Lost lips in 10th century Dravidian: An acoustic-perceptual path to debuccalization in Kannada

Abstract

Kannada is unique among South Dravidian languages for its development of *h* from older *p*, whereas languages like Tamil retain it. This sound change is particularly notable because it originated from a bilabial stop, whereas debuccalization to *h* more commonly follows from a spirant. Tuttle (1929) hypothesized that Ka. $p > h$ occurred because of a chain shift initiated by the earlier fortition of Ka. *v* to *b*—with speakers aspirating short-lag *p* to distinguish it from the phonetically similar *b*. In this paper I first examine Tuttle's hypothesis in an artificial word-learning experiment in Tamil, which does not contrast voicing in onsets, showing that Tamil-speaking children do not further aspirate *p* when contrasting a new word with *b*. I then offer an alternative explanation for Kannada debuccalization, arguing that the aerodynamics of short-lag *p* result in a low-intensity burst, mimicking the burst-less *h*, which is misperceived by the listener. In two perception experiments I show that short-lag *p* is confusable with *h* in both clear and noisy listening conditions. Thus, the *seeds* of debuccalization in Kannada may not necessarily be found in phonological pressures in its history, but rather in aerodynamic constraints and their acoustic-perceptual consequences.

Keywords— historical phonology, sound change, debuccalization, burst amplitude, acoustic-perceptual salience, artificial word learning, Kannada, Tamil

1 Introduction

Perhaps the most recognizable feature of the modern Kannada (South Dravidian) lexicon, for speakers of south Indian languages, is the presence of *h* in words where its neighbors, such as Tamil and Malayalam, have *p*:

Ta.	paːl	pɪnnaːl	pire	puː	poː
Ma.	paːl	pɪnnɪl	pira	puːʋu	poːgu
Ka.	haːlu	hɪnd̪e	hire	huːʋu	hoːgu
	'milk'	'behind'	'gourd'	'flower'	'go'

The debuccalization of Kannada *p* (in both word-initial and intervocalic positions)[1] likely began around the 10$^{\text{th}}$ century, as evidenced in the vast and extensive epigraphic record (in Sanskrit, Prakrit, Kannada, and Tamil) of the Kannada-speaking regions of southern India[2]. *H*-forms appear widely distributed geographically (i.e., not limited to a particular area) in prose inscriptions mostly describing land grants and local histories, and reflecting nuances of regional pronunciations, while *p*-forms are retained in poetic verse venerating rulers or presenting official decrees, and tended to be more conservative in form and structure. The 11$^{\text{th}}$-12$^{\text{th}}$ century reflects a period of free variation between *p*- and *h*-forms in the epigraphic record, but by the 14$^{\text{th}}$ century *h* had nearly displaced *p* (Narasimhia, 1941).

While debuccalization is a relatively common phenomenon in the phonological histories many unrelated languages (O'Brien, 2012), the nature of the sound change in Kannada is particularly noteworthy. The *typical* path for obstruents losing their oral place features and changing manner, especially in prosodically strong word-initial position, involves an intervening frication stage with an oral constriction of some sort: either *h* developing from an earlier fricative (e.g., $\chi > h$ in Middle Chinese) (Pulleyblank, 1984) or via aspiration with an intermediate fricative stage (e.g., Proto-Indo-European $*b^h olh_3$-*yom* > Latin *folium* > Occitan *huelha* 'leaf'). Debuccalization in Kannada, however, does not follow either of these paths, developing from an unaspirated bilabial plosive. There are a few examples of debuccalization similar to the Kannada change, but either stem from unattested reconstructions or with an intervening fricative stage, which is assumed due to correspondences in similar languages. For example, Proto-Indo-European **p* is realized as Armenian *h* (Beekes, 2003). Rotuman and Sa'a *h* is a reflex of Proto-Eastern-Oceanic **p*, but correspondences with related Oceanic

[1]Although the debuccalization occurred in both positions, *p* has been reintroduced in most intervocalic cases, leaving the word-initial cases the glaringly obvious reflex of the change in the modern language.

[2]A conservative estimate of the number of Kannada inscriptions is *at least* 10,000 dating from the 5$^{\text{th}}$-15$^{\text{th}}$ century CE and made on granite and copper plates.

languages (e.g., Tongan, Samoan) suggest an intervening f stage ($*p$ > ($*f$) > h)(Greenhill and Clark, 2011). Perhaps the most comparable situation to the Kannada change is found in Japanese, where there is more concrete evidence for the phonological history. Modern Japanese has h corresponding to Old Japanese p. While at first blush this presents a very similar change, ancillary evidence suggests that the phonetic nature of the older form was not accurately captured. Early modern (16$^{\text{th}}$-17$^{\text{th}}$ century) Portuguese transcriptions of p are rendered as *f*, while a 17$^{\text{th}}$ century Korean textbook of Japanese uses the Hangul letter *ph* or *hw* (representing labial frication) (Miyake, 2013).

Kannada presents a unique problem, as we can be fairly certain about the phonetic characteristics of the older (p) sound, which remains unchanged in related and geographically proximal southern Dravidian languages like Tamil, Toda, and Malayalam. The epigraphic record of Kannada, a linguistically rich historical reference for gleaning information about its phonological history, does not reveal a frication stage in the history of p debbuccalization, as in Japanese. Rather Kannada inscriptions move directly from p to h (e.g., *piriya* 'big, older' appears in 950CE, and as *hiriya* in 1060CE)(Narasimhia, 1941). Importantly, the glyph representing the aspirated bilabial stop (p^h) (even in its earliest form, the *Kadamba* script) would have been known and available to scribes who very often rendered Old Kannada (and pre-Old Kannada) alongside Sanskrit and Prakrits (Indo-Aryan), which, unlike Dravidian, have a contrastive voiced/voiceless, unaspirated/aspirated plosive series (Krishnamurti, 2003). If there were an aspiration/frication stage in the history of Kannada p we might have expected the epigraphic record to reveal as much—yet nowhere does the inscriptional evidence show p^h for earlier p. While I am not suggesting that speakers definitively did not fricate/aspirate p prior to the debuccalization change, the evidence from epigraphy and phonological correspondences with very closely related languages calls into question the fortition of p. Despite the lack of historical evidence, however, early 20th-century philologists nonetheless posited an intermediary stage marked by aspiration.

1.1 Extant explanations for the debuccalization

Subbaiya (1909) suggests that Kannada debuccalization was catalyzed by contact with Marathi, an Indo-Aryan language bordering the Kannada-speaking region to the north. Middle Indo-Aryan had, by the 4th-5th century debuccalized older Indo-Aryan *bh-* to *h-*. Subbiya notes that in Kannada (under the influence of Marathi), "*p* seems to have at first become aspirated as *ph* and then changed to *h*." If Subbaiya's hypothesis were correct, we would expect a geographically constrained distribution of the early *h* forms, which is not what the epigraphic record shows.[3] Tuttle (1929) offers a more linguistic explanation for the Kannada change. He describes, what we would now call, a *push-chain* triggered by an earlier sound change in pre-Old Kannada, *v*[4] > *b* (cf. Ta. *va-* > Ka. *ba-* 'come'). Tuttle suggests that the stopping of *v* to *b* exerted phonological pressure on the voiceless unaspirated *p*, which is presumably acoustically similar to *b*. He writes, "In order to make the difference clearer, many persons strengthened *p* to *ph*, which later developt [*sic*] thru *f* to *h*"(Tuttle, 1929, p.154). Both Tuttle (1929) and Subbaiya (1909) assume that Kannada *p* went through an aspiration stage before the debuccalization exhibited in the modern language and Tuttle's speculation remains the only linguistic explanation for the sound change offered in the literature.

In the remainder of this paper I unpack the nature of Kannada debuccalization: first examining Tuttle's hypothesis directly in an artificial word-learning experiment (§3), and then ask whether debuccalization to *h* in word-initial position *necessarily* follows from an

[3]According to Narasimhia (1941), the earliest attestation of *p* debuccalization occurs in an inscription dated to 910 CE from Śriraṅgapatnaṁ, in the southern Kannada-speaking region, far from where Marathi would have been used (*Epigraphia Carnatica, Vol. 3, Seringapatam Taluq, 134*)(Rice, 1896).

[4]Likely a voiced labio-velar [w] or labio-dental [ʋ] rather than a voiced labio-dental fricative (Keane, 2004)

intermediate stage of aspiration/frication. That is, does debuccalization in Kannada result from an accommodation in *production* or are the seeds of the change found elsewhere, such as the inherent ambiguity of the speech signal (Ohala, 1981), in this case the acoustic phonetic properties of unaspirated bilabial plosives (§4). I offer an alternative explanation for the sound change, and suggest that the aerodynamic-acoustic properties of *p* lend themselves to being misperceived by the listener in the direction of *h* (§4.1,4.2).

1.2 Acoustic-perceptual motivations for sound change

The literature has long recognized speech acoustics and listener perception as providing a motivation for certain sound changes (Sweet, 1888, p.17). Sweet identifies *internal* sound changes as being either *organic* (resulting from "tendencies of the organs of speech themselves") or *acoustic* (resulting from "the acoustic quality of the sounds themselves, as when *f* is substituted for þ, by defective imitation"). Ohala (1981) developed this idea further and identified the variety of ways the less-than-ideal transmission channel between the speaker and the listener can lead to sound change. Importantly, Ohala's work centered the focus of sound change on the task of the listener and her unpacking the inherently ambiguous speech signal. Ambiguity in the signal can be introduced by the properties of the vocal tract itself. Ohala argues that an internal motivation for sound change is the listener's failure to reconstruct the speaker's intended speech, which has been distorted by the nature of vocal tract dynamics. For example, the reinterpretation by the listener of small but systematic perturbations of fundamental frequency following voiced consonants, resulting from either aerodynamic pressures or vocal-cord tension differences in voicing (Hombert, Ohala, and Ewan, 1979), is widely accepted as a source of linguistic tone.

Blevins (2004) formalizes sound change that follows from acoustic similarities between actual and perceived utterances as CHANGE (versus other types of sound change: CHANCE and CHOICE). A classic example of acoustic similarity leading to CHANGE is θ to f in languages like English (Cockney), Italian (Veneto), and Arabic (Tunisian, Bahraini), and Athabaskan (Blevins, 2019). Here the similarity of the noise spectra of the two dental fricatives (Jongman, Wayland, and Wong, 2000) leads listeners to reconstruct the unintended utterance. Indeed a number of sound changes in the world's languages are motivated by phonetic processes that leave the acoustic signal susceptible to misperception, especially in conditions that reflect the inherently noisy conditions of the communicative process. For example, the merging of velar nasal onsets with alveolar nasals (e.g., Thao, Malagasy, Hawaiian, Tahitian) is likely the consequence of the acoustic overlap between the nasals in perceptually relevant F2/F3 space (Narayan, 2008). Similarly, the cross-linguistic tendency for velar stops in high-front vowel environments to palatalize have been shown to result from articulatory constraints leaving an impression on the acoustic signal, which is then misperceived by the listener (Chang, Plauché, and Ohala, 2001; Guignard Guion, 1998). Another way listeners' perception might be implicated in sound change is when acoustic cues are reweighted, reflecting variation in the signal. For example, in the historical change from vowel-nasal stop sequences to a distinctive nasalized vowel, covariation in the acoustic signal between vowel nasalization and nasal murmur duration can lead to perceptual variation in how much weight is assigned to coarticulatory information. The change from VN > $\tilde{V}$ results from the reweighting of the consonantal nasal gesture as being vocalic (Beddor, 2009).

I will argue that debuccalization in Kannada aligns closely with these and other examples of sound change, where ambiguity in the speech signal—stemming from natural constraints of the vocal apparatus and articulatory gestures—leads to the misperception of an intended sound as a similar one, which then becomes phonologized in the language.

The following section examines the push chain hypothesis for debuccalization in Kannada, ultimately concluding that it is not supported by synchronic experimental results. I then move on to testing an acoustic-perceptual salience hypothesis. Two experiments lend support to the idea that debuccalization in Kannada stems from listener-driven misperception of an acoustically weak signal.

2 Reconstructing the linguistic milieu of 10th century south India

2.1 Modern Tamil as a proxy for Old Kannada

Deciphering the linguistic landscape of ancient south India, or any ancient language setting for that matter, is a challenging task. The best an experimental historical phonologist can do is approximate the conditions in which a sound change occurred using modern languages that have comparable characteristics to the ancient ones. In this case, modern Indian Tamil was used as a proxy for the stage of Kannada (*halegannada* or Old Kannada) that was spoken when the debuccalization change was first evident in the epigraphic record. Tamil is one of the more conservative South Dravidian languages. It lacks phonemic voicing and retains most of phonological characteristics of the parent, Proto-Dravidian, language (Burrow and Emeneau, 1961; Zvelebil, 1972). The plosive place-of-articulation inventory of modern Tamil is comparable to Old Kannada, with bilabial, dental, retroflex, and velar places of articulation. Unlike modern Tamil, however, Old Kannada contrasted the full complement of voiced and voiceless, unaspirated and aspirated plosives, as well as the glottal fricative, in order to accommodate Indo-Aryan borrowings. Modern Tamil borrows heavily from Sanskrit and English, but unlike Kannada, aspiration and voicing are not represented in the orthography. Sanskrit and English borrowings into Tamil with voiced onset plosives are typically produced with prevoicing (negative or lead voice-onset time)(Lisker and Abramson, 1964), while aspiration disappears altogether. Like Old Kannada, word-initial plosives in Dravidian-origin words exhibit no phonological voicing contrast in modern Tamil (Keane, 2004). While *h* as a phoneme is absent from the oldest Tamil and Proto-Dravidian[5], it exists as a marginal phoneme in modern Tamil in order to accommodate non-Dravidian borrowings. The introduction of *h* likely occurred very early in the linguistic history of the southern India given the extensive and prolonged contact between Dravidian and Indo-Aryan (Sridhar, 1981).

3 Recreating the push-chain hypothesis

Tuttle (1929) suggested that debuccalization of *p* in Kannada was precipitated by the change from *v* to *b* (Fig.1). Although he does not specify *why* this stopping change affected the debuccalization of *p*, we can presume that Tuttle understood that the phonetic character (articulatory and acoustic) of *b* was comparable enough to the unaspirated *p* of Old Kannada to contribute to speakers disambiguating the two. But articulatory similarity is not necessarily the case in languages with a two-way voicing contrast where voiceless is implemented with very little (short-lag) aspiration (e.g., Dutch). Rather than the voiced counterpart being produced with short-lag aspiration as in English /b/, the voiced counterpart to the short-lag voiceless plosive is typically produced with prevoicing. Voiced stops in Dutch that are

[5]Krishnamurti (2003) suggests that a laryngeal fricative **H* to account for certain phonological phenomena described by the earliest Tamil grammarians.

$$\begin{array}{cccc} v \longrightarrow b & & b & b \\ p & p \dashrightarrow & {}^{*}p^{h} \dashrightarrow & h \end{array}$$

Figure 1: The push-chain hypothesis of debuccalization in Kannada.

produced *without* prevoicing are more likely to be misidentified as voiceless (Van Alphen and Smits, 2004), suggesting that Tuttle's presumption that *p* (unaspirated) and *b* (without prevoicing) are potentially confusable is well founded.

An artificial word-learning task was designed to test Tuttle's (1929) (push-chain) hypothesis. Naive listener/speakers of Tamil, in this case children, whose native word-initial plosives are produced with short-lag voice onset, were taught new words with initial voicing, thereby creating a minimal pair contrast. Monolingual Tamil-speaking children were used as listener/speakers for the task because it was thought that their lexicon would not be as influenced by voiced-initial English loan words as adult Tamil speakers. In asking Tamil-speaking children to contrast *p/b*, this experiment served as an approximation, and rudimentary test, of Tuttle's hypothesis that the introduction of *b* forced "many persons to strengthen *p* to *ph*." To ensure participants maximized the *p/b* contrast, they were instructed to speak as though they were teaching new words to Tamil speakers. Further, productions were made in three ambient noise listening conditions as way to induce Lombard effects (Junqua, 1996) in the speakers. The Lombard effect, the phenomenon whereby speakers increase vocal "effort" in the presence of ambient noise, was thought to enhance any laryngeal timing adjustments speakers might make in the minimal pair task.

3.1 Experiment 1: Testing Tuttle (1929)

3.1.1 Methods

Fifteen monolingual Tamil-speaking and reading children (7 girls and 8 boys, aged between 9-12 years) participated in the minimal pair word-production task. The participants were recorded individually by a Tamil-speaking research assistant in quiet schoolhouse rooms in three villages around Madurai, Tamil Nadu, India.[6]

The experiment was divided into three parts. In part A, participants' baseline productions of *p*- and *k*-initial words were recorded. Pictures representing the ten common picturable *p*- and *k*-initial words were assembled and individually shown to the participant on a slide deck and the participant was asked "What is this picture?"[7]. Their single-word response (repeated for a total of two productions per word) was recorded using sound software via a head-mounted microphone. In part B participants were taught to associate a nonsense image[8] with a *b*-initial word (described above)(e.g., "[buli]" presented with a picture of an aardvark) and then asked produce the new *b*-initial word. These recordings served as a baseline for *b* recordings. Part C was the minimal-pair production task. Each nonce *b*-initial word learned in part B was paired with the same nonsense object image again in Part C. Nonce words were paired with nonsense images and the auditory model of the new word was again presented to the participant over headphones three times. Participants were told that they were to teach this new word to a Tamil speaker using the sentence frame: "Don't say [new *b*-initial word], Say [real *p*-initial Tamil word]." The target *b*- and *p*-initial words were

[6]Approved ethics protocols were followed for informed consent. Participants received a pen/pencil/notebook set for their participation.

[7]*k*-initial words were recorded for Experiment 2.

[8]These "nonsense" images were in fact images of objects (flora and fauna) that would be unfamiliar to the pre-teen Tamil-speaking participants.

not displayed using Tamil orthography, but only pictured. Figure 2 shows a sample prompt for the artificial minimal-pair production task.

Figure 2: Sample minimal-pair word production prompt. Participants were instructed to read pictures as words “Do not say PICTURE1, say PICTURE2.” In this example: [buli enru sollaːd̪ɨ, puli enru solluŋgaːl].

Participants recorded each set of minimal pairs three times, once in three different listening conditions. The first production was made with no noise in the participant’s headphones. The second and third productions were made with 70dB and 75dB pink noise (respectively) in the participant’s headphones.

While both voice-onset time and release burst energy in the mid-frequency band (taken as difference in amplitude between F1 of the post-consonantal vowel the maximum amplitude in the F2/F3 band of the release burst)(Stevens, Manuel, and Matthies, 1999) were measured for the initial CV of each target word production, only voice-onset times were used in the analysis of the experimental results.

3.1.2 Speech materials

Speech materials, which served as model *b*-initial words, were recorded in a sound proof booth by a female native speaker of standard South Indian Tamil whose second language was Indian English. The speaker recorded ten common *picturable* Tamil *p*-initial words:[pʌɖəgɨ] “boat,” [paːnai] “pot,” [paːrai] “rock,” [pajːan] “boy,” [pʌləm] “fruit,” [pugai] “smoke,” [puli] “tiger,” [puːnai] “cat,” [pɛɳ] “girl,” [pʌɳəm] “money.” She also recorded *b*-initial Indian English words with the same CV onsets (differing in voicing) as the Tamil words (e.g., Ta. [**pu**li] ‘tiger,’ InEn. [**bu**li] ‘bully’; Ta. [**pʌ**ɳəm] ‘money,’ InEn. [**bʌ**ni] ‘bunny,’ etc.). Ten minimal pairs were created by replacing the initial CV of the Tamil *p*-initial words with the initial CV of the Indian English *b*-initial words. Specifically, two glottal cycles (lead VOT) of closure voicing, release transient, and four vocalic cycles from the *b*-initial Indian English words replaced the release burst (transient and aspiration) and first four vocalic cycles of the of original *p*-initial Tamil words. The Indian English words were naturally produced with lead VOT (different from Canadian and other Englishes where /b-/, when produced in isolation, has a very short positive VOT). Four glottal cycles of the vowel were chosen so as to provide low-frequency microprosodic cues for listeners in order that they hear the new word as acoustically different from Tamil *p*-initial words(Whalen, Abramson, Lisker, and Mody, 1993). In this way, the ten *p*-initial Tamil words had ten *b*-initial counterparts (10 minimal pairs, differing in initial voicing and onset fundamental frequency, e.g., [puli]-[**b**uli])(Fig. 3).

3.1.3 Results

Baseline. Baseline productions of known words(Part A) had voice-onset times consistent with the literature on place-of-articulation effects (e.g., Lisker and Abramson, 1964), with velars being longer than bilabials [Mean $\mathrm{VOT_k}$=0.029s (*SD*=0.01s); Mean $\mathrm{VOT_p}$=0.017s

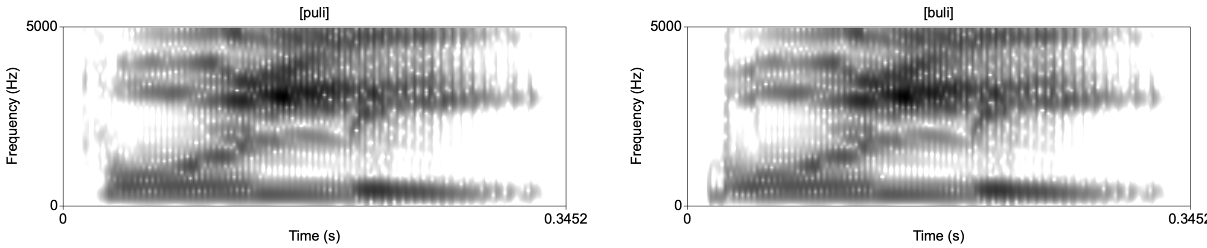

Figure 3: Original [puli] and modified [buli] created by replacing [pu] (Ta.) with [bu] (InEn.). Two pre-release glottal cycles, release, and four glottal cycles of [bu-] from InEn. [buli] 'bully' replaced the onset and first four glottal cycles of the original [puli] token to create the new target word, [buli].

(SD=0.013s)($p < 0.005$)][9]. Baseline productions of b in newly learned words had mostly lead (negative) VOTs (Mean VOT_b=-0.04s, SD=0.014).

Minimal pair production. Figure 4A shows the distribution of raw VOT values for p and b in the minimal pair production task. In order to account for individual variation in VOT production (Allen, Miller, and DeSteno, 2003) a measure representing deviation of individual's VOT in the minimal pair task from their baseline VOT was calculated. An average VOT_p for each participant was calculated from the baseline (Part A) task. This value was subtracted from an individual's p- productions in the minimal pair task to arrive at a "VOT difference" measure which was then used in the analysis. Likewise, the average VOT_b for each participant was calculated from baseline b- productions from Part B and subtracted from b- productions in the minimal pair task to arrive at an individualized VOT difference measure. VOT difference for p and b are shown in Figure 4B.

A mixed effects regression model of VOT difference values as a function of Voicing and Noise (with random intercepts for Speaker and Word) showed no significant of effect of Noise, suggesting that speakers did not adjust their productions to produce Lombard

[9]A mixed effects model, with random intercepts for participant, word, and repetition, were fit for VOT as a function of place of articulation (β =-0.011, SE=0.003 t=-3.94).

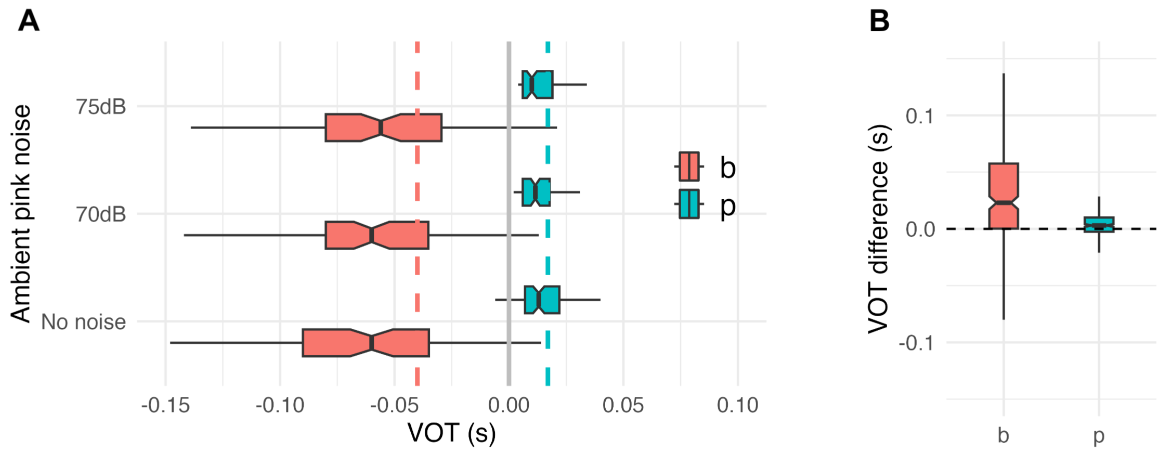

Figure 4: **A**: VOT (represented by notched boxplots) for *p* and *b* in the minimal pair task in three listening conditions, with dashed lines representing the mean VOT in baseline productions. **B**:VOT difference for *b* and *p* productions in the minimal pair task.

speech as listening conditions became more adverse. Removing the NOISE variable from the model showed that the VOT difference for productions of *b* was significantly different from productions of *p* in the minimal pair task (β = -0.04, *SE*= 0.008, *t*= -4.81). This suggests that the magnitude of change from baseline for productions of *b* was significantly greater than for *p*. Figure 4B shows a boxplot of VOT difference for *b* and *p* production in the minimal pair task.

3.2 Discussion

This experiment tested Tuttle's (1929) hypothesis that Tamil speakers strengthened word-initial *p* due to the shift of *v* to *b* in Old Kannada. Tuttle's reasoning suggests that in strengthening *p* via aspiration, speakers maximize the acoustic-perceptual salience of the contrast with *b*. Modern Tamil was used as a stand-in for Old Kannada in the study, as it shares similar voicing features with older South Dravidian languages, particularly the absence of voicing contrasts in word-initial stops. Tamil-speaking children, rather than adults, were recruited for the experiment, as they were thought to be less influenced by voiced word-initial voicing from English borrowings into Tamil. Participants were taught to associate newly-learned *b*-initial words with uncommon objects and instructed to contrast them with common *p*-initial words as if teaching them to a Tamil speaker.

Tamil speakers did not aspirate known *p*-initial words when contrasting them with the *b*-initial words. Rather, Tamil speakers produced longer prevoicing in the minimal pair task than in their baseline (isolated) productions of *b*. Tamil speakers' production of *p* were unchanged even in noisy ambient listening conditions. This result suggests that Tuttle's (1929) hypothesis, that Kannada's debuccalization of *p* results from a push chain initiated by the change of *v* to *b*, may not be phonological event catalyzing the sound change. That speakers increased voicing during consonant closure (prevoicing) rather than increasing aspiration is consistent with the cross-linguistic developmental literature showing a bias in productive mastery against long-lag VOT (Jakobson, 1941; Macken and Barton, 1980). Some research has shown that closure voicing precedes voiceless unaspirated (short-lag) production (e.g., Gierut and Dinnsen, 1986). Further, the phonology literature has detailed the reasons (including positional neutralization, distribution within languages, commonality cross linguistically, etc.) as to why unaspirated stops might be considered an *unmarked* feature in phonological systems (e.g., Lombardi, 2018, but see Vaux and Samuels, 2005).

Another way to view the results of Experiment 1 is through the lens of *dispersion* in sound systems (Liljencrants and Lindblom, 1972), which suggests that contrastive sounds in languages are very often situated in acoustic space in a way that maximizes their perceptual distinctiveness. Short-lag aspiration is constrained in its temporal malleability. Voice-onset time is positively correlated with post-consonantal vowel duration, but in short-lag stops there is a limit to how long aspiration can be. For example, in a study of speech rate in Tamil (which systematically varied according to vowel duration), speakers' VOT for *p* was maximally ~30ms at slowest speech rates (vowel duration ~400ms), while VOT in long-lag p^h in Canadian English fell between 80-150ms at the slowest speech rates (Narayan, 2023). Positively extending VOT for *p* would disrupt the Tamil laryngeal (short-lag) category, but utilizing the negative VOT space maintains the integrity of the basic laryngeal category while accommodating the newly learned one. Tamil speakers' extend prevoicing to maximize the acoustic-perceptual salience of the *b/p* contrast in a way consistent with the predictions of dispersion theory.[10]

Lastly, a possible interpretation of the Experiment 1 results suggests that speakers, under the *focused* condition of producing a minimal pair contrast, are modifying the *new* sound category (extending prevoicing in *b*) instead of the existing category (extending aspiration in *p*)–that is, /p/ for Tamil speakers has phonetic specifications for its implementation, whereas /b/ does not. Such an interpretation weakens the evidence against Tuttle's hypothesis, as one could argue that the artificial nature of the minimal pair task may have led speakers to change the new category but does not necessarily preclude modification of the existing category in more naturalistic settings.

Despite an argument that absence of evidence of *p* fortition in this single artificial minimal pair experiment is not evidence that Old Kannada speakers did not aspirate *p*, a strong reading of Tuttle's (1929) hypothesis lacks definitive support, or at the least suggests that there may be other forces at play in Kannada debuccalization. I now explore an alternative explanation, one that draws on language-internal articulatory and aerodynamic

[10]While the prevoiced/short-lag vs. long-lag VOT is likely perceptually privileged for psychoacoustic reasons(Benkí, 2005), it is here overridden by the the phonological integrity afforded by the prevoiced vs. short-lag contrast for Tamil speakers.

constraints in short-lag VOT stops and their perceptual effects. The explanation capitalizes on the fact that release bursts for *p* have less amplitude (relative to the following vowel) than release bursts at other places of articulation. In Experiment 1 (Part A) children's release bursts for *p* and *k* were measured (Fig. 5) showing this difference. Consistent with previous literature (Narayan, 2023; Stevens et al., 1999) children's production of *p* had release bursts which were ~ 20dB lower than their *k* bursts in the mid frequency (F2/F3) range. In the next section I detail the reasons behind this pattern and follow up with two experiments examining the perceptual consequences for listeners that would provide the seeds for the debuccalization change in Kannada.

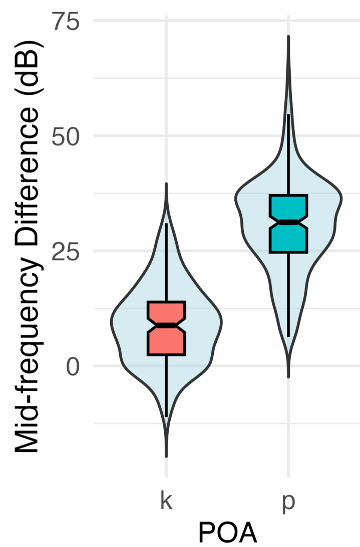

Figure 5: Box- and violin-plots for burst Mid-frequency difference in Tamil-speaking children's' production of word-initial *p* and *k*. Mid-frequency difference (MidDiff) is the difference between the maximum amplitude (dB) in the F2/F3 region in the transient portion the release burst and the amplitude (dB) of F1 in the following vowel. The greater the MidDiff, the lower the amplitude of the mid-frequency region of the release burst (see Narayan, 2023).

4 An aero-acoustic source for Kannada debuccalization

Plosives in initial position in modern Tamil (and other South Dravidian languages other than Kannada) lack contrastive voicing (Krishnamurti, 2003), which is predictable in intervocalic contexts. Word-initial plosives are produced with short-lag voice-onset time (VOT), with well-known variation according to place of articulation (velars having the longest VOT for example). The phonological oro-laryngeal timing requirement results in an aerodynamic scenario different from languages with long-lag VOT/aspirating languages (e.g., Hindi, North American English).

In order for voicing to sustain in CV syllables, there must be a pressure differential between the sub-glottal (P_{sub}) and oral (P_{o}) cavities, causing vocal-fold vibration via the Bernoulli effect (Catford, 1977). Upon release of the oral constriction, positive P_{o}, decreases until equalizing with ambient pressure (P_{amb}). P_{sub} being greater than P_{amb} results in airflow through the glottis affecting voicing. For short-lag stops, like in Tamil, voicing initiation occurs very soon after the release of the oral constriction (~10-25ms) (Narayan, 2023). Therefore, for voicing to occur so quickly in short-lag plosives, P_{o} would be expected to be sufficiently low such that on release of the oral constriction, P_{o} is rapidly equalized with P_{amb}. Indeed this is the case with voiced (short-lag) plosives in English, which have an oro-laryngeal timing comparable to the short-lag (voiceless) plosives, and have a lower P_{o} during closure than their long-lag counterparts (Arkebauer, Hixon, and Hardy, 1967; Malécot, 1970).

A prediction we can make then, given the aerodynamic requirements for short-lag plosives, is that their burst amplitude is lower than that of long-lag plosives. The release burst (transient release and aspiration noise) reflects both P_{o} as well as the flow geometry of the oral channel. Burst amplitude (relative to the following vowel) has been shown to be correlated with the relative timing of the consonant release and the onset of vocalization. Indeed the burst amplitudes of English long-lag stops [p^ha] and [k^ha] were on average 8dB higher than Tamil short-lag [pa] and [ka] in the mid-frequency band (F2/F3) and 7dB higher in the high-frequency band (> 3-3.5kHz) (Narayan, 2023), *i.e.,* long-lag stops are louder than short-lag stops.

Burst amplitude and overall spectral envelope also reflect the *place* of the oral constriction. Spectral prominences, reflecting the resonating cavity anterior the constriction, are in the high frequency band (3.5-7kHz) for alveolars, and the lower band (<3.5kHz) for velars. Bilabials, with no anterior resonating cavity, show no spectral prominences in their release bursts (Stevens et al., 1999). Bilabials have a lower burst amplitude across the spectrum than posterior places of articulation (Narayan, 2023; Stevens et al., 1999). There are a few reasons why this might be the case. Oral-pharyngeal cavity volume during bilabial production (from lips to glottis) is naturally greater than the cavity behind lingually articulated consonants, with the most posterior consonants (velar, uvular) having the smallest cavity volumes. Supposing the volume of air pressurizing the oral cavity is the same in both bilabial and velar stops, we would expect P_o for bilabials to be less than for velars simply due to the former's greater cavity volume (Boyle's law). Further, Maddieson (2013) argues that the large oral cavity behind the lips allows for the "greatest volume of air to flow between the open vocal folds from the lungs into the mouth cavity before that cavity becomes fully pressurized and the flow stops," thereby resulting low pressure burst when the stop is released. Passive expansion of the oral cavity can further contribute to decreased P_o, which may be more readily implemented in bilabials than in velars owing to the larger amount of oral surface area exposed to the impinging air pressure (Ohala and Riordan, 1979). Finally, the posterior portion of the tongue mass moves more slowly in a lowering gesture ([k]) than does the lower lip in an opening gesture ([p]) (Klatt, 1975; Löfqvist and Gracco, 1994,9). The resulting cross-sectional area of the channel created when releasing a lingual consonant is less than that of the opening lips. We would then expect the velocity of pressurized air released with tongue lowering to be greater than the velocity of air escaping the labial orifice (Bernoulli's principle).

Short-lag bilabial stops are therefore prone to lower burst amplitudes for aerodynamic reasons intrinsic to their oro-laryngeal timing and their place of articulation. I propose that the weak release burst of *p* reduces its perceptual salience as an acoustic landmark for place of articulation (Maddieson, 2013; Ohde and Stevens, 1983). However, place perception relies on a constellation of cues including *both* burst energy concentration and spectral characteristics at the onset of vocalism. Especially in non-high vowel contexts, bilabials have shorter F2 transitions than lingually articulated stops (Kewley-Port, 1982). Longer formant transitions are generally discriminated better than shorter transitions (Elliott, Hammer, Scholl, Carrell, and Wasowicz, 1989), suggesting that bilabials might be more susceptible to misperception of the rapid spectral movement in the vocalism. The following experiments test the burst amplitude perception hypothesis. Discrimination (Experiment 2) and confusion tasks (Experiment 3) were performed in varying noise conditions. Masking noise (multi-talker speech babble) was used to determine how salient the release burst is to *p* perception while stressing the vocalic cues to place. It was hypothesized that the effect on perception of the naturally produced low amplitude *p* burst would be most evident when listening conditions were challenged, that is, the broadband transient noise of the bilabial burst would be easily masked by noise while other places of articulation, with their higher amplitude bursts, would be less affected by varying listening conditions.

4.1 Experiment 2: AX discrimination task

4.1.1 Methods and stimuli

The AX (“same-different”) task was conducted online in Tamil Nadu, India. The experiment employed a between-subjects design with three speech noise conditions (15dB, 10dB, and 5dB SNR). During each trial, listeners determined whether two syllables were the same or different by selecting the corresponding on-screen button. All instructions were provided in Tamil.

Speech stimuli were recorded by a male speaker of standard Indian Tamil in a sound-proof room in the Phonetics Studio at XXX University. The speaker recorded multiple tokens of CV syllables (C=[p,t̪,ʈ, k, h], V=[a,i,u]). A single token of each syllable was selected such that all syllables generally matched in duration and overall pitch contour. Table 1 gives the relevant acoustic characteristics of the speech syllable materials: total duration, duration of burst and aspiration, the amplitude of the burst relative to the amplitude of F1 in both the mid- and high-frequency bands (see Narayan, 2023, for measurement details), the fundamental frequency of the vowel immediately (20ms) following the release, and F1 and F2 at the CV transition.

CV stimuli were then fully crossed yielding 120 AX trials with a 400ms interstimulus interval. Multi-talker babble was added to these AX trials, resulting in three groups varying in signal-to-noise ratio: 15dB, 10dB, and 5dB. Multi-talker babble was constructed from a high-quality recording of Tamil speech banter (8-10 males)[11]. A five-second sample was time-reversed in order to remove any word or phrase level information. For reference, figure 6 shows spectrograms of three syllables, [pa], [ka], [ha] with and without multi-talker babble

[11] The multi-talker babble was recorded at a staged (mock) cockfight for a film production in a rural suburb of Chennai, Tamil Nadu, India.

	Total (s)	Asp (s)	MidDiff (dB)	F0 (Hz)	F1 (Hz)	F2 (Hz)
pa	0.372	0.011	32	122	537	1080
t̪a	0.360	0.022	21	124	462	1452
ʈa	0.354	0.021	22	124	334	1523
ka	0.361	0.029	7	125	448	1538
ha	0.395	0.075	29	124	634	1130
pi	0.384	0.022	28	127	281	2233
t̪i	0.421	0.031	15	130	268	2283
ʈi	0.385	0.016	21	131	301	2208
ki	0.422	0.055	15	127	234	2426
hi	0.441	0.053	31	131	202	2358
pu	0.413	0.030	40	130	265	809
t̪u	0.410	0.032	28	131	236	1531
ʈu	0.427	0.016	22	134	270	1548
ku	0.412	0.064	22	137	259	824
hu	0.466	0.074	51	133	207	679

Table 1: Acoustic properties of Tamil CV stimuli: Total duration of the syllable (s), aspiration duration (s), mid-frequency band difference in burst amplitude (dB) (see text), fundamental frequency (F0) of the first 20ms of vowel (Hz), F1 and F2 at the CV transition (Hz).

(10dB SNR).

Sixty four Tamil speakers participated[12] online in the AX task in a between-subjects design: 20 in the 15dB SNR condition, 24 in the 10dB SNR condition and 20 in the 5dB SNR condition.[13] All participants were located in Tamil Nadu, India, and all instructions were provided in Tamil. Participants were instructed to click "same" or "different" buttons on screen after the presentation of each trial. The experimental task proceeded after five practice trials where feedback (correct or incorrect) was given. Participants were given a

[12] All recruiting and testing protocols were approved by XXX University's research ethics board. Participants were paid 300 Indian Rupees.

[13] No attempt was made to control the audio presentation hardware or headphones on the listener side. Participants were told to use headphones on recruitment and in instructions for the online study. Participants had to pass a headphones test, which tested audibility of in- and out-of-phase pure tones, in order to proceed with the study.

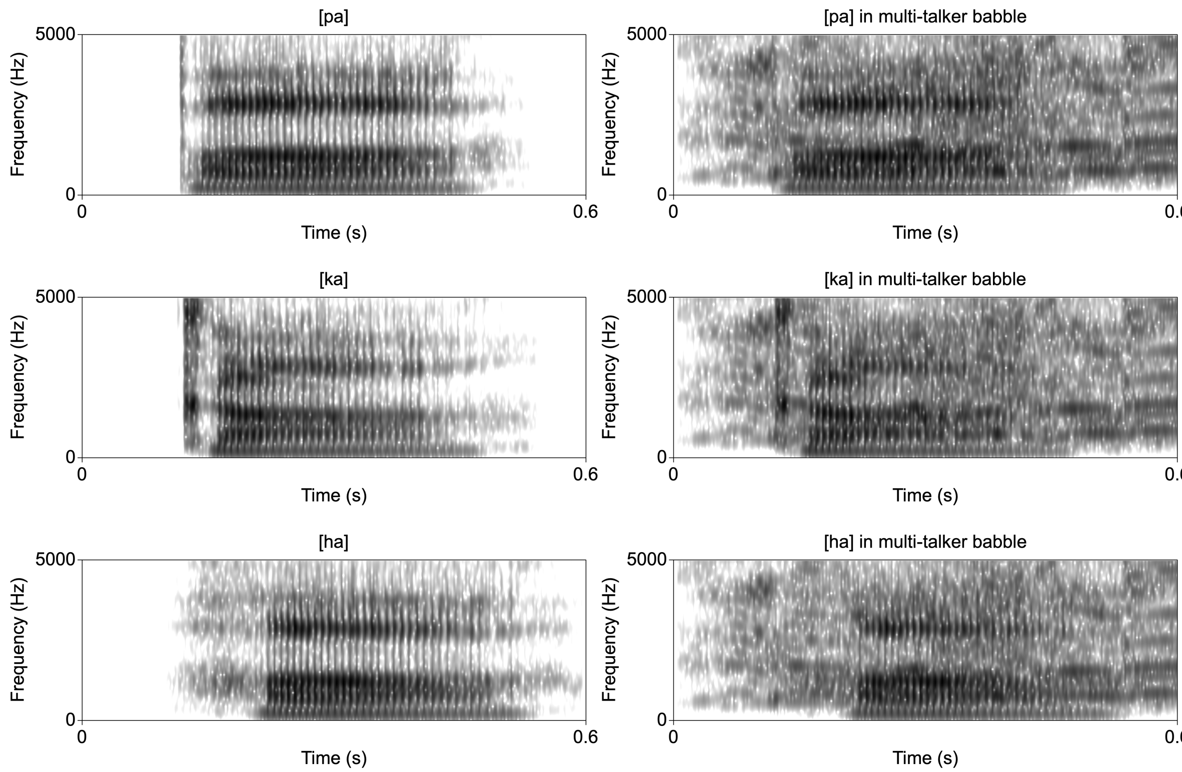

Figure 6: [pa], [ka], and [ha] tokens without multi-talker babble noise (left column), and with 10dB SNR multi-talker babble noise (right column).

self-paced break after the first 60 trials.

4.1.2 Results and discussion

Accuracy on all *different* trials was analyzed. The order of stimuli within an AX trial was not considered, with both orders collapsed into a single Contrast condition (e.g., [pV]-[hV] and [hV]-[pV] trials were coded as *p-h*). Figure 7 shows listeners' discrimination performance in three Noise conditions.

For analysis, the data were subset according to place-of-articulation contrasts (bilabial, dental, retroflex, velar). In order to determine whether accuracy on stop-h contrasts (the

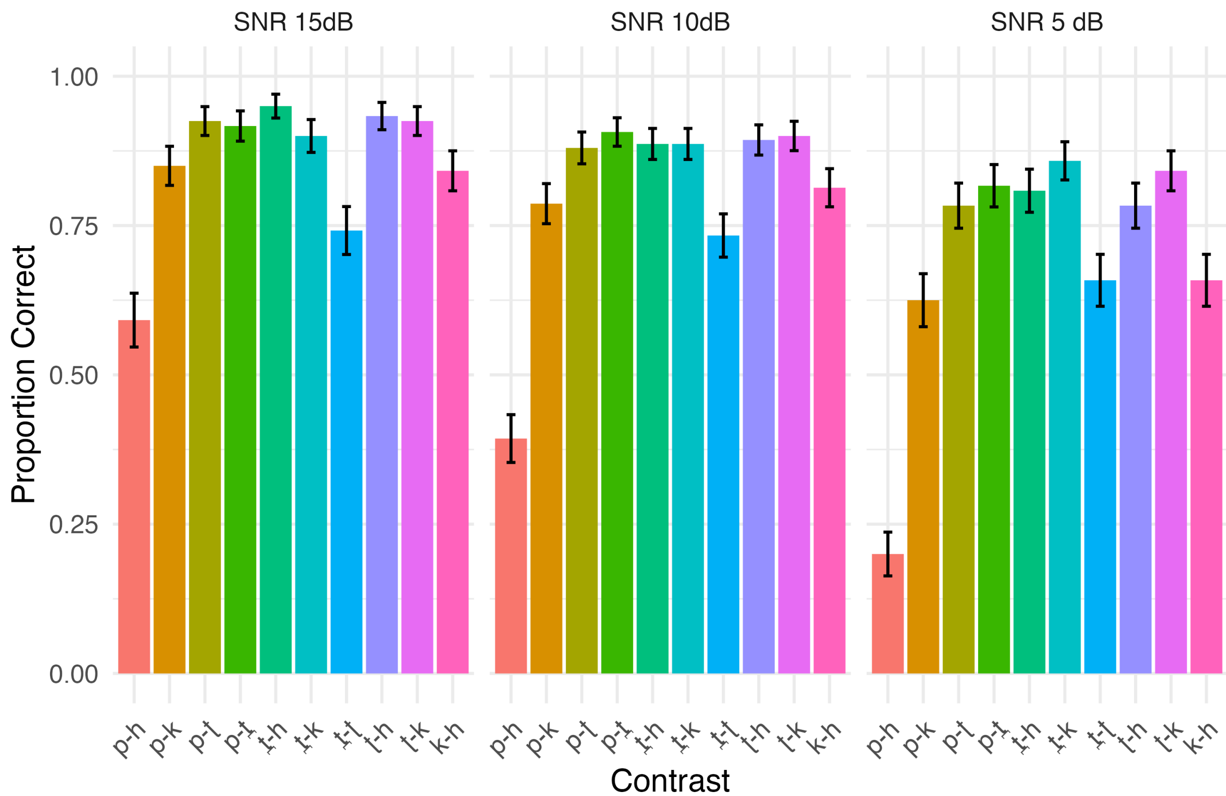

Figure 7: AX discrimination accuracy of all consonant contrasts by Tamil speakers (n=25/condition) in 15dB, 10dB, and 5dB SNR conditions. Each bar represents accuracy in both C_1- C_2 and C_2-C_1 pairs (e.g., "p-h" includes [pV]-[hV] and [hV]-[pV] trials). Error bars represent standard error.

reference condition) differed from stop-stop contrasts, mixed effects models were fit to each place-of-articulation subset.[14] Model statistics are given in Appendices A-D. Each model shows an effect of CONTRAST. Table 2 shows a summary of the main effect of CONTRAST on accuracy.

BILABIAL CONTRASTS. (Appendix A) Accuracy on bilabial contrasts decreased with increasing noise. Accuracy on the *p-h* contrast was significantly lower than other bilabial

[14]Using the `lme4` package, accuracy was modeled as a function of NOISE CONDITION crossed with CONTRAST, with random intercepts for PARTICIPANT and VOWEL context.

Contrast pair

Reference pair	**-p**	**-t̪**	**-ʈ**	**-k**
p-h		p-h < p-t̪***	p-h < p-ʈ***	p-h < p-k***
t̪-h	t̪-h ≈ t̪-p		t̪-h > t̪-ʈ***	t̪-h ≈ t̪-k
ʈ-h	ʈ-h ≈ ʈ-p	ʈ-h > ʈ-t̪***		ʈ-h ≈ ʈ-k
k-h	k-h ≈ k-p	k-h ≈ k-t̪	k-h < k-ʈ*	

Table 2: Summary of main effects of CONTRAST in four mixed effects models of discrimination accuracy between reference pairs (*stop-h*) and *stop-stop* pairs. Model statistics are given in Appendices A-D. $^{***}p < 0.001$; $^{*}p < 0.05$

contrasts. There was no significant interaction between CONTRAST and NOISE for bilabial contrast pairs.

DENTAL CONTRASTS. (Appendix B) Accuracy on dental contrasts was affected by noise only in the SNR 5dB condition. Accuracy on the *t̪-h* contrast was similar to all other t̪-stop contrasts except *t̪-ʈ*, which was significantly lower. There was an interaction between NOISE and CONTRAST suggesting that accuracy on the *t̪-ʈ* contrast was significantly lower in the 5dB SNR condition than in the 15dB SNR condition.

RETROFLEX CONTRASTS. (Appendix C) Accuracy on retroflex contrasts in the 5dB SNRcondition was significantly lower than in the 15dB SNR condition. As with the dental contrasts, accuracy on the *ʈ-h* contrast was similar to other ʈ-stop contrasts except the *t̪-ʈ* contrast, which was significantly lower. Accuracy on the *t̪-ʈ* contrast was significantly lower in the 5dB SNR condition than in the SNR 15dB condition.

VELAR CONTRASTS. (Appendix D) Only the 5dB SNR affected accuracy on the velar contrasts, which was significantly lower than accuracy in the 15dB SNR noise condition. Accuracy on the k-h pair was significantly lower than *k-ʈ* in the 5dB SNR condition than in the other noise conditions.

The pattern of accuracy in the AX task suggests that, consistent with the burst amplitude perception hypothesis of debuccalization, the *p*-*h* contrast is disproportionately affected by multi-talker babble noise relative to *p* in contrast with other places of articulation. Other places of articulation, when in contrast with *h*, showed accuracy comparable to contrasts with other stops, suggesting that perception of the *p*-*h* contrast is different from other contrasts for Tamil speakers. Interestingly, discrimination of the dental-retroflex contrast is less accurate than other contrasts, though not nearly as poor as the *p*-*h* contrast in increasing noise, which falls below chance in 10dB SNR noise condition and below 25% in the 5dB SNR noise condition.

While Experiment 2 demonstrated that listeners' diminished sensitivity to the acoustic differences between *p* and *h* it does not tell us about the possible inception of the debuccalization change. Rather, Experiment 2 provides psychoacoustic evidence that [pV] and [hV] are perceptually similar, but does necessarily reveal whether listeners would identify *p* as *h*. Given that Kannada speakers in the 10th century already had a *h* phoneme via Sanskrit and Prakrit borrowings, it is important to ask how readily listeners would identify [pV] as [hV]. As mentioned earlier, /h/ exists as a *marginal* phoneme in Tamil, with its own grapheme, primarily to accommodate Sanskrit and English loanwords. So in the following identification task, Tamil-speaking adults were again recruited for participation.

4.2 Experiment 3: Consonant confusion

Experiment 3 asked two questions, do listeners identify *p* for *h* (Part A), and if so, is burst amplitude implicated in the misperception of *p* (Part B). The predictions for Experiment 3 are that Tamil-speaking listeners' identification of *p* would follow closely the results of Experiment

2–with *p* being identified as *h* more than other places of articulation. Experiment 3 also asks whether varying vowel context affects the perceptual salience of *p*. F2 transitions into the vowel following the release of a consonant also provide place-of-articulation information. Since labials have low F2 transition onsets, we might expect F2 information to be more acoustically salient for labials followed by [i] (characterized by high F2) than by [u] or [a] (lower F2). In this case, then, confusions with *h* should be lower for [pi] than for [pa] or [pu], a prediction following from the assumption that weak burst information together with weak F2 information results in high error rates. This pattern would further strengthen the interpretation that weak burst information is responsible for misidentification of *p* as *h*. When the burst amplitude of *p* is amplified (Part B), the prediction is that error rates with *h* would be more comparable to other places of articulation.

To the best of my knowledge, the literature on consonant confusions in noise (e.g., Miller and Nicely, 1955 and subsequent) does not include *h* as a response option so the predictions for Experiment 3 are guided solely by discrimination results and the theoretical aero-acoustic motivation detailed in § 4.

4.2.1 Methods and stimuli

Part A was five-alternative forced choice identification task with CV stimuli presented in two blocks. The first block was presented without any added babble noise ("clean") and the second with babble noise at 10dB SNR ("noise"). Part B was a pseudo-replication of Part A differing only in the [pV] stimuli, which had burst noise amplified by 12dB relative to the naturally produced [pV] stimuli in Part A. Each block had 45 trials (five places of articulation [p,t̪,ʈ, k, h], in three vowel contexts [a,i,u], repeated three times).

Fifty Tamil-speaking participants completed the study, 27 in Part A and 25 in Part B.

Participants were instructed in Tamil to press one of the five buttons on screen (each with a Tamil grapheme representing one of the five onset consonants) matching the "beginning sound" of the audio stimuli. Trials advanced when a button press was made. Participants were given a self-paced break after the first (clean) block of 45 trials, after which the second (noise) block commenced. All other details of the online presentation were identical to Experiment 2. Audio stimuli for Part A were the same as in Experiment 2. For Part B the broadband release transient of the burst in all *p* tokens was amplified by 12dB, bringing the overall amplitude of the burst in line with burst amplitudes for the lingually articulated consonants.

4.2.2 Results and discussion

Part A. Figure 8 shows confusion matrices for clean and noise (10dB SNR) conditions across the vowel contexts. Overall accuracy was 79.2% in the clean condition and 57.8% in the noise condition. Between the clean and noise blocks correct identification of retroflex, dental, velar, and glottal places dropped an average of 14%, while correct identification of bilabial place dropped 59%. Approximately 14% (34/243) of *p*-initial tokens in the clean block were identified as *h*, while 50% (121/243) of all *p*-initial tokens were identified as *h* in the noise condition. Interestingly, the addition of babble noise increased the identification of *h* as *p*, though not nearly as dramatic as the identification of *p* as *h*.

The data were next analyzed by following vowel context. Figure 9 shows confusion matrices in clean and noise conditions for each vowel context subset of the data. In the clean condition the pattern of confusions for the individual vowel contexts was comparable to the combined data shown in Figure 8, with the correct identification of the signal except for the retroflex and bilabial, with the latter having most confusions in the [-u] context. The bilabial was most identified as *h* in the [-u] context, followed by [-i]. The patterns of error

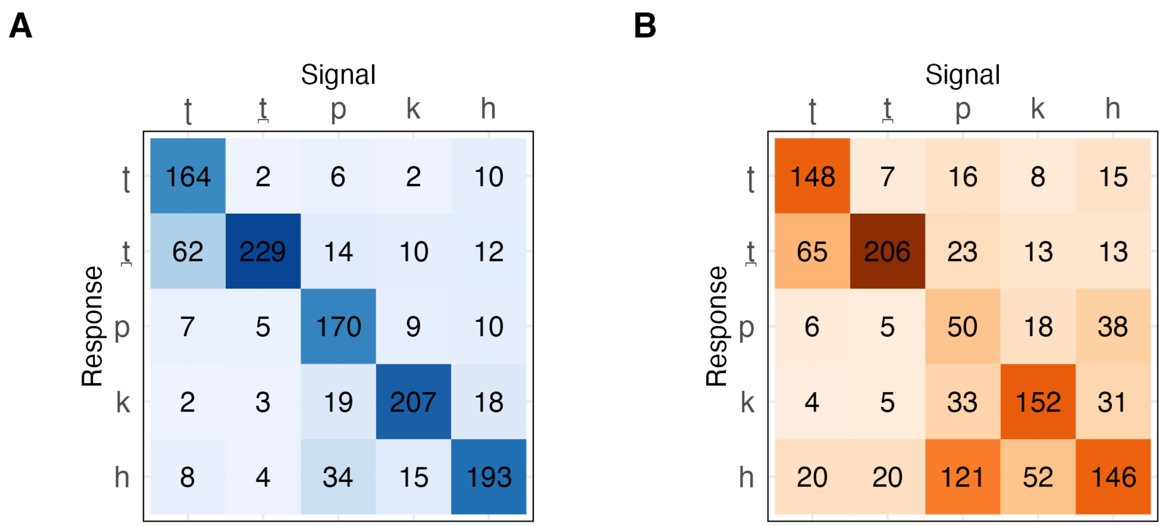

Figure 8: Confusion matrices (count values) for consonant identification tasks clean (A) and noise (B) listening conditions.

by vowel context in the noise condition was similar to the combined data except for the [-u] context where retroflex was less often identified as dental. In the noise conditions *p* was most identified as *h* in [-u] and [-a] contexts (54% and 65%, respectively).

The error rates by vowel context in the noise condition confirm the prediction that *p* would be most often identified as *h* in back vowel contexts due to the similarity in their post-release F2 characteristics. The steeper rise in F2 in high vowel contexts presumably serves as a more reliable acoustic cue to place (than the shallow rise in low vowel contexts) especially when burst characteristics are naturally weak and further masked by noise. Although the epigraphic evidence in Kannada does not reflect a debuccalization of *p* to *h* constrained by vowel context at any stage during the sound change, the pattern of errors by vowel context in Part A suggests that the weak burst characteristics of *p* contribute to its misperception.

Part B further explores the weak burst hypothesis by replicating the methods of Part A for all places of articulation except bilabial stops, which had their transient burst noise

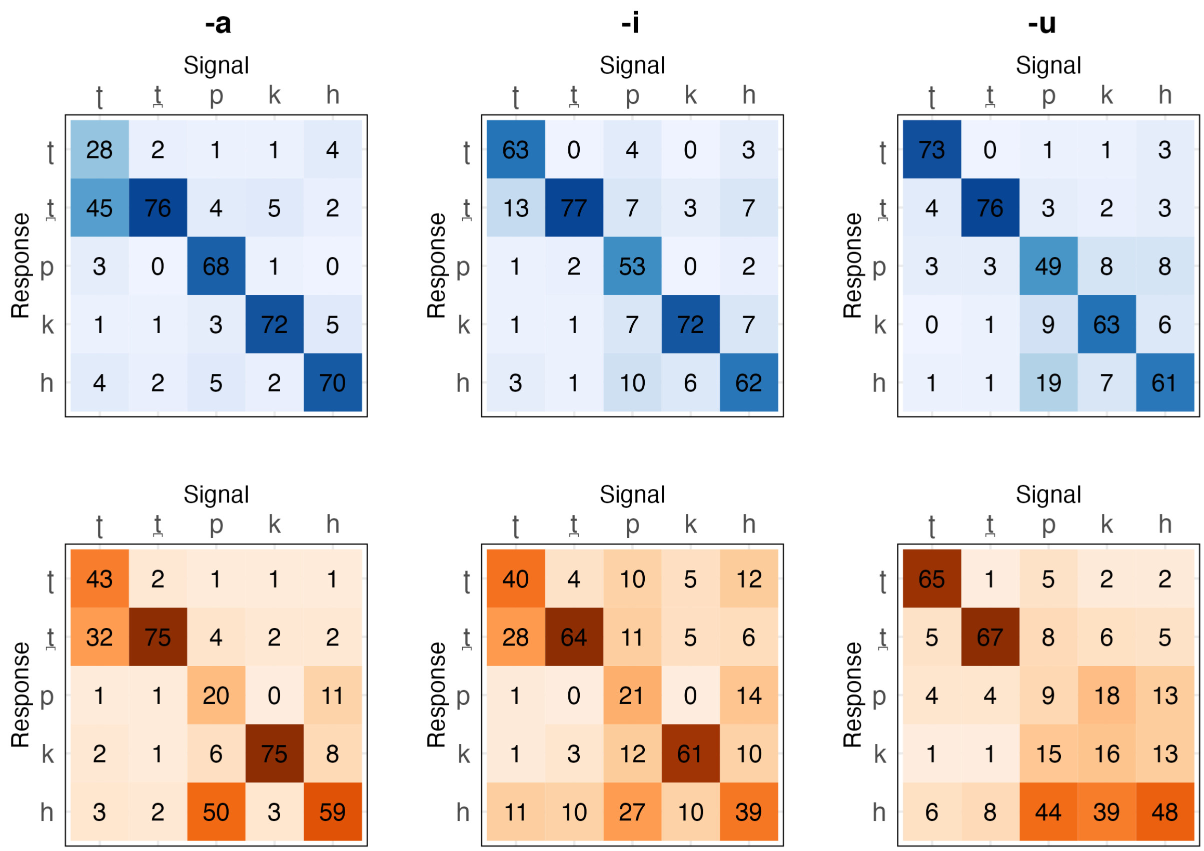

Figure 9: Confusion matrices (count values) for consonant identification tasks clean (top row) and noise (bottom row) conditions as subset according to the following vowel ([-a], [-i], [-u]).

spectra amplified by 12dB.

Part B. Figure 10 shows confusion matrices for consonant identification where *p* tokens had burst noise amplified by 12dB in clean and noise (10dB SNR) conditions across the vowel contexts. Perceptual confusions in Part B follow closely those in Part A, except for *p*-initial syllables, which showed reduced confusion with *h*. Consistent with predictions based on the weak burst hypothesis, in the noise condition, *p* being identified as *h* dropped from 50% in Part A to 13% in Part B (β=3.39, SE=0.7, z=4.82) which is more in line with confusions with other places of articulation.

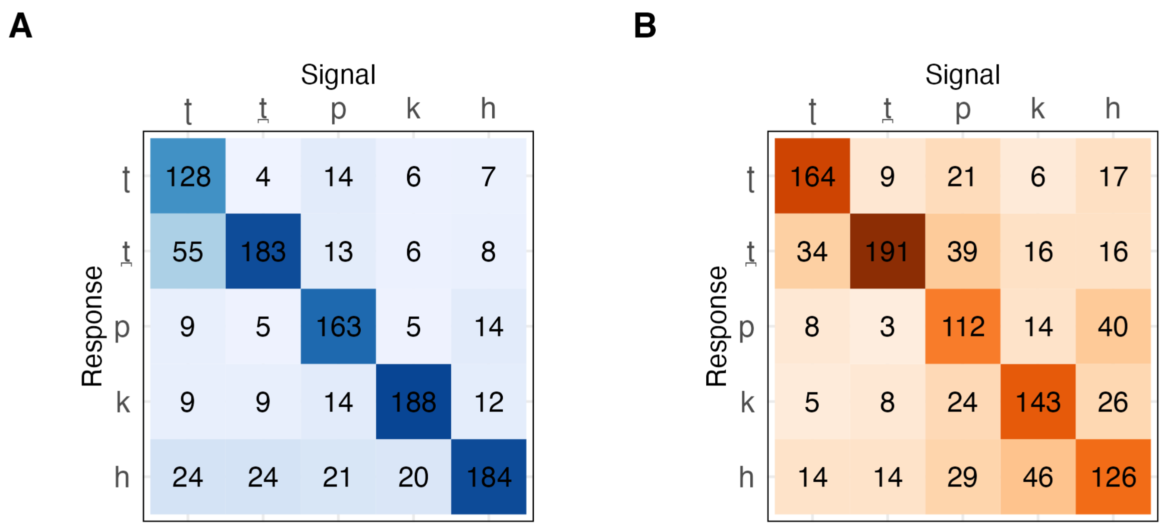

Figure 10: Confusion matrices for Exp.3 Part B (p-burst amplified) tasks in clean (A) and noise (B) listening conditions.

It is likely the case confusions with other places of articulation (such as [t̪] and [k]) increased due to listeners overall unfamiliarity with perceptual discontinuity created by the high burst amplitude p. Crucially, p being misidentified as h was less likely when the p-burst was amplified, thereby implicating the weak p burst as a source of the misperception.

5 General Discussion

The goal of this research program was to critically examine extant theories and offer a new explanation for why p debuccalized to h in Old Kannada, a problem that has remained elusive in the Dravidian linguistics literature. The most plausible theory was proposed by Tuttle(1929) who viewed the change as the consequence of another change in Kannada, $v > b$. Tuttle suggested that this stopping change catalyzed the debuccalization of p via what we now call a "push chain." Tuttle hypothesized that b exerts phonological pressure on p which then strengthens (via aspiration)–fortified $p(>^*p^h)$ then lenited to h in a way comparable to other debuccalization changes in the world's languages that follow from frication. Experiment 1 directly tested this hypothesis using modern Tamil as a proxy for Old Kannada, as they share similar voicing characteristics and phoneme inventory. Monolingual Tamil-speaking children were taught new *b*-initial words and asked to contrast them with common *p*-initial words forming minimal pairs. Speakers in Experiment 1 did not aspirate *p*, but rather extended prevoicing in the *b*-initial words. While not definitive (given that the field/laboratory study was only an approximation of the phonological scenario of 10th C. Kannada) the results of Experiment 1 nonetheless cast doubt on Tuttle's hypothesis.

The study then moved to language-internal acoustic-perceptual dynamics as the potential source of the debuccalization. I argued that short-lag stops, and bilabials in particular, are naturally prone to having quiet release bursts due to their articulatory and aerodynamic constraints, which leads to burst noise becoming a weak perceptual cue to place of articulation. So, while burst noise and F2 transition information together provide the listener place information, *p* can be argued to provide less salient place information than other voiceless oral plosives. I hypothesized that when the noisy oral/aural transmission channel is stressed, the perception of *p* would be disproportionately affected relative to other places of articulation that have more robust bursts. Speech-like babble noise was added to natural CV stimuli in discrimination and identification tasks. With increasing noise in the signal Tamil-speaking listeners' discrimination accuracy of the *p*-*h* contrast decreased precipitously. Likewise, noise dramatically affected identification of [pV] tokens, which were confused with [hV] 50% of the time. In noisy conditions, the effect of vowel context reinforced the weak burst hypothesis. Bilabials, which have low F2 transition onsets, were hypothesized to be better identified (absent the burst cue) in high F2 vowel contexts (front vowels), as the low to high transition would be a better cue to place than when the bilabial is followed by a low F2 back vowel. Consistent with this hypothesis, listeners made twice as many misidentifications of [pu] and [pa] than they did [pi]. When the burst noise of *p* was amplified, listeners' confusion with *h* dropped and was more in line with other places of articulation, thus implicating the burst as a source of the misperception.

One issue that remains unclear, and that will be followed up in future research, is whether the misperception of *p* results solely from listeners missing place characteristics in the burst, or whether there is variation in how listeners weight the acoustic cues to place in the burst in clean conditions was accurate but less than other places of articulation, and discrimination in low-noise conditions only slightly above chance levels. Once the auditory channel is under adverse conditions, however, perception of *p* is dramatically affected. These results can be interpreted as suggesting that listeners weight post-release spectral characteristics in ideal listening conditions more than they do the weak burst cues to place. This weighting then becomes unsuccessful when the listening conditions deteriorate. The interplay between listeners weighting of the burst versus the F2 transitions may shed light on the underlying perceptual mechanisms contributing to the change.

While acoustic-perceptual dynamics have the potential to catalyze debuccalization, we might ask whether the structure of the Old Kannada sound system supported or biased the change. Blevins (2009) argues that pre-existing categories in a language can prime or bias the direction of a sound change. In cases of $\theta > f$, languages such as Cockney English, Veneto Italian, and Tunisian and Bahraini Arabic all had phonemic /f/ in their early stages. But a pre-existing category need not be present in the language for a listener-sourced sound change to occur, e.g., Rotuman (Blevins, 2019). In the case of Old Kannada, speakers would have had a *h* category from extensive contact and borrowing from Sanskrit, so the change of a mostly Dravidian *p*-initial vocabulary to *h*-initial would not have been a pernicious development in to the lexicon.

Finally, we may also take into account extra-linguistic and structural factors that influence articulatory and aerodynamic constraints on burst amplitude. Recently speech rate has been implicated as an actuating factor in certain types of lenition changes in languages (e.g., deletion of intervocalic stops)(Priva and Gleason, 2018). Indeed increased speech rate results in reduced burst amplitudes in Tamil (Narayan, 2023) (though the effect is not as dramatic as with long-lag stops in English). When coupled with results in the developmental literature showing Tamil-speaking mothers have syllable rates in infant-directed speech comparable to their adult-directed rates, which is itself faster than languages like Korean and Tagalog (Narayan and McDermott, 2016), we arrive at an acquisition scenario where the acoustic evidence for *p* to a child's developing sound system reflects weak place characteristics on account of both natural constraints on bilabial short-lag stops as well as fast speech. Future research into the inception of the debuccalization change in Kannada might examine whether Tamil-learning infants discriminate [pV]-[hV] as they do other places of articulation. Failure to discriminate the contrast would suggest acoustic-perceptual fragility plays an important role in the sound change (cf. Narayan, 2013).

6 Conclusions

This paper sought to answer why *p* debuccalized to *h* without evidence of an intervening frication stage in Old Kannada. The results of one production experiment and two perception experiments suggest that 1) speakers of a language with short-lag VOT stops do not readily aspirate when contrasting them with acoustically similar stops, and 2) we can explain the debuccalization of *p* in Kannada by appealing to the natural articulatory and aerodynamic constraints of short-lag bilabial stops and their acoustic consequences. This work suggests that aspiration or frication is not a necessary step for *p* to debuccalize to *h*.

The motivation and results in this paper align closely with the phonetics literature suggesting that certain sound changes originate from listener misperception. As Ohala, Beddor, Blevins, and others have argued, increased likelihood of confusability can result from natural features of the articulatory system which have the potential to be reinterpreted and phonologized by listeners. I have argued that debuccalization in Old Kannada follows this path: constraints on articulation have aerodynamic and acoustic consequences resulting in an acoustic signal that is perceptually equivalent with another sound in the phonology. In this way, bilabial place was *lost* to speakers of Old Kannada and substituted with an essentially placeless consonant–the perceptually equivalent *h*, which was already available in the sound system.

Although this study focused on the phonetic underpinnings of a typologically rare example of debuccalization, I would expect that natural articulatory and aerodynamic constraints would result in the warping of acoustic-perceptual salience in ways that are reflected in other sound changes as well. For example, the directions of change undergone by retroflex sibilants in languages like Taiwanese Mandarin (where it is merging with the dental sibilant) or Hindi (where it has merged with the alveo-palatal sibilant) likely reflects listeners' perceptual weighting of different acoustic characteristics (e.g., center of gravity, formant transition onset) of the sibilant, which is prone to variation in articulation (Luo, 2020; Tiede, Chen, and Whalen, 2019). By closely examining the relationship between articulation, acoustics, and perception, we might arrive at solutions to problems in historical sound change that may have otherwise defied explanation.

7 Appendix A

	Estimate (*SE*)	z
(Intercept)	0.54(0.53)	1.01
$\text{SNR}_{10\text{dB}}$	−1.05(0.48)	-2.20^{*}
$\text{SNR}_{5\text{dB}}$	−2.30(0.52)	-4.45^{***}
$\text{Contrast}_{\text{p-k}}$	1.88(0.37)	5.04^{***}
$\text{Contrast}_{\text{p-ʈ}}$	2.83(0.45)	6.30^{***}
$\text{Contrast}_{\text{p-t̪}}$	2.67(0.44)	6.19^{***}
$\text{SNR}_{10\text{dB}} \times \text{Contrast}_{\text{p-k}}$	0.45(0.48)	0.93
$\text{SNR}_{5\text{dB}} \times \text{Contrast}_{\text{p-k}}$	0.54(0.50)	1.08
$\text{SNR}_{10\text{dB}} \times \text{Contrast}_{\text{p-ʈ}}$	0.33(0.56)	0.59
$\text{SNR}_{5\text{dB}} \times \text{Contrast}_{\text{p-ʈ}}$	0.57(0.58)	0.99
$\text{SNR}_{10\text{dB}} \times \text{Contrast}_{\text{p-t̪}}$	0.79(0.57)	1.40
$\text{SNR}_{5\text{dB}} \times \text{Contrast}_{\text{p-t̪}}$	0.97(0.57)	1.68

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Mixed effects logistic model of AX accuracy of p-C contrasts by Tamil speakers in three listening conditions. Reference conditions are the 15dB SNR Noise condition and the ‘p-h’ Contrast condition.

8 Appendix B

	Estimate (SE)	z
(Intercept)	3.53(0.55)	6.41***
SNR_{10dB}	−0.61(0.64)	−0.95
SNR_{5dB}	−1.79(0.63)	−2.87**
$\text{Contrast}_{\text{t̪-p}}$	−0.59(0.54)	−1.11
$\text{Contrast}_{\text{t̪-k}}$	−0.82(0.52)	−1.56
$\text{Contrast}_{\text{t̪-ʈ}}$	−2.17(0.48)	−4.5***
$\text{SNR}_{\text{10dB}} \times \text{Contrast}_{\text{t̪-p}}$	0.87(0.68)	1.27
$\text{SNR}_{\text{5dB}} \times \text{Contrast}_{\text{t̪-p}}$	0.65(0.64)	1.03
$\text{SNR}_{\text{10dB}} \times \text{Contrast}_{\text{t̪-k}}$	0.82(0.66)	1.24
$\text{SNR}_{\text{5dB}} \times \text{Contrast}_{\text{t̪-k}}$	1.23(0.64)	1.92
$\text{SNR}_{\text{10dB}} \times \text{Contrast}_{\text{t̪-ʈ}}$	0.75(0.60)	1.24
$\text{SNR}_{\text{5dB}} \times \text{Contrast}_{\text{t̪-ʈ}}$	1.25(0.58)	2.16*

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Mixed effects logistic model of AX accuracy of t̪-C contrasts by Tamil speakers in three listening conditions. Reference conditions are the 15dB SNR Noise condition and the ‘t̪-h’ Contrast condition.

9 Appendix C

	Estimate (*SE*)	z
(Intercept)	3.15(0.47)	6.70^{***}
SNR_{10dB}	−0.35(0.59)	−0.57
SNR_{5dB}	−1.60(0.58)	-2.77^{**}
$\text{Contrast}_{\text{ʈ-p}}$	−0.14(0.51)	−0.27
$\text{Contrast}_{\text{ʈ-t̪}}$	−1.8(0.44)	-4.13^{***}
$\text{Contrast}_{\text{ʈ-k}}$	−0.14(0.51)	−0.27
$\text{SNR}_{\text{10dB}} \times \text{Contrast}_{\text{ʈ-p}}$	−0.02(0.64)	−0.04
$\text{SNR}_{\text{5dB}} \times \text{Contrast}_{\text{ʈ-p}}$	0.14(0.61)	0.22
$\text{SNR}_{\text{10dB}} \times \text{Contrast}_{\text{ʈ-t̪}}$	0.41(0.56)	0.73
$\text{SNR}_{\text{5dB}} \times \text{Contrast}_{\text{ʈ-t̪}}$	1.06(0.54)	1.98^{*}
$\text{SNR}_{\text{10dB}} \times \text{Contrast}_{\text{ʈ-k}}$	0.22(0.65)	0.34
$\text{SNR}_{\text{5dB}} \times \text{Contrast}_{\text{ʈ-k}}$	0.57(0.62)	0.93

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 5: Mixed effects logistic model of AX accuracy of t̪-C contrasts by Tamil speakers in three listening conditions. Reference conditions are the 15dB SNR Noise condition and the 't̪-h' Contrast condition.

10 Appendix D

	Estimate (*SE*)	*z*
(Intercept)	2.58(0.85)	3.04^{**}
SNR_{10dB}	−0.17(0.55)	−0.30
SNR_{5dB}	−1.58(0.56)	-2.82^{**}
$\text{Contrast}_{\text{k-p}}$	0.09(0.41)	0.22
$\text{Contrast}_{\text{k-t̪}}$	0.70(0.44)	1.60
$\text{Contrast}_{\text{k-ʈ}}$	1.09(0.47)	2.33^{*}
$\text{SNR}_{\text{10dB}} \times \text{Contrast}_{\text{k-p}}$	−0.33(0.53)	−0.63
$\text{SNR}_{\text{5dB}} \times \text{Contrast}_{\text{k-p}}$	−0.30(0.52)	−0.59
$\text{SNR}_{\text{10dB}} \times \text{Contrast}_{\text{k-t̪}}$	0.11(0.58)	0.19
$\text{SNR}_{\text{5dB}} \times \text{Contrast}_{\text{k-t̪}}$	0.90(0.58)	1.55
$\text{SNR}_{\text{10dB}} \times \text{Contrast}_{\text{k-ʈ}}$	10.10(0.61)	−0.16
$\text{SNR}_{\text{5dB}} \times \text{Contrast}_{\text{k-ʈ}}$	0.33(0.60)	0.56

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 6: Mixed effects logistic model of AX accuracy of k-C contrasts by Tamil speakers in three listening conditions. Reference conditions are the 15dB SNR Noise condition and the 'k-h' Contrast condition.

References

Allen, J. Sean, Joanne L. Miller, and David DeSteno. 2003. "Individual talker differences in voice-onset-time." *The Journal of the Acoustical Society of America* 113:544–552.

Arkebauer, Herbert J, Thomas J Hixon, and James C Hardy. 1967. "Peak intraoral air pressures during speech." *Journal of Speech and Hearing Research* 10:196–208.

Beddor, Patrice Speeter. 2009. "A coarticulatory path to sound change." *Language* pp. 785–821.

Beekes, Robert S.P. 2003. "Historical phonology of Classical Armenian." *Armeniaca: Comparative notes* pp. 133–211.

Benkí, José R. 2005. "Perception of VOT and first formant onset by Spanish and English speakers." In *Proceedings of the 4th International Symposium on Bilingualism*, pp. 240–248. Cascadilla Press Somerville, MA.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge University Press.

Blevins, Juliette. 2009. "Structure-preserving sound change: A look at unstressed vowel syncope in Austronesian." In *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, pp. 33–49. Pacific Linguistics.

Blevins, Juliette. 2019. "Deconstructing markedness in sound change typology: Notes on $\theta >$ f and f $> \theta$." *Perspectives on language structure and language change* pp. 107–122.

Burrow, Thomas and Murray Barnson Emeneau. 1961. *A Dravidian Etymological Dictionary*. Oxford, Clarendon Press.

Catford, J.C. 1977. *Fundamental Problems in Phonetics*. Indiana U. Press, Bloomington.

Chang, Steve S., Madeline C. Plauché, and John J. Ohala. 2001. "Markedness and Consonant Confusion Assymetries." In *The Role of Speech Perception in Phonology*, edited by E. Hume and K. Johnson. Academic Press.

Elliott, Lois L., Michael A. Hammer, Margo E. Scholl, Thomas D. Carrell, and Jan M. Wasowicz. 1989. "Discrimination of rising and falling simulated single-formant frequency transitions: Practice and transition duration effects." *The Journal of the Acoustical Society of America* 86:945–953.

Gierut, Judith A and Daniel A Dinnsen. 1986. "On word-initial voicing: Converging sources of evidence in phonologically disordered speech." *Language and Speech* 29:97–114.

Greenhill, Simon J and Ross Clark. 2011. "POLLEX-Online: The Polynesian lexicon project online." *Oceanic Linguistics* 50:551–559.

Guignard Guion, Susan. 1998. "The role of perception in the sound change of velar palatalization." *Phonetica* 55:18–52.

Hombert, Jean-Marie, John J. Ohala, and William G. Ewan. 1979. "Phonetic explanations for the development of tones." *Language* pp. 37–58.

Jakobson, Roman. 1941. "Kindersprache, Aphasie, und allgemeine Lautgesetze. Uppsala, Almqvist & Wiksell." *Eng. Tr. Child Language, Aphasia and Phonological Universals. The Hague, Mouton* .

Jongman, Allard, Ratree Wayland, and Serena Wong. 2000. "Acoustic characteristics of English fricatives." *The Journal of the Acoustical Society of America* 108:1252–1263.

Junqua, Jean-Claude. 1996. "The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex." *Speech Communication* 20:13–22.

Keane, Elinor. 2004. "Tamil." *Journal of the International Phonetic Association* 34:111–116.

Kewley-Port, Diane. 1982. "Measurement of formant transitions in naturally produced stop consonant–vowel syllables." *The Journal of the Acoustical Society of America* 72:379–389.

Klatt, Dennis H. 1975. "Voice onset time, frication, and aspiration in word-initial consonant clusters." *Journal of Speech and Hearing Research* 18:686–706.

Krishnamurti, Bhadriraju. 2003. *The Dravidian Languages*. Cambridge University Press.

Liljencrants, Johan and Björn Lindblom. 1972. "Numerical simulation of vowel quality systems: The role of perceptual contrast." *Language* pp. 839–862.

Lisker, Leigh and Arthur S. Abramson. 1964. "A cross-language study of voicing in initial stops: Acoustical measurements." *Word* 20:384–422.

Löfqvist, Anders and Vincent L. Gracco. 1994. "Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context." *Phonetica* 51:52–67.

Löfqvist, Anders and Vincent L. Gracco. 1997. "Lip and jaw kinematics in bilabial stop consonant production." *Journal of Speech, Language, and Hearing Research* 40:877–893.

Lombardi, Linda. 2018. *Laryngeal features and laryngeal neutralization*. Routledge.

Luo, Shan. 2020. "Articulatory tongue shape analysis of Mandarin alveolar–retroflex contrast." *The Journal of the Acoustical Society of America* 148:1961–1977.

Macken, Marlys A and David Barton. 1980. "The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants." *Journal of Child Language* 7:41–74.

Maddieson, Ian. 2013. "Voicing and Gaps in Plosive Systems (v2020.4)." In *The World Atlas of Language Structures Online*, edited by Matthew S. Dryer and Martin Haspelmath. Zenodo.

Malécot, André. 1970. "The lenis-fortis opposition: its physiological parameters." *The Journal of the Acoustical Society of America* 47:1588–1592.

Miller, George A and Patricia E Nicely. 1955. "An analysis of perceptual confusions among some English consonants." *The Journal of the Acoustical Society of America* 27:338–352.

Miyake, Marc Hideo. 2013. *Old Japanese: A phonetic reconstruction*. Routledge.

Narasimhia, A.N. 1941. *A Grammar of the Oldest Kanarese Inscriptions*. Number 1 in Studies in Dravidian Philology. University of Mysore.

Narayan, Chandan R. 2008. "The acoustic–perceptual salience of nasal place contrasts." *Journal of Phonetics* 36:191–217.

Narayan, Chandan R. 2013. "Developmental perspectives on phonological typology and sound change." In *Origins of Sound Change: Approaches to phonologization*, edited by A.C.L. Yu, pp. 128–146. Oxford University Press Oxford.

Narayan, Chandan R. 2023. "Speaking rate, oro-laryngeal timing, and place of articulation effects on burst amplitude: Evidence From English and Tamil." *Language and Speech* 66:851–869.

Narayan, Chandan R. and Lily C. McDermott. 2016. "Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations." *The Journal of the Acoustical Society of America* 139:1272–1281.

O'Brien, Jeremy. 2012. *An experimental approach to debuccalization and supplementary gestures*. Ph.D. thesis, University of California, Santa Cruz.

Ohala, John J. 1981. "The listener as a source of sound change." *Parasession on Language and Behaviour, Chicago Linguistic Society* .

Ohala, John J. and Carol J. Riordan. 1979. "Passive vocal tract enlargement during voiced stops." *The Journal of the Acoustical Society of America* 65:S23–S23.

Ohde, Ralph N. and Kenneth N. Stevens. 1983. "Effect of burst amplitude on the perception of stop consonant place of articulation." *The Journal of the Acoustical Society of America* 74:706–714.

Priva, Uriel Cohen and Emily Gleason. 2018. "The role of fast speech in sound change." In *CogSci*.

Pulleyblank, Edwin G. 1984. *Middle Chinese: A study in historical phonology*. UBC Press.

Rice, B. Lewis (ed.). 1896. *Epigraphia Carnatica*, volume 3 of *Epigraphia Indica*. Bangalore: Mysore Archaeological Department. Reprint: New Delhi, Archaeological Survey of India, 1987.

Sridhar, S.N. 1981. "Linguistic convergence: Indo-Aryanization of Dravidian languages." *Lingua* 53:199–220.

Stevens, Kenneth N., Sharon Y. Manuel, and Melanie Matthies. 1999. "Revisiting place of articulation measures for stop consonants: Implications for models of consonant production." In *Proceedings of the International Congress of Phonetic Sciences*, pp. 1117–1120.

Subbaiya, K.V. 1909. "A primer of Dravidian phonology." *The Indian Antiquary* 38:188–200.

Sweet, Henry. 1888. *A history of English sounds from the earliest period: with full word-lists*, volume 11. Clarendon Press.

Tiede, Mark, Wei-rong Chen, and Douglas H Whalen. 2019. "Taiwanese Mandarin sibilant contrasts investigated using coregistered EMA and ultrasound." *Proceedings of ICPhS 2019* pp. 427–431.

Tuttle, Edwin H. 1929. "Dravidian researches." *The American Journal of Philology* 50:138–155.

Van Alphen, Petra M. and Roel Smits. 2004. "Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing." *Journal of Phonetics* 32:455–491.

Vaux, Bert and Bridget Samuels. 2005. “Laryngeal markedness and aspiration.” *Phonology* 22:395–436.

Whalen, Douglas H., Arthur S. Abramson, Leigh Lisker, and Maria Mody. 1993. “F0 gives voicing information even with unambiguous voice onset times.” *The Journal of the Acoustical Society of America* 93:2152–2159.

Zvelebil, Kamil V. 1972. “Initial plosives in Dravidian.” *Lingua* 30:216–226.