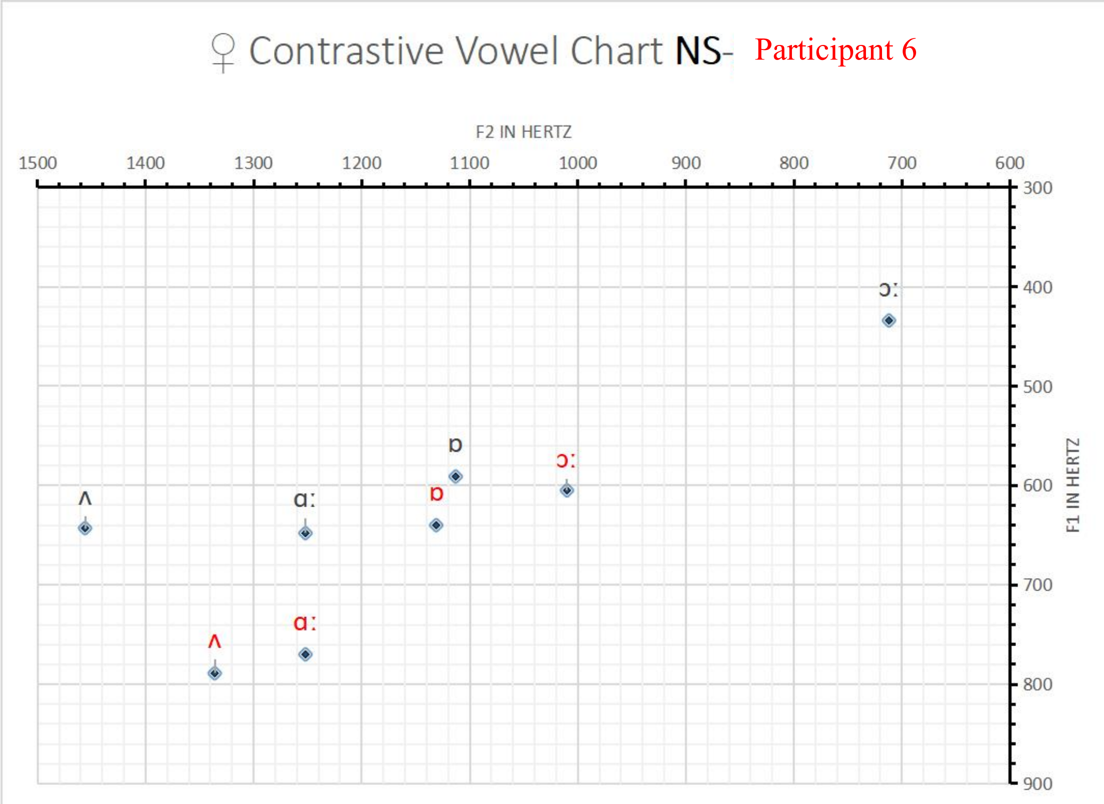

♂ Contrastive Vowel Chart NS - Participant 1
F2 IN HERTZ
1400
1300
1200
1100
1000
900
800
700
600
250
350
450
550
650
750
850
950
F1 IN HERTZ
ɔː
ɔː
ɒ
ɒ
ɑː
ʌ
ɑː
ʌ


♀ Contrastive Vowel Chart NS- Participant 4
F2 IN HERTZ
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
300
400
500
600
700
800
900
F1 IN HERTZ
ɔː
ɔː
ɒ
ʌ
ɑː
ɒ
ɑː
ʌ

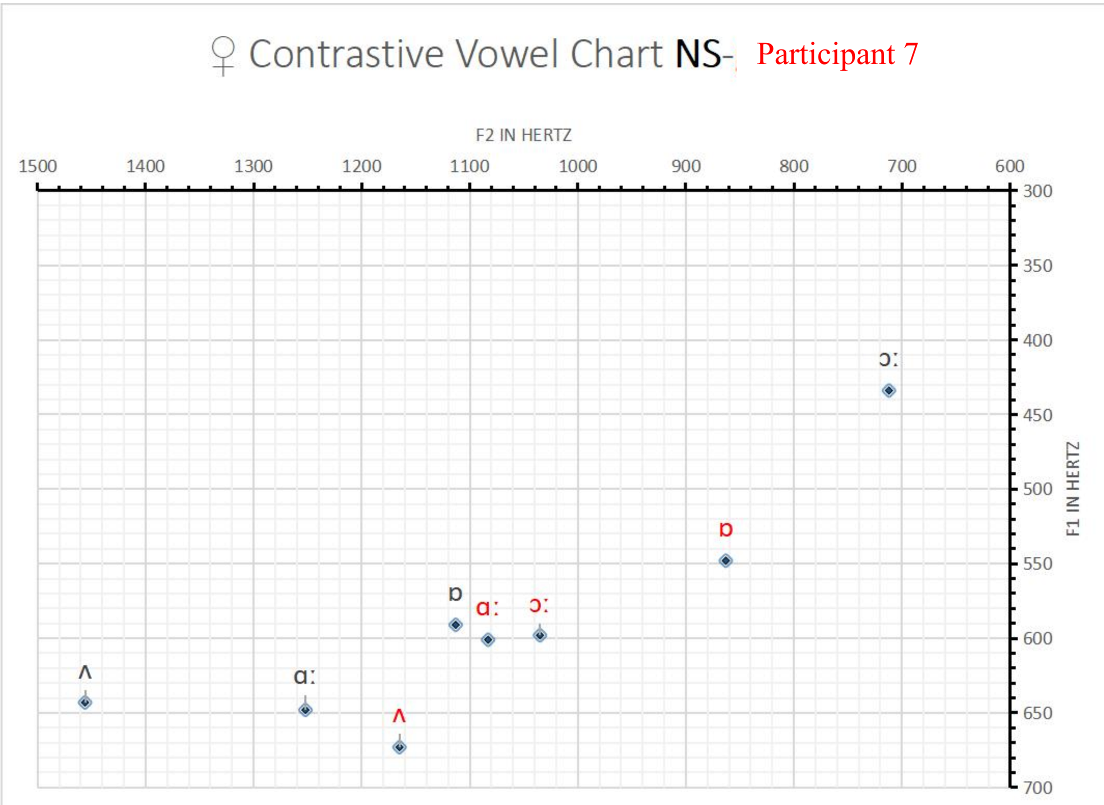

♂ Contrastive Vowel Chart NS - Participant 2
F2 IN HERTZ
1300
1200
1100
1000
900
800
700
600
250
350
450
550
650
750
850
950
F1 IN HERTZ
ɔː
ɒ
ɔː
ɒ
ɑː
ʌ
ʌ
ɑː


♂ Contrastive Vowel Chart NS - Participant 3
F2 IN HERTZ
1400
1300
1200
1100
1000
900
800
700
600
250
350
450
550
650
750
850
950
F1 IN HERTZ
ɔː
ɔː
ɒ
ɒ
ɑː
ʌ
ʌ
ɑː

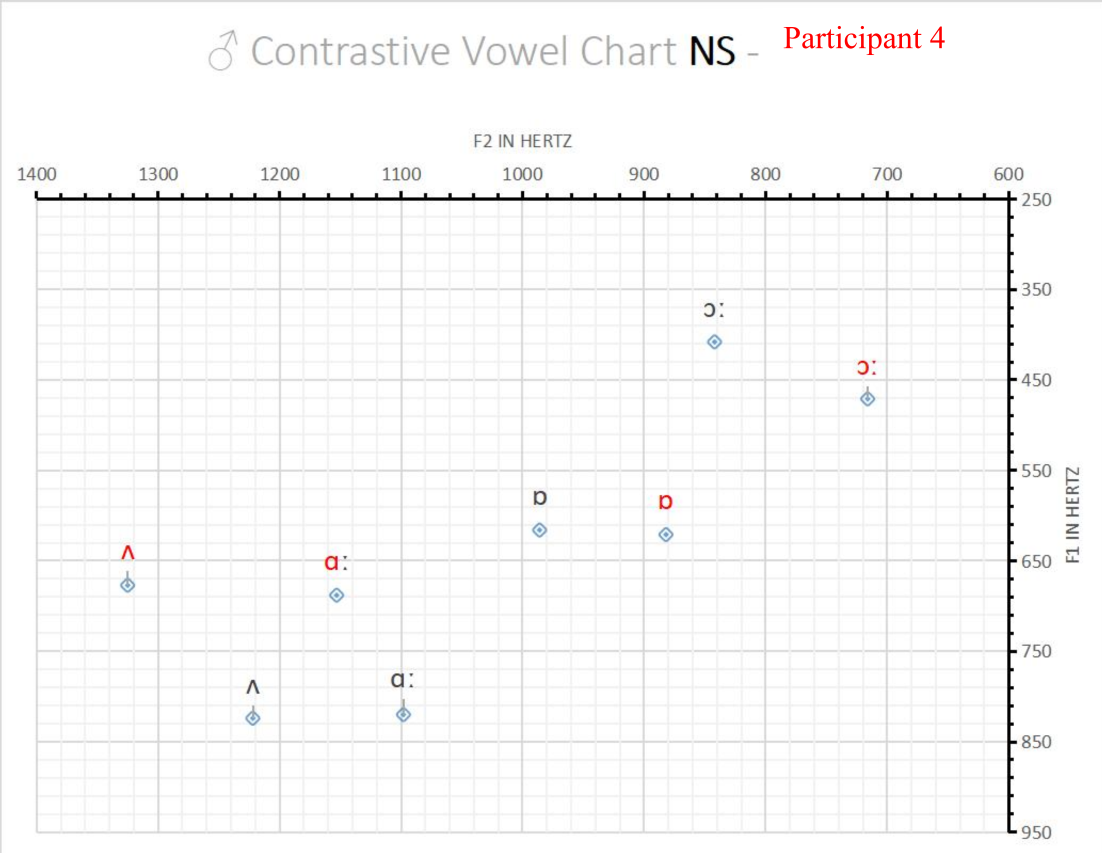

♀ Contrastive Vowel Chart NS- Participant 5
F2 IN HERTZ
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
300
400
500
600
700
800
900
F1 IN HERTZ
ɔː
ɒ
ʌ
ɑɒ
ɔː
ʌ
ɑː



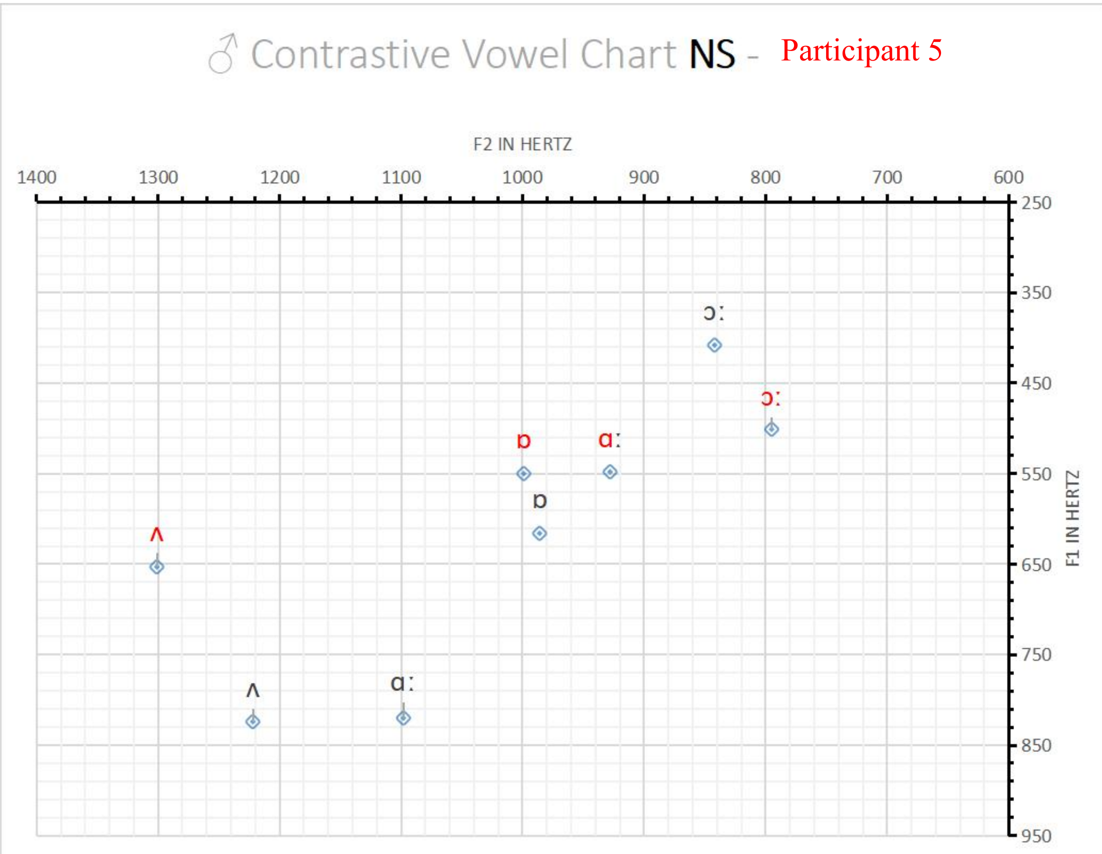
♀ Contrastive Vowel Chart NS- Participant 6
F2 IN HERTZ
1500
1400
1300
1200
1100
1000
900
800
700
600
300
400
500
600
700
800
900
F1 IN HERTZ
ɔː
ɒ
ɔː
ʌ
ɑː
ɒ
ɑː
ʌ



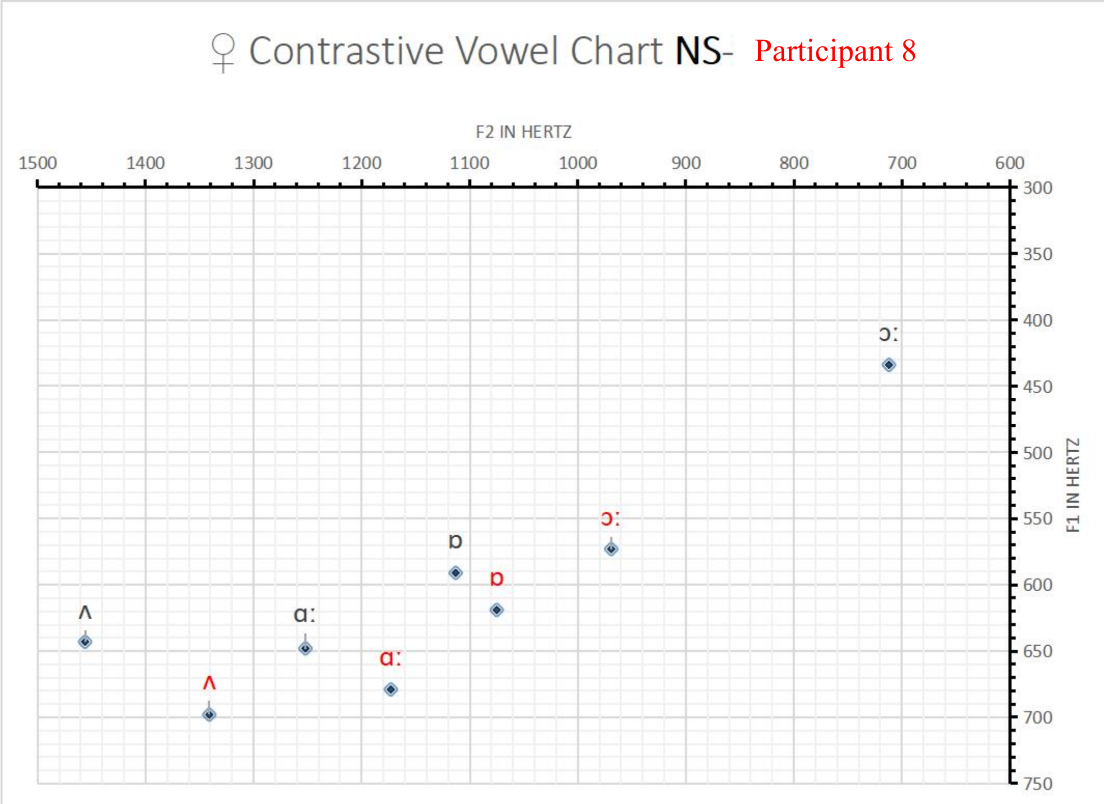
♀ Contrastive Vowel Chart NS- Participant 7
F2 IN HERTZ
1500
1400
1300
1200
1100
1000
900
800
700
600
300
350
400
450
500
550
600
650
700
F1 IN HERTZ
ɔː
ɒ
ɒ
ɑː
ɔː
ʌ
ɑː
ʌ

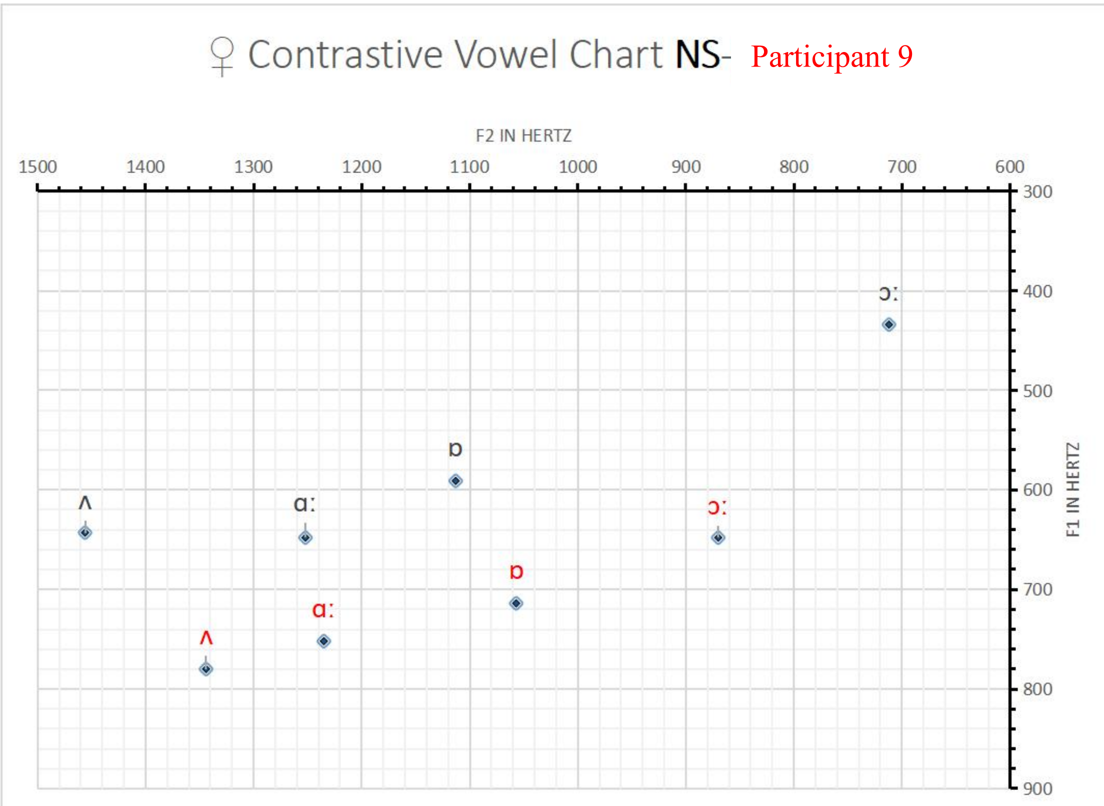


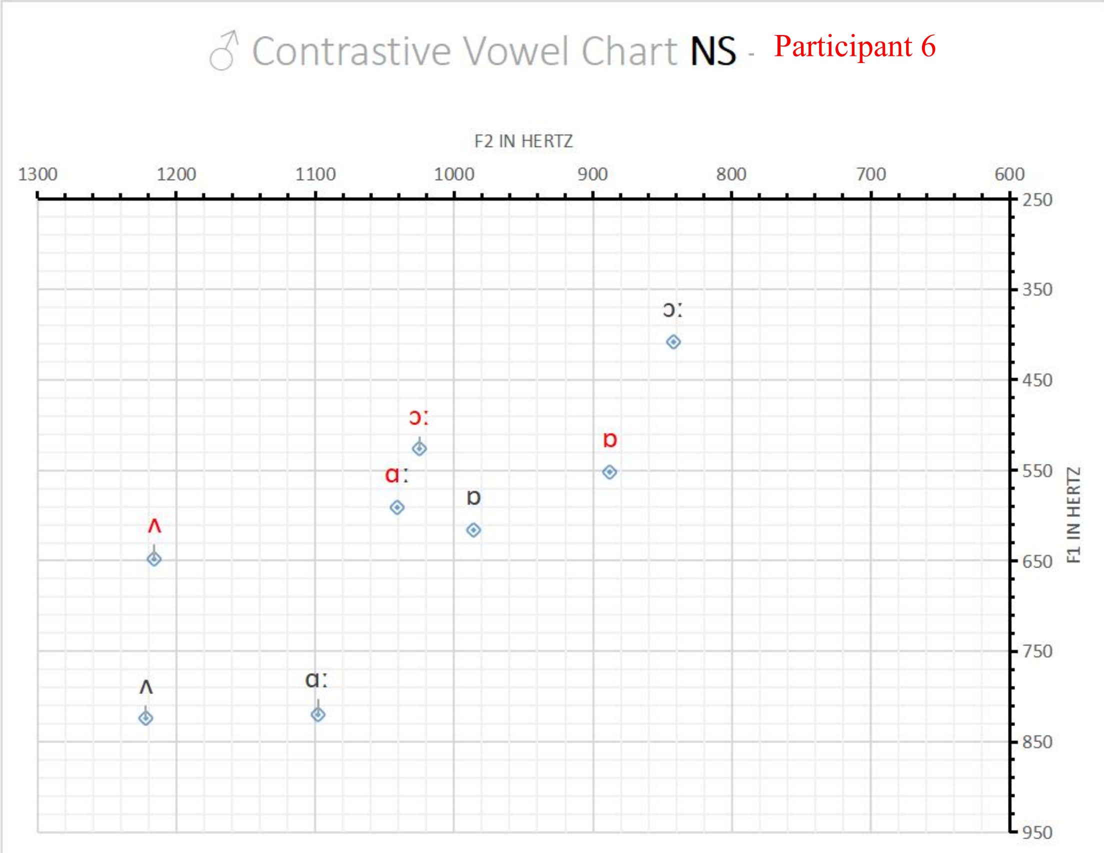
♂ Contrastive Vowel Chart NS - Participant 4
F2 IN HERTZ
1400
1300
1200
1100
1000
900
800
700
600
250
350
450
550
650
750
850
950
F1 IN HERTZ
ɔː
ɔː
ɒ
ɒ
ʌ
ɑː
ʌ
ɑː



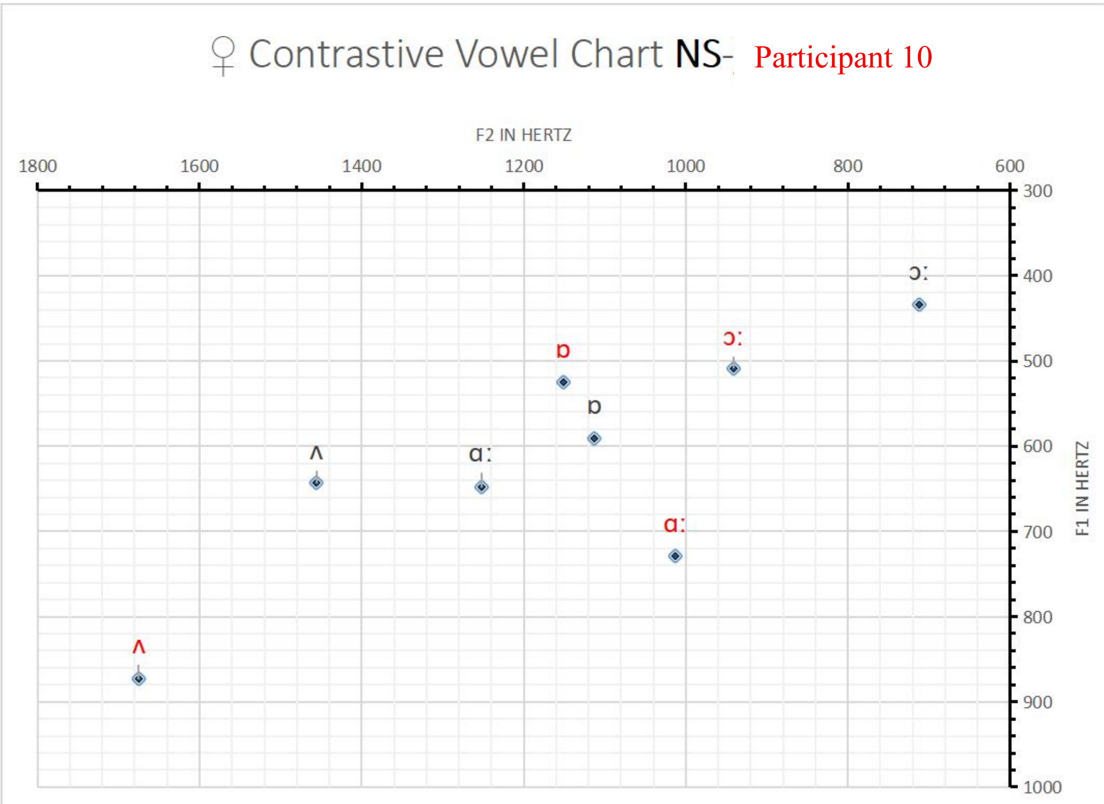
♂ Contrastive Vowel Chart NS - Participant 5
F2 IN HERTZ
1400
1300
1200
1100
1000
900
800
700
600
250
350
450
550
650
750
850
950
F1 IN HERTZ
ɔː
ɔː
ɒ
ɑː
ɒ
ʌ
ʌ
ɑː


♀ Contrastive Vowel Chart NS- Participant 8
F2 IN HERTZ
1500
1400
1300
1200
1100
1000
900
800
700
600
300
350
400
450
500
550
600
650
700
750
F1 IN HERTZ
ɔː
ɒ
ɒ
ɔː
ʌ
ɑː
ɑː
ʌ

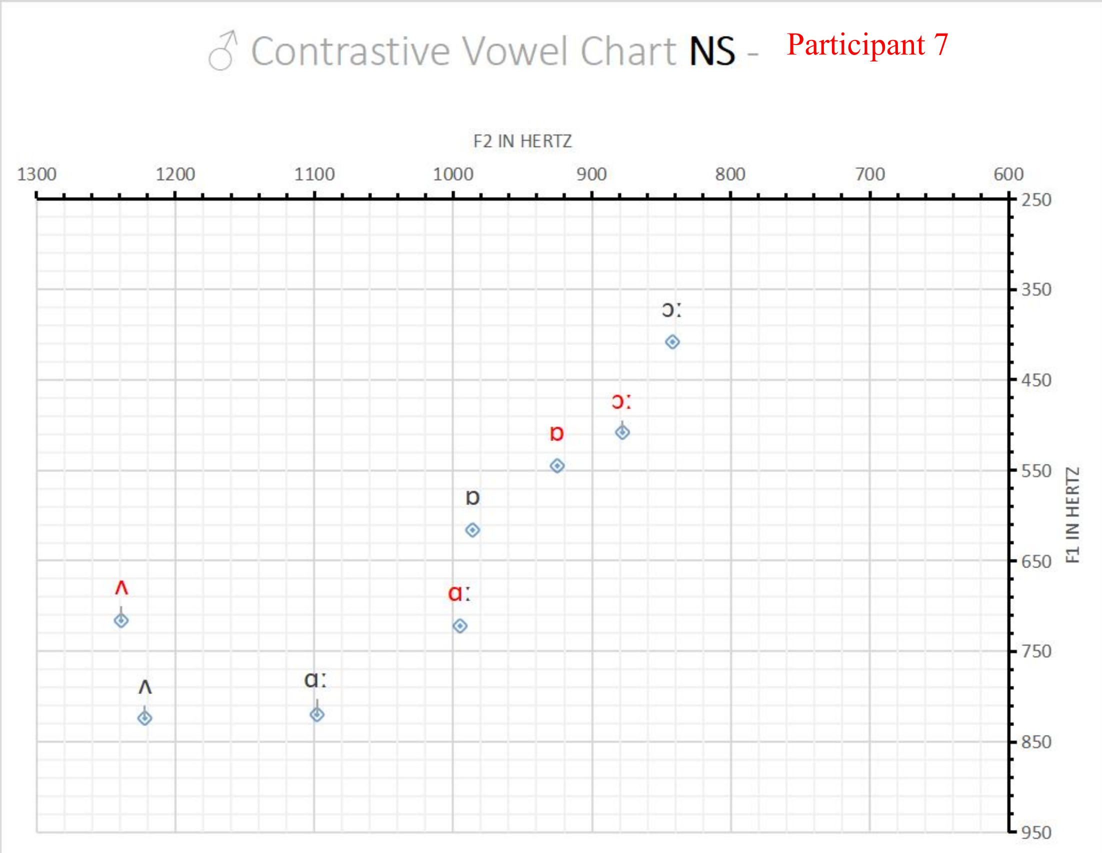

♀ Contrastive Vowel Chart NS- Participant 9
F2 IN HERTZ
1500
1400
1300
1200
1100
1000
900
800
700
600
300
400
500
600
700
800
900
F1 IN HERTZ
ɔː
ɒ
ʌ
ɑː
ɔː
ɒ
ɑː
ʌ



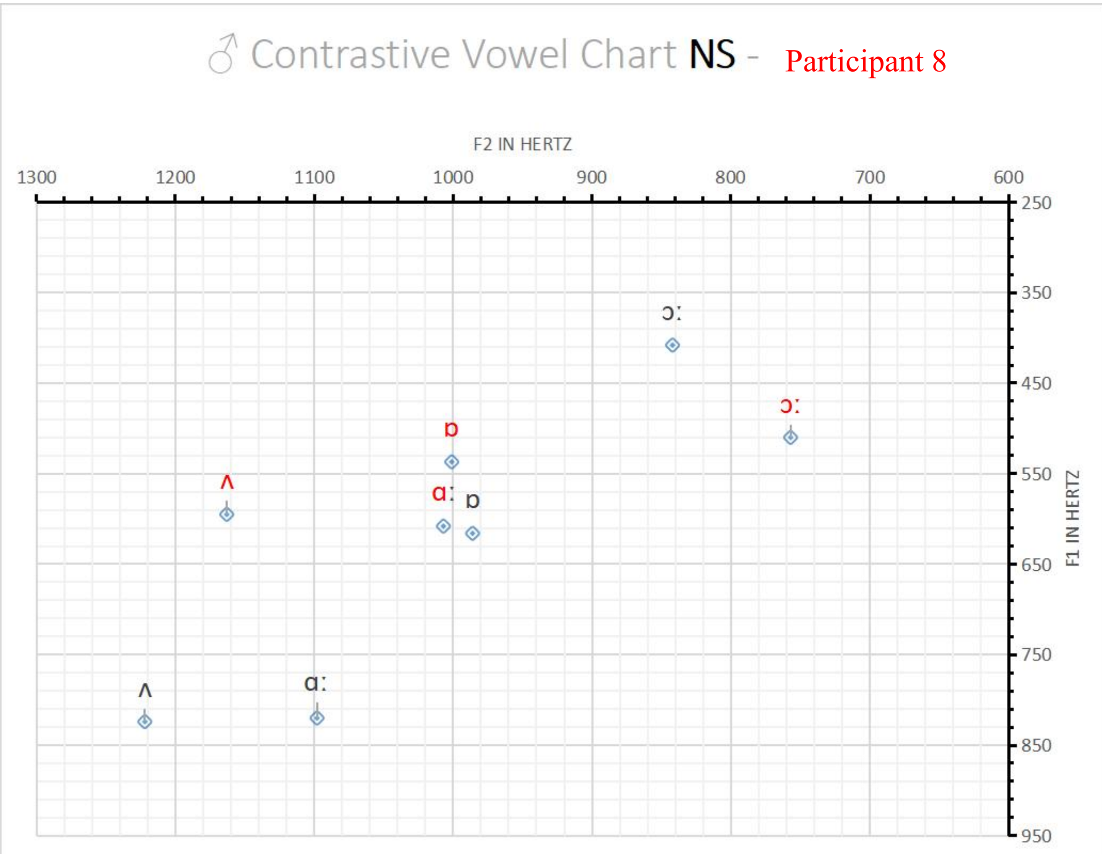
♂ Contrastive Vowel Chart NS - Participant 6
F2 IN HERTZ
1300
1200
1100
1000
900
800
700
600
250
350
450
550
650
750
850
950
F1 IN HERTZ
ɔː
ɔː
ɒ
ɑː
ɒ
ʌ
ʌ
ɑː



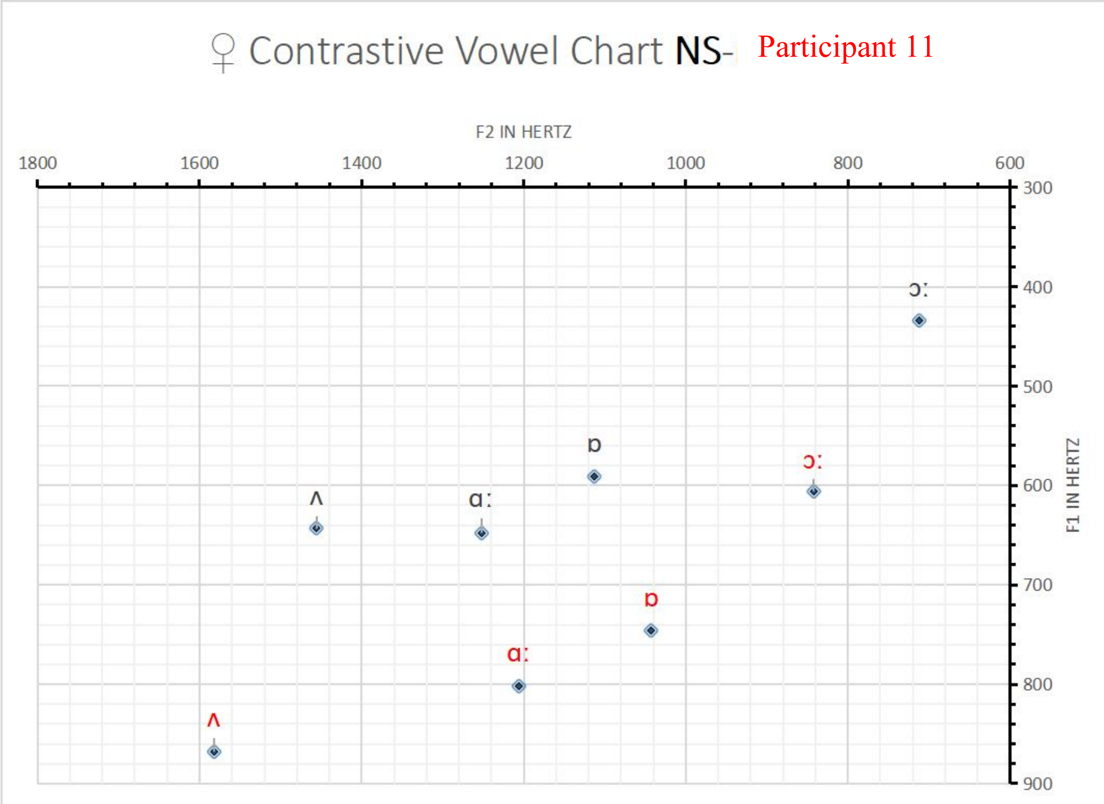
♀ Contrastive Vowel Chart NS- Participant 10
F2 IN HERTZ
1800
1600
1400
1200
1000
800
600
300
400
500
600
700
800
900
1000
F1 IN HERTZ
ɔː
ɔː
ɒ
ɒ
ʌ
ɑː
ɑː
ʌ



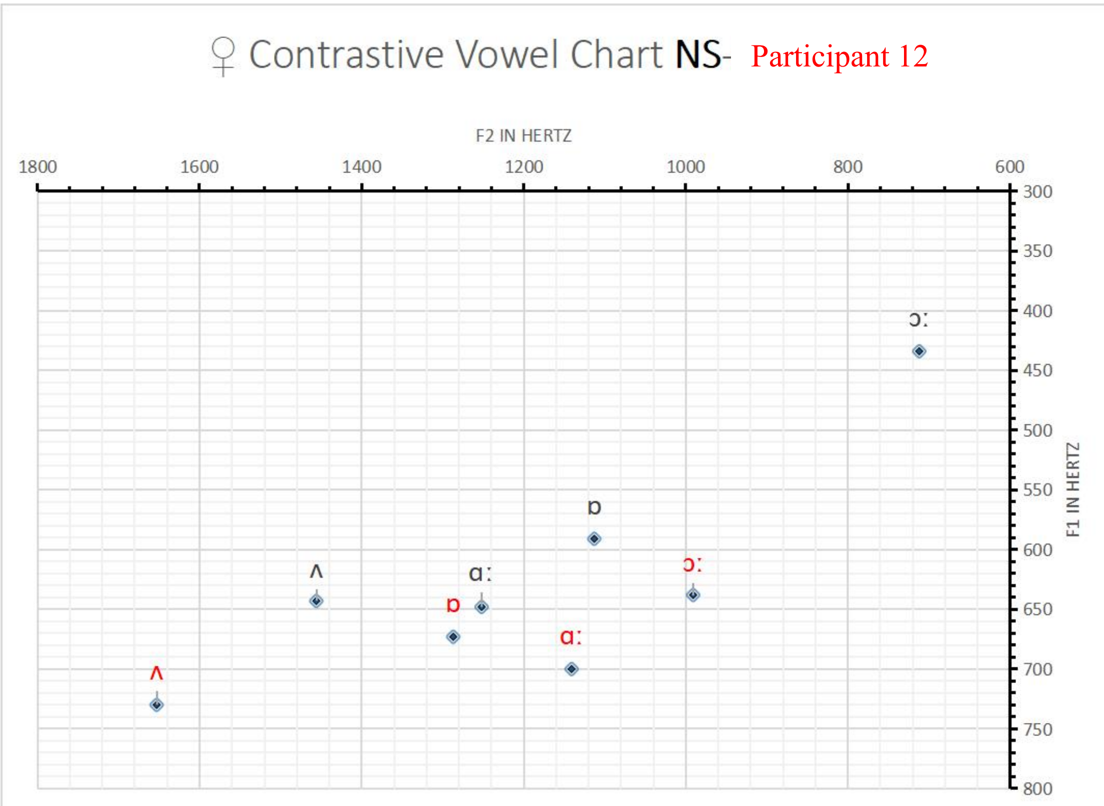
♂ Contrastive Vowel Chart NS - Participant 7
F2 IN HERTZ
1300
1200
1100
1000
900
800
700
600
250
350
450
550
650
750
850
950
F1 IN HERTZ
ɔː
ɔː
ɒ
ɒ
ʌ
ɑː
ʌ
ɑː



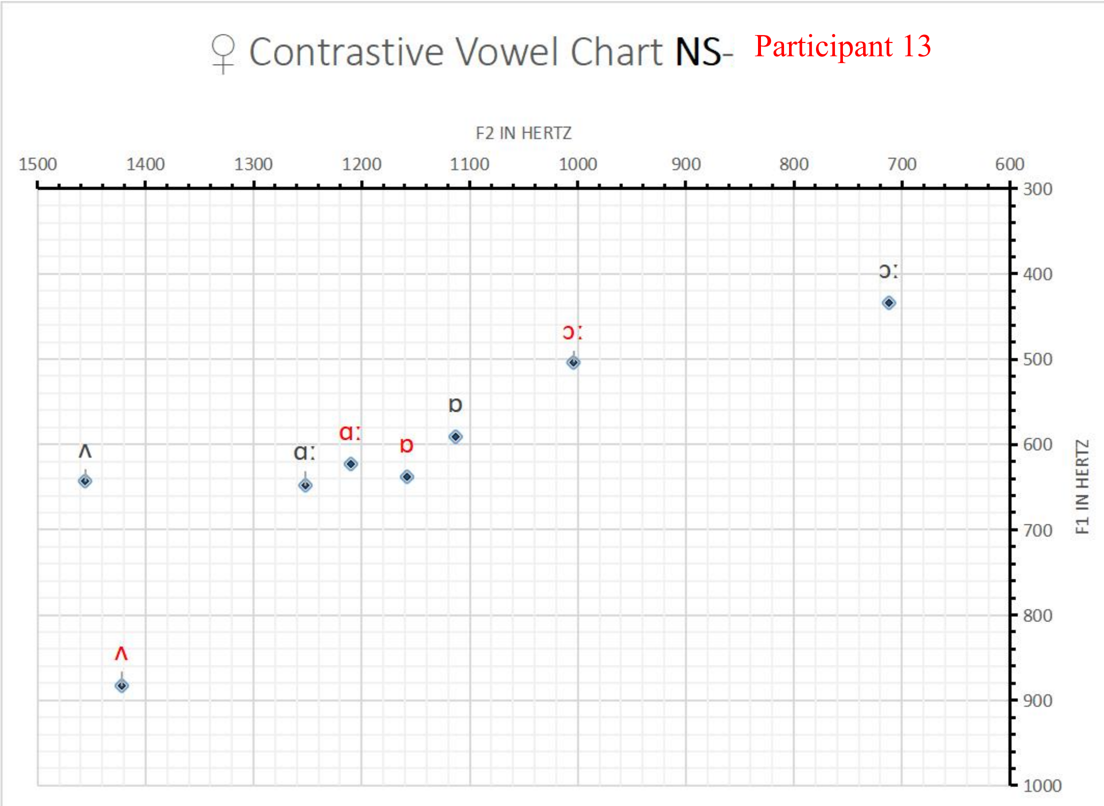
♂ Contrastive Vowel Chart NS - Participant 8
F2 IN HERTZ
1300
1200
1100
1000
900
800
700
600
250
350
450
550
650
750
850
950
F1 IN HERTZ
ɔː
ɔː
ɒ
ʌ
ɑː
ɒ
ʌ
ɑː



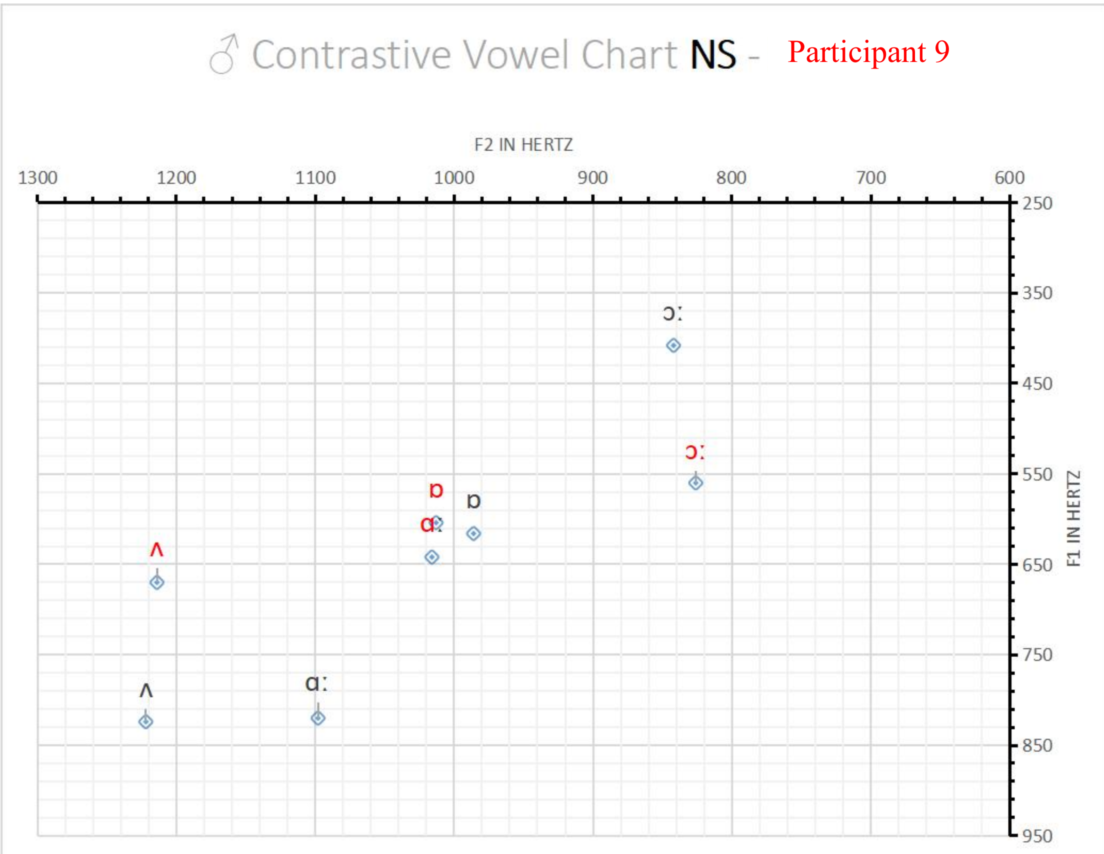
♀ Contrastive Vowel Chart NS- Participant 11
F2 IN HERTZ
1800
1600
1400
1200
1000
800
600
300
400
500
600
700
800
900
F1 IN HERTZ
ɔː
ɒ
ɔː
ʌ
ɑː
ɒ
ɑː
ʌ



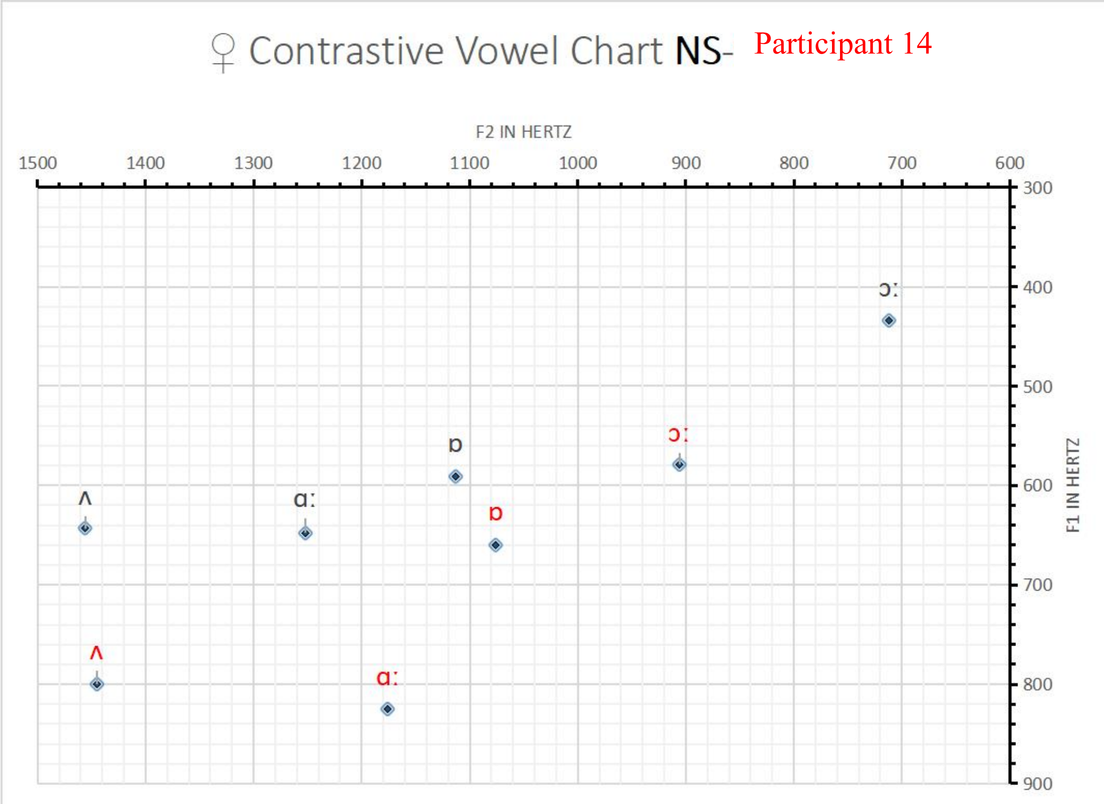
♀ Contrastive Vowel Chart NS- Participant 12
F2 IN HERTZ
1800
1600
1400
1200
1000
800
600
300
350
400
450
500
550
600
650
700
750
800
F1 IN HERTZ
ɔː
ɒ
ɔː
ʌ
ɑː
ɒ
ɑː
ʌ



♀ Contrastive Vowel Chart NS- Participant 13
F2 IN HERTZ
1500
1400
1300
1200
1100
1000
900
800
700
600
300
400
500
600
700
800
900
1000
F1 IN HERTZ
ɔː
ɔː
ɒ
ɑː
ɒ
ʌ
ɑː
ʌ



♂ Contrastive Vowel Chart NS - Participant 9
F2 IN HERTZ
1300
1200
1100
1000
900
800
700
600
250
350
450
550
650
750
850
950
F1 IN HERTZ
ɔː
ɔː
ɒ
ɒ
ɑː
ʌ
ʌ
ɑː



♀ Contrastive Vowel Chart NS- Participant 14
F2 IN HERTZ
1500
1400
1300
1200
1100
1000
900
800
700
600
300
400
500
600
700
800
900
F1 IN HERTZ
ɔː
ɒ
ɔː
ʌ
ɑː
ɒ
ʌ
ɑː

Appendix 2

Pretest and Posttest word lists (perception and production)

Pretest Production	Pretest Perception	Correct Answer	Perception Training	Posttest Perception	Correct Answer	Posttest Production
born	Bobble-bubble	bobble	See next page	Balled-bold	bold	barge
bought	Bodge-barge	bodge		Bog-bug	bog	bark
Bubble	Bog-bug	bug		Boss-bus	boss	bog
Calm	Bot-bought	bought		Butter-barter	butter	butter
Cart	Cod-card	card		Chalk-chock	chalk	chalk
Caught	Come-calm	calm		Chum-charm	Chum	clock
Come	Cop-cup	cop		Cough-cuff	cough	court
Cop	Court-cot	cot		Dawn-don	dawn	dart
Cup	Dock-duck	duck		Dork-dock	dork	guard
Cut	Dork-dock	dock		Doug-dog	Doug	heart
Dark	God-guard	guard		Hop-harp	harp	jog

Words in red appear at least twice over the course of the experiment.

Dock	Hug-hog	hug		Lock-luck	luck	jug
Fox	Lock-luck	luck		Port-pot	port	lung
God	Luck-lark	lark		Potty-party	potty	model
Golf	Shot-shut	shot		Robber-rubber	rubber	muddle
Hard	Snot-snort	snot		Rob-rub	rob	nut
Hoard	Sock-suck	suck		Shock-shark	shark	shock
Hut	Start-stut	stut		Shot-shut	shut	shot
Long	Walk-wok	wok		Taught-tart	tart	sob
Not	Wart-what	wart		Warned-wand	warned	sort
Short						sought
Spot						swan
Stock						sworn
Stuck						Talk
What						walk

Perception Training

Items appear in a fixed order (though this is mainly due to E-Babylab)

Perception Training	Correct Answer
Balled-bold	balled
Bobble-bubble	bobble

Borg-bog	bog
Born-barn	barn
Boss-bus	bus
Bot-Bart	Bart
Bot-but	but
But-bot	bot
Calf-cuff	calf
Cart-cut	cart
Cob-cub	cub
Collar-colour	collar
Cot-cut	cot
Cough-cuff	cuff
Caught-cot	caught
Cup-carp	carp
Dot-dart	dot
Duck-dark	duck
Forks-fucks	forks
Fox-fucks	fox
Hoard-hod	hoard
Hod-hud	hud
Hot-hut	hot

Knot-nut	knot
Lodge-large	lodge
Muck-mark	muck
Port-pot	pot
Pup-Pop	pup
Rob-rub	rub
Sawed-sod	sawed
Shop-sharp	shop
Sob-Saab	Saab
Spot-sport	sport
Stalk-stock	stalk
Stalk-stuck	stuck
Stomp-stump	stump
Stutter-starter	stutter
Taught-tart	taught
Warp-whop	warp
Wars-was	was

Experiment material

I designed the experiment to feature a total of: 46x /ɒ/ 35x /ʌ/ 28x /ɔ:/ 21x /a:/

List of all collected minimal pairs:

Out of the 170 words listed below I used a total of 120 for the experiment

/ɔ:/	/ɒ/	/ʌ/	/ɑ:/
hoard	hod	hud	hard
bored	bod	but	Bart
caught	cot	cut	cart
forks	fox	fucks	
talk	tock		
taught	tot		tart
wars	was		
wart	what		
bought	bot		
chalk	chock		
court	cot		
short	shot		
sport	spot		
stalk	stock	stuck	stark

walk	Wok		
warned	wand		
balled	bold		
cord	cod		
dork	dock	duck	dark
dawn	don		
naught	not		
port	pot		
roared	rod		
shorn	shone		
sort	sot		
bauble	bobble		
borg	bog		
wrought	rot		
snort	snot		
sawed	sod		
sought	sot		
sworn	swan		
ward	wad		

warp	whop		
	cough	cuff	calf
	hog	hug	
	jog	jug	
	golf	gulf	
	Pop	pup	
	boss	bus	
	lock	luck	lark
	knot	nut	
	cop	cup	carp
	cob	cub	
	sock	suck	
	hot	hut	heart
	dog	Doug	
	shot	shut	
	rob	rub	
	long	lung	
	pot	putt	
	bog	bug	

	robber	rubber	
	stomp	stump	
	collar	colour	
	model	muddle	
born			barn
		come	calm
		stut	start
		putt	part
		stutter	starter
		chum	charm
		buck	bark
		muck	mark
		butter	barter
	bodge		barge
	cod		card
	clock		clerk
	dot		dart
	God		guard
	hop		harp

	lodge		large
	mock		mark
	potty		party
	sob		Saab
	shock		shark
	shop		sharp

Appendix 3

Instructions to the experiment (for participants)

This was sent to all potential participants.

Liebe Teilnehmerinnen und Teilnehmer,

mein Name ist Fabian Neumann und ich bin Student an der Universität Tromsø in Norwegen. Im Rahmen meiner Masterarbeit im Fach Linguistik führe ich eine Studie durch.

Wenn Sie diese Zeilen lesen, scheinen Sie die Teilnahme an meiner linguistischen Studie in Erwägung zu ziehen. Dafür möchte ich mich ganz herzlich bedanken!

Für das Experiment benötige ich Muttersprachler/innen des Deutschen, welche weiterhin über gute Englischkenntnisse verfügen. Ich bin daran interessiert, wie Deutsche die englische Sprache wahrnehmen und anwenden. Es geht ausschließlich um das Erkennen und Produzieren einzelner englischer Wörter (KEINE Grammatikaufgaben, KEINE komplexen Hörverstehensaufgaben o.Ä.). Ihre durch einen Klick ausgewählten Antworten zeichne ich auf. Ebenso werde ich Sie bitten, einzelne Wörter laut vorzulesen, sodass diese aufgezeichnet werden können. Dafür müssen Sie Ihrem Browser gestatten, auf das Mikrofon Ihres Endgerätes zuzugreifen.

Die Auswahl an Wörtern reicht von sehr gebräuchlichen Vokabeln bis hin zu ein paar „Exoten". Die Befragung wird anonymisiert durchgeführt. Sollten Sie ein mobiles Gerät wie

Handy oder Tablet verwenden, dann benutzen Sie dieses bitte im horizontalen Modus (quer). Um aussagekräftige Ergebnisse zu erhalten ist ein Vorgehen in drei Runden vorgesehen. Runde 1 besteht aus drei Teilen; Teil a,b und c. Diese sind am gleichen Tag, nacheinander zu absolvieren. (ca. 25 Minuten).

Teil a: https://www.babylex.uiocloud.no/8ae81beb-3e21-4dc4-808a-696e6f07b0b5/information/

Teil b: https://www.babylex.uiocloud.no/517be42b-88aa-4e17-b543-05932f87e7d3/information/

Teil c: https://www.babylex.uiocloud.no/5e9b62e3-f4ba-4463-84a0-ad4bc961a195/information/

Runde 2 besteht aus nur einem Teil (ca. 10 Minuten):

https://www.babylex.uiocloud.no/093117a7-cf0f-4475-a3fe-4e692c9e6d54/information/

Runde 3 besteht wieder aus drei Teilen; Teil d, e und f. Diese sind wieder nacheinander zu absolvieren (ca. 25 Minuten).

Teil d: https://www.babylex.uiocloud.no/8f521dd1-7c86-4046-bedb-3fccdc106379/information/

Teil e: https://www.babylex.uiocloud.no/62ad0142-668e-4395-a556-a4886f4bf8f5/information/

Teil f: https://www.babylex.uiocloud.no/154d26e5-4274-4a0d-8fcd-68b5ea791183/information/

Ich bitte Sie, die drei Runden des Experiments innerhalb einer Woche zu bearbeiten, dabei jedoch NICHT mehr als eine Runde pro Tag zu absolvieren. Kopieren Sie die Links einfach in Ihren Browser. Die Studie startet dort automatisch. Detaillierte Anweisungen finden Sie im Experiment, unmittelbar vor Beginn der Befragung.

Haben Sie Interesse an den Ergebnissen dieser Untersuchung oder einer Visualisierung Ihrer Aufnahmen, schicken Sie mir gerne eine Nachricht an: fabian.n92@gmail.com. Unter Angabe Ihrer „Kennung“ (siehe Experiment) kann ich Ihnen Ihre Daten zukommen lassen.

Ich möchte mich ganz herzlich bei allen Teilnehmerinnen und Teilnehmern für Ihr Zeitinvestment bedanken! Nach erfolgreicher Teilnahme ist eine Verbesserung Ihrer aktiven und passiven Fertigkeiten der englischen Sprache nicht ausgeschlossen ;)

Dear participants,

My name is Fabian Neumann, and I am a student at the University of Tromsø in Norway. As part of my master's thesis in linguistics, I am conducting a study. If you are reading this, it seems you are considering participating in my linguistic study, and for that, I sincerely thank you!

For the experiment, I need native speakers of German who also have good English skills. I am interested in how Germans perceive and use the English language. The study involves recognizing and producing individual English words only (NO grammar tasks, NO complex listening comprehension tasks, etc.). Your selected responses by clicking will be recorded. Additionally, I will ask you to read out individual words aloud, which will be recorded. For this, you will need to allow your browser to access the microphone on your device.

The selection of words ranges from very common vocabulary to a few "exotic" ones. The survey will be conducted anonymously. If you are using a mobile device such as a phone or tablet, please use it in landscape mode (horizontal). To obtain meaningful results, the study is divided into three rounds. Round 1 consists of three parts: parts a, b, and c, which should be completed consecutively on the same day (approximately 25 minutes).

Part a: https://www.babylex.uiocloud.no/8ae81beb-3e21-4dc4-808a-696e6f07b0b5/information/

Part b: https://www.babylex.uiocloud.no/517be42b-88aa-4e17-b543-05932f87e7d3/information/

Part c: https://www.babylex.uiocloud.no/5e9b62e3-f4ba-4463-84a0-ad4bc961a195/information/

Round 2 consists of only one part (approximately 10 minutes):

https://www.babylex.uiocloud.no/093117a7-cf0f-4475-a3fe-4e692c9e6d54/information/

Round 3 again consists of three parts: parts d, e, and f, which should be completed consecutively (approximately 25 minutes).

Part d: https://www.babylex.uiocloud.no/8f521dd1-7c86-4046-bedb-3fccdc106379/information/

Part e: https://www.babylex.uiocloud.no/62ad0142-668e-4395-a556-a4886f4bf8f5/information/

Part f: https://www.babylex.uiocloud.no/154d26e5-4274-4a0d-8fcd-68b5ea791183/information/

Please complete the three rounds of the experiment within one week, but do not complete more than one round per day. Simply copy and paste the links into your browser. The study will start automatically. Detailed instructions are provided in the experiment, just before the survey begins.

If you are interested in the results of this study or in receiving a visualization of your recordings, please feel free to send me a message at fabian.n92@gmail.com. Please include your "identifier" (see the experiment), and I can provide you with your data.

I would like to express my sincere thanks to all participants for investing your time! Successful participation may lead to an improvement in your active and passive English language skills ;)

(Google Translate)