The adaptation of English coronals into Indian languages: A perception-based OT account*

Maxim Dauenhauer
ACLC, University of Amsterdam

This paper formalizes the different adaptations of the English coronal segments /t, d, n, l, θ, ð/ into Indo-Aryan languages. This is done by assuming that the same cue constraints that are used in native perception also apply in loanword adaptation/second-language acquisition. First, the adaptation of English dental stops as retroflex are explained with a higher-ranking of burst-related cues compared to formant transitions. Second, in order to account for the mapping of the English two-way laryngeal contrast onto the typical Indo-Aryan four-way laryngeal contrast, cue constraints regarding prevoicing, aspiration, pitch and burst amplitude are all assumed to be unranked. The formalization provided in this paper has important implications for future perception experiments, mainly a greater importance of burst-related cues than formant transitions in the perception of the native dental-retroflex contrast.

1 Introduction

Retroflexion is an areal feature of South Asian languages (Bhat 1973), stretching not only across northern India, where Indo-Aryan languages are spoken predominantly, but also in Dravidian languages in the south of India. This means that most Indo-Aryan languages such as Hindi, Marathi or Punjabi have at least a two-way coronal stop contrast, usually consisting of a dental series and and a retroflex series (e.g., Ohala 1994 for Hindi, Dhongde and Wali 2009 for Marathi, Bhatia 2013 for Punjabi).[1]

Due to India's colonial history, many English words have entered South-Asian languages as loanwords and many Indians speak English as an L1, L2 or Ln. I am assuming that both loanword adaptation and second-language acquisition

* This paper was written under the supervision of Silke Hamann. I would like to thank her for her guidance and vital feedback for all parts of the paper, as well as checking all tableaux and proofreading the whole text. I would also like to thank an anonymous reviewer for their feedback, especially regarding the distinction between the terms amplitude, intensity and loudness.

[1] Most Dravidian languages have a similar two-way coronal stop contrast—with the notable exception of Malayalam, which preserves the historic Dravidian three-way coronal stop contrast (Asher 2013). This paper will focus on Indo-Aryan languages in order to limit its scope and because there is more data available for them. Nevertheless, it is plausible that the analysis presented here could be extended to Dravidian languages.

(at least at the onset of learning) are governed by the same perceptual processes—i.e., first language phonological perception (Boersma & Hamann 2009). In both cases, a foreign word is filtered through the L1 listeners' perception grammar. In the case of second-language acquisition, this will likely change over the course of learning, as many listeners at least partly acquire some of the L2-sound system, often generating a new grammar that lies somewhere in between their L1 and native speakers of their L2. However, at the onset of learning, we can assume that a speaker with no exposure to the L2 (and perhaps little experience with learning foreign languages) starts out only with their L1-perception grammar, making it the same process as loanword adaptation. Therefore, a distinction between these two would be unnecessary; however, most of the data presented henceforth will come from loanword adaptation.

When speakers of languages that have a two-way coronal stop distinction adopt English alveolars, they have to make a decision: Do they adapt them as dentals or retroflexes? The answer seems to be very clear, namely that Indo-Aryan languages unanimously adapt the English alveolar stops /t/ and /d/ as retroflex rather than dental (e.g., Ohala 1983 for Hindi), while they choose dental adaptations for sonorants (e.g., Dhongde and Wali 2009 for Marathi). For instance, English *taxi* is adapted as /ʈæksi/ and English *doctor* is adapted as /ɖɑkʈər/ in Hindi (Ohala 1983: 169, 172). On the other hand, the English alveolar nasal in *line* is adapted as /læ̃n̪/ in Punjabi (Mahmood et al. 2011: 239), rather than */læ̃ɳ/. Furthermore, most Indo-Aryan languages have a four-way laryngeal contrast, unlike the two-way laryngeal contrast found in English. It is therefore interesting that English /d/ and /t/ are adapted as /ɖ/ and /ʈ/, rather than /ʈ/ and /ʈʰ/, for instance.

It should furthermore be noted that many Indians are native speakers of English and acquired Indian English from the start, meaning that their input would already provide a retroflex realization during the acquisition process. However, this was not yet the case when they were initially introduced to English, so this paper can in a way be seen of modeling the adaptation process of that time. This is adequate because the relevant aspects of the phonology of Indo-Aryan languages had already long been established and are by no means a recent development. On top of that, loanwords from Western English increasingly enter Indo-Aryan languages nowadays. These recent loans are also adapted in the same way as the old ones including by people who do not speak English at all. This is unsurprising considering the aforementioned stability of Indo-Aryan phonology in the relevant aspects (see also Section 1.1).

This paper will thus aim at giving an explanation for why English alveolar stops are adapted as they are, as well as the adaptation of English coronals into Indo-Aryan languages more broadly (see Section 1.2 for a more precise explanation of the goals of this paper).

1.1 The phonology of Indo-Aryan languages

As an example, Table 1 shows part[2] of Marathi's coronal inventory, which is similar to that of many Indo-Aryan languages. As a result of a mixture between innovation and borrowings from Dravidian languages, Indo-Aryan languages almost unanimously distinguish a dental from a retroflex place of articulation (PoA) in a systematic fashion, with very few exceptions (such as Assamese; Cardona and Luraghi 2018: 383, which only has one coronal place of articulation). The exact inventory varies from language to language; Hindi-Urdu, for instance, arguably[3] lacks a phonemic retroflex nasal /ɳ/ (Ohala 1983), which languages like Marathi or Punjabi contrast with dental /n̪/ (Bhatia 2013; Dhongde & Wali 2009). Marathi, on the other hand, does not distinguish the taps /ɽ/ and /ɽʱ/ from the voiced stops /ɖ/ and /ɖʱ/, as Hindi-Urdu does (Ohala 1994). However, the plosives in lines 1–4 in Table 1 are distinguished in the vast majority of Indo-Aryan languages and also form the focus of the present analysis.

Table 1: Marathi's dental and retroflex plosives and nasals (Dhongde & Wali 2009: 11).

		Dental	**Retroflex**
Voiced plosive	unaspirated	d̪	ɖ
	aspirated	d̪ʱ	ɖʱ
Voiceless plosive	unaspirated	t̪	ʈ
	aspirated	t̪ʰ	ʈʰ
Nasal	unaspirated	n̪	ɳ
	aspirated	n̪ʱ	ɳʱ

1.2 Theoretical framework and goals

The present paper will work within the framework of bidirectional phonology and phonetics (BiPhon), which distinguishes a phonetic form (indicated by square brackets) from a phonological surface form (SF, indicated by slashes), the mapping of which is governed by the interaction between structural constraints and cue constraints (see, e.g., Boersma 2009).[4]

[2] Marathi also has several affricates and other sonorants which are not displayed because they are not relevant for the adaptations discussed here.

[3] Ohala (1994) mentions that in Hindi, a retroflex nasal can occur before a homorganic consonant and has been included as a phoneme by some authors.

[4] In the BiPhon model there is a third, higher level (called underlying form UF); however, the distinction between UF and surface form (SF) is not relevant for this analysis and therefore I will use slashes to indicate a collapsed SF/UF.

The only perception experiment that — to the best of my knowledge — exists about the adaptation of English alveolars into Indo-Aryan languages is Ohala (1978), who found that American English alveolar stops were perceived as retroflex by Hindi speakers 91% of the time. This is almost as accurate as their perception of native retroflex stops as retroflex (96%) and even slightly more reliable than their categorization of Hindi dentals as dental (89%). Ohala (1978) reasons that, unlike in production, retroflexes might be less marked that dentals in perception. She also speculates that Hindi retroflex stops might be acoustically more similar to (American) English alveolar stops. While these explanations do acknowledge the relevance of perception, they invoke notions such as acoustic similarity and markedness, which are unlikely to be accessible information to an individual speaker of a language and thus unrealistic. Rather, I argue in the present paper that loanword adaptation and naïve second-language acquisition can be explained entirely with the native perception grammar and should therefore be treated the same.

This approach eliminates the need for proposing different mechanisms for native and loanword phonology, as has been done in the past. For instance, LaCharité and Paradis (2005) do not consider the role of perception when it comes to storing loanwords in the lexicon and Kang (1996) proposes a different constraint ranking for native Korean grammar and loanword adaptation (see Boersma and Hamann 2009, Section 6.1, for a more detailed review). Furthermore, I will assume that the adaptation process takes place already in perception rather than production. This is done by some authors such as Silverman (1992) and Yip (1993); however, they still employ loanword-specific constraints. This can be avoided by assuming 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.

As an example for loanword-specific explanations that have been put forth, Pandharipande (1981) claims that English loanwords in Marathi are not fully nativized and thus treated differently compared to Persian and Sanskrit borrowings and native words. They propose that differences in phonotactic constraints arise from the speakers' attitudes towards the three donor languages, even though not much evidence is provided for this rather speculative claim. The formalization provided here will refrain from such ad-hoc explanations.

There are two main motivations for formalizing the adaptation of English coronals into Indian languages. First, there do not exist any OT formalizations of this phenomenon yet, which is a gap that the present paper will attempt to fill. Second and perhaps more importantly, formalizations enable us to make predictions for future experiments. For instance, if certain phonetic cues turn out to be more important in the formalization (indicated by a higher ranking in the constraint hierarchy), this result can serve as the basis for a perception experiment testing the

Table 2: English alveolar stops, dental fricatives and selected alveolar sonorants and their adaptation into various Indo-Aryan languages.

#	English input	Indian adaptation	Source
1	t	ʈ	e.g. Hindi (Koshal 1978; Ohala 1983),
	d	ɖ	Punjabi (Hussain 2014)
2	θ	t̪ʰ	e.g. Marathi (Dhongde & Wali 2009)
	ð	d̪	e.g. Marathi for English /ð/ (Dhongde & Wali 2009),
3	n	n̪	e.g. Punjabi for English /n/ and /l/
	l	l	(Mahmood et al. 2011)

integration of different competing cues (see Section 4.2 for a discussion of the predictions resulting from the present formalization).

The main processes that this paper will attempt to account for are illustrated in Table 2. These adaptations are interesting because, at first glance, there seems to be some inconsistency between the different groups of sounds and their adaptations in #1–3. English voiced and voiceless alveolar stops are commonly adapted as retroflex (#1), but the same does not hold for the English alveolar sonorants in #3, which are adapted as dental or alveolar. Furthermore, the English voiceless dental fricative /θ/ is commonly adapted as the aspirated dental stop /t̪ʰ/ (#2), whereas its voiced counterpart does not receive aspiration and can therefore be grouped with the sounds in #3.

Lastly, it should be noted that I am aiming at providing an analysis that can account for these general processes that are largely observed across most Indo-Aryan languages. This decision was made for both practical and theoretical reasons: First, for most of these languages, there is not much acoustic data available (see Section 2), so expanding the scope to include Indo-Aryan languages in general means that more acoustic data becomes relevant; and second, the patterns observed do seem to reflect some overarching mechanism that seems to be shared in most Indo-Aryan languages, so limiting the scope here would miss this potentially crucial generalization. Nonetheless, it is important to keep in mind that these languages are still a very heterogeneous group and that the formalization will naturally be more adequate to describe those languages from which the particular acoustic data was drawn.

2 Acoustic Descriptions

In order to give an accurate formalization of the perception process of the phenomena described in the previous section, it is necessary to provide a characterization of the acoustic details of the relevant sounds in Indian languages as well as English, enabling the formulation of phonetically grounded constraints.

2.1 Place of articulation

This section will outline acoustic data regarding the retroflex vs. dental stop contrast commonly made in Indo-Aryan languages (Section 2.1.1), followed by a brief characterization of English alveolar stops, which does not differentiate more than one coronal PoA for stops (Section 2.1.2).

2.1.1 Dental and retroflex stops in Indo-Aryan languages

Generally speaking, most of the acoustic differences between retroflex and dental stops arise from their apical (retroflex) and laminal (dental) articulations, respectively (Arsenault 2006: 1). The main cue typically associated with retroflexes crosslinguistically is a low third formant (F3) transition, as opposed to the more mid F3 transition of dentals (Hamann 2003: 63). For instance, Verma and Chawla (2003) measured F3 transition values of very roughly 2500 Hz for retroflex stops and 3000 Hz for dental stops in Hindi, although these values are quite heavily influenced by vowel context. The second formant (F2), on the other hand, is described by Hamann (2003) as a less reliable cue. Some studies report a reduced difference between F2 and F3 for retroflex stops as a distinctive factor in Hindi, Punjabi and Nepali (Agrawal et al. 2019; Hussain et al. 2017).[5]

However, formant transitions are far from the only cues relevant for distinguishing retroflex stops from dental ones that have been reported in the literature (see Hussain et al. 2017 for a quite exhaustive summary). For lack of a better term, I will refer to this collection of cues as ‘burst-related cues’, even though some of them (such as VOT) do not necessarily fit this definition very well. These cues include, but may not be limited to: 1^{st}–4^{th} spectral moments (i.e., center of gravity (CoG), variance (although I use the standard deviation, SD), skewness and kurtosis), burst duration, closure duration, total stop duration, voice onset time (VOT) and burst intensity.

Starting with burst intensity, this cue has been claimed to be an important differentiating factor (Verma & Chawla 2003), with retroflex stops having a higher burst intensity than dental stops.

VOT is not only relevant for the four-way laryngeal contrast of most Indo-Aryan languages (see also Table 1 again), but it seems to slightly vary by place of articulation (Benguerel & Bhatia 1980). For instance, Hussain et al. (2017) mentions a VOT of around 18 ms for plain dental stops and a slightly shorter

[5] It should be noted that an F3–F2 difference is not the same as individual F2/F3 measures and Hussain et al. (2017) does not provide individual formant data. F3–F2 differences are somewhat of a simplification and worse to interpret; however, Hussain et al. (2017)’s data still provides an interesting implication for the present formalization later in this section.

Table 3: Overall tendencies of several cues associated with laminal (dental) and apical (retroflex) stops across languages (Benguerel & Bhatia 1980; Hamann 2003; Hussain 2014; Hussain et al. 2017; Samudravijaya 2003; Verma & Chawla 2003).

	dental	**retroflex**
F2	high	high
F3	mid	low
VOT	longer	shorter
Burst duration	longer	shorter
Closure duration	longer	shorter
Total stop duration	longer	shorter
Spectral CoG	higher	lower
Spectral SD	higher	lower
Spectral skewness	negative	positive
Spectral kurtosis	negative	positive
Burst intensity	lower	higher

VOT of 9 ms for plain retroflex stops in Punjabi. When it comes to distinguishing dental from retroflex stops in Punjabi, burst characteristics and temporal cues interact with each other, in addition to influences from vowel environments and positional factors (Hussain et al. 2017). VOT was reported to be the only cue that consistently differentiated dental from retroflex stops in all vowel and positional contexts.

Hussain et al. (2017) furthermore found other cues to be significantly different between the two PoAs in some contexts, one of which was spectral SD in all vowel environments except /e_e/. Similarly, when Punjabi speakers were asked to produce English alveolar stops, they produced these with a significantly different 1^{st} and 2^{nd} spectral moment compared to their dental stops, which was similar to their pronunciation of native Punjabi retroflex stops (Hussain 2014).

Lastly, durational measures other than VOT also seem important. Similar to spectral SD above, total stop duration was reported to be significantly different for Punjabi speakers word-medially (Hussain et al. 2017). Furthermore, Punjabi speakers also produced English alveolar stops with a significantly different burst duration from their dental stops, similar to their articulation of native retroflex stops (Hussain 2014). For Hindi, longer burst and closure durations have been found for voiced and voiceless dental stops compared to retroflex ones (Samudravijaya 2003). More specifically, closure durations of dental stops in Punjabi seem to be around 110–130 ms as opposed to that of retroflex stops with around 90–110 ms; similarly, total stop duration was reported to be around 130–150 ms for dental stops and around 110–130 ms (Hussain et al. 2017: 4530).

Table 3 summarizes how different cues correlate with dental and retroflex stops. This does not necessarily mean that every study referenced in this section found a significant difference between the two PoAs. Crucially, Hussain et al. (2017) reports that F3–F2 differences were not a reliable differentiating factor in some front-vowel contexts for Punjabi, which they note corroborates results from Gujarati (Dave 1977). This, combined with the fact that some burst-related cues have been shown to quite reliably distinguish retroflex stops from dental stops, has important implications for the formalization in Section 3.

2.1.2 Alveolar stops in English

Due to the fact that English speakers do not have to differentiate between different coronal PoAs when it comes to plosives, there is considerable articulatory variation. For instance, Dart (1998) recorded four different PoAs (apical, upper apical,[6] apicolaminal, laminal) in the articulation of alveolar stops in Pacific American English.[7] Thus, any perceptual cues in English would likely also exhibit a lot of variation. I will still try to give a rough overview of the relevant acoustic cues in English.

Generally speaking, alveolar stops are characterized by a more mid F3 transition (Hamann 2003), which is more similar to dental stops than retroflex ones (see again Table 3). F2 and F3 transitions in English are very roughly around 1800 Hz and 2900 Hz, respectively (judging by the spectrogram for *dad* in Ladefoged and Disner 2012: 50)

Furthermore, alveolar stops in Canadian English (CE) have been reported to be loud, have a high mean burst frequency, low spectral SD and positive kurtosis (Sundara 2005). Out of these measurements, only the high mean burst frequency (= CoG) patterns with dentals in Table 3, the others are more similar to retroflexes.

As for burst duration, 9–26 ms have been reported for American English (AE) /d/ and over 100 ms for burst + aspiration duration of AE /t/ (Jongman et al. 1985: 243). Compared to the measurements by Hussain et al. (2017) mentioned in the previous section, this is again more similar to retroflex stops rather than dental stops in Punjabi.

In sum, the acoustic data summarized here indicates strongly that the F3 transition of English alveolar stops would favor a dental interpretation by Indian listeners, whereas several burst-related cues such as burst/closure duration,

[6] Dart (1998: 75) defines upper apical as “where it is not the rim but the upper surface of the apex which has made contact.”

[7] I could not find relevant acoustic data on British English varieties, which would have been more desirable in this case. However, I am also not aware of any crucial differences between the dialects described in this section and British English. What is more, contact between English and Indian languages happened long before any good acoustic data was available, so the actually relevant acoustic data simply does not exist.

intensity/amplitude/loudness[8] and some spectral characteristics seem to favor a retroflex interpretation. Taken together with Hussain et al.'s (2017) aforementioned claim that Punjabi speakers might rely more on burst-characteristics than formant transition due to the unreliability of the latter in high-vowel environments, this could explain the unanimous adaption of English alveolar stops as retroflex in Indo-Aryan languages. The same conclusion, i.e. that cues other than formant transitions are more important, is also reached by Arsenault (2006), although they explain this phenomenon with underspecified phonological features rather than optimality-theoretic constraints.

To the best of my knowledge, there exists only one perception experiment on this topic, namely a small-scale study by Sinha and Rao (1984), and their results corroborate the hypothesis that formant transitions are not the most important cues. The authors manipulated English CV tokens such as /da/ by removing either the burst or the formant transitions. They found that to listeners from many different Indian languages, the burst was more informative than formant transitions when it comes to categorizing English /d/ correctly (with the other options being /b/ and /g/). While this experiment did not ask participants to classify English /d/ as either retroflex or dental, the fact that burst cues were more useful to Indian listeners goes hand in hand with the data of Hussain et al. (2017).

2.2 Voicing and aspiration

As illustrated in Table 1, Indo-Aryan languages typically have a four-way laryngeal contrast; i.e., they differentiate between plain voiced, plain voiceless, voiced aspirated and voiceless aspirated stops. English, on the other hand, only employs a two-way laryngeal contrast, between phonologically voiced /b, d, g/ and voiceless /p, t, k/. The following sections will describe which cues have been identified as relevant for each of these two contrasts.

2.2.1 The four-way laryngeal contrast in Indo-Aryan languages

Prevoicing (duration), aspiration (duration) and pitch (F0) have been found to be important cues for the laryngeal contrast in languages such as Hindi-Urdu (e.g., Schertz and Khan 2020), Marathi (Dmitrieva & Dutta 2019) and Nepali (Schwarz et al. 2019, although only for prevoicing and aspiration).[9] Schertz and Khan

[8] Verma and Chawla (2003) refer to amplitude and Sundara (2005) refers to intensity. While these are not identical, different authors usually measure and report either one or the other and sometimes even mix them up. Since they are interrelated and perceptually both manifest as loudness in the end, a distinction will not be necessary for this paper. Instead, I will always refer to this measure as intensity, as it is more closely related to how humans perceive loudness.

[9] H1–H2, i.e., the difference in amplitude of the first and second harmonics, was also relevant specifically to distinguish the voiced aspirated stops, but since this is not a cue relevant for English and the native grammar can easily be modeled without it, I will disregard this measure here.

(2020) conclude that, both in production and perception, all of these cues are used, rather than one being significantly more important than the other. This has important implications for the constraint ranking in the formalization (see Section 3). Similarly, Dmitrieva and Dutta (2019) claim that VOT, which acoustically manifests as the presence and/or duration of prevoicing and aspiration, is not sufficient to distinguish the four-way laryngeal contrast in Marathi. Thus, other cues such as F0 are necessary.

Aspirated stops are characterized by a very long VOT, whereas plain voiceless stops have a VOT slightly above 0 (Benguerel and Bhatia 1980, recorded VOT values of 119 ms for aspirated stops and 15 ms for plain voiceless stops in Hindi.) Furthermore, voiceless stops are consistently realized with a higher pitch on the following vowel than voiced stops (e.g. Schertz and Khan 2020 for Hindi and Urdu, Dmitrieva and Dutta 2019 for Marathi).

2.2.2 The two-way laryngeal contrast in English

English only has a two-way laryngeal contrast, distinguishing plain voiceless /d/ [d̥] from aspirated /t/ [tʰ] (at least in word-initial position, which is what I will assume for the formalization in Section 3). VOTs in English seem to be 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. Furthermore, /t/ is usually realized with a more mid F0, compared to the lower F0 of /d/ (Lehiste & Peterson 1961; Ohde 1984). The same tendency is found in many Indo-Aryan language (see previous section) and is likely a universal tendency which has been linked to physiological causes, namely the position of the hyoid bone and the position of the larynx (Ohde 1984).

2.3 English dental fricatives

Since English does not contrast different manners of articulation at the (inter-) dental PoA, they are quite free to vary. Interestingly though, /ð/ seems to vary much more than /θ/ with regards to duration, intensity, voicing and place and manner of articulation (Smith 2013), with only 55% of /θ/ and 29% of /ð/ tokens being realized as fricatives by AE speakers in Smith (2013: 121)'s study. /ð/ was also realized as a nasal (23.4%), as a stop (18.4%) and as an approximant or flap (15.4%) (Smith 2013: 121). Similarly, /θ/ also exhibited substantial (but less) variation, being realized as a stop (15.9%) and as a voiced fricative (10.1%). The authors furthermore conclude that neither duration, intensity or voicing are reliable cues to distinguish the two dental fricatives from each other in AE. Nevertheless, it seems that /ð/ tends to be more voiced than /θ/, as would be expected.

3 Formalization

In order to show that loanword/L2 perception is filtered through the same perceptual mechanism as native words, a native perceptual grammar has to be proposed first. Assuming a maximal coronal stop inventory such as the one displayed for Marathi in Table 1, Indo-Aryan listeners have to make a categorization for place (dental vs. retroflex), voicing (voiced vs. voiceless) and aspiration (aspirated vs. non-aspirated) when it comes to mapping phonetic input to their L1 phonology. Since different cues apply for different positions, I will assume that the sound in question occurs word-initially for the formalization part, ensuring consistency.

3.1 Adapting English alveolars: place of articulation

In this section, I will show how the adaptation of English alveolar stops as retroflex in Indo-Aryan languages can be explained by formalizing the native grammar used for the perception of the dental/retroflex distinction and then providing the English acoustic input to the same native grammar.

3.1.1 Modeling the native place contrast

The phonological distinction between a dental and a retroflex PoA is commonly made with the features [+/–anterior] and/or [+/–distributed]. I will arbitrarily refer to dentals with /+dist/ and to retroflexes with /–dist/, written in slashes to indicate the fact that features are purely phonological, in line with BiPhon notation (see, e.g., Boersma and Hamann 2009).

I concluded Section 2.1.1 with the observation that formant transitions might not be a reliable cue for differentiating dental and retroflex stops, whereas some other burst-related cues might be more useful in this regard. Specifically, retroflex stops are characterized by a shorter, louder and less noisy (= shorter VOT) burst (see Section 2.1.1 again for a closer examination of the data). It is, however, quite difficult to discern which of these cues are relevant to which degree. Furthermore, including VOT here is difficult because of its important role in the laryngeal contrast. Thus, while it is likely that VOT plays some role for PoA as well, I will refrain from directly referring to it here since there are also other cues that are likely sufficient for the place distinction. Similarly, some spectral cues have been identified to play a role in this distinction (see Section 2.1.1 again) and these too could be included in the collective constraints posited below, but will be left out for simplicity.

Since untangling all of these cues is not really necessary for the formalization and it is difficult to tell which constraints end up being crucial, it should be sufficient to summarize them in the following constraints: *[*apical burst*]/+dist/, which assigns a violation if a short and loud release burst is mapped to a segment with a feature specification of /+dist/ in the surface form (SF); and *[*laminal* *burst*]/–dist/, which assigns a violation if a long and quiet release burst is mapped to a segment with a feature specification of /–dist/ in the SF. The labels 'apical burst' and 'laminal burst' were chosen because the cues they summarize likely stem from their apical/laminal articulation (Arsenault 2006). However, despite the name, these constraints still only refer to the acoustic input rather than any articulatory gestures. Due to the aforementioned reliability of these cues, the new constraints can be assumed to be high-ranked, while their counterparts *[*apical burst*]/–dist/ and *[*laminal burst*]/+dist/ are very-low-ranked and will therefore not be displayed in the following tableaux.

Furthermore, since formant transitions have been reported to be a less reliable cue (see Section 2.1.1 again), we need to posit lower-ranked constraints that refer to formant transitions, such as *[2500 Hz]/+dist/, which militates against perceiving a second formant transition of 2500 Hz as a /+dist/ segment; and *[3000 Hz]/–dist/, which militates against perceiving a second formant transition of 3000 Hz as a /–dist/ segment. Here, 2500 Hz and 3000 Hz are rough placeholders for the typical low F3 of retroflex stops and the typical mid F3 of alveolar and dental stops, based on the data in Section 2. Again, their very-low-ranked counterparts *[2500 Hz]/–dist/ and *[3000 Hz]/+dist/ will not be displayed.

These four constraints are sufficient to model the native place contrast of Indo-Aryan languages, displayed in Tableaux 1 and 2. The typical Indo-Aryan four-way laryngeal contrast will be covered in a later section, which is why /d̪/ and /ɖ/ will be shown as example candidates for a dental or retroflex interpretation, respectively.

[{*laminal burst*}, F3 = 3000 Hz]	*[*apical burst*] /+dist/	*[*laminal burst*] /–dist/	*[2500 Hz] /+dist/	*[3000 Hz] /–dist/
☞ /d̪/				
/ɖ/		*!		*

Tableau 1: Indian perception of native /d̪/, implemented with a laminal burst and a mid F3 transition.

Crucially, while the losing candidate violates both a burst-related constraint and a constraint related to formant transitions, the decision is already made after the first violation. However, these cues could come into play again if the input is in some way obscured that leaves only the formant transitions discernible (such as noisy speech), rendering these constraints still necessary for the native grammar.

It is worth noting that all perception tableaux in this section work bidirectionally; that is, the same constraints that work in perception also work in production, with an SF as an input and several phonetic realizations as output candidates.

[{*apical burst*}, F3 = 2500 Hz]	*[*apical burst*] /+dist/	*[*laminal burst*] /–dist/	*[2500 Hz] /+dist/	*[3000 Hz] /–dist/
/d̪/	*!		*	
☞ /ɖ/				

Tableau 2: Indian perception of native /ɖ/, implemented with an apical burst and a low F3 transition.

This is illustrated in Tableau 3, with the example of the phonetic implementation of native /d̪/. It also shows that constraints that do not come into play in perception (in this case formant transitions) are still relevant in production. This kind of production tableau will not be shown again for any of the following perception tableaux because they will always work analogously to Tableau 3.

/d̪/	*[*apical burst*] /+dist/	*[*laminal burst*] /–dist/	*[2500 Hz] /+dist/	*[3000 Hz] /–dist/
☞ [{*laminal burst*}, F3 = 3000 Hz]				
[{*apical burst*}, F3 = 3000 Hz]	*!			
[{*laminal burst*}, F3 = 2500 Hz]			*!	
[{*apical burst*}, F3 = 2500 Hz]	*!		*	

Tableau 3: Indian phonetic implementation of native /d̪/.

3.1.2 Perception of English alveolar stops as retroflex

After having established a constraint ranking for the native grammar, we can now plug in an English alveolar stop to model its perception by Indian listeners. As indicated in Section 2.1.2, English stops largely have properties similar to Indian retroflex ones, such as a higher intensity and short burst duration. On the other hand, they have a mid F3 transition, as is typical for alveolar stops. Thus, I will consider the input for an unspecified English alveolar stop to be realized as an apical burst with an F3 transition of roughly 3000 Hz.

As Tableau 4 shows, the constraints referring to the burst override the lower-ranked constraints referring to formant transitions, leading to the correct winning candidate, i.e., a retroflex perception.

[{*apical burst*}, F3 = 3000 Hz]	*[*apical burst*] /+dist/	*[*laminal burst*] /–dist/	*[2500 Hz] /+dist/	*[3000 Hz] /–dist/
/d̪/	*!			
☞ /ɖ/				*

Tableau 4: Indian perception of English /d/, implemented with an apical burst and a mid F3 transition.

3.1.3 Extending the analysis to sonorants

As shown in Table 2, we also have to account for the fact that those languages which also employ the two-way coronal place contrast for sonorants (such as Marathi, which distinguishes dental /n̪/ from retroflex /ɳ/ and alveolar /l/ from retroflex /ɭ/) adapt English alveolar /n/ and /l/ as dental/alveolar rather than retroflex. For this, we first have to show that the constraint ranking established in the previous sections also represents an accurate perception grammar of the native place contrast for nasals in the languages that make such a distinction. This is shown in tableaux 5 and 6, using nasals as an example (the same exact analysis holds for the lateral approximants, too). Crucially, nasals do not have a loud release burst like plosives do, rendering the two high-ranked constraints inapplicable. Now, the formant transitions become relevant to correctly perceive the native dental and retroflex nasal, with the characteristic mid and low F3 transitions of dentals and retroflexes (Hamann 2003: 61).[10]

[{*nasal sonorant*}, F3 = 3000 Hz]	*[*apical burst*] /+dist/	*[*laminal burst*] /–dist/	*[2500 Hz] /+dist/	*[3000 Hz] /–dist/
☞ /n̪/				
/ɳ/				*!

Tableau 5: Native perception of e.g. Marathi /n̪/, implemented as a nasal sonorant with a mid F3 transition.

Plugging in English alveolar /n/, realized as a nasal sonorant with the typical mid F3 of alveolars,[11] we get a tableau identical to Tableau 5, now displaying the

[10]Hamann (2003) actually mentions F3 values of 2499 Hz for /n̪/ and 2129 Hz for /ɳ/, citing Busby (1979). However, these values a) stem from data from Australian aboriginal languages rather than Indo-Aryan languages and b) the values 3000 Hz and 2500 Hz are only placeholders for a mid and low F3, respectively, which is why I will keep them as such. To see how continuous formant values can be modeled with cue constraints in OT, refer to Boersma (2009).

[11]I am ignoring any formants throughout the nasal here (as opposed to formant transitions) since they are not necessary and usually faint anyway.

[{*nasal sonorant*}, F3 = 2500 Hz]	*[*apical burst*] /+dist/	*[*laminal burst*] /–dist/	*[2500 Hz] /+dist/	*[3000 Hz] /–dist/
/n̪/				*!
☞ /ɳ/				

Tableau 6: Native perception of e.g. Marathi /ɳ/, implemented as a nasal sonorant with a low F3 transition.

Indian perception of English /n/, implemented as a nasal sonorant with a mid F3 transition. Thus, the ranking *[*apical burst*]/+dist/, *[*laminal burst*]/–dist/ ≫ *[2500 Hz]/+dist/, *[3000 Hz]/–dist/ is both warranted by the literature (Hussain et al. 2017, see also Section 2.1.1 again)—although perception experiments with manipulated stimuli would be even more desirable here—and produces the correct outputs for the perception of native and English plosives and nasals.

3.2 Adapting English alveolars: voicing and aspiration

Having modeled the perception process regarding PoA, let us now turn to how the English two-way laryngeal contrast is mapped to the common four-way Indo-Aryan laryngeal contrast.

3.2.1 Modeling the native laryngeal contrast

For the native laryngeal contrast, I will assume here that the voiced stops /d̪/, /d̪ʱ/, /ɖ/ and /ɖʱ/ share the phonological feature /+voice/, as opposed to their /–voice/ counterparts /t̪/, /t̪ʰ/, /ʈ/ and /ʈʰ/. Similarly, I will assume that the aspirated stops /d̪ʱ/, /t̪ʰ/, /ɖʱ/ and /ʈʰ/ share the feature /+asp/, while /d̪/, /t̪/, /ɖ/ and /ʈ/ are /–asp/.

In order to accurately model the perception process, we have to assume detailed, non-abstract auditory representations of the native sounds as inputs to the tableaux. In line with the notation employed by Boersma and Hamann (2009), I am roughly transcribing the relevant sounds as [̫ᵈā], [_ᵗá], [̫ᵈʰ̫ː ā] and [_ᵗʰːá], corresponding to the phonemes /d̪/, /t̪/, /d̪ʱ/ and /t̪ʰ/, respectively (I will be ignoring PoA in this section because it has been modeled in the previous section). Since I am assuming a word-initial position, the stop consonants are followed by the placeholder vowel [a], which carries either a high pitch “ ́ ” after a voiceless stop or a mid pitch “ ̄ ” after a voiced stop. Furthermore, “ ʰː ” stands for a long, non-sibilant, noisy (i.e., non-periodic) bit, as it commonly occurs in aspirated stops in Indo-Aryan languages. Similarly, “ ̫ ” stands for the periodic ‘murmur’ of voiced stops during closure (as well as during the aspirated portion for voiced-aspirated stops, indicated as “ ʰ̫ː ”), whereas “_” refers to silence during the closure. Lastly, “ ᵈ” and “ ᵗ” indicate a presumably fortis and lenis dental burst (in dB), respectively (this will become relevant in the next part, see Section 3.2.2).

As mentioned in Section 2.2.1, the common Indo-Aryan four-way laryngeal contrast is typically made with several cues, of which at least prevoicing, aspiration (noisiness) and F0 are relevant here. Since there is no concrete evidence for any of these cues being more important, we can now propose several equally-high-ranked constraints: First, *[$^{h:}$]/–asp/ militates against interpreting a long noisy portion after the burst as a non-aspirated segment on the surface form. Second, *[^{h}]/+asp/ prevents a short noisy bit after the burst from being interpreted as aspirated and *[*no noise*]/+asp/[12] does the same for a burst without any noise. Conversely, the symmetrically opposite constraints *[$^{h:}$]/+asp/, *[^{h}]/–asp/ and *[*no noise*]/–asp/ would then be very low-ranked.

Similar to the constraints referring to aspiration, *[̰]/–voice/ would militate against interpreting periodicity as a voiceless segment and *[_]/+voice/ would disfavor a voiced interpretation of an absence of periodicity.[13] Again, the counterparts *[̰]/+voice/ and *[_]/–voice/ can be assumed to be very low-ranked.

Lastly, two more constraints that refer to F0 need to be proposed: *[¯]/–voice/ militates against interpreting a mid pitch as voiceless and *[´]/+voice/ militates against interpreting a high pitch as voiced. This reflects the tendency of voiced stops to be realized with a mid pitch on the following vowel, in contrast to the high pitch following voiceless stops. Again, their counterparts *[¯]/+voice/ and *[´]/–voice/ can be assumed to be very low-ranked.

Together, these constraints can successfully account for the production and perception of the four-way laryngeal contrast common in many Indo-Aryan languages. This is illustrated in the perception Tableaux 7-10 (ignoring PoA). The constraints that have been assumed to be very low ranked (but must exist due to the nature of OT) are not shown in the tableaux to ease illustration.

Thus, the unranked constraints *[$^{h:}$]/–asp/, *[^{h}]/+asp/, *[*no noise*]/+asp/, *[¯]/–voice/, *[´]/+voice/, *[̰]/–voice/, *[_]/+voice/ produce an accurate native perception (and production) grammar of voicing/aspiration in Indo-Aryan languages with the maximal laryngeal contrast.

[12]These constraints are analogous to those employed for Korean by Boersma and Hamann (2009).

[13]Similar to the constraints regarding aspiration noise, these constraints are modeled after Boersma and Hamann (2009). However, their constraints have a narrower definition, only referring to periodicity and silence during closure before a burst. Since not only stops but also fricatives will become relevant in the present formalization, these constraints will refer to the presence or absence of periodicity generally.

[̰ᵈā]	*[ʰː] /–asp/	*[ʰ] /+asp/	*[*no noise*] /+asp/	*[¯] /–voice/	*[´] /+voice/	*[̰] /–voice/	*[_] /+voice/
☞ /d̪/							
/t̪/				*!		*!	
/d̪ʱ/			*!				
/t̪ʰ/			*!	*!		*!	

Tableau 7: Indian perception of native /d̪/, implemented as [̰ᵈā], i.e., voicing before a burst without noise, followed by a vowel with a mid pitch.

[_ᵗá]	*[ʰː] /–asp/	*[ʰ] /+asp/	*[*no noise*] /+asp/	*[¯] /–voice/	*[´] /+voice/	*[̰] /–voice/	*[_] /+voice/
/d̪/					*!		*!
☞ /t̪/							
/d̪ʱ/			*!		*!		*!
/t̪ʰ/			*!				

Tableau 8: Indian perception of native /t̪/, implemented as [_ᵗá], i.e., silence before a burst without noise, followed by vowel with a high pitch.

[̰ᵈʰ̰ːā]	*[ʰː] /–asp/	*[ʰ] /+asp/	*[*no noise*] /+asp/	*[¯] /–voice/	*[´] /+voice/	*[̰] /–voice/	*[_] /+voice/
/d̪/	*!						
/t̪/	*!			*!		*!	
☞ /d̪ʱ/							
/t̪ʰ/				*!		*!	

Tableau 9: Indian perception of native /d̪ʱ/, implemented as [̰ᵈʰ̰ːā], i.e., voicing before a very noisy burst, followed by a murmured vowel with a mid pitch.

[_ᵗʰːá]	*[ʰː] /–asp/	*[ʰ] /+asp/	*[*no noise*] /+asp/	*[¯] /–voice/	*[´] /+voice/	*[̰] /–voice/	*[_] /+voice/
/d̪/	*!				*!		*!
/t̪/	*!						
/d̪ʱ/					*!		*!
☞ /t̪ʰ/							

Tableau 10: Indian perception of native /t̪ʰ/ implemented as [_ᵗʰːá], i.e. silence before a very noisy burst, followed by a vowel with a high pitch.

3.2.2 Mapping English alveolar stops to the native voicing/aspiration system

The native grammar established in the previous section can now be applied to English alveolar stops in order to explain why they are commonly adapted as voiced and plain voiceless rather than aspirated (still disregarding PoA, which was addressed in Section 3.1 and will become relevant again in the following section). We can assume auditory representations such as [_$^{\mathrm{th}}$á] and [_$^{\mathrm{d}}$ā] as the input to the perception tableaux, corresponding to the English phones /t/ and /d/. Here, "$^{\mathrm{h}}$" stands for the moderately long noisy bit that is commonly found in word-initial voiceless stops in English. Plugging in the auditory form of English voiceless /t/ (see Tableau 11), we can see how the native grammar can easily account for its adaptation as plain voiceless /ʈ/ rather than /ʈ$^{\mathrm{h}}$/ or even /ɖ$^{\mathrm{ɦ}}$/ since the noisy part of the English stop is simply not long enough for an Indo-Aryan listener to be perceived as aspirated. Furthermore, the cues referring to pitch and prevoicing prevent a voiced interpretation.

[_$^{\mathrm{th}}$á]	*[$^{\mathrm{hː}}$] /–asp/	*[$^{\mathrm{h}}$] /+asp/	*[*no noise*] /+asp/	*[¯] /–voice/	*[´] /+voice/	*[̰] /–voice/	*[_] /+voice/
/ɖ/					*!		*!
☞ /ʈ/							
/ɖ$^{\mathrm{ɦ}}$/		*!			*!		*!
/ʈ$^{\mathrm{h}}$/		*!					

Tableau 11: Indian perception of English word-initial /t/ implemented as [_$^{\mathrm{th}}$á], i.e. silence before a moderately noisy burst, followed by a vowel with a high pitch.

Once we plug in the auditory form of the English voiced stop /d/, we run into a problem, though. As shown in Tableau 12, the previously established grammar now produces two optimal candidates: /ɖ/ and /ʈ/. This is the result of the conflicting cues in English (for an Indian listener) regarding F0 and prevoicing. English /d/ usually has no prevoicing in word-initial position, favoring a voiceless interpretation, while it is also accompanied by a lower pitch, favoring a voiced interpretation. Thus, we must assume one of two things: either, there actually *is* a ranking of the displayed cues as opposed to the previous assumption that they are all unranked (specifically, we would have to assume that cues referring to prevoicing are lower-ranked than the rest) or there is another crucial cue that we have not considered thus far. The former conclusion is not really substantiated by the current literature since all cues have been reported to be important for this contrast (see Section 2.2.1 again). The latter, on the other hand, might be quite plausible, especially if we consider which other cues are relevant for the laryngeal contrast in English that Indian listeners might be picking up on. While there is not too much data to work with, burst intensity (in dB) has been suggested to be relevant for the English voicing contrast, with English /t/ usually being louder than English /d/ (Jongman et al. 1985: 245). I have already been indicating this in all tableaux with the symbols "^{t}" and "^{d}", standing for a fortis and lenis burst, respectively. There is similarly little data for Indo-Aryan languages in this regard, but Polka (1991: 2974–2976) reports lower mean intensitys for voiced /d̪/ and /ɖ/ compared to their voiceless equivalents by their one speaker of Hindi.

[_^{d}ā]	*[$^{\text{hː}}$] /–asp/	*[^{h}] /+asp/	*[*no noise*] /+asp/	*[¯] /–voice/	*[´] /+voice/	*[̰] /–voice/	*[_] /+voice/
☞ /ɖ/							*
☞ /ʈ/				*			
/ɖ^{ɦ}/			*!				*!
/ʈ^{h}/			*!	*!			

Tableau 12: Indian perception of English word-initial /d/ implemented as [_^{d}ā], i.e. silence before a burst without noise, followed by a vowel with a mid pitch.

If we assume that burst intensity is indeed another relevant cue, we can propose two more constraints *[^{d}]/–voice/ and *[^{t}]/+voice/ (and again, their very low-ranked counterparts). Crucially, we cannot assume that these new constraints are equally ranked with the others since there is not enough data compared to the cues that have been employed so far, whose importance has been well-established. If the two new constraints are added as low-ranked constraints to the native perception grammar, none of the tableaux handling the perception of native stops (i.e., Tableaux 7-10) would change, since the decision is always already made before these constraints would come into play. The same goes for Tableau 11, shown as Tableau 13 with the new constraints.

Tableau 14, on the other hand, now produces only one optimal candidate compared to Tableau 12: /ɖ/. Thus, if burst intensity is taken into account—even if low-ranked—it is able to explain the perception of English /d/ as /ɖ/ rather than /ʈ/ in a way that is consistent with the native grammar, albeit with little acoustic data as backup. There are also other potential explanations for this issue outside the scope of the present analysis (see Section 4.3).

[_ᵗʰá]	*[ʰː] /–asp/	*[ʰ] /+asp/	*[*no noise*] /+asp/	*[¯] /–voice/	*[´] /+voice/	*[̰] /–voice/	*[_] /+voice/	*[ᵈ] /–voice/	*[ᵗ] /+voice/
/ɖ/					*!		*!		*
☞ /ʈ/									
/ɖʱ/		*!			*!		*!		*
/ʈʰ/		*!							

Tableau 13: Indian perception of English word-initial /t/ implemented as [_ᵗʰá], i.e. silence before a moderately noisy burst, followed by a vowel with a high pitch (revised).

[_ᵈā]	*[ʰː] /–asp/	*[ʰ] /+asp/	*[*no noise*] /+asp/	*[¯] /–voice/	*[´] /+voice/	*[̰] /–voice/	*[_] /+voice/	*[ᵈ] /–voice/	*[ᵗ] /+voice/
☞ /ɖ/							*		
/ʈ/				*!					*
/ɖʱ/			*!				*!		
/ʈʰ/			*!	*!					*

Tableau 14: Indian perception of English word-initial /d/ implemented as [_ᵈā], i.e. silence before a burst without noise, followed by a vowel with a mid pitch (revised).

3.3 *Adapting English dental fricatives: synthesizing place, voicing and aspiration*

We have now seen that the adaptation of English alveolar stops and sonorants into Indo-Aryan languages can successfully be handled by the native grammar, both regarding PoA and voicing/aspiration. These processes did not interact with each other since entirely different constraints are relevant for them and were thus not displayed together in one tableau. Once we turn to English fricatives, though, we can display all relevant constraints in one tableau, since their number is slightly reduced. All those constraints referring to a burst—namely *[*apical burst*]/+dist/, *[*laminal burst*]/–dist/, *[ᵈ]/–voice/ and *[ᵗ]/+voice/—can be ignored now due to the lack of a burst in fricatives. Furthermore, I am not aware of any literature reporting F0 values on vowels following English dental fricatives, which is why the constraints *[¯]/–voice/ and *[´]/+voice/ will also be disregarded in this section.[14] The constraints *[$^{\text{hː}}$]/–asp/, *[$^{\text{h}}$]/+asp/ and *[*no noise*]/+asp/, on the other hand, are still relevant here due to the potential noisiness of fricatives. The same goes for constraints referring to formant transitions, i.e., *[2500 Hz]/+dist/ and *[3000 Hz]/–dist/, and the constraints regarding voicing *[̼]/–voice/ and *[_]/+voice.

We need to posit one more constraint, however, because we now also have fricative candidates to consider. Most Indo-Aryan languages do not have a large fricative inventory, often only distinguishing /s/ and /ʃ/ at the coronal PoA such as Marathi (Dhongde & Wali 2009). Some languages additionally distinguish /z/, such as Hindi (Ohala 1994) or Punjabi (Bhatia 2013) (although in both languages this depends on the speaker) and some only have /ʃ/ like Bengali (Klaiman & Lahiri 2018). Crucially, these are all sibilants (/+stri/) and can thus easily be ruled out by a constraint like [$^{\text{hː}}$]/+stri/, which militates against interpreting non-sibilant noise in the auditory input as a /+stri/ segment on the SF. This constraint is likely high-ranking in the native grammar since Indian listeners will frequently encounter non-sibilant noise in the form of aspiration and have to classify this accurately as such. However, since such a ranking is not necessary for this formalization and other rankings are theoretically possible, I will again assume no ranking of this constraint. Other potential non-coronal fricative candidates such as /f/ will be disregarded here, but could also be ruled out by constraints referring to formant transitions, which will be highly different from coronal formant transitions.

Choosing an auditory form for English dental fricatives is difficult due to their aforementioned high degree of variation. However, the data mentioned in Section 2.3 does suggest that English /θ/ is usually an actual voiceless fricative, while English /ð/ is more likely to be a voiced approximant of some kind. Furthermore, we can assume the characteristic mid F3 of dentals, signified here again with the placeholder 3000 Hz (visual inspection of the spectrograms in Ladefoged and Disner 2012: 56 confirms this). Lastly, since the constraints referring to noise in the input require durational information, I will assume a duration of roughly 160 ms for /θ/ (Jongman et al. 2000: 1260), which would be classified as long noise under the present definition of the constraints *[$^{\text{hː}}$]/–asp/ and *[$^{\text{h}}$]/+asp/. Thus, using the same constraints from earlier sections and the additional high-ranked [$^{\text{hː}}$]/+stri/, we can now plug in these auditory representations of English dental fricatives, shown in Tableaux 15 and 16. I will again assume no ranking

[14]There is evidence for a general tendency of English fricatives to exhibit the same pattern as English stops, namely a higher pitch on vowels following voiceless fricatives compared to voiced ones (Hanson 2009). However, there is no data in this regard for dental fricatives specifically. Thus, if we did want to extrapolate from these results and assume that this also holds for dental fricatives, we could include the constraints *[¯]/–voice/ and *[´]/+voice/ after all. In any case, since English dental fricatives would then pattern the same way as English *and* Indo-Aryan stops, the output of Tableaux 15 and 16 would not change.

here because the only rankings established so far involve burst-related constraints, which are not relevant here.

[{*long, non-sibilant, non-periodic noise*}, F3 = 3000 Hz]	[hː] /+stri/	*[hː] /–asp/	*[ʰ] /+asp/	*[*no noise*] /+asp/	*[2500 Hz] /+dist/	*[3000 Hz] /–dist/	*[‿] /–voice/	*[_] /+voice/
/s/	*!							
/d̪/		*!						*!
/t̪/		*!						
/d̪ʱ/								!*
☞ /t̪ʰ/								
/ɖ/		*!				*!		*!
/ʈ/		*!				*!		
/ɖʱ/						*!		*!
/ʈʰ/						*!		

Tableau 15: Indian perception of English /θ/ implemented as a voiceless (i.e., non-periodic) fricative with a mid F3 transition.

Thus, we see that the constraints [hː]/+stri/, *[hː]/–asp/, *[ʰ]/+asp/, *[*no noise*]/+asp/, *[2500 Hz]/+dist/, *[3000 Hz]/–dist/, *[‿]/–voice/ and *[_]/+voice/ are able to accurately model the adaptations of English /θ/ and /ð/ as /t̪ʰ/ and /d̪/, initially displayed in Table 2. The fact that English /θ/ is usually implemented as an actual fricative, whereas /ð/ is not, ends up being the crucial difference that causes one to be adapted as aspirated and not the other.

4 Discussion and conclusion

This section will start by giving a short summary of the results of the formalization in the previous section (Section 4.1), followed by its implications for future research (Section 4.2) and rounding off with potential alternative explanations outside the realm of OT (Section 4.3).

4.1 Summary of results

The main goal of this paper, which was to provide a perception-based OT account of the adaptation of English coronals into Indo-Aryan languages, was achieved.

[{*periodic approximant*}, F3 = 3000 Hz]	*[^{hː}] /+stri/	*[^{hː}] /–asp/	*[^{h}] /+asp/	*[*no noise*] /+asp/	*[2500 Hz] /+dist/	*[3000 Hz] /–dist/	*[̰] /–voice/	*[_] /+voice/
/s/	*!						*	
☞ /d̪/								
/t̪/							*!	
/d̪^{ɦ}/				*!				
/t̪^{h}/				*!			*	
/ɖ/						*!		
/ʈ/						*!	*	
/ɖ^{ɦ}/				*!		*!		
/ʈ^{h}/				*!		*!	*	

Tableau 16: Indian perception of English /ð/ implemented as voiced (i.e., periodic) approximant with a mid F3 transition.

For place of articulation, collective constraints referring to burst duration, noisiness and intensity being ranked above formant transitions ended up being crucial.

As for aspiration and prevoicing, the initial collection of unranked constraints referring to aspiration noise, prevoicing and pitch was able to account for the native grammar as well as the adaptation of English /t/ as /ʈ/, but produced two optimal candidates for English /d/. At this point, more data would have been desirable, but the best explanation seemed that burst intensity, even if low-ranked, would become crucial in the categorization of English /d/ as /ɖ/. This explanation, albeit an ad hoc one, is still consistent with the native grammar, especially considering that burst intensity was also important for PoA (as part of the constraints referring to a laminal/apical burst), meaning that an Indo-Aryan listener would likely be accustomed to paying attention to this cue.

With this perception grammar, the auditory input of the English syllable /ta/, realized as [_thá] will be perceived, stored and subsequently produced as shown in (1). Analogously, (2) shows the same process for English /da/, realized as [_^{d}ā]. This illustrates that the adaptation only happens in perception, with the production process being identical to the native one.

(1) [_thá] → /ʈa/ → |ʈa| → /ʈa/ → [_^{ʈ}á][15]

(2) [_^{d}ā] → /ɖa/ → |ɖa| → /ɖa/ → [̰^{ɖ}ā]

[15]The symbols "^{ʈ}" and "^{ɖ}" indicate an apical burst with an F3 transition of around 3000 Hz here, as is typical of Indo-Aryan retroflex stops.

Finally, the grammar established up to this point was also able to accurately account for the adaptation of English dental fricatives, with only one more additional constraint to rule out sibilant fricatives, which had not been considered thus far. Just as with stops, constraints referring to aspiration, voicing *and* formant transitions were crucial for the adaptation of fricatives. Specifically, the typically more noisy fricative-like realization of /θ/ results in an aspirated interpretation by Indian listeners, whereas the same is not true for the usually more approximant-like /ð/.

To summarize, the full (perception) grammar[16] established here consists of mostly unranked cue constraints referring to burst-related cues, formant transitions, aspiration noise, voicing, pitch levels and burst intensities. Out of these constraints, the only ranking that was strictly necessary in the present formalization was *[*apical burst*]/+dist/, *[*laminal burst*]/–dist/ » *[2500 Hz]/+dist/, *[3000 Hz]/–dist/. Furthermore, *[ᵈ]/–voice/ and *[ᵗ]/+voice/ were presumed to be lower-ranked on the basis of the literature, but this is not strictly necessary for either the native grammar or the English adaptations to be formalized correctly. Thus, the final grammar could be expressed as consisting of the two pairs of ranked constraints and many more ‘floating constraints’, which are not ranked with respect to any of the other constraints:

{ *[*apical burst*]/+dist/ *[*laminal burst*]/–dist/ }
»
{ *[2500 Hz]/+dist/ *[3000 Hz]/–dist/ }

{ *[ʰː]/–asp/ *[ʰ]/+asp/ *[*no noise*]/+asp/
*[̄]/–voice/ *[´]/+voice/ *[̰]/–voice/
*[_]/+voice/ *[ᵈ]/–voice/ *[ᵗ]/+voice/
*[ʰː]/+stri/ }

Figure 1: Final constraint ranking of the native Indo-Aryan perception grammar regarding coronal stops and sonorants. Strictly ranked constraints are displayed on the left, unranked or ‘floating’ constraints on the right.

It is arguably unusual for an OT analysis to have such few ranked constraints. The reason for this is likely that this formalization employed very detailed acoustic constraints, many of which go in the same direction. For instance, both prevoicing and pitch cues achieve exactly the same thing, namely categorizing sounds such as [̰ᵈā] as voiced /d̪/ while categorizing [_ᵗá] as voiceless /t̪/. Therefore, all candidates will have the same violations of these constraints for any native input, making it impossible to tell if any of them are higher-ranked than the other without any further data. When there was acoustic data suggesting a specific ranking such as in the case of low-ranking *[ᵈ]/–voice/ and *[ᵗ]/+voice/, this ranking was

[16]Of course, the actual full perception grammar would also consist of structural constraints, but these were not employed here because phonotactics were not relevant for this topic.

assumed but could still not be proven. Perception experiments, on the other hand, would constitute much better evidence for proposing specific rankings for different cues, but there are unfortunately none that are relevant to this topic yet. Thus, the following section will provide a few predictions that could be verified with perception experiments that employ manipulated stimuli.

4.2 Predictions for future research

One motivation for the present formalization was that it enables predictions for future experimental research. As indicated above, there was only one hard ranking, namely that of burst-related cues vs. formant transitions for the retroflex vs. dental stop contrast. A prediction that follows from this is that in a perception experiment, burst-related cues would be more important than formant transitions. More specifically, this could be tested by cross-splicing tokens of retroflex and dental CV articulations in an Indo-Aryan language at the onset of voicing. This method would yield stimuli that consist of a mismatch of burst and formant transitions. Crucially, regardless of whether listeners make use of the duration of the burst, the duration of the closure, spectral characteristics, duration of the (small) aspirated portion, burst intensity, etc., these cues would all be contained in the part before the onset of voicing. This mimics the constraints of *[*apical burst*]/+dist/ and *[*laminal burst*]/–dist/, which were themselves a collection of several potentially relevant cues. Once voicing and thus the vowel starts, there will then be (different) formant transitions, depending on the PoA of the stop. Thus, a listener who is confronted with, say, a laminal burst and a mid F3 transition, has to make a decision about how to categorize the mismatched input. If they choose the interpretation that is consistent with the burst-related cues, i.e., a dental one, then we have evidence that these cues are indeed more important in the native perception, thereby explaining the adaptation of English stops as retroflex.

Notably, I mentioned in Section 2.2.2 that one perception study did find evidence for burst-related cues being more important than formant transitions, but they used English stimuli rather than native ones and furthermore did not ask their participants to make a retroflex vs. dental categorization (Sinha & Rao 1984). Regardless, they provide some grounds for formulating similar experiments such as the one proposed above.

The story looks different when it comes to predictions regarding the four-way laryngeal contrast since Schertz and Khan (2020) have already conducted a relatively exhaustive perception experiment in this regard with Hindi/Urdu listeners. While their study does provide some insights—namely that prevoicing and aspiration “do much of the work in partitioning the four-way contrast overall, as expected” (Schertz & Khan 2020: 14), but that F0 is nevertheless important as well—there is still more work needed on this topic. On the one hand, their findings could be interpreted as a lower ranking of F0 cues compared to prevoicing and aspiration cues. Such a ranking would pose a problem for the present analysis since it would falsely predict that word-initial English /d/ would be adapted as /ʈ/ rather than /ɖ/. On the other hand, their main analysis only featured the binary comparison of /b^{ɦ}/ vs. /p, b, p^{h}/, meaning that the contrast that would be relevant for this paper (plain voiceless vs. plain voiced) was not captured here. Thus, there is still more research to be done to find out how exactly this contrast is made, perhaps with another experiment akin to the one proposed a few paragraphs above.

The present analysis furthermore assumed no ranking between prevoicing, aspiration and F0 constraints, meaning that it is rather difficult to verify if this is actually the case since one would essentially be predicting a null-result. However, an experiment where F0 *is* found to be less relevant than the other two would certainly pose a problem for this analysis, potentially causing some revisions to be necessary.

Lastly, the analysis proposed here furthermore hinges on another cue, namely burst intensity, in order to predict the correct English adaptation. Unfortunately, very little data is available in this regard, none of which comes from perception experiments. In this case, however, the formalization works regardless of whether intensity cues are assumed to be low-ranked or unranked with regards to the other three. On the other hand, it is also theoretically possible that F0-related cues are lower-ranked and intensity cues are actually on the same level as prevoicing and aspiration. The former would have some evidence in the literature (see previous paragraph), while the latter would have little to no basis. Again, perception experiments that disentangle these cues would be necessary to shed more light on this topic.

4.3 Other word positions and explanations

As mentioned in Section 2.1.2, English voiceless stops are usually aspirated word-initially, but the aspirated portion is shorter than that of a /+asp/ stop in Indo-Aryan languages. While English voiced stops usually have a VOT of around 0 word-initially, they do tend to be voiced intervocalically. This fact would be missed by the present analysis due to the fact that it was limited to word-initial position. However, we could hypothetically also propose that this cue is picked up on by Indian listeners when it *is* present, although this would be outside the scope of this paper because it wold involve processes on a higher level than the phonetics-phonology interface, which was discussed here.

In such an analysis, one would then argue that—at a later stage in the perception process, namely at lexicalization—there is another high-ranking constraint that disfavors mapping two phonetically distinct sounds in the input (English [_thá] and English [_^{d}ā]) to the same segment in the SF (Indian /ʈ/). This is not necessarily a very satisfying explanation, though, especially since it would require the assumption that the mapping [_thá] → /ʈ/ occurs before [_^{d}ā] → /ɖ/. Alternatively, one could also assume an initially ‘wrong’ mapping of English /d/[_^{d}ā] to /ʈ/ in initial position and a ‘correct’ mapping of English /d/[a̰^{d}ā] in medial position, resulting in the two allophones /ʈ/ and /ɖ/ of English |d| on the SF. Then, one could propose that during the mapping to the UF, the ‘wrongly’ perceived SFs are turned into ‘correctly’ stored UFs. For this to work properly, one would potentially have to rely on pairs of clearly derived words such as *do* and *redo*, where the |d| is once in initial and once in medial position, meaning that there is a clear indication present that these two ‘d-like’ sounds belong to the same UF due to their shared lexical origin. This last step could be formalized by employing lexical constraints which link a UF to semantic representations (this is also part of the BiPhon model, see, e.g. Boersma and Hamann 2009: 27).

Lastly, one could argue that—at least nowadays—there is potential influence of orthography, providing a clear link between segments despite them having different phonetic implementations. In such an account, one would posit the following: in cases like English word-initial [_^{d}ā] /d/, cue-constraints that favor a plain retroflex adaptation /ʈ/ due to the lack of prevoicing in English are overruled by orthographic constraints that require written <d> to be adaptated as voiced /ɖ/. However, when Indo-Aryan speakers first made contact with English, they naturally were not yet able to read it, rendering this explanation unsatisfactory at least for the original adaptations.

To conclude, the formalization in this paper is the best fit considering the available data compared to the more ad-hoc nature of other explanations mentioned above. Nevertheless, more acoustic studies for specific cues and especially perception experiments would be desirable to further support or update the account presented here.

5 References

Agrawal, S. S., Bansal, S., & Sinha, S. (2019). Comparative acoustic-phonetic analysis of retroflex consonants of some Indian languages. *Proceedings of the 23rd International Congress on Acoustics*.

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

Asher, R. (2013). *Malayalam*. Routledge.

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

Bhat, D. N. S. (1973). Retroflexion: An areal feature. *Working Papers on Language Universals*, *13*, 27–67.

Bhatia, T. (2013). *Punjabi*. Routledge.

Boersma, P. (2009). Cue constraints and their interactions in phonological perception and production. In P. Boersma & S. Hamann (Eds.), *Phonology in Perception* (pp. 55–110). Mouton de Gruyter.

Boersma, P., & Hamann, S. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & L. W. Wetzels (Eds.), *Loan Phonology* (pp. 11–58). John Benjamins.

Busby, P. A. (1979). *A classificatory study of phonemic systems in Australian aboriginal languages*. Master's thesis, Australian National University.

Cardona, G., & Luraghi, S. (2018). Indo-aryan languages. In B. Comrie (Ed.), *The World's Major Languages* (pp. 383–389). Routledge.

Dart, S. N. (1991). Articulatory and acoustic properties of apical and laminal articulations. *UCLA Working Papers in Phonetics*, *79*, 1–155.

Dart, S. N. (1998). Comparing French and English coronal consonant articulation. *Journal of Phonetics*, *26*(1), 71–94.

Dave, R. (1977). Retroflex and dental consonants in Gujarati. a palatographic and acoustic study. *Annual Report of the Institute of Phonetics University of Copenhagen*, *11*, 27–155. https://doi.org/10.7146/aripuc.v11i.131664

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

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

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America*, *125*(1), 425–441. https://doi.org/10.1121/1.3021306

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

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

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252–1263. https://doi.org/10.1121/1.1288413

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

Klaiman, M. H., & Lahiri, A. (2018). Bengali. In B. Comrie (Ed.), *The World's Major Languages* (pp. 427–446). Routledge.

Koshal, S. (1978). *Conflicting phonological patterns: A study in the adaptation of English loan words in Hindi*. Bahri Publications.

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

Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants* (3rd ed.). Blackwell Publishing.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America*, *33*(4), 419–425. https://doi.org/10.1121/1.1908681

Mahmood, R., Hussain, Q., & Mahmood, A. (2011). Phonological adaptations of English words borrowed into Punjabi. *European Journal of Social Sciences*, *22*(2), 234–245.

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.

Ohala, M. (1994). Hindi. *Journal of the International Phonetic Association*, *24*(1), 35–38. https://doi.org/10.1017/s0025100300004990

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America*, *75*(1), 224–230. https://doi.org/10.1121/1.390399

Pandharipande, R. (1981). On the nativization of lexicon: The case of Marathi. https://doi.org/10.1515/ling.1981.19.9-10.987

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America*, *89*(6), 2961–2977. https://doi.org/10.1121/1.400734

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

Schertz, J., & Khan, S. (2020). Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu. *Journal of Phonetics*, *81*, 100979. https://doi.org/10.1016/j.wocn.2020.100979

Schwarz, M., Sonderegger, M., & Goad, H. (2019). Realization and representation of Nepali laryngeal contrasts: Voiced aspirates and laryngeal realism. *Journal of Phonetics*, *73*, 113–127. https://doi.org/10.1016/j.wocn.2018.12.007

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

Sinha, A. N., & Rao, G. K. (1984). Release burst and formant transition as acoustic cues for English stop consonants perception by native speakers of Indian languages. *IETE Journal of Research*, *30*(6), 231–233. https://doi.org/10.1080/03772063.1984.11453275

Smith, B. (2013). An acoustic analysis of voicing in American English dental fricatives. *60*, 117–128.

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

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.

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