UNIVERSITY OF AMSTERDAM
Faculty of Humanities

The influence of orthography on second language perception: an experiment with Marathi listeners of German and Dutch

Maxim Dauenhauer

Student number: 12993050

MA Thesis

Supervisor: Dr. Silke Hamann

July 9, 2025

Contents

Abstract

English /d/ and /t/ are unanimously adapted into Indo-Aryan languages as retroflex voiced /ɖ/ and retroflex voiceless /ʈ/, respectively. Previous research (Dauenhauer 2025) argued that these adaptations happen due to specific cue-constraint rankings in the L1-perception grammar. This perception experiment investigates these claims by presenting naïve listeners of the Indo-Aryan language Marathi with stimuli from German, which is phonetically very similar to English, and Dutch, which unlike English is a voicing language with dental stops. One key methodological innovation is the elicitation of an orthographic answer, thereby directly inferring participants' categorical perception as opposed to their production. Stimuli were first presented only in audio form, followed by a second block where the spelling of each stimulus was provided.

German coronal stops are perceived as more retroflex than Dutch ones by Marathi listeners in this experiment. Additionally, the ambiguity of the German stops regarding Marathi's four-way laryngeal contrast was largely resolved for many participants by providing orthographic input. This could indicate that orthography might also have played a role when English coronal stops were adapted into Marathi. This study provides evidence for perception based accounts that rely not simply on acoustic similarity, but on a language- and speaker-specific ranking of cue-constraints and orthographic constraints.

Keywords: perception experiment, second-language acquisition, phonetics-phonology interface, Indo-Aryan languages

1 Introduction

Like most Indo-Aryan languages, Marathi has a two-way coronal stop contrast, consisting of a retroflex and a dental stop series (Dhongde & Wali 2009). Due to India's colonial history, many

English words have entered Marathi and many other Indo-Aryan languages as loanwords, and many L1 Marathi speakers also speak (Indian) English quite proficiently, or even natively.

When speakers of a language like Marathi adopt English alveolar stops, they have to make a decision: do they adapt them as dental or as retroflex? The answer seems to be very clear, namely that Indo-Aryan languages such as Marathi unanimously adapt the English alveolar stops /t/ and /d/ as retroflex rather than dental (e.g., Ohala 1983 for Hindi). For instance, English *taxi* is adapted as [ʈæksi] and English *doctor* is adapted as [ɖɔkʈər] in Marathi[1]. Like most Indo-Aryan languages, Marathi also has a four-way laryngeal contrast, unlike the two-way laryngeal contrast found in English. It is thus interesting that English /d/ and /t/ are adapted as /ɖ/ and /ʈ/ in Marathi, rather than /ʈ/ and /ʈʰ/, for instance.

In Dauenhauer (2025), I provided an OT-formalization explaining these adaptations into Indo-Aryan languages. The underlying assumption was that both loanword adaptation and second language acquisition (at least at the onset of learning) are governed by the same perceptual processes — namely, first language phonological perception (Boersma & Hamann 2009). This made it unnecessary to distinguish between these two processes — an assumption which is also crucial for the present experiment.

In this paper, I present a follow-up empirical study on the basis of the results of my formalization, using one particular Indo-Aryan language: Marathi. Unfortunately, it is difficult to investigate how English is perceived by Marathi listeners nowadays since contact between English and Marathi was made over 200 years ago and many speakers nowadays simply learn Indian English rather than Western varieties of English. Thus, one would never be sure whether a certain adaptation is due to listeners' perception or because they simply learned a given sound as is. Therefore, this paper investigates how Marathi listeners perceive the coronal stops of two languages which they have no exposure to: German and Dutch. These languages are interesting insofar as one of them, German, essentially has the same alveolar stop system as English, whereas the Dutch stop system has important phonetic differences in terms of both both place of articulation (PoA) and laryngeal contrast (i.e., the way that a phonological voicing distinction is made phonetically; see Section 1.2).

[1] Examples adapted from Arsenault (2006) for Hindi.

Lastly, the present experiment investigates how orthography interferes with the second-language perception process, which has long been neglected in second-language acquisition (SLA) theories (Zhou & Hamann 2024). Due to the established unanimous adaptation of English alveolar stops as retroflex, Marathi listeners are used to the Roman letters <d> and <t> referring to retroflex sounds, which could potentially interfere with their perception of the foreign stops. Analogously, <d> and <t> also always represent plain and voiced sounds in Marathi and Indian English, never aspirated sounds, which could also potentially influence speakers' perceptions and subsequent adaptations of the corresponding English sounds. The present experiment aims to show this by first presenting Dutch and German stimuli only in auditory form, followed by a block where orthography is presented alongside audio.

The remainder of this section is organized as follows: Section 1.1 will outline some general information on Marathi's phonology; Section 1.2 will provide the relevant acoustic details of German, Dutch and Marathi; Section 1.3 will give some background on the influence of orthography on perception and Section 1.4 will present the research questions and their embedding in phonological theory.

1.1 The phonology of Marathi

Table 1 shows Marathi's coronal stop inventory, which is similar to that of many Indo-Aryan languages. As a result of a mixture between innovation and borrowings from Dravidian languages, almost all Indo-Aryan languages like Marathi systematically distinguish a dental from a retroflex PoA. One of the very few exceptions is Assamese, for instance (Cardona & Luraghi 2018: 383), which only has one coronal place of articulation. Thus, while the exact inventory and its exact phonetic implementation can vary from language to language, the inventory displayed in Table 1 is quite universal across Indo-Aryan languages, making insights gained from this experiment likely to be generalizable across many of them.

		Dental	Retroflex
Voiced plosive	unaspirated	d̪	ɖ
	aspirated	d̪ʱ	ɖʱ
Voiceless plosive	unaspirated	t̪	ʈ
	aspirated	t̪ʰ	ʈʰ

Table 1: Marathi’s coronal plosive inventory (Dhongde & Wali 2009: 11).

1.2 Acoustics of German, Dutch and Marathi stops

As mentioned before, investigating Marathi speakers’ naïve perception of English would be quite problematic since even people with limited English knowledge would be used to many English loanwords having a retroflex adaptation. Thus, this experiment uses stimuli from German, which happens to be very similar to English in the relevant aspects: German tends to have alveolar (e.g., Kohler 1976) rather than dental stops and it is an aspiration language rather than a voicing language (e.g., Jessen 1999). In other words, German contrasts plain voiceless stops with aspirated stops in word initial position. This can differ in other positions, however both Dauenhauer (2025) and the present experiment focus only on phrase/word-initial stops.

In contrast, Dutch coronal stops are different in both PoA and laryngeal contrast: Dutch tends to have dental stops (Jongman et al. 1985) and is a voicing language (e.g., Lisker and Abramson 1964); i.e., it contrasts prevoiced stops with plain aspirated stops. Table 2 summarizes these differences.

	/d/	**/t/**
German/English	[d̥]	[tʰ]
Dutch	[d̪]	[t̪]

Table 2: Phonetic realizations of the two coronals stops in German, English and Dutch.

Unlike Marathi, neither Dutch nor German contrasts more than one coronal PoA for stops, meaning that there is quite some space for variation (see, e.g., Dart 1998 for English). Nevertheless, the canonical coronal PoAs are dental for Dutch and alveolar for English and German. It should also be mentioned here that due to articulatory factors, German /d/ has a tendency to be more retracted than its voiceless counterpart /t/ (Hamann & Fuchs 2010), which would naturally make it sound more retroflex to a Marathi speaker. Therefore, it will be important for the analysis to distinguish whether the stimulus contained /t/ or /d/. Similar effects could be possible for Dutch as well.

The specific cues related to dental, alveolar and retroflex articulations are numerous and with complex interactions. The principal distinguishing cue between dental/alveolar and retroflex stops is typically the third formant (F3), which is mid for the former group and quite low for the latter (Hamann 2003: 63). However, it is likely that any combination of spectral cues, burst intensity and duration, and even voice onset time (VOT) are important for the dental/retroflex distinction in Indo-Aryan languages such as Marathi (see Dauenhauer 2025 for a summary and values for these cues). Dutch coronal stops should be similar to Marathi dental stops due to their (canonically) identical PoA. However, just like English stops, German stops are ambiguous for Marathi ears as their PoA (canonically) lies in between dental and retroflex.

As for VOT measurements, Marathi is reported to have an average VOT of -115 ms for voiced stops, 11 ms for plain stops and 73 ms for aspirated stops (averaged across PoA by Dmitrieva and Dutta 2019: 212; data from Lisker and Abramson 1964). English tends to have a VOT of around 50 ms–90 ms for /t/ (Sundara 2005, p. 1030; Dart 1991, p. 88) and slightly above 0 (Sundara 2005: 1030) for /d/ for North American English. Similarly, Jessen (1999: 86) reports an average VOT of 74 ms for phrase-initial voiceless stops and 21 ms for phrase-initial voiced stops in German. Note that these numbers differ for intervocalic positions, where /d/ may become fully voiced in both languages. For Dutch, on the other hand, Lisker and Abramson (1964: 392) report an average VOT of -80 ms for /d/ and 15 ms for /t/.

Comparing these figures, Dutch coronal stops are very similar to Marathi /d̪/ and /t̪/, although Marathi voiced stops have an even longer prevoicing duration. The values for aspiration in Marathi are somewhat puzzling though, as the average VOT reported by Lisker and Abramson (1964) and Dmitrieva and Dutta (2019) for Marathi aspirated stops is quite similar to the aspiration duration of German /t/ reported by Jessen (1999), namely around 75 ms. This goes against the well-established principle of auditory dispersion (Boersma & Hamann 2008), which results from the pressure for languages to keep their phonological categories maximally distinct. This usually leads to languages with a three-way laryngeal contrast having longer aspiration durations than those with a two-way laryngeal contrast. Additionally, comparing VOT measures between studies is very tricky since they are quite susceptible to speech rate differences and the VOT for Marathi aspirated stops can range from 50 ms to 150 ms (Dmitrieva & Dutta 2019: 222). Thus, German /t/ lies either in between Marathi's plain and aspirated stops or is comparable to Marathi's aspirated stops, whereas German /d/ should be quite similar to plain voiceless stops in Marathi (see Section 1.4 for the specific predicted interpretations).

1.3 The role of orthography

Orthography has been shown to interfere with the second-language perception process (Giannakopoulou et al. 2013; McGuire 2014; Zhou & Hamann 2020), ranging from hindering target-like acquisition to having no effect at all (see Bassetti et al. 2015 for a review). For instance, Zhou and Hamann (2020) showed that when spelling is presented alongside an audio clip, some naïve Mandarin Chinese speakers were more likely to imitate the European Portuguese tapped /ɾ/ with a sound corresponding to that spelling in their own orthography. Specifically, some Mandarin listeners imitated this Portuguese sound as [ɹ], but only when they were presented with the corresponding Portuguese orthography. This happened likely due to their association of this Chinese sound with the letter <r>, which is used to represent it in Pinyin (the most common Chinese romanization system).

Zhou and Hamann (2020) also found substantial individual differences for this effect, meaning that certain participants were more prone to this orthography-effect, while others did not exhibit the effect at all. This suggests different individual weightings of information from auditory cues versus orthographic cues (see also Zhou and Hamann 2024 for a formalization of individual differences and orthographic influence in SLA). It will thus be interesting to see whether such listener-types also emerge in the present experiment.

Lastly, research about orthographic effects in SLA usually focuses on cases where one L1 category covers two L2 categories (this is true for all studies cited in this section). This is natural, as learning how to successfully make a non-native contrasts is important for SLA-learners. However, if we want to gain a holistic understanding of L2 perception, we should also consider the opposite case; i.e., having to decide which of several L1 categories to use for one L2 category. The present experiment does just that, as it investigates the mapping of a two-way coronal contrast to only one coronal category; as well as the mapping of a four-way laryngeal contrast to a two-way laryngeal contrast.

1.4 Theoretical framework and research questions

Just like Dauenhauer (2025), the present paper will work within the framework of bidirectional phonology and phonetics (BiPhon, Boersma 2011). In this formalization, I argued that the loanword adaptation/SLA process can be explained entirely with the native perception grammar. Specifically, this means that perception is already phonological and that the same constraints are used in the comprehension process as in production, which are two important tenets of the BiPhon model (for a more in-depth theoretical discussion, see Dauenhauer 2025). The present paper aims to provide empirical evidence for these assumptions.

Evidence for a perception-based adaption is provided by Ohala (1978), who found that English alveolar stops were perceived as retroflex by Hindi speakers 91% of the time. While these findings provide an important basis for the present experiment, this study was quite small-scale with only seven participants and no statistical analysis. Thus, the present paper presents — to the best of my knowledge — the second only perception experiment investigating the adaptation of alveolar stops into Indo-Aryan languages. Ohala (1978) proposes several explanations for her findings, invoking concepts like markedness and acoustic similarity. While she does acknowledge the relevance of perception, listeners cannot possibly have access to what is acoustically similar and what is marked. Acoustic similarity continues to be used by models such as the Perceptual Assimilation Model (e.g., Best and Tyler 2008, Kogan and Mora 2022) or the SLM-r (Flege et al. 2021).

Rather than relying solely on acoustic similarity, this paper assumes that foreign words are adapted by being filtered through the L1-perception mechanism. Showing this would serve as evidence against accounts that propose loanword/L2-specific treatments outside of the native grammar (e.g., LaCharité and Paradis 2005, Kang 1996, Silverman 1992, Yip 1993; again see Dauenhauer 2025).

Thus, the first research question/aim of the present experiment is to show that foreign coronal stops are already adapted into Marathi in perception, by using German as a proxy to (historical) English and comparing it to Dutch. Dutch was chosen here because it differs from German and English in both relevant regards, namely PoA and laryngeal contrast. In principle, any other language that fulfills these criteria would be applicable here. Similarly, any Indo-Aryan language other than Marathi would have been applicable for this experiment and Marathi was chosen for purely practical reasons to do with participant recruitment.

As a result of the specific phonetic properties of the coronal stops in each language, the adaptations in Table 3 are expected. It shows that retroflex stops are predicted for German due to the alveolar PoA of German stops — just like English. For Dutch, dental adaptations are expected, which should be a straightforward adaptation since both Marathi and Dutch have dental stops.

	d	**t**
Marathi interpretation of **German**	/ɖ/	/ʈ/
Marathi interpretation of **Dutch**	/d̪/	/t̪/

Table 3: Hypothesized interpretations of the German and Dutch coronal stops by Marathi listeners.

Since the present experiment aims to tap into the phonological categorization made by Marathi listeners, one methodological innovation compared to Zhou and Hamann (2020) will involve eliciting an orthographic response rather than a spoken one. Crucially, the Marathi script (Devanagari) makes a distinction between retroflex and dental characters. Thus, by asking participants to write down their perception of the German/Dutch words in their own script, we can directly infer which distinct category they perceive the sounds as, rather than asking them to produce a replication that could exist anywhere on a number of acoustic spectra. Another problem with articulatory responses is that there exists the possibility that listeners *are* able to perceive the foreign sound correctly, but subsequently fail to reproduce it accurately. The present methodology eliminates this added layer of production.

Additionally, the present experiment investigates whether the hypothesized dental interpretation of Dutch /d/ and /t/ can be overridden by orthography due to participants being used to <d> and <t> representing retroflex sounds in Indian English. This can be considered a sub-question of the first research question and is certainly a more exploratory aspect of this paper, especially seeing as Marathi is nowadays also commonly spelled in Roman, where <d> and <t> can represent both dental and retroflex stops. This is a crucial difference to Zhou and Hamann (2020) because Pinyin <r> only stands for Mandarin /ɻ/.

Regarding the laryngeal contrast, both German and Dutch stops are expected to be adapted as prevoiced and plain, even though German voiceless stops are aspirated and German voiced stops are phonetically plain voiceless. However, Marathi aspirated stops might or might not have an even longer aspirated portion than German voiceless stops (as discussed in Section 1.2) due to Marathi's four-way laryngeal contrast. Therefore, German /t/ might not be aspirated enough for a Marathi speaker to perceive it as aspirated (see Dauenhauer 2025 for Indo-Aryan languages; Boersma and Hamann 2009 for Korean).

German /d/ might also be tricky. Due to its lack of prevoicing, it might very well be interpreted as plain voiceless /ʈ/, rather than /ɖ/. This is *not* predicted by Dauenhauer (2025), where the difference in intensity between English /d/ and /t/ ended up being crucial for English /d/ to be perceived as Marathi voiced /ɖ/. This would work analogously for German, leading to the predictions in Table 3. However, if German /d/ *is indeed* perceived as plain voiceless rather than voiced, this allows for another interesting hypothesis. In Dauenhauer (2025), one potential alternative explanation for the voiced adaptation of English /d/ in Indo-Aryan languages was orthography. Specifically, the letter <d> suggests that the sound should be voiced, even if it is often realized as plain in English and German. Therefore, if a listener in the present experiment opts for a voiceless transcription of German /d/, this might be overridden by the orthography condition, which clearly suggests that voicing should be present. Similarly, if German /t/ *is* interpreted as aspirated by some speakers, this interpretation could be overridden by orthography. Again, individual variation is likely here since orthography effects tend to affect different people to different degrees (Zhou & Hamann 2020). Thus, the second research question of this experiment is how German (and Dutch) coronal stops will be interpreted regarding Marathi's complex laryngeal contrast and if a potential plain perception of German /d/ can be overridden by orthography to be perceived as voiced.

In summary, the present experiment aims to answer two main research questions, each with one sub-question regarding orthography:

RQ1. Are German and Dutch /d/ and /t/ perceived as retroflex or dental by Marathi listeners?

  RQ1a. Can their interpretation be overridden by providing orthographic input?

RQ2. Are German and Dutch /d/ and /t/ perceived as voiced, prevoiced or aspirated by Marathi listeners?

  RQ2a. Can their interpretation be overridden by providing orthographic input?

The respective predictions for the experiment are thus as follows:

P1. German /d/ and /t/ will be perceived as more retroflex than Dutch /d/ and /t/.

  P1a. Orthography will increase retroflex adaptations.

P2. German and Dutch /d/ will be perceived as voiced; German and Dutch /t/ will be perceived as voiceless. If there is any language difference, German stops would be more likely to be perceived as aspirated (for /t/) or plain (for /d/).

  P2a. If there are aspirated responses for /t/ or plain responses for /d/, orthography will reduce those responses.

2 Methods

In order to answer the questions laid out in the previous section, a web-experiment was set up, which was approved by the Ethics Committee of the University of Amsterdam, Faculty of Humanities (project number: FGW-5133).

2.1 Participants

29 native speakers of Marathi with no prior experience with or exposure to German and Dutch were recruited, of which 4 were excluded due to various reasons (see below), resulting in 25 remaining participants. All participants reported that Marathi was their first language and that they were fluent in English. Out of the non-excluded participants, the mean age was 27.4 years with a range of 17–67 years and a standard deviation of 10.78 years. 8 of them went to an English-medium school, 15 to a Marathi-medium school and 2 to a Semi-English medium school, which are terms referring to the main language of instruction at their school.[2] Participants were recruited via contacts at the Linguistics Department at Deccan College Pune and personal contacts/social media. No payment was offered.

Three participants were excluded because they exhibited a lexical effect to different degrees. One participant thought that the experiment required them to write down the closest *lexical word* in Marathi. This was evident because they indicated this in the post-hoc questionnaire (see Section 2.3) and all of their answers were indeed real Marathi words. Similarly, another participant reported in the post-hoc questionnaire that they sometimes went by which words were similar to Marathi, which is why they were also excluded, even if not all of their answers were real words in Marathi. This was done because for the words that *were* lexical words, there was no way of knowing whether they were deliberately chosen or whether their lexical status was simply coincidental. The third participant was excluded because they likely also exhibited a lexical effect, but did not indicate so in the post-hoc questionnaire. Specifically, there were two instances where they added a syllable at the beginning of a word, which made them into lexical words in Marathi. Those instances were German /toːn/ being transcribed as *mithūn*[3] 'gemini (astrological sign)' and German /tuːn/ being transcribed as *kuṭhūn* 'from where'. These examples illustrate that this lexical effect directly impacted the chosen PoA for the target sound (dental in the former case; retroflex in the latter, indicated by the subdot). P20 initially seemed to misunderstand the experiment because their first

[2] A Semi-English medium school is typically a school in which science and math are taught in English, with the rest of the subjects being taught in Marathi.

[3] Transliterations are given according to the IAST (International Alphabet of Sanskrit Transliteration, Transliteration Committee 1895).

few answers bore very little acoustic resemblance to the stimuli, such as transcribing German *des* as *ho* ‘yes’ or German *moos* as *undir* ‘mouse, rat’. They likely did this because they perceived these stimuli as the English words *yes* and *mouse*. However, after the first three tokens, their answers seemed perfectly logical and similar to the stimuli, which is why the participant as a whole was kept. Lastly, another participant was excluded because they translated many of the stimuli when their written form was the same or similar to an English word into the corresponding Marathi word. For instance, they transcribed German *die* as *maraṇ* ‘dying, to die’.

2.2 Stimuli

The stimuli consisted of 40 target words and 40 fillers, all of which had the form CV(V) or CV(V)C. Half of both the target and filler words was German and the other half was Dutch. Within each language, again half of the target words started with /d/ and the other half with /t/. The target words contained no /d/ or /t/ in the coda and the fillers contained none whatsoever. Across all stimuli, sounds that were judged ‘difficult’ for Marathi speakers because they would not have any similar sounds in their inventory were largely avoided in order to avoid participants spending a lot of time thinking about things that would not be analyzed regardless. These sounds included for instance front rounded vowels such as /y/. Other than that, the target consonants occurred in a variety of vowel environments in both languages, so that roughly a similar amount of front/back and high/low vowels would be present in each list. Table 4 shows an overview of all items. A full list of stimuli can be found in Appendix A.

<table>
<tr><th></th><th>German</th><th>Dutch</th></tr>
<tr><td>/t/</td><td>10</td><td>10</td></tr>
<tr><td>/d/</td><td>10</td><td>10</td></tr>
<tr><td>Filler</td><td>20</td><td>20</td></tr>
<tr><td rowspan="2">Total</td><td>40</td><td>40</td></tr>
<tr><td colspan="2">80</td></tr>
</table>

Table 4: Number of stimuli within each condition.

Stimuli were recorded with a Neumann TLM103 cardioid condenser microphone in a sound-proof booth by two native speakers of German and Dutch, respectively, yielding four sets of recordings. The German speakers will be referred to as DE1 and DE2 and the Dutch speakers as NL1 and NL2 for the remainder of this paper. For each language, speaker 1 was male and speaker 2 was female. The number and gender of the speakers was chosen this way to avoid participants associating one particular language with one speaker or gender. This could potentially have led to them choosing a default option for each language or speaker/gender. Having more than one speaker also made any influence of speaker-specific pronunciations less likely. The German speakers were from Göppingen and Emlichheim, respectively; and the Dutch speakers grew up in Groningen and Wormer, respectively. All speakers' accents were judged to be a standard accent of their respective language. The stimuli were extracted from the recordings and their peak scaled to 0.99 using Praat (Boersma & Weenink 2025). This ensured equal loudness across speakers and tokens.

For the two Dutch speakers, there were five /d/ tokens of which none of the three repetitions recorded had proper prevoicing. These five tokens were *duim* and *deel* spoken by NL1 and *diep*, *duim* and *doel* spoken by NL2. This of course has implications for the hypotheses regarding participants' perception of the laryngeal contrast in Dutch (see Section 3 and 4). Specifically, it could potentially skew the results of /d/ tokens to be perceived as more plain on average than they would have been if all tokens had proper prevoicing. It should be noted, though, that prevoicing is not the only cue that listeners could make use of here. The devoiced /d/ tokens still had an audibly less intense release-burst compared to the same speaker's /t/ tokens, which is something that Marathi speakers could feasibly pick up on (Dauenhauer 2025). For this reason and to ensure that the stimuli stayed balanced, no recordings were excluded.

2.3 Procedure

The experiment was designed using the software Experiment Designer (Vet 2025). Participants could perform the experiment via a shared link that could be accessed with any browser using a non-mobile device (such as a laptop). After being informed about the research and giving their consent to participate, they were asked about their language background, age and which type of school they went to. Specifically, they were asked whether Marathi was their first language and whether they were fluent in English; if either of these two questions were answered with ‘no’, the experiment immediately stopped for them and they were excluded from the experiment. The question regarding their education gave the options ‘English-medium’, ‘Marathi-medium’, ‘Semi-English’ and ‘other (please specify)’. This information was collected for potential post-hoc analysis since it is feasible that participants’ age and the type of their education might have affected their degree of exposure to Western English and English spelling, which would be relevant for the orthography-related aspects of this paper.

At the beginning of each trial, half of the tokens in each condition were randomly assigned to block one (= auditory only) or block two (= auditory + orthography). For both blocks, participants were instructed to provide a Marathi transcription in Devanagari that most closely matched the audio that they heard. In the orthography condition, all words were displayed entirely in lowercase despite the typical capitalization of nouns in German to not give away the language of the stimulus. Tokens were also balanced across speakers, ensuring that two out of the five tokens per condition and block were spoken by one speaker and three by the other speaker of the respective language. Each audio file could be repeated an infinite number of times by pressing the ‘Replay’ button. Participants could take as long as they wanted for their answer. An example screenshot from the experiment is provided in Figure 1.

After the main experiment, participants could optionally answer four post-hoc questions. This way, they had the chance to explain their reasoning for what they did, which might also lead to interesting insights. The four questions were as follows:

1. Were there sounds that you had a hard time writing down in Devanagari?
2. Do you think the written words changed or influenced your answers compared to when no writing was present?
3. Why did you write down the words containing <t> and <d> the way you did in Devanagari? (for instance, why did you choose ट rather than ड, त, द, ठ, थ, etc.)[4]
4. Do you have any other thoughts that you would like to share?

[4] These are the Devanagari characters for Marathi’s non-breathy coronal plosives, as displayed in Table 1.

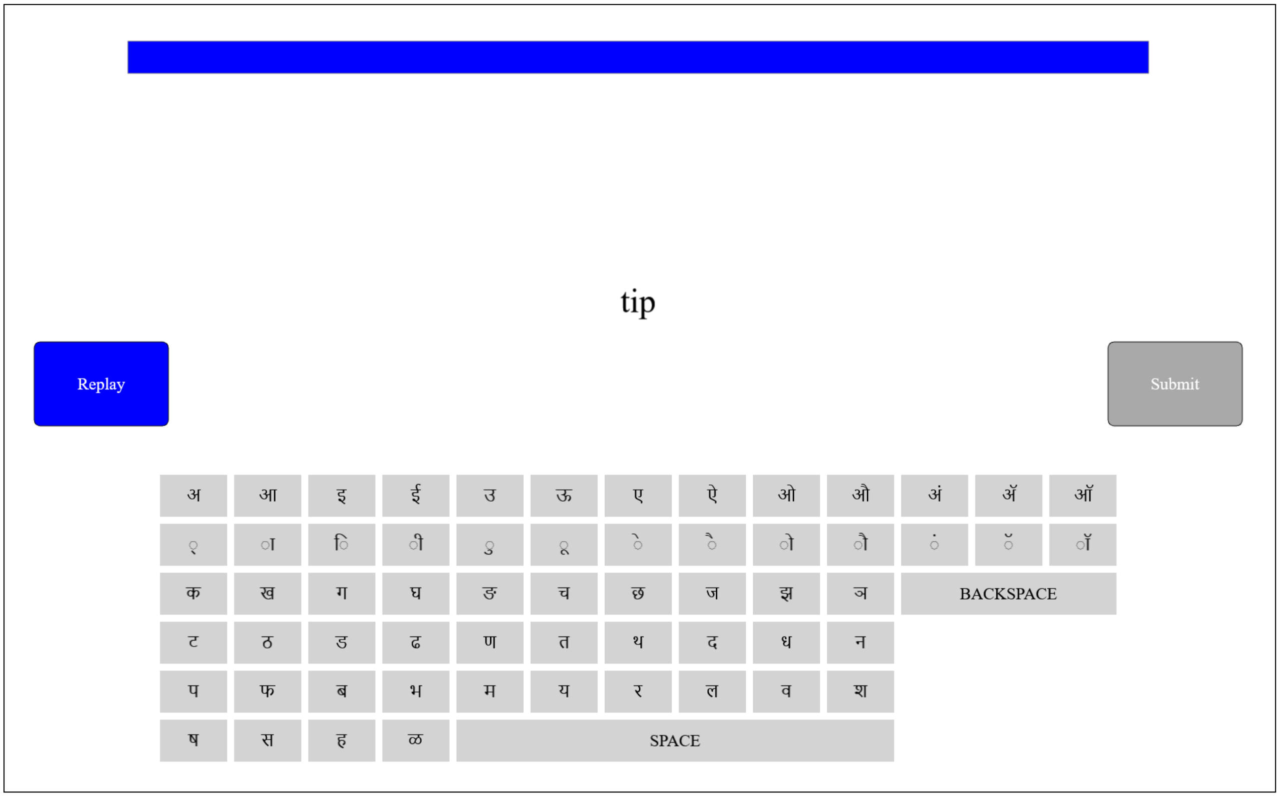

Figure 1: Example screenshot from a random item within block two (auditory + orthography).

2.4 Analysis

Statistical analysis was done in R (R Core Team 2020). For the first research question, a logistic mixed-effects regression was fit using the R-packages 'lme4' (Bates et al. 2015) and 'lmerTest' (Kuznetsova et al. 2017). The maximally complex model that still converged was selected using the 'buildmer' package (Voeten 2025), which does so by performing backwards elimination. For optimization, the "bobyqa" optimizer was used. Additionally, a diagonal covariance matrix for the random effects was assumed in order to make the model converge. The original formula used is displayed in (1).

(1) answerPlace $\sim$ language $*$ voicing $*$ orthography $+$ (language $*$ voicing $*$ orthography $||$ participant) $+$ (voicing $*$ orthography $||$ stimulus)

Here, *answerPlace* stands for a binary variable with either a retroflex (coded as 1) or a dental (coded as 0) adaptation; *language* is a binary variable containing the language of the stimulus; *voicing* is a binary variable containing whether the stimulus was phonologically voiced or voiceless in the stimulus language (i.e., /d/ or /t/); and *orthography* is a binary variable containing whether the stimulus was presented in the first block, i.e., auditory only, or in the second block, i.e., with orthography. After parameter selection, the model in (2) ended up being used, only containing a random intercept for participant and speaker.

(2) answerPlace $\sim$ language $*$ voicing $*$ orthography $+$ (1 || participant) $+$ (1 || stimulus)

The first hypothesis regarding PoA would be supported by a significant effect of *language*, where Dutch stimuli are more likely to be perceived as dental than German ones and vice versa (note that the fillers were not included in any analysis). The first sub-hypothesis regarding orthography would be supported by a significant effect of *orthography* (see Section 4 for more details). The factor *voicing* does not relate to any research question, but is included nonetheless because of the potential effects of voicin on PoA mentioned in Section 1.2. Additionally, the original model contained a random intercept and slopes for the random effects *participant* and *stimulus*. For each of the binary predictors, orthogonal contrast coding was applied; i.e., one level was coded as -0.5 and the other as +0.5.

For the second research question, a cumulative link mixed model was fit using the 'ordinal' package (Christensen 2018). In (3), *answerVoice* represents an ordinal factor with plain responses being in between voiced and aspirated responses. This model was chosen because at the very least phonetically speaking, there seems to be an inherent order of the possible answer values. Fitting a model with an unordered, categorical dependent variable would have failed to capture this reality. The model converged in its maximally complex form, rendering parameter selection unnecessary. Otherwise, the formula for this model was identical to the first. Breathy answers were dropped (n = 2) due to their ambiguous nature of having both prevoicing and aspiration. Again, the original formula used is displayed below:

(3) answerVoice $\sim$ language $*$ voicing $*$ orthography $+$ (language $*$ voicing $*$ orthography || participant) $+$ (voicing $*$ orthography || stimulus)

If Marathi listeners indeed pick up on the aspiration of German /t/ tokens and the lack of prevoicing of German /d/ tokens, we would expect a significant effect of language. If this could then be overridden by orthography, we would again expect a significant interaction between language and orthography.

3 Acoustic Analysis of the Stimuli

In order to ensure that the recorded stimuli did not substantially differ from their canonical properties outlined in Section 1.2, a small-scale phonetic analysis was conducted. The current section will outline this analysis.

All target words of all speakers were annotated in Praat (Boersma & Weenink 2025) for the following time points: the start of the burst of [d/t], the onset of prevoicing in the case of Dutch [d], the onset of voicing of the following vowel and the end of the following vowel. Figures 2 and 3 illustrate examples of these annotations. All timepoints were moved to the nearest zero crossing except the onset of the burst, as this would have affected the boundary too much in many cases.

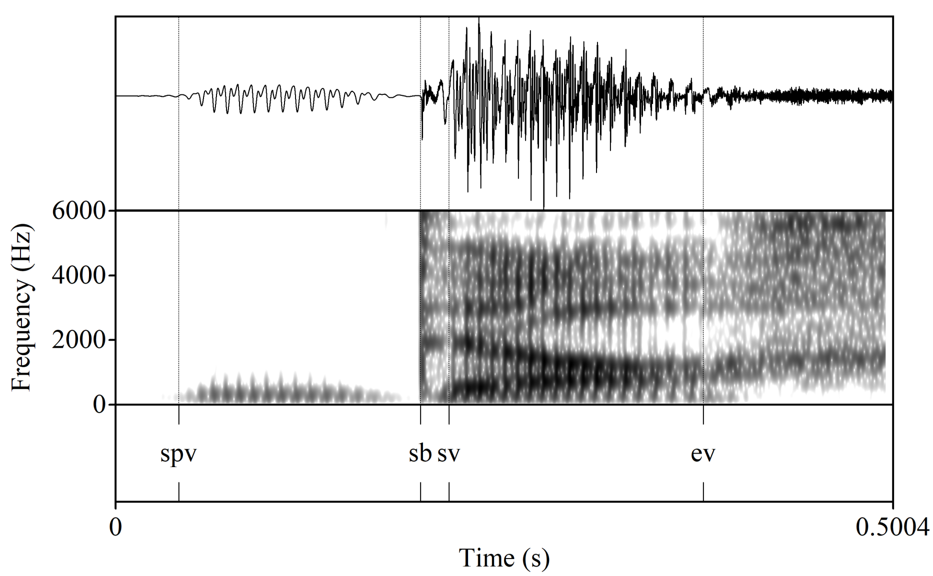

Figure 2: Boundaries drawn for the Dutch stimulus *dag*, spoken by speaker NL1. *spv* = start prevoicing, *sb* = start burst, *sv* = start voicing of following vowel and *ev* = end of following vowel.

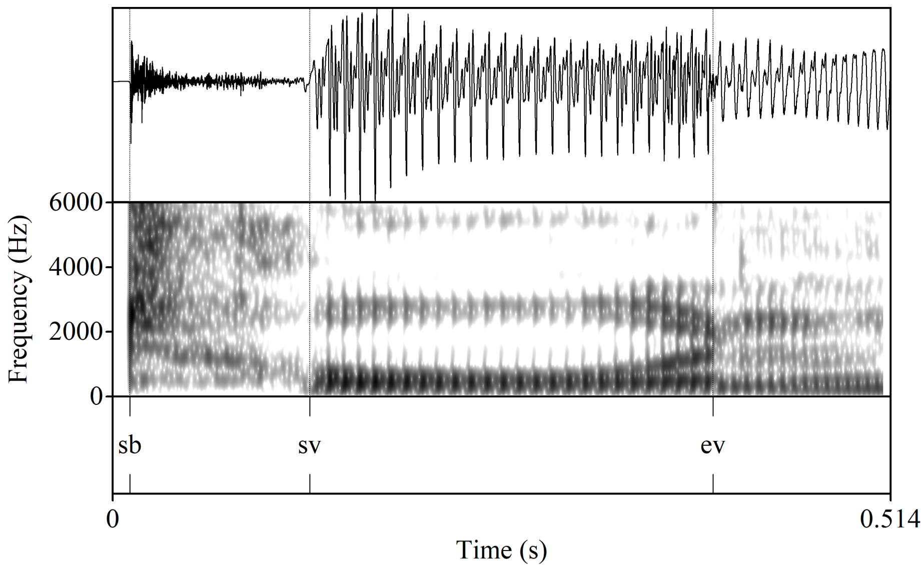

Figure 3: Boundaries drawn for the German stimulus *Ton*, spoken by speaker DE1. *sb* = start burst, *sv* = start voicing of following vowel and *ev* = end of following vowel.

For the start of the burst, the main annotating criterion was the spectrogram. The start of voicing was set at the onset of clear periodicity in the waveform and the end of the vowel was set where clear first and second formants faded out.

After annotation, the following measurements were extracted via a Praat script (which can be found in Appendix D): VOT was calculated as the difference between the onset of (pre-)voicing and the start of the burst; the four spectral moments (center of gravity (CoG), standard deviation (SD), skewness and kurtosis) were calculated within the first 5 ms[5] after the burst, with the standard setting of power = 2.0 in Praat; intensity was calculated by dividing the power within the first 5 ms of the burst by the power of the following vowel.[6] This was done because humans perceive burst intensity relative to the intensity of the surrounding speech signal. Lastly, the mean second (F2) and third formant (F3) were calculated within the first 5 ms after the onset of voicing of the following vowel. F2 and F3 were only calculated for *plain tokens*, i.e., Dutch /t/ and German /d/. This was done because measuring formant transitions after aspiration — which would have been necessary

[5] The original value chosen here was 10 ms, but some bursts were too short for this threshold.
[6] This method of measurement was chosen after PC with Paul Boersma.

in the case of German /t/ — is rather difficult due to the long aspirated portion, making the F2/F3 values not really transitioning anymore. Therefore, Dutch /t/ and German /d/ tokens were chosen for this analysis, as they have a very comparable VOT (see Figure 4). Tables with more detailed descriptive statistics of the measurements can be found in Appendix B. Burst duration was not measured since the aspirated portion in tokens with an even slightly positive VOT was difficult to distinguish from the burst, especially in /Ci/ tokens. Measurements for all spectral moments of the token *dom* by speaker NL2 were removed because the burst duration was under 5 ms, making the measurements faulty.

	dental	**retroflex**
Burst intensity	lower	higher
F2	high	high
F3	mid	low
Spectral CoG	higher	lower
Spectral SD	higher	lower
Spectral skewness	negative	positive
Spectral kurtosis	negative	positive

Table 5: Overall tendencies of several cues associated with laminal (dental) and apical (retroflex) stops across Indo-Aryan languages (Benguerel & Bhatia 1980; Hamann 2003; Hussain 2014; Hussain et al. 2017; Samudravijaya 2003; Verma & Chawla 2003). Table adapted from Dauenhauer (2025: 8).

Table 5 summarizes how Marathi speakers likely use the measured cues to make their dental vs. retroflex distinction in their own production/perception. We can now compare these tendencies to the measurements for the stimuli of the present experiment. The expectation here would be that the Dutch stimuli pattern more like Marathi dentals and that the German stimuli are more similar to Marathi retroflexes. All measurements that were taken are illustrated in Figure 4.

The acoustic measurements largely confirm the assumptions made about Dutch and German regarding this experiment. As for VOT, which is the main relevant factor for the second research question regarding Marathi's laryngeal contrast, the four speakers largely align with the measurements reported in the literature (see Section 1.2). This is illustrated in Figure 4 (top left), where

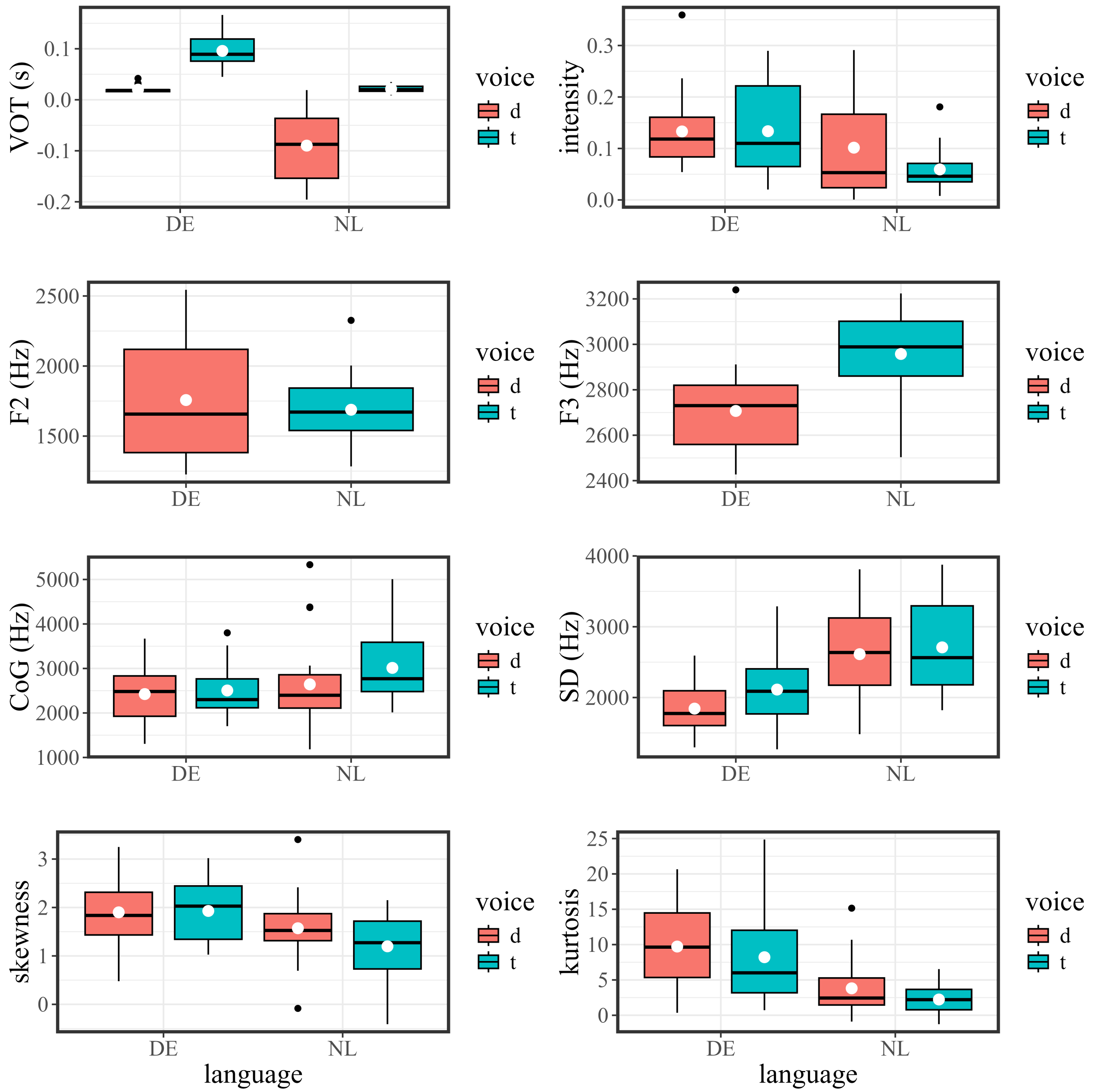

Figure 4: Boxplots for each acoustic measurement for the phonemes /d/ and /t/ in German and Dutch. For F2 and F3, only German /d/ and Dutch /t/ were measured. Horizontal lines indicate the median, white dots indicate the mean and black dots show outliers. Note that intensity does not have a unit here as it is calculated by dividing the power in the burst by the power in the vowel.

the German speakers clearly make a plain vs. aspirated distinction, compared to Dutch speakers making a prevoiced vs. plain distinction. The VOT values of German /d/ and Dutch /t/ are nearly identical, hovering around 20 ms.

Next, intensity is relevant for both the laryngeal contrast and the place contrast in Marathi. Regarding the laryngeal contrast, the results are actually counter to what one would assume. For both languages, both /t/ and /d/ essentially have the same median intensity, but the range of Dutch /d/ extends much higher, leading to a higher mean intensity. This of course is contrary to the expectation that fortis bursts would be louder than lenis bursts (e.g., Jongman et al. 1985: 245 for English).[7] As for the place contrast, the intensity results do pattern as would be expected (according to Table 5), with the German stops being louder on average than the Dutch stops. This likely stems from the more apical articulation of alveolar stops compared to dental stops.

As would be expected from Table 5, F2 values between the German speakers' /d/ and the Dutch speakers' /t/ were very comparable, although there was more variation for the German speakers. F3, however, which is a major cue differentiating dental and retroflex stops (Hamann 2003), was more substantially different. The German speakers' supposedly alveolar stops had a considerably lower F3 transition compared to the Dutch speakers' supposedly more dental stops, as would be expected. It should be noted that these values are very much subject to influence from the respective vowel environment, making them only somewhat reliable. However, this should be mitigated by the fact that only plain tokens were measured and the very small time window of 5 ms after the onset of voicing.

The results regarding the four spectral moments on average all correspond to the expectations, with the biggest difference between the German and Dutch stimuli being for kurtosis. It should be noted, though, that according to Table 5, we should expect a negative skewness and kurtosis for dental stops, but Dutch stops on average had positive values for both, albeit lower than German. This shows that while the Dutch stops were probably more dental than the German ones on average, they are likely not *as dental* as those found in Marathi. This is only logical since Dutch does not have to distinguish them with another coronal PoA, allowing for more variation and fewer 'truly' dental articulations. This will become important when discussing the results regarding PoA (see Section 5.1).

[7] This strange pattern for intensity was observed equally for the original threshold of 10 ms after the burst, meaning that this is not simply an artifact of a very small measurement window.

To conclude this section, while there are certainly a few oddities in these measurements — mostly the unexpected pattern of intensity between phonologically voiced and voiceless sounds — the stimuli largely correspond to the previously made assumptions and are thus suitable for this experiment. Nevertheless, language and voicing differences are far from clear-cut for most of the cues, meaning that we should expect considerable variation in how participants interpret these cues.

4 Results

This Section summarizes the results of the web-experiment. For the entire analysis, 21 tokens out of 1000 total target tokens were excluded as participants transcribed the target sound with a non-coronal character, usually either a velar or bilabial stop. The remainder of this section will first summarize the the results regarding PoA (Section 4.1), followed by the results for the laryngeal contrast (Section 4.2). Lastly, the post-hoc analyses of the questionnaire data will be explained in Section 4.3.

4.1 Place of articulation

Figure 5 illustrates overall results for the first research question regarding place of articulation. First, it is immediately clear that the German stimuli were overwhelmingly perceived as retroflex. The Dutch stimuli, on the other hand, are far less retroflex on average; however, more than half of the Dutch stimuli were still perceived as retroflex.

Table 6 shows the output of the model fit for place of articulation. As Figure 5 suggested, there was a significant effect of language ($p < 0.001$), with participants being 1.77 times more likely to interpret the German stimuli as retroflex compared to Dutch. While Figure 5 suggests the predicted orthography effect for both languages — but a bigger one for Dutch — neither the overall orthography effect ($p = 0.08$) nor the interaction between language and orthography (p = 0.94) were statistically significant. None of the other effects, including interactions, were significant.

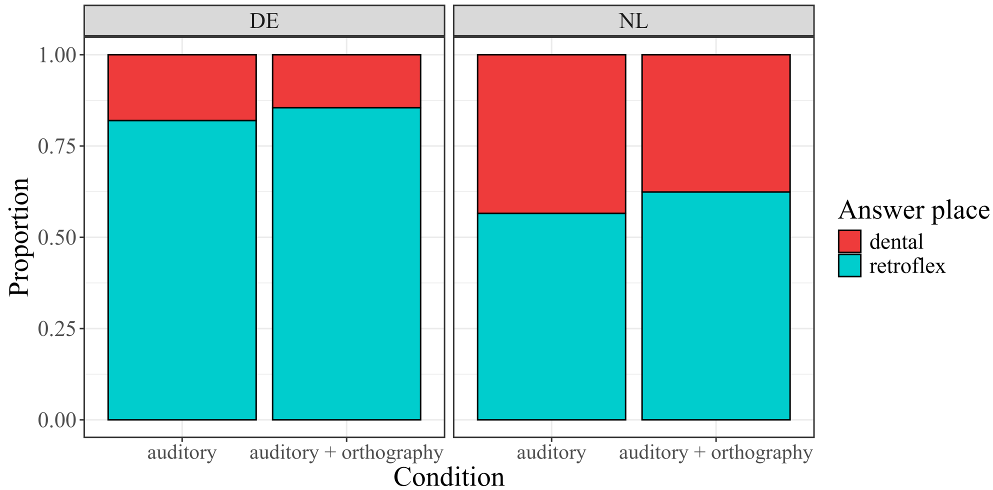

Figure 5: Proportion of retroflex and dental responses by language and whether orthography was present or not.

	Odds ratio	SE	z	p	Sig.
intercept	3.46	0.22	5.67	< 0.001	***
lang-NL+DE	4.75	0.26	6.07	< 0.001	***
voice-d+t	1.60	0.25	1.86	0.062	.
orthography-yes+no	0.74	0.17	-1.76	0.079	.
lang-NL+DE:voice-d+t	1.94	0.51	1.31	0.19	
voice1:orthography-yes+no	1.35	0.34	0.88	0.38	
lang-NL+DE:orthography-yes+no	0.98	0.34	-0.072	0.94	
lang-NL+DE:voice-d+t:orthography-yes+no	0.58	0.68	-0.81	0.42	

Table 6: Output of the maximally possible Glmer for place of articulation. Displayed are estimated odds (ratios), standard error, z and p-value for both decision thresholds and all fixed effects including interactions. The target variable was coded as 1 for retroflex and 0 for dental. Values with '-' are coded as -0.5 and values with '+' are codes as +0.5. 'Sig.' indicates degree of significance.

As expected, there was considerable variation between participants, which is illustrated in Figure 6. Here, we can clearly see that some speakers seem to exhibit a strong orthography effect, such as P03 or P11. Both of these participants used a roughly equal number of retroflex transcriptions as dental transcriptions for German, but they switched to using only retroflex ones once orthography was present. Other participants were not as susceptible to orthography, with some even exhibiting a reverse orthography effect. For instance, P18 changed their answers for the German stimuli from mostly retroflex to 100% dental when orthography was provided. P21 also only gave dental answers for Dutch with orthography, although they already almost exclusively interpreted the Dutch stimuli as dental without orthography.

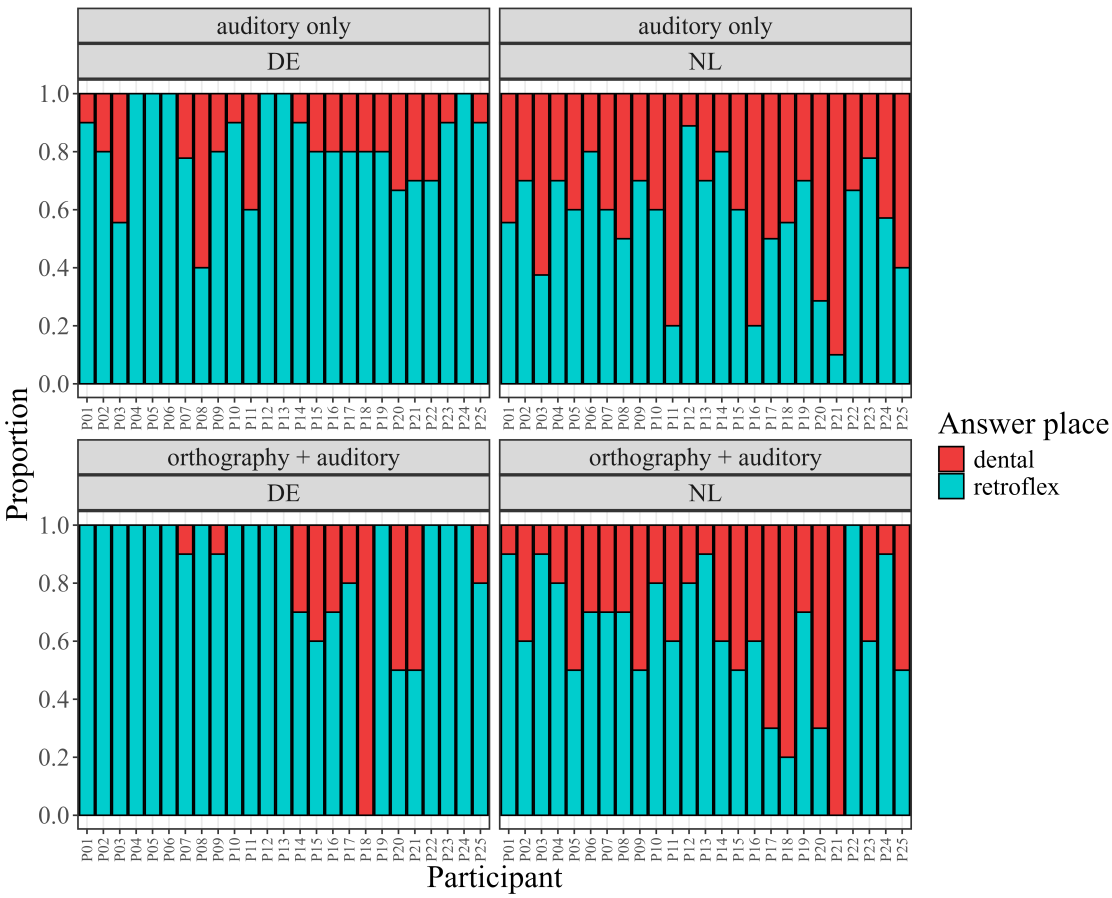

Figure 6: Proportion of retroflex and dental responses split by individual participants and language/condition.

4.2 Laryngeal contrast

Figure 7 illustrates overall results for the second research question regarding the four-way laryngeal contrast of Marathi. The graph shows that most Dutch stimuli were perceived as predicted, with most /t/ tokens being perceived as plain voiceless and most /d/ tokens as voiced. In both cases, some tokens were classified as the opposite category, which is not surprising considering that some /d/ tokens were not properly prevoiced by the two Dutch speakers and some /t/ tokens had a considerable aspiration time (see again Section 3). These confusions were reduced quite a bit when orthography was provided, as would be expected since the letter <d> suggests voicing and the letter <t> voicelessness.

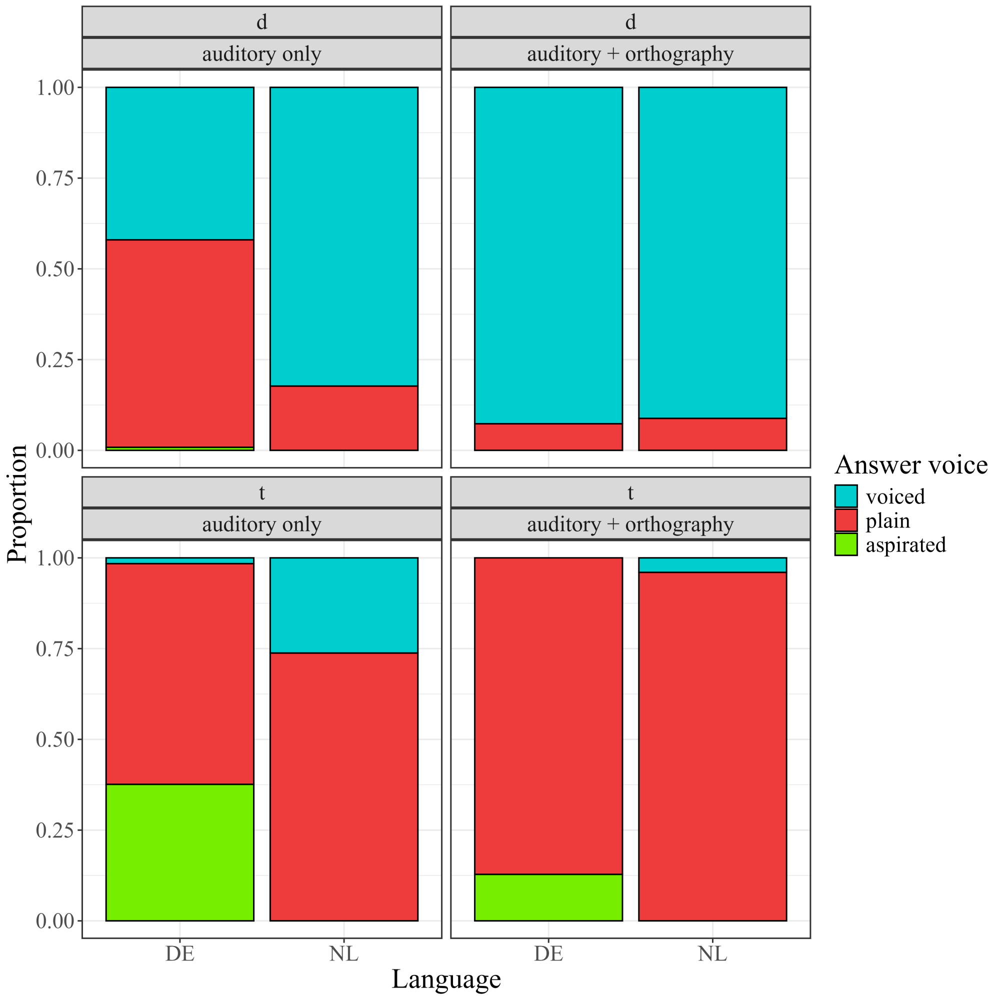

Figure 7: Proportion of voiced, plain and aspirated responses by language and whether orthography was present or not.

The five Dutch /d/ tokens without proper prevoicing were *duim* and *deel* spoken by NL1 and *diep*, *duim* and *doel* spoken by NL2. To check whether this influenced the results, Figure 8 plots voicing answers for each of the Dutch /d/ stimuli. For NL1, the only two stimuli that were ever perceived as not voiced were indeed those two without prevoicing. Similarly, the three tokens without prevoicing for NL2 were often perceived as plain, although there are a few other tokens that were perceived as plain as well. Thus, while absence or presence of prevoicing does not fully account for the variation within the answers for Dutch /d/, it seems to explain a large part of it.

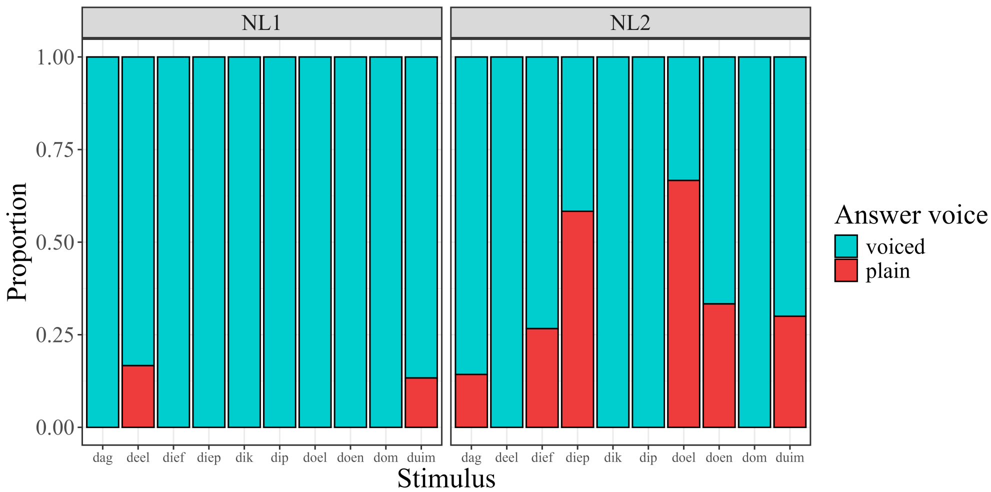

Figure 8: Voicing response for Dutch /d/ tokens by stimulus and speaker.

For German, the story is quite different. Many German /t/ tokens were perceived as aspirated, although the more common interpretation was plain. This makes sense, as German /t/ has a similar aspiration duration to Marathi aspirated plosives, although seemingly not all tokens are aspirated enough to be consistently perceived as such. Similarly, many /d/ tokens were classified as plain, most likely due to their VOT being similar to Marathi plain stops, although many were also classified as properly voiced. In the latter case, participants likely picked up on cues other than VOT, such as intensity or spectral characteristics. To summarize, participants as a group did not show a clear preference as to the voicing status of the German tokens. Again, this all changed when orthography was provided: Figure 7 clearly shows that most of this uncertainty went away, making the answers for German quite closely resemble those for Dutch, albeit with a few leftover aspirated responses. Those remaining aspirated answers are in fact quite interesting, as they suggest that while participants were clearly influenced by orthography, they did not blindly copy the voicing of the letters, in which case we could expect no remaining aspirated responses.

The output of the model for the laryngeal contrast is displayed in Table 7. In this output, the effect size for each predictor is the odds ratio of moving one category to the right on a spectrum from voiced over plain to aspirated. Thus, odds ratios above 1 indicate movement towards the aspirated end of the spectrum, while odds ratios of below 1 indicate movement towards the voiced end of the spectrum.

	Odds (ratio)	SE	z	p	Sig.
threshold voiced|plain	0.27	0.28	-4.68	< 0.001	***
threshold plain|aspirated	795.17	0.53	12.49	< 0.001	***
voice-d+t	499.30	0.50	12.49	< 0.001	***
lang-NL+DE	14.60	0.42	6.39	< 0.001	***
orthography-yes+no	2.65	0.21	4.64	< 0.001	***
lang-NL+DE:orthography-yes+no	19.89	0.43	7.03	< 0.001	***
voice-d+t:lang-NL+DE	23.84	0.83	3.83	< 0.001	***
voice-d+t:orthography-yes+no	0.10	0.42	-5.52	< 0.001	***
voice-d+t:lang-NL+DE:orthography-yes+no	3.12	0.84	1.36	0.17	

Table 7: Output of the maximally possible CLMM for the laryngeal contrast. Displayed are estimated odds (ratios), standard error, z and p-value for both decision thresholds and all fixed effects including interactions. The target variable was coded as 0 for voiced, 1 for plain and 2 for aspirated. Values with '-' are coded as -0.5 and values with '+' are codes as +0.5. 'Sig.' indicates degree of significance.

All predictors and two-way interactions were highly significant, with only the three-way interaction being non-significant. The individual significant effect for voicing of the stimulus, language and orthography ($p < 0.001$) is not very helpful regarding the research question, as we would naturally expect /t/ tokens across both languages to be categorized towards the aspirated end. The individual effect for orthography ($p < 0.001$) — with tokens without orthography being 2.65 times more likely to be categorized towards the aspirated end — is also not very interesting by itself, as it is averaged across both languages. Interestingly, though, there was also an individual effect of language ($p < 0.001$), indicating that German tokens on average were more likely to be perceived towards the aspirated end.

Furthermore, there was a significant interaction ($p < 0.001$) between language and orthography, indicating that the language difference was higher without orthography compared to with orthography. In other words, the presence of orthography caused the answers for both languages to converge, as was shown in Figure 7. There was also a significant interaction ($p < 0.001$) of voice and language. This is very logical, as it simply means that the tendency of /t/ to be perceived as more towards the aspirated end of the spectrum compared to /d/ was stronger for German tokens. Lastly, there was a significant interaction ($p < 0.001$) between voice and orthography. This means that the /d/-/t/ difference was larger when orthography was present. In other words, participants kept the two categories apart better with orthography, whereas they blurred together more without orthography.

It should again be noted that there was considerable variation between participants, as illustrated by Figure 9. First, inspection of Figure 9 (left) shows that even without orthography, there were very few participants who gave all three possible answers in a given condition. In fact, the only participants who did so were P15 and P20, who gave exactly one voiced answer for German /t/, which is certainly unexpected. Other than that, all participants fall somewhere on a spectrum between perceiving a German /t/ as aspirated or plain, with both extremes being attested. P01, for instance, interpreted all German /t/ tokens as aspirated, whereas P02 perceived all of them as plain. Similarly, all participants fall somewhere between strongly favoring plain and strongly favoring a voiced reponse for German /d/, Dutch /d/ and Dutch /t/. Looking at Figure 9 (right), essentially all participants exhibited some sort of orthography effect, with many of them always opting for the same answer in each panel on the right half. Still, some participants were less prone to this than others, such as P04, who still perceived more than half of German /d/ tokens as plain or P14 and P17, who perceived more than half of German /t/ tokens as aspirated. Again, it is striking that some participants merely reduced e.g. their number of aspirated responses for German /t/ when orthography was present rather than never giving an aspirated answer entirely.

4.3 Post-hoc

This section will outline the results of the post-hoc analysis which included analyzing the data collected in the two questionnaires before and after the experiment. Quantitative results regarding the former will be laid out in Section 4.3.1, while qualitative results will be summarized in Section 4.3.2.

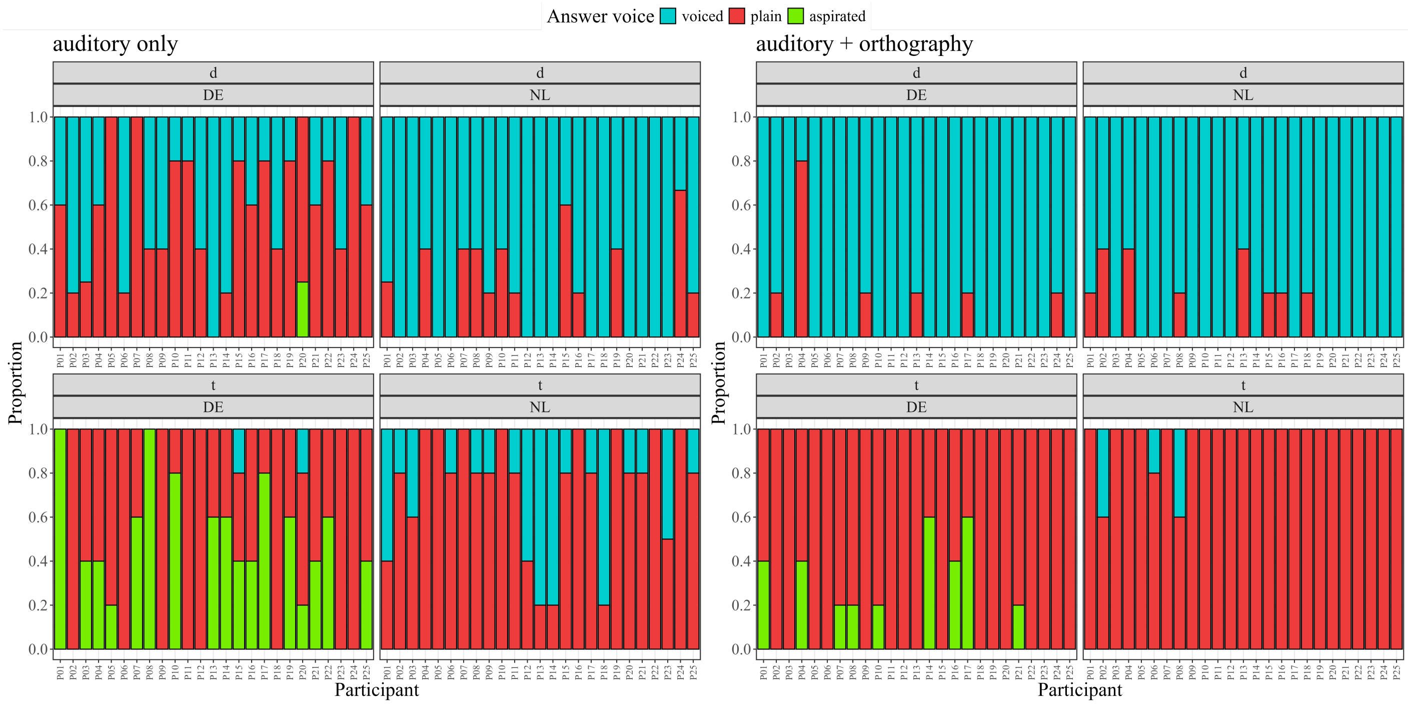

Figure 9: Proportion of voiced, plain and aspirated responses split by participants and language/voicing in the auditory only condition (left) and in the auditory + orthography condition (right).

4.3.1 Statistical analysis of questionnaire variables

The pre-experiment questionnaire collected data regarding participants age and school (Marathi, English or Semi-English medium, see Section 2.3). Age was treated as a continuous variable and school as 3-level factor, with the following two contrasts: The first contrast compared Marathi to English medium by coding them as -0.5 and 0.5, respectively, with Semi-English being coded as 0. Second, Semi-English was compared to the average of Marathi and English medium, by coding it as +2/3 and the other two as -1/3, respectively.

Additionally, manual inspection of the data suggested potential influence from surrounding vowels. Therefore, two additional variables were created: vowel frontness and vowel height. As the names suggest, these were two binary variables encoding whether a vowel is front/back and high/low. The low vowels, namely Dutch /aː/, Dutch /ɑ/ and German /ä, äː/ were coded as neither front nor back, as any front/back difference is much weaker at the bottom of the vowel triangle. For height, any high vowel (both tense and lax) was coded as high, while any other vowel height (including high-mid) was coded as non-high. Again, orthogonal contrast coding was applied. The four additional factors were included in both maximal models without interactions. Parameter selection was done using the ‘buildmer’ package as outlined in Section 2.4 (using the “nlminb” optimizer for the CLMM), resulting only in random intercepts for both participant and stimulus. Only significant effects will be reported in this section; the full output of the post-hoc models can be found in Appendix D.

For place of articulation, only the first school contrast was a significant post-hoc predictor ($p = 0.032$), with participants who went to a Marathi medium school being 2.40 times as likely to give a retroflex answer. Vowel frontness was barely not significant ($p = 0.064$), but the directionality of the directionality of the effect is quite interesting here. Namely, stimuli with a front vowel following the /t/ or /d/ were 1.68 times more likely to be interpreted as retroflex. This is of course entirely counter-intuitive, as we would expect coarticulatory effects to drag the consonant towards the front, making it more similar to a dental rather than a retroflex. For the laryngeal contrast, the only significant post-hoc predictor was age ($p = 0.015$). Specifically, adding one year of age increased the likelihood to step from voiced to plain or from plain to aspirated by 3%. The post-hoc findings will be discussed in Section 5.3.

4.3.2 Qualitative results of post-hoc questionnaire

Most participants gave at least minimal responses to the post-hoc questionnaire. 18 participants (72%) reported to have had some difficulty transcribing the German and Dutch stimuli, while only 4 (16%) reported to have had none and 3 (12%) gave no answer. This could indicate that participants were aware that the input does not properly match their own phonological categories; however, very few gave concrete examples of difficult sounds, none of which included the target sounds.

18 participants (71%) reported that orthography changed or influenced their answers to varying degrees. Their answers ranged from “absolutely” or “yes definitely. I caught some subtle pronunciations when writing was present” to “little bit, not much” or “to a small extent”. 5 participants (20%) reported to not have been influenced by orthography, stating things such as “no, because I went by the pronunciation sounds in both parts”. Despite this, 3 out of these 5 participants (P03, P09 and P15) arguably exhibited some orthographic influence, judging by Figure 9. P15, for instance, used all three possible voicing answers for German /t/, but only ever used plain answers

once orthography was present. Furthermore, some participants reported something along the lines of the spelling not matching the pronunciation of the stimuli. This indicates a lack of familiarity of German and Dutch spelling rules, which is of course expected for naïve listeners. This lack of familiarity, however, did not seem to prevent participants from being affected by the orthography.

The answers regarding the third question, which specifically asked participants to explain the reasoning for their transcriptions of the target sounds, were quite uninteresting. Many participants indicated that they simply went by which Marathi sound they perceived to be the most similar to the input. P01 wrote that they "could not make out the voicing or devoicing, and the aspirations", indicating difficulty classifying the voicing of the stimuli despite some apparent familiarity with linguistic concepts. P21 reported that they tried imitating each sound and chose their transcription based on what they thought they were saying. In a way, this means that this participant's data is production data; nevertheless, the input was first filtered through their perception and they were still forced to categorize their answer as they had to give an orthographic response.

None of the participants reported anything related to the target stimuli in the last question, where they could share any remaining thoughts. They did report struggling with transcribing other sounds. Thus, we have no indication that any participant caught on to the purpose of the experiment. All in all, the post-hoc questionnaire provided little extra insight other than aiding with participant exclusion and providing certain sanity checks.

5 Discussion

Similar to Section 4, this section will first discuss the results regarding PoA (Section 5.1), followed by discussing the results for the laryngeal contrast (Section 5.2) and the post-hoc analysis (Section 5.3).

5.1 Place of articulation

The main hypothesis of the first research question regarding place of articulation was strongly supported by the results. Due to the highly significant group effect, we have strong evidence to conclude that Marathi listeners perceive the German stimuli as more retroflex than the Dutch stimuli. However, it was found that more than half of Dutch tokens were still perceived as retroflex rather than dental. While this does not fully correspond to the hypothesized dental interpretation of the Dutch stimuli, this is still not very surprising. As mentioned in Section 3, the Dutch stimuli did not exhibit a negative skewness or kurtosis, albeit a less positive one than the German stimuli. This suggests that the two Dutch speakers used a more dental articulation than the two German speakers, but not as dental as we might expect a Marathi speaker to use. This is again logical, as Marathi speakers are subject to the pressure of keeping the retroflex and dental categories as distinct as possible.

It should also be noted that no conclusions can be made about the general population of German/Dutch speakers, as no random effects were included for the speakers due to their small number (n = 4). Therefore, all results can only be generalized to the average of the two speakers per language. However, the point of this paper is not really to show how Marathi speakers interpret German and Dutch sounds *in general*, but rather to relate differences in the acoustic input to differences in perception. Regarding this, Section 3 clearly showed not only that there were clear acoustic differences between the Dutch and the German speakers, but also that these differences favor certain interpretations over others to Marathi ears. While we cannot say which cues were actually picked up on by the participants and which ones were not, the differences between cues as a whole seemed to have been crucial for the listeners' categorizations. Specifically, the German stimuli contained cues that favored a retroflex interpretation by Marathi listeners while the Dutch stimuli favored a more dental interpretation (at least compared to the German stimuli). The present experiment thus provides evidence for perception-based accounts of loanword adaptation. As the main methodological innovation of this experiment was the elicitation of an orthographic response compared to a spoken one, the present results reflect Marathi listeners' perception *directly*.

Dauenhauer (2025) concluded with the prediction that formant transitions must be less important in the Marathi production and perception grammar than burst-related cues when it comes to making a dental vs. retroflex distinction. This is corroborated by the results of Sections 3 and 4: The German stimuli corresponded well to Indo-Aryan retroflexes for the burst-related cues (in this case intensity and the four spectral measures), but less so when it comes to formant transitions. While the median F3 of the German stimuli was considerably closer to Marathi retroflexes than the F3 of the Dutch stimuli, an F3 of around 2700 Hz is still far higher than the F3 of proper retroflexes, which can be as low as 1800 Hz (Hamann 2003: 61). This particular circumstance thus presents an issue for those accounts that still employ the notion of acoustic similarity (on top of their theoretical issues such as being cognitively unrealistic, see again Section 1.4). An account such as BiPhon Optimality Theory (OT), however, is easily compatible with these results, as it assumes that the same cue constraints that Marathi speakers use in order to produce their 2-way coronal contrast are also used in reverse order when perceiving and attempting to categorize foreign input. In fact, these results are even predicted by such an account, if we assume the aforementioned higher ranking of burst-related constraints compared to formant transitions. The Dutch stimuli, on the other hand, favored a dental interpretation more so than the German stimuli, although they were still ambiguous at least in the burst. This last fact is again crucial; if Marathi listeners relied exclusively on formant transitions — with F3 being commonly cited as the most distinctive cue of retroflex consonants (Hamann 2003) — then we would expect the Dutch stimuli to have been interpreted as fully dental due to their very high F3. This clearly did not happen and thus also provides evidence for the ranking proposed in Dauenhauer (2025). While this ranking was established in part considering Indo-Aryan adaptations of English coronals, it was modeled to be the *L1* perception grammar of Marathi listeners, thereby necessarily being able to account for input from other languages.

It should furthermore be noted that even individual speakers did not fully stick to one adaptation for each of the German and Dutch phonemes. This makes sense in a naïve context, where the listeners have had no exposure to the L2. In an actual case of language contact/acquisition, we would expect unanimous adaptations for each phoneme, as we do indeed observe when English /t/ and /d/ are unanimously adopted as retroflex by Indo-Aryan speakers. With more exposure, learners would start associating the same category in the L2 with the same sound and subsequently adjust their constraint ranking via some sort of learning algorithm (see, e.g., the ‘Gradual Learning Algorithm’ by Boersma et al. 1997)

The orthography-related part of the first research question was not supported, on the other hand. Even though the predicted effect seemed to be applicable for a few individual participants, we have no evidence for such an effect on a group-level. One potential factor that might have hindered Marathi speakers from associating <t> and <d> with a retroflex PoA is the fact that in commonly used romanizations of Marathi and Hindi (e.g., for texting), these letters can stand for both dental and retroflex sounds.

5.2 Laryngeal contrast

The second hypothesis regarding the mapping of Dutch and German stops onto the Marathi four-way laryngeal contrast was strongly supported. Dutch stops seem to be quite easy for Marathi listeners to map onto their laryngeal categories, as they are confronted with largely unambiguous cues. Dutch /d/ was relatively unanimously interpreted as voiced, likely due to the clear prevoicing present in most of the stimuli (with the caveat of some stimuli lacking prevoicing as discussed in Section 4.2). Similarly, Dutch /t/ was quite unanimously interpreted as plain, confirming the first hypothesis at least for the Dutch stimuli.

The story is quite different for the German stimuli, which Marathi listeners seem to have some difficulty categorizing. Section 1.2 mentioned measurements that indicated a similar aspiration time for German /t/ and Marathi aspirated stops, which would go against the principle of auditory dispersion. The present results confirm that on a group level, Marathi listeners are able to pick up on the aspiration sometimes, whereas the German aspiration duration is often not enough for them to categorize a given sound as phonologically aspirated. On an individual level, however, it became apparent that there exists some sort of spectrum between people who consistently interpret German /t/ as aspirated and those who consistently interpret it as plain, with many people falling somewhere in the middle of this spectrum. The same is true for German /d/, which was interpreted anywhere from consistently voiced to consistently plain. These results suggest individual differences in cue-ranking between participants, resulting in different classification thresholds.

A prediction that results from these findings would be that those people who are more ‘liberal’ with their classification of aspiration also have a smaller aspiration duration in their own production. This follows from the assumption that the same cue-constraint ranking is active in both the perception and production direction (= bidirectionality).

Unlike for place of articulation, the orthography-related hypothesis for the laryngeal contrast was strongly supported. It seems that the considerable variation caused by the ambiguity of the German cues vanishes once orthography is provided. This makes sense, as the letters <d> and <t> clearly indicate voicing and plain voicelessness, respectively. All Marathi speakers with any English proficiency would be aware of this, as this is the associated pronunciation of these letters in Indian English. Aspiration is only ever present in Indian English when <h> is present in writing, namely for the dental fricative /θ/ which is commonly adapted as /t̪ʰ/ into Indian English. The individual differences that were observed for this orthography effect thus corroborate the results of Zhou and Hamann (2020), although in the present case we can in fact speak of orthographic perception rather than production.

These results have implications for the formalization presented in Dauenhauer (2025). First, it is very important to note that these implications come with a big disclaimer, namely that the results from this paper do not directly tell us anything about historical adaptations of English, as no historical data was used and the stimuli were not in English. Nevertheless, we have good reason to believe that adaptations for German and English would behave very similarly due their strong phonetic resemblance. That being said, the formalization presented in Dauenhauer (2025) concluded that — in word-initial position — auditory cues were enough to result in a voiced adaptation of English /d/ and a plain adaptation of English /t/. The results of the present experiment cast doubt on this analysis. One alternative explanation which the author discussed was influence from the same phoneme in other positions. For instance, English /d/ tends to become properly voiced between vowels and derivational connections such as between *do* and *redo* would make it apparent that these acoustically different sounds are indeed the same phonological category. Another possible explanation which was discussed, however, was orthography. The present experiment does provide evidence for this idea of multimodal cue integration. In optimality-theoretic terms, this would mean assuming very high-ranked constraints such as *<d>/-voice/, *<d>/+asp/, *<t>/+voice/ and *<t>/+asp/, using the formalism provided in Hamann and Colombo (2017). These constraints would militate against perceiving and producing the letter <d> as plain or aspirated and the letter <t> as voiced or aspirated (with similar constraints ruling out any breathy adaptations). For certain listeners, these constraints are so high ranking that they are never violated. For others, they are ranked much lower, allowing for the occasional interpretation which violates these constraints in favor of auditory cue constraints (assuming an account which allows for evaluation noise). For yet other listeners, these constraints might be so low-ranked that they rarely or never end up being crucial, resulting in (almost) purely auditory-based adaptations (see Zhou and Hamann 2024 for modeling individual differences combined with orthographic effects).

5.3 Post-hoc

Section 4.3 showed that some post-hoc variables turned out to be significant, which will be discussed here. First, school (i.e., language of instruction) ended up being a significant predictor for place of articulation, indicating that Marathi listeners who went to a Marathi medium school perceive the stimuli overall as more retroflex than those who went to an English medium school. The most obvious explanation for this effect is probably that the degree of exposure to Western English might differ between these two groups. This could be the case because the latter group's higher English proficiency might increase their consumption of Western media. Plenty of research has shown that exposure to foreign accents can lead to a relatively quick adaptation to that accent (e.g., Bradlow and Bent 2008; Clarke and Garrett 2004). It is thus entirely feasible that increased exposure to Western English might have affected Marathi listeners' perception grammar, subsequently reducing their retroflex perceptions. One could also imagine that people who went to an English medium school might be more aware of stereotypes surrounding their accents, causing them to adapt their production and therefore also perception grammar for English, which they then applied to the novel stimuli. However, this explanation is purely speculative.

Furthermore, there was a barely not significant effect for place of articulation, namely vowel frontness. This was strange because front vowels made a retroflex response more likely. There exists no obvious explanation for the directionality of this counter-intuitive effect, but it is important to note that the stimuli were largely balanced in this regard. Therefore, even if there had been a significant frontness effect, it would have impacted both languages and voicing conditions somewhat equally.

Lastly, age was the only significant post-hoc predictor in the model for the laryngeal contrast, with younger speakers being slightly more likely to give a response towards the aspirated end of the 'answer spectrum'. However, the effect size was incredibly small, making it quite negligible. If we attempt to interpret this effect nonetheless, this could again be a tiny effect of exposure to Western English. Younger speakers would likely have such a higher exposure, making them used to hearing aspirated /t/s and plain /d/s, which in turn might have slightly affected their perception grammar.

5.4 Limitations

Despite the interesting findings of the present study, they naturally come with some limitations. First, it was already mentioned in Section 5.1 that the present results cannot be generalized to the entire population of Dutch and German speakers, which does not really pose an issue as this was never the point of this experiment. However, it is still important to keep in mind that with other speakers these results might have looked differently, as there can always be unforeseen speaker-specific differences. This is mitigated by the fact that the acoustic analysis of the stimuli in Section 3 showed that the recordings corresponded largely to the expectations and thus fulfilled their role. Nevertheless, Figure 8 in Section 4.2 showed that a prevoicing difference between speakers NL1 and NL2 directly impacted the results, albeit in a non-crucial manner.

Next, there will always remain a possibility for undetected lexical effects having influenced the results. An effort was made to exclude any participants that exhibited confirmed or strongly suspected lexical influence as outlined in Section 2.1, but it is entirely possible that some non-excluded participants were influenced by the lexical status of their answers without reporting so and without detection by the author.

For all (sub-) research questions, substantial individual variation was observed. However, this was only done via visual inspection of figures, rather than a sophisticated statistical approach. Future research should therefore attempt to quantify these individual differences, for example by using cluster analysis (Voeten 2021).

Lastly, there are of course always other unaccounted variables which might have influenced the results in unforeseen and undetected ways. For instance, the experiment contained no linguistic background questionnaire, meaning that participants' English proficiency and dominance might have varied considerably, just like they might have had experience with learning foreign languages other than German or Dutch. Still, the important factors were controlled for, namely that they all shared the same L1 (Marathi), were proficient in English and had no exposure to or experience with German and Dutch.

6 Conclusion

The present experiment provided the first methodologically and statistically driven perception experiment investigating the perception of (denti-)alveolar stops into an Indo-Aryan language — Marathi. One key methodological innovation was the elicitation of an orthographic answer, which forced participants to give categorical (= phonological) answers rather than a continuous phonetic one, which would be the case if production had been elicited.

It was found that Marathi listeners largely categorize German alveolar stops as retroflex, while they did not perceive Dutch dental stops as clearly dental or retroflex. Furthermore, this experiment showed that Dutch /d/ and /t/ are perceived as voiced and plain voiceless, respectively, by Marathi listeners. On the other hand, German /t/ was perceived as either aspirated or plain and German /d/ as plain or voiced. This uncertainty regarding German — which stems from the ambiguity of German stops regarding the Marathi four-way laryngeal contrast — was significantly mitigated by providing orthographic input. This happened due to the association of the letters <d> and <t> with voicing and plain voicelessness, respectively, in Indian English or Marathi romanization.

This study provided evidence that phonological adaptations in second language acquisition happen in perception and that orthography plays a major role in this process. More specifically, cues that are relevant for L1-perception and production are used in the same manner to categorize novel, foreign input. Orthographic cues then come to interact with auditory ones, leading to multimodal cue integration. This effect highly varies between individuals, ranging from very strong to not present at all.

Acknowledgements

I would like to thank Silke Hamann for her supervision and feedback throughout the entire process of this thesis, Dirk Vet for implementing the experiment, Paul Boersma for his input on the acoustic analysis, Cesko Voeten for his help with the statistical analysis and Manjari Ohala for providing a scan of her 1978 paper. Lastly, I would like to thank Sayuri Deokar for proofreading, participant recruitment and native speaker input, without which the analysis would not have been possible.

References

Arsenault, P. (2006). The adaptation of English alveolar stops in Telugu and Hindi. *Handout from the Montreal-Ottawa-Toronto (MOT) Phonology Workshop*.

Bassetti, B., Escudero, P., & Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics*, *36*(1), 1–6. https://doi.org/10.1017/S0142716414000393

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benguerel, A.-P., & Bhatia, T. K. (1980). Hindi stop consonants: An acoustic and fiberscopic study. *Phonetica*, *37*(3), 134–148. https://doi.org/10.1159/000259987

Best, C. T., & Tyler, M. D. (2008). Nonnative and second-language speech perception: Commonalities and complementarities. In *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.07bes

Boersma, P., et al. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, *21*, 43–58.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. In *Bidirectional Optimality Theory* (pp. 33–72). John Benjamins.

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, *25*(2), 217–270. https://doi.org/10.1017/S0952675708001474

Boersma, P., & Hamann, S. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & L. W. Wetzels (Eds.), *Loanword Phonology* (pp. 11–58). John Benjamins. https://doi.org/10.1075/cilt.307.02boe

Boersma, P., & Weenink, D. (2025). Praat: doing phonetics by computer [Computer program]. Version 6.4.31, retrieved 3 May 2025 from http://www.praat.org.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Cardona, G., & Luraghi, S. (2018). Indo-Aryan languages. In B. Comrie (Ed.), *The world's major languages* (pp. 383–389). Routledge.

Christensen, R. H. B. (2018). Cumulative link models for ordinal regression with the R package ordinal. *Submitted in J. Stat. Software*, *35*, 1–46.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Dart, S. N. (1991). Articulatory and acoustic properties of apical and laminal articulations. *UCLA Working Papers in Phonetics*, *79*, 1–155. https://escholarship.org/uc/item/52f5v2x2

Dart, S. N. (1998). Comparing French and English coronal consonant articulation. *Journal of phonetics*, *26*(1), 71–94. https://doi.org/10.1006/jpho.1997.0060

Dauenhauer, M. (2025). The adaptation of English coronals into Indian languages – a perception based OT-account. *Linguistics in Amsterdam*, *16*(1), 1–29. https://dspace.uba.uva.nl/handle/record/55899

Dhongde, R. V., & Wali, K. (2009). *Marathi*. John Benjamins Publishing.

Dmitrieva, O., & Dutta, I. (2019). Acoustic correlates of the four-way laryngeal contrast in Marathi. *Phonetica*, *77*(3), 209–237. https://doi.org/https://doi.org/10.1159/000501673

Flege, J. E., Aoyama, K., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r) applied. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 84–118). Cambridge University Press Cambridge. https://doi.org/10.1017/9781108886901.002

Giannakopoulou, A., Uther, M., & Ylinen, S. (2013). Phonetic and orthographic cues are weighted in speech sound perception by second language speakers: Evidence from Greek speakers of English. *Proceedings of Meetings on Acoustics*, *20*(1). https://doi.org/10.1121/2.0000206

Hamann, S. (2003). *The phonetics and phonology of retroflexes [Doctoral dissertation]*. LOT Press.

Hamann, S., & Colombo, I. E. (2017). A formal account of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. *Natural Language & Linguistic Theory*, *35*, 683–714. https://doi.org/10.1007/s11049-017-9362-3

Hamann, S., & Fuchs, S. (2010). Retroflexion of voiced stops: data from Dhao, Thulung, Afar and German. *Language and Speech*, *53*(2), 181–216. https://doi.org/10.1177/0023830909357159

Hussain, Q. (2014). Adaptation of English alveolars as retroflexes in Indo-Aryan languages. *Perception-Production Studies and Corpus-based Approaches in Second Language Phonetics and Phonology Workshop*.

Hussain, Q., Proctor, M., Harvey, M., & Demuth, K. (2017). Acoustic characteristics of Punjabi retroflex and dental stops. *The Journal of the Acoustical Society of America*, *141*(6), 4522–4542. https://doi.org/10.1121/1.4984595

Jessen, M. (1999). *Phonetics and phonology of tense and lax obstruents in German*. John Benjamins Publishing Company. https://doi.org/10.1075/sfsl.44

Jongman, A., Blumstein, S. E., & Lahiri, A. (1985). Acoustic properties for dental and alveolar stop consonants: A cross-language study. *Journal of Phonetics*, *13*(2), 235–251. https://doi.org/10.1016/s0095-4470(19)30738-7

Kang, H. (1996). English loanwords in Korean. *Studies in Phonetics, Phonology and Morphology*, *2*, 21–47. https://doi.org/10.1017/s0952675710000114

Kogan, V. V., & Mora, J. C. (2022). The effects of individual differences in native perception on discrimination of a novel non-native contrast. *Laboratory Phonology*, *24*(1). https://doi.org/10.16995/labphon.6431

Kohler, K. (1976). Die Instabilität wortfinaler Alveolarpiosive im Deutschen: eine elektropalatographische Untersuchung. *Phonetica*, *33*(1), 1–30.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

LaCharité, D., & Paradis, C. (2005). Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic inquiry*, *36*(2), 223–258. https://doi.org/10.1162/0024389053710666

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

McGuire, G. (2014). Orthographic effects on phonetic cue weighting. *Proceedings of the 15th Australasian International Conference on Speech Science and Technology*, 201–204.

Ohala, M. (1978). Conflicting expectations for the direction of sound change. *Indian Linguistics*, *39*(1-4), 25–28.

Ohala, M. (1983). *Aspects of Hindi phonology*. Motilal Banarsidass Publishers.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Samudravijaya, K. (2003). Durational characteristics of Hindi stop consonants. *Proc. 8th European Conference on Speech Communication and Technology (Eurospeech 2003)*, 81–84. https://doi.org/10.21437/Eurospeech.2003-56

Silverman, D. (1992). Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology*, *9*(2), 289–328. https://doi.org/10.1017/s0952675700001627

Sundara, M. (2005). Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *The Journal of the Acoustical Society of America*, *118*(2), 1026–1037. https://doi.org/10.1121/1.1953270

Transliteration Committee. (1895). Tenth International Congress of Orientalists, Held at Geneva. *The Journal of the Royal Asiatic Society of Great Britain and Ireland*, 879–892. http://www.jstor.org/stable/25207765

Verma, R., & Chawla, P. (2003). Comparative analysis of Hindi retroflex and dental CV syllables and their synthesis. *Proceedings of the Workshop on Spoken Language Processing*, 163–167.

Vet, D. J. (2025). Experiment Designer [Computer program]. Version 2025.03, retrieved 20 April 2025 from https://www.fon.hum.uva.nl/dirk/ed.php.

Voeten, C. C. (2021). Individual Differences in the Adoption of Sound Change. *Language and Speech*, *64*(3), 705–741. https://doi.org/10.1177/0023830920959753

Voeten, C. C. (2025). Package buildmer. https://doi.org/10.32614/CRAN.package.buildmer

Yip, M. (1993). Cantonese loanword phonology and optimality theory. *Journal of East Asian Linguistics*, *2*(3), 261–291. https://doi.org/10.1007/bf01739135

Zhou, C., & Hamann, S. (2020). Cross-Linguistic Interaction Between Phonological Categorization and Orthography Predicts Prosodic Effects in the Acquisition of Portuguese Liquids by L1-Mandarin Learners. *INTERSPEECH 2020*, 4486–4490. https://doi.org/10.21437/Interspeech.2020-2689

Zhou, C., & Hamann, S. (2024). Modelling the acquisition of the Portuguese tap by L1-Mandarin learners: A BiPhon-HG account for individual differences, syllable-position effects and orthographic influences in L2 speech. *Glossa: a journal of general linguistics*, *9*(1). https://doi.org/10.16995/glossa.9692

Appendix A — Stimuli

DE target	DE filler	Nl target	NL filler
Ton	Baum	taak	lip
Tal	Bock	toon	kool
Tip	Maus	tas	pil
Tag	Wal	til	hal
taub	Schaf	toen	lam
Teig	Pin	taal	vol
tief	Sieb	tof	sap
Tee	Sieg	tien	kip
Tisch	Gas	top	pan
tun	Laub	tuin	was
doof	See	dief	koe
Ding	Lob	diep	soep
das	viel	dom	bak
den	Saal	doen	vaak
die	Moos	doel	bal
dumm	Haus	dik	veel
des	Lamm	deel	zoon
Dieb	Ball	dip	naam
Dock	Kinn	dag	mok
doll	Fass	duim	zeep

Table 8: Full list of stimuli by language and condition.

Appendix B — Acoustic measurements

	mean	sd	median	min	max
VOT	0.020	0.0080	0.018	0.011	0.042
intensity	0.13	0.073	0.12	0.054	0.36
CoG	2424.47	626.40	2482.36	1306.04	3670.73
SD	1843.32	358.52	1774.64	1296.11	2592.39
skewness	1.90	0.74	1.84	0.48	3.25
kurtosis	9.72	5.98	9.64	0.35	20.66
F2	1757.25	434.04	1657.07	1226.25	2543.98
F3	2706.80	214.10	2730.26	2427.34	3239.88

Table 9: Acoustic measurements for German /d/.

	mean	sd	median	min	max
VOT	-0.090	0.078	-0.087	-0.20	0.019
intensity	0.10	0.10	0.053	0.0010	0.29
CoG	2642.57	1048.54	2397.65	1183.31	5332.05
SD	2612.30	592.16	2636.10	1482.12	3811.17
skewness	1.57	0.74	1.52	-0.084	3.40
kurtosis	3.81	3.97	2.44	-0.89	15.15

Table 10: Acoustic measurements for Dutch /d/.

	mean	sd	median	min	max
VOT	0.10	0.029	0.089	0.045	0.17
intensity	0.13	0.084	0.11	0.020	0.29
CoG	2505.50	577.49	2300.45	1703.59	3800.81
SD	2112.41	561.53	2088.49	1268.54	3287.90
skewness	1.93	0.67	2.03	1.03	3.02
kurtosis	8.21	6.70	6.00	0.72	24.86

Table 11: Acoustic measurements for German /t/.

	mean	sd	median	min	max
VOT	0.021	0.0070	0.020	0.0090	0.035
intensity	0.059	0.042	0.046	0.0080	0.18
CoG	3012.91	752.23	2769.42	2013.43	5006.63
SD	2708.71	637.76	2563.21	1820.95	3876.52
skewness	1.20	0.70	1.27	-0.41	2.15
kurtosis	2.23	2.17	2.21	-1.26	6.53
F2	1688.05	271.44	1671.37	1283.75	2326.07
F3	2957.62	198.63	2988.76	2503.41	3223.99

Table 12: Acoustic measurements for Dutch /t/.

Appendix C — Output of post-hoc models

	Odds (ratio)	SE	z	p	Sig.
intercept	5.17	0.54	3.05	0.002	**
lang-NL+DE	5.48	0.27	6.19	< 0.001	***
school-EN+MA	2.40	0.41	2.14	0.032	*
school-E/M+SE	3.12	0.72	1.59	0.11	
vowel_frontness-front+back	0.60	0.28	-1.85	0.064	.
voice-d+t	1.63	0.30	1.64	0.10	.
orthography-yes+no	0.66	0.30	-1.38	0.17	
age	0.99	0.018	-0.38	0.71	
vowel_height-hi+lo	1.01	0.27	0.044	0.96	
lang-NL+DE:voice-d+t	1.87	0.52	1.21	0.23	
voice-NL+DE:orthography-yes+no	1.34	0.36	0.82	0.41	
lang-NL+DE:orthography-yes+no	0.82	0.36	-0.54	0.59	
lang-NL+DE:voice-d+t:orthography-yes+no	0.50	0.72	-0.97	0.33	

Table 13: Output of the posthoc Glmer for place of articulation. Formula: answerPlace $\sim$ language $*$ voicing $*$ orthography $+$ age $+$ school $+$ vowel frontness $+$ vowel height $+$ (1 $||$ participant) $+$ (1 $||$ stimulus). See Section 4.3.1 for more details.

	Odds (ratio)	SE	z	p	Sig.
voiced|plain	0.12	0.48	-4.40	< 0.001	***
plain|aspirated	364.55	0.64	9.26	< 0.001	***
voice-d+t	515.65	0.49	12.72	< 0.001	***
lang-NL+DE	15.58	0.41	6.63	< 0.001	***
orthography-yes+no	2.61	0.21	4.57	< 0.001	***
age	0.97	0.014	-2.44	0.015	*
school-EN+MA	0.68	0.32	-1.22	0.22	
school-E/M+SE	0.39	0.55	-1.71	0.088	.
vowel_frontness-front+back	1.53	0.32	1.30	0.19	
vowel_height-hi+lo	0.70	0.30	-1.17	0.24	
lang-NL+DE:orthography-yes+no	20.25	0.42	7.09	< 0.001	***
voice-d+t:lang-NL+DE	25.03	0.81	3.97	< 0.001	***
voice-d+t:orthography-yes+no	0.10	0.42	-5.48	< 0.001	***
voice-d+t:lang-NL+DE:orthography-yes+no	3.20	0.83	1.39	0.16	

Table 14: Output of the posthoc CLMM for the laryngeal contrast. Formula: answerVoice $\sim$ language $*$ voicing $*$ orthography $+$ age $+$ school $+$ vowel frontness $+$ vowel height $+$ (1 $||$ participant) $+$ (1 $||$ stimulus). See Section 4.3.2 for more details.

Appendix D — Praat script

```
path$ = "Stimuli recordings/"

form Set measurement threshold
    positive: "threshold", "0.005"
endform


#get file list
files = Create Strings as file list: "list", path$ + "*.TextGrid"
numberOfTotalFiles = Get number of strings

#set no of files to analyze
numberOfFiles = numberOfTotalFiles

#create results table
results = Create Table with column names: "results", 0, {"speaker", "lang",
... "gender", "word", "VOT", "CoG", "SD","skewness", "kurtosis", "intensity",
... "F2", "F3"}


for fileNumber to numberOfFiles
    selectObject: files
    textgridFile$ = Get string: fileNumber
    textgrid = Read from file: path$ + textgridFile$

    #number of words in word tier
    numberOfWords = Get number of intervals: 2

    soundFile$ = textgridFile$ - ".TextGrid" + ".wav"
    sound = Read from file: path$ + soundFile$

    #scale sound
    Scale peak: 0.99

    speakerName$ = textgridFile$ - ".TextGrid"

    #get gender of participant
    if speakerName$ = "NL1"
        gender$ = "Male"
        lang$ = "NL"
    elsif speakerName$ = "DE1"
        gender$ = "Male"
        lang$ = "DE"
    elsif speakerName$ = "NL2"
        gender$ = "Female"
        lang$ = "NL"
    else
        gender$ = "Female"
        lang$ = "DE"
```

```
endif

#set formant ceiling
if gender$ = "Male"
    formantCeiling = 5000
elsif gender$ = "Female"
    formantCeiling = 5500
else
    exitScript("Gender not specified.")
endif

#create formant object
selectObject: sound
formantObject = To Formant (burg): 0, 5, formantCeiling, 0.025, 50

for word to numberOfWords
    selectObject: textgrid
    word$ = Get label of interval: 2, word

    if word$ <> ""
        startWord = Get start time of interval: 2, word
        endWord = Get end time of interval: 2, word

        firstPoint = Get nearest index from time: 3, startWord
        lastPoint = Get nearest index from time: 3, endWord

        for point from firstPoint to lastPoint
            point$ = Get label of point: 3, point

            #abbreviations: sv = start vowel, sb = start burst, ev = end vowel
            ... , spv = start prevoicing
            spv_point = 0
            if point$ = "sv"
                sv_point = point
            elsif point$ = "sb"
                sb_point = point
            elsif point$ = "ev"
                ev_point = point
            elsif point$ = "spv"
                spv_point = point
            else
                exitScript("Wrong point label.")
            endif

        endfor

        @measure: speakerName$, lang$, gender$, word$, sv_point, sb_point,
        ... ev_point, spv_point

    endif

endfor

removeObject: sound, textgrid
```

```
    removeObject: formantObject

endfor

removeObject: files

#saving
selectObject: results
Save as comma-separated file: "stimuli_results.Table"

procedure measure: .speakerName$, .lang$, .gender$, .word$, .sv_point, .sb_point,
... .ev_point, .spv_point

    #get timepoints
    .sv_time = Get time of point: 3, .sv_point
    .sb_time = Get time of point: 3, .sb_point
    .ev_time = Get time of point: 3, .ev_point

    #VOT
    if .spv_point
        .spv_time = Get time of point: 3, .spv_point
        .vot = .spv_time - .sb_time
    else
        .vot = .sv_time - .sb_time
    endif


    #Spectral moments
    selectObject: sound
    .burst = Extract part: .sb_time, .sb_time + threshold, "rectangular", 1.0, "no"
    .burst_spectrum = To Spectrum: "no"
    .cog = Get centre of gravity: 2.0
    .sd = Get standard deviation: 2.0
    .skewness = Get skewness: 2.0
    .kurtosis = Get kurtosis: 2.0
    removeObject: .burst, .burst_spectrum

    #intensity
    selectObject: sound
    .power_vowel = Get power: .sv_time, .ev_time
    .power_burst = Get power: .sb_time, .sb_time + threshold
    .intensity = .power_burst / .power_vowel

    #F2 & F3
    selectObject: formantObject
    .f2 = Get quantile: 2, .sv_time, .sv_time + threshold, "hertz", 0.5
    .f3 = Get quantile: 3, .sv_time, .sv_time + threshold, "hertz", 0.5

    #storing
    selectObject: results
    Append row
    .row_no = Get number of rows
    Set string value: .row_no, "speaker", .speakerName$
    Set string value: .row_no, "gender", .gender$
```

```
    Set string value: .row_no, "lang", .lang$
    Set string value: .row_no, "word", .word$
    Set numeric value: .row_no, "VOT", .vot
    Set numeric value: .row_no, "CoG", .cog
    Set numeric value: .row_no, "SD", .sd
    Set numeric value: .row_no, "skewness", .skewness
    Set numeric value: .row_no, "kurtosis", .kurtosis
    Set numeric value: .row_no, "intensity", .intensity
    Set numeric value: .row_no, "F3", .f3
    Set numeric value: .row_no, "F2", .f2
endproc
```