

RESEARCH

Bilingual phonology in dichotic perception: A case study of Malayalam and English voicing

Sayantan Mandal¹, Catherine T. Best², Jason Shaw³ and Anne Cutler^{2,4,5}¹ Concordia University, Montreal, CA² The MARCS Institute, Western Sydney University, AU³ Department of Linguistics, Yale University, New Haven, US⁴ Max Planck Institute for Psycholinguistics, Nijmegen, NL⁵ ARC Centre of Excellence for the Dynamics of Language, AUCorresponding author: Sayantan Mandal (sayantan.mandal@mail.concordia.ca)

Listeners often experience cocktail-party situations, encountering multiple ongoing conversations while tracking just one. Capturing the words spoken under such conditions requires selective attention and processing, which involves using phonetic details to discern phonological structure. How do bilinguals accomplish this in L1-L2 competition? We addressed that question using a dichotic listening task with fluent Malayalam-English bilinguals, in which they were presented with synchronized nonce words, one in each language in separate ears, with competing onsets of a labial stop (Malayalam) and a labial fricative (English), both voiced or both voiceless. They were required to attend to the Malayalam or the English item, in separate blocks, and report the initial consonant they heard. We found that perceptual intrusions from the unattended to the attended language were influenced by voicing, with more intrusions on voiced than voiceless trials. This result supports our proposal for the feature specification of consonants in Malayalam-English bilinguals, which makes use of privative features, underspecification and the “standard approach” to laryngeal features, as against “laryngeal realism”. Given this representational account, we observe that intrusions result from phonetic properties in the unattended signal being assimilated to the closest matching phonological category in the attended language, and are more likely for segments with a greater number of phonological feature specifications.

Keywords: dichotic listening; speech perception; voicing; distinctive feature theory; underspecification; perceptual assimilation

1 Introduction

Listening to speech in the real world involves continuous detection and processing of speech signals in a situation where conditions are not ideal. Rarely, if ever, do listeners receive speech signals in isolation, and the environment almost inevitably contains noise as well as speech from other people, both of which are possibly irrelevant to the attended speech stream. Competing speech signals vie for the listener’s attention, so that the listener’s task includes isolating a single target utterance as well as locating meaningful linguistic units, processing them for a meaningful message and interpreting the total message to arrive at its semantic content. In that process, both *linguistic knowledge*, and the representations of that subset of linguistic knowledge that help listeners abstract semantic concepts from speech signals, i.e. *phonological knowledge*, are called upon. For bilingual listeners, the task is further complicated by the availability of more than one set of linguistic/phonological knowledge, especially given the evidence that speech input activates for such listeners the linguistic units of each language (Grosjean 1988; Spivey & Marian

1999; Weber & Cutler 2004). For all listeners, processing speech requires a combination of both linguistic and non-linguistic mental/cognitive and motor functions; for bilingual listeners the linguistic functions are potentially increased to encompass the phonological grammars of two languages. It is only by a combination of all these functions that active speech perception, and then comprehension, is made possible.

One function that listeners must use, in order to limit the amount of data forwarded for processing, is *selective attention*, i.e., focusing on particular stimuli while ignoring others. Cherry's (1953) landmark study of selective attention in speech examined how listeners track certain conversations while tuning others out (now known as the "cocktail party effect"). In his experiments, two auditory messages were presented simultaneously, one to each ear (i.e., dichotically), and participants were asked to attend to and repeat back one of them. The monolingual English participants were able to do this easily, but most interestingly, when asked about the content of the other message, they were unable to say anything about it. Cherry found that even when contents of the unattended message were suddenly switched (such as changing from English to German mid- message, or suddenly playing backwards) very few participants noticed. If the speaker of the unattended message switched from male to female (or vice versa), however, or if the unattended message was swapped with a 400-Hz tone, the change was always noticed. Cherry's findings have been often replicated, with similar results holding for lists of words and musical melodies. The task, now called dichotic listening, was widely adopted in studies of brain lateralization in language processing. A right ear advantage is typically found for dichotic perception of consonants, which is interpreted as reflecting a left brain hemisphere superiority in phonological processing (Kimura 1961a; b; 1967; Liberman et al. 1967). That right ear advantage is more reliable in right than left handers.

In this study, we investigated the interaction among the general cognitive ability of selective attention, physical properties of the speech signal, and the linguistic functions involved in phonological decisions. Our participants were bilinguals with two phonologically quite different languages, allowing us to examine the ability to selectively attend to one set of phonological knowledge rather than another, while our task was modeled on Cherry's classic dichotic selective attention paradigm and its use in studies of hemispheric asymmetries in speech perception. Specifically, we explored how competing speech signals are mapped onto constitutive linguistic knowledge, using dichotic listening. Our experiments were conducted using second language (L2) dominant early sequential bilinguals whose first language (L1) is Malayalam and whose L2 is Australian English, and we tested their perception of consonants carefully selected for their phonological properties.

In what follows, we describe this study and its outcome. Section 2 provides background information on the Malayalam and English consonants of interest and their respective phonetic and phonological properties. This section outlines a feature-based proposal for the consonants. Sections 3 and 4 report the experimental design and the results, respectively. Section 5 provides a discussion of the results and Section 6 briefly concludes.

2 English and Malayalam target consonants

Our experimental design focused on the labial stops and fricatives (obstruents) of Malayalam and English. Malayalam, sometimes referred to as Kairali, is an Indic language of the Dravidian family, speculated to be philologically related to 6th Century Sen-Tamil (Middle-Tamil) (Asher 1985). For labial obstruents, Malayalam phonology employs a four-way voicing-aspiration contrast with stops, but it lacks fricatives at the labial place of articulation. English is a Germanic language that has a two-way laryngeal contrast between its two labial stops /p, b/ and its two labio-dental fricatives /f, v/. Table 1 illustrates the

Table 1: Distribution of stops and fricatives in English and Malayalam. Gray cells in a given language indicate phonological gaps in that language.

Labial Obstruents			English	Malayalam
STOPS	Voiceless	Unaspirated	/p/	/p/
		Aspirated	*	/p ^h /
	Voiced	Unaspirated	/b/	/b/
		Aspirated		/b ^h /
FRICATIVES	Voiceless		/f/	
	Voiced		/v/	

complementary gaps in the phonological inventories of English and Malayalam labials. English has fricatives, Malayalam lacks them; Malayalam makes full use of a 4-way stop voicing × aspiration contrast, English uses only a 2-way laryngeal contrast.

For the purposes of description, we have characterized English in Table 1 according to what Hall (2001) calls the “standard approach” to laryngeal features. This approach maintains that the phonological features capturing two-way laryngeal contrasts are the same in languages that realize the contrast primarily in terms of voicing, e.g., /p/ vs. /b/, and those that realize it primarily in terms of aspiration, e.g., /p/ vs. /p^h/. The standard approach goes back to at least to Lisker & Abramson (1964), is adopted by Chomsky & Halle (1968), and is explicitly argued for by Keating (1984) and Lombardi (2018). There is also a substantial phonological tradition that treats voicing distinctions in aspiration languages as featurally distinct from voicing distinctions in true voicing languages. Hall (2001) dates this tradition back to Jakobson (1949). In more contemporary studies, it has been referred to as “laryngeal realism” (Honeybone 2005) and has been considered justifiable on both phonetic and phonological grounds by a number of researchers (Iverson & Salmons 1995; 1999; 2003; Jessen & Ringen 2002; Kehrein & Golston 2004; Petrova et al. 2006; Iverson & Ahn 2007; Kager et al. 2007; Beckman et al. 2013). A key aspect of our interest in Malayalam-English bilinguals is that in contrast to the substantial debate over the voicing feature in Germanic languages, such as English, as opposed to Romance languages with true voicing contrasts, laryngeal specification in Malayalam is rather uncontroversial, since features for voicing and aspiration ([spread glottis]) are fully crossed. We return to the issue of laryngeal phonology in our discussion, in light of our dichotic listening results.

One motivation for investigating the interaction of phonetic features of stops and fricatives by this population is prompted by the fact that native Malayalam speakers often replace the labiodental fricatives (which don’t exist in Malayalam) of English loan-words with one of the aspirated Malayalam bilabial stops. Adaptations such as pronouncing the English word ‘Venice’ [venis] as [b^henis] are quite common for these speakers. This suggests that at some level of representation /v/ and /b^h/ are in correspondence. Our experiment investigates whether bilingual listeners of Malayalam (native language, or L1) and English (second language, or L2) are able to switch attention from one language to the other in a dichotic task, ignoring stimuli presented in the unattended language (opposite ear). To the extent that they are unable to ignore the distractor stimuli, we ask how dichotic presentations of stimuli will be mapped onto phonological representations in each language.

We anticipate that both phonetic and phonological similarity of the consonants in Table 1 may play a role in intrusions. Acoustic distinctions between voiced and voiceless stops

and fricatives can be captured by the measure we focus on in the present study, the ratio of periodic (harmonic) to aperiodic noise, also referred to as *harmonics to noise ratio* (HNR), as measured during the consonantal period. Voiced fricatives have a higher HNR value (expressed in dB) than voiceless fricatives. For stops, the ratio should again be higher for voiced than voiceless ones. However, because stops involve much more rapidly changing acoustic properties within a shorter time window than fricatives (brief release burst followed by pre-vocalic voicing, silence or aspiration) the magnitude of the HNR difference between voiced and voiceless stops is smaller than seen in fricative voicing contrasts. Moreover, aspiration, which involves turbulent airflow, would also be expected to affect the ratio, thus lowering HNR scores for aspirated relative to unaspirated stops. In the following section, we present HNR comparisons for the stops and fricatives used in our experiment; the comparisons confirm the coarse-grained expectations stated here.

To make concrete how phonological factors may also structure patterns of intrusions, we have constructed a feature-based proposal for our bilingual population. The proposal makes use of privative feature theory and underspecification (Archangeli 1988; Steriade 1995; Golston 1996; Lahiri & Reetz 2010). This approach aids us in relating the phonological notion of markedness to our experimental data, as we can define *markedness* as the number of phonological features required to represent a speech sound. Unmarked sounds tend to be typologically more frequent in the world's languages. Specifically, the presence of a marked sound in a language tends to imply the presence of an unmarked counterpart. This is true of voiced and voiceless stops. The presence of voiced stops in the inventory of a language almost always implies the presence of voiceless stops, but not necessarily the other way around (Maddieson 1984). Unmarked sounds are also sometimes argued to be either perceptually salient and/or articulatorily simple (Kenstowicz 1994), although empirical evidence supporting this claim is incomplete. If unmarked sounds are perceptually more salient and thus easier to process, then one would expect that such ease of processing would make the unmarked voiceless stops in our experiments harder to ignore, leading to more intrusions when they occur in the unattended language (and ear).

On the other hand, unmarked sounds have, on our proposal, fewer phonological features, as we have assumed that predictable information is unspecified. Information can be considered predictable in one of two ways. First, information is predictable if it is redundant or allophonic: the aspiration of the bilabial stop at onset of the word *pan* /pæn/ in English is predictable (because in stress-initial syllable position, voiceless stops are always aspirated in English). Second, this predictability means it is possible to leave one value of the feature blank in underlying representations. If the feature is not specified, e.g. [+F], then it must be [-F], by default, i.e., a *privative* feature. The approach that defines predictability in this latter fashion is known as radical underspecification, and it makes a different prediction for our experiment than that made by the assumption that unmarked features are perceptually salient. Specifically, it could be that we observe more intrusions from segments that have a greater number of feature specifications.

Our feature-based proposal, shown in Table 2, makes use of three features to differentiate the labial stops and fricatives relevant to this study. The Malayalam voiceless stop /p/ is underspecified for all three features, which is indicated by gray shading. In our experimental design this Malayalam consonant is paired with its dichotic competitor, the English voiceless fricative /f/, which is specified for one feature ([continuant]). In a similar fashion, in the voiced stimulus pairing, we find that the Malayalam voiced stop /b/ bears a single specification ([voice]), whereas its competitor, the English voiced fricative /v/, bears two specifications ([voice] and [continuant]), the latter again being the feature that differentiates the English from the Malayalam item.

Table 2: Privative feature matrix contrasting the stimuli-initial consonants. Specified features are indicated by \checkmark while gray cells marked indicate underspecified features.

Phoneme	Privative Features		
	[voice]	[sp. glottis]	[continuant]
/p/			
/p ^h /		\checkmark	
/b/	\checkmark		
/b ^h /	\checkmark	\checkmark	
/f/			\checkmark
/v/	\checkmark		\checkmark

The proposal follows from privative feature matrices and radical underspecification in the following way. The voiceless stops lack voicing and are assumed to be the default specification for stops; hence they do not require an underlying specification for [voice]. The feature [continuant] does not apply to stops, and [spread glottis] is, likewise, not an essential for unaspirated stop articulation, but is only required to express an aspiration contrast (where applicable) between voice-matched stops. Further, the voiceless fricative /f/ lacks specifications for [voice] and [spread glottis], and only requires [continuant], which captures the turbulence of continuous air-flow through a narrow constriction somewhere in the vocal tract that characterizes a fricative. This phoneme contrasts with the voiced fricative only in respect of the latter's specification for the feature [voice].

By making explicit our phonological proposal for contrasts within our bilingual population, we can evaluate how phonological specification may relate to patterns of intrusion in the dichotic selective attention task. Specifically, we investigate bilinguals' ability to attend selectively to one of their languages, while attempting to tune out the other, asking whether such selective attention tasks are marked by significant numbers of intrusions of phonetic properties from the simultaneously presented item in the unattended language. If intrusions are indeed observed, we ask whether they can be explained by acoustic properties, e.g., HNR, and/or how phonological features are distributed across the bilingual representational space.

3 Experiment

We configured our methodological choices so as to enable us to create an environment where listeners receive simultaneous auditory input from both of their languages, in separate ears, but are required to attend to only one language and attempt to ignore the other language.

3.1 Stimuli

The audio target and response choice stimuli, displayed in Table 3, were 'CVCV nonce words with the phonological properties of Malayalam and English but meaningless in both languages. They were designed with attention to syllable structure and consonant realizations in stress-initial positions of the two languages (see Table 1 above). For English there were two labio-dental fricatives contrasting in [voice], /f/ and /v/; for Malayalam, there were two bilabial stops contrasting in [voice], /p/ and /b/. In the dichotic trials, nonce words from English always began with one of the two fricatives /f, v/ in the initial stressed syllable, while nonce words from Malayalam always began with one of the two unaspirated stops /p, b/ in the initial stressed syllable. While the participants in the study only ever heard Malayalam unaspirated stops /p, b/, and English fricatives /f, v/,

Table 3: The dichotic nonce stimuli in each language, in which the stress-initial consonants are the target and distractor items in each respective language-attend condition.

English	Malayalam
[ˈfata]	[ˈpata]
[ˈvata]	[ˈbata]

we also recorded and calculated the HNR of every labial phoneme from Malayalam and English that were provided as a target response choice (as we explain below, there were also coronal response choices included as distractors; these however are irrelevant for the HNR analysis).

To record our stimulus materials, we recruited a Malayali-Australian male bilingual (age: 27 years) from the Sydney Malayali-Australian community, who produced all stimuli for the experiment including those with the English labial stops and the Malayalam aspirated stops, which were not used in the dichotic selective attention experiment but were measured for HNR comparisons. He was born in Australia to Malayalam-speaking parents, and acquired Malayalam as his L1 in the home within the first few years of life. However, all of his formal education was in Australia in English, thus he was a fluent speaker of Australian English, which had become his dominant language (L2-dominant), although he used Malayalam regularly and remained fluent in his L1 as well. We recorded him at MARCS Institute, Western Sydney University in the anechoic chamber using a Roland UA 25-EX sound card on a Lenovo Thinkpad laptop running Windows 7. He produced 10 or more tokens of each of the eight nonce stimulus types (the four Malayalam stops; the two English stops and two English fricatives) in citation form, with a constant intonation contour. We selected the 8 tokens of each category that were best-matched in duration, loudness, and pitch to use as the audio stimuli in the dichotic task (described in Procedure), as well as in the HNR analyses we conducted (described next).

In order to ascertain the acoustic phonetic nature of the phonological differences represented in our stimulus materials, we derived HNR scores for all tokens of each labial phoneme presented either as auditory stimuli in the dichotic task (English /f, v/; Malayalam /p, b/) or as response choices in the task (English /p, b, f, v/; Malayalam /p, p^h, b, b^h/). For the stops, we calculated HNR over the temporal window starting with stop-release and ending at vowel onset. For the pre-voiced Malayalam /b/, we excluded the pre-voicing temporal window from the measure. For the fricatives we measured the entire duration, starting with beginning of frication and ending at vowel onset.

A $2 \times 2 \times 2$ repeated-measures ANOVA with Language (English and Malayalam), Voicing (voiced and voiceless) and Turbulence was conducted on the HNR scores. The Turbulence factor refers to whether there is a narrow articulatory constriction at some location in the vocal tract that results in airflow turbulence and acoustic noisiness (i.e., fricatives and aspirated stops), or whether vocal tract lacks such a constriction and the articulation thus lacks turbulence (unaspirated stops). The HNR ANOVA revealed significant effects of Language (higher HNR in English than Malayalam), $F(1, 7) = 5182.55$, $p < 0.01$; Turbulence (higher HNR for fricatives and aspirated stops than for unaspirated stops), $F(1,7) = 2937.52$, $p < 0.01$; and Voicing (higher HNR scores for voiced than voiceless items overall), $F(1,7) = 5148.75$, $p < 0.01$). Interactions also appeared between Language and Turbulence, $F(1, 7) = 38.72$, $p < 0.01$, Language and Voice, $F(1, 7) = 7027.91$, $p < 0.01$, Turbulence and Voice, $F(1, 7) = 8391.52$, $p < 0.01$, as well as Language, Voice and Turbulence, $F(1, 7) = 720.16$, $p < 0.01$. Overall, we found that while English phonemes display higher HNR values on average, across languages the more turbulent phonemes (English fricatives and Malayalam phonologically aspirated

stops) have much higher HNRs for voiced than voiceless consonants, whereas phonemes with low turbulence (phonologically unaspirated stops in both languages) instead show slightly higher HNRs for voiceless items than voiced ones. However, average HNR values are much lower for voiceless phonemes that are low in turbulence than for the turbulent voiced phonemes. Accordingly, the audio stimuli used in the dichotic selective attention perceptual experiment (Malayalam /p, b/, English /f, v/) can be arranged in the following hierarchy, ranging from highest to lowest HNR values: English /v/ > Malayalam /b/ > English /f/ > Malayalam /b/. These data will inform our discussion of the dichotic listening results in the General Discussion section of the paper.

In the dichotic task (see *Procedure*), the Malayalam-English (or English-Malayalam) pairs of audio stimuli were always matched for voicing, and contrasted only in terms of aperiodicity in the signal (as reflected by HNR measures of the consonant and vowel portions of the initial consonants of the nonce words – see Table 4). In phonological terms, this reduces to a contrast simply between presence/absence of the feature [continuant] in each trial (see Table 2). In the Malayalam-attend condition, the Malayalam unaspirated stops were the target items and the simultaneously-presented English initial fricatives in the opposing ear were the distractors. The converse was true for the English-attend condition.

3.2 Participants

Thirteen participants (seven female; age range 18–45 years) took part in the dichotic listening study, recruited from the Malayali-Australian community via flyers posted in churches, schools, and other community activity locations. Like the stimulus speaker, all were adult Australians with a Malayali heritage, and were born in Australia, but acquired Malayalam as their L1 from their family in the home within the first few years of life. However, their formal education being in Australia, all were fluent speakers of Australian English, with all segmental and supra-segmental qualities relevant to the current design. All participants completed a language-background questionnaire, which confirmed their fluent bilingual Malayalam-L1/ English-L2 dominant language status. None had any auditory or speech impairment, and all were right-handed, as handedness is known to correlate with the lateralization of phonological processing. All gave voluntary consent for participation. One interesting observation early on in participant recruitment was that while all our participants were verbally fluent in both English and Malayalam, they were literate only in English. This was taken as a confirmation that the participants were L2-dominant in English, while remaining fluent bilinguals.

3.3 Procedure

Stimuli were presented to the bilingual listeners via Sennheiser M2200X isolated headphones, using a Lenovo Thinkpad laptop computer and a Roland UA 25-EX audio card. They were presented dichotically, with the listener being instructed to attend to a given language in a given ear in 4 blocks of trials, one each for Malayalam-attend right ear and

Table 4: Mean HNR scores in dB for English and Malayalam stimulus consonants. Standard deviations are provided in parentheses.

Stimuli	/p/	/p ^h /	/b/	/b ^h /	/f/	/v/
ENGLISH	4.237 (0.149)		6.398 (0.218)		0.596 (0.085)	16.852 (0.376)
MALAYALAM	3.469 (0.068)	3.103 (0.132)	-2.959 (0.275)	5.314 (0.086)		

Malayalam-attend left ear, and the corresponding two blocks for English-attend. Within each block, there were 128 trials. These consisted of 8 tokens of each of the four stimulus categories (shown in Table 3) combined with one of two distractors. On each trial, one 'CVCV item from each language set (matched for voice quality, speaking rate and pitch) was presented to the two ears simultaneously. The two items presented on any trial differed in manner but matched in the voicing of the initial consonant in the first (stressed) syllable (see Table 3). All participants completed four blocks in which they attended to L1 and to L2, each in the right ear vs. the left ear, with order of blocks counterbalanced across participants.

For the task, the participants were required to listen to the simultaneous stimuli presented dichotically. A set of pictorial representations of actual Malayalam and English content words were displayed on a monitor for the participant to select from. The pictures included words beginning with the target consonants, which were all labial, e.g. *pot* for /p/, *ball* for /b/, etc., as well as coronal consonants, which were never the correct response choices and served as foils. The total set of 14 consonant response choices per language were given, represented by the pictures: for English /p, b, f, v, s, ʃ, t, d, m, n, w, z, tʃ, dʒ/; for Malayalam /p, p^h, b, b^h, s, ʃ, t, d, m, z, c, c^h, j, j^h/). An example of a picture response wheel for each language is provided in Figure 1 for each language. The orthographic labels in the picture are presented here for the readers' aid and were not included in the actual task. The circular arrangement of the items was selected in order to not bias responses towards any particular option. In order to accommodate all target consonants (labials) and distractor consonants (coronals), two response wheels per language were required. Pictorial representations were selected based on the target consonants in the language that the participants were required to track and identify. Thus, the initial consonant in the name of the picture, for example Malayalam /p/ in /paava/ "doll" in the picture of a rag-doll, was required to match the target consonant /p/ in the attended Malayalam nonce word /pala/. The participants were instructed to click on the image whose name began with the same consonant as the word they heard in their attended ear. We used pictures as response choices instead of the orthographic form of the phones that formed the response choice set because the participants were not literate in Malayalam. We provided the listeners with a whole range of coronal consonants as possible response

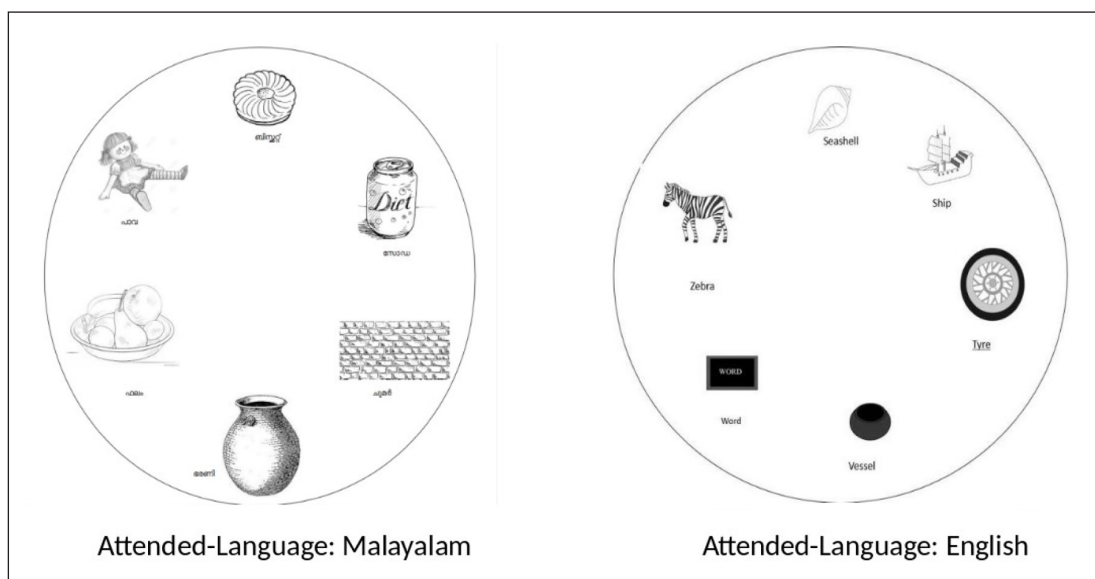


Figure 1: Examples of response-choice wheels.

choices in order to ensure that participants have the phonological freedom to select an emergent perceptually assimilated form without being restricted a priori by the available response choices.

4 Results

We report our results in terms of accuracy (Table 5) and in terms of intrusions (Figure 2). A response was counted as accurate when the picture corresponding to the initial consonant in the attended ear was selected. For example, in the Malayalam-attend condition, responses of /p/ to [pata] (with [fata] in the unattended ear) and of /b/ to /bata/ (with [vata] in the unattended ear) ear were counted as correct responses. An intrusion was defined as an incorrect response influenced by the stimulus played in the unattended ear. To continue with the same example of the Malayalam-attend condition, responses of /p^h/ (to [pata] with [fata] in the unattended ear) and /b^h/ (to /bata/ with [vata] in the unattended ear) were counted as intrusions.

Table 5: Mean percent correct responses for English and Malayalam.

LANGUAGE	VOICING			
	Voiced		Voiceless	
	EAR			
	Right	Left	Right	Left
English	69.7	64.4	83.7	75.5
Malayalam	66.8	63.0	83.7	75.0

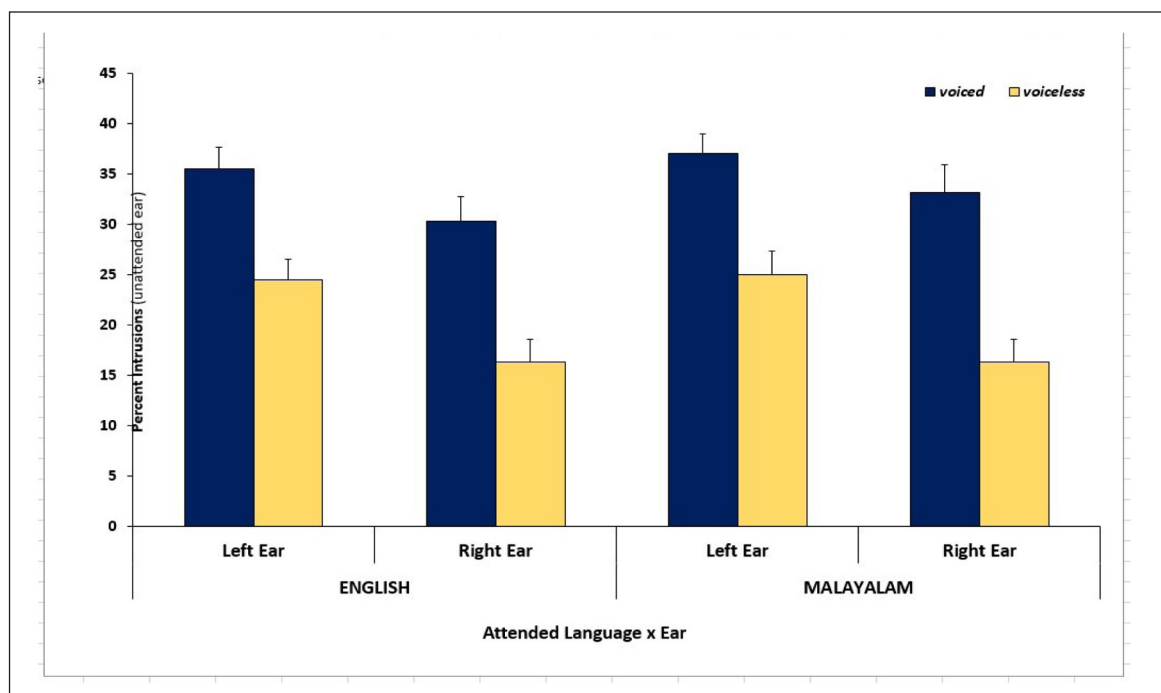


Figure 2: Intrusions from unattended language/ear by language, ear and target consonant type.

Our design also allowed for other incorrect responses. For example, a participant could have, in principle, responded /p^h/ to [bata] with [vata] in the unattended ear or could instead have chosen one of the coronal response options. These types of errors were extremely rare. Incorrect responses in which a listener chose a consonant that differed in voicing from the target stimulus were non-existent. As a consequence, the percentage of intrusions (Figure 2) is very nearly the complement of the percentage of accurate response (Table 5).

Table 5 provides the mean percent correct for each cell in the design. Accuracy ranged from 63.0% to 83.7%. Consistent with past work using the dichotic listening paradigm (Kimura 1961; 1967; Liberman et al. 1967), the means suggest a tendency for accuracy to be higher when the target was presented to the right ear than when it was presented to the left ear. There was also a tendency for accuracy to be higher for voiceless targets than for voiced targets. The error patterns by voicing and ear (left/right) were similar regardless of the target language.

Since there were no incorrect responses involving a mismatch in voicing, e.g., selecting /p^h/ to [bata] or /b^h/ for [pata] in the case of the Malayalam-attend condition or selecting /p/ for [vata] or /b/ for [fata] in the English-attend condition, the pattern of intrusions largely mirrors the accuracy patterns in Table 5. The intrusions are the issue of core interest to the present study, and are presented by condition in Figure 1. Visual inspection indicates that intrusion rates were lower when the target stimulus was presented in the right ear and lower for voiceless stimuli than for voiced stimuli. This pattern was consistent across language conditions.

To evaluate the statistical significance of these observations, we ran a $2 \times 2 \times 2$ repeated-measures ANOVA on the rate of intrusions, with the factors attended language (English, Malayalam), ear of attended language (right, left) and target voicing (voiced, voiceless). The ANOVA revealed statistically significant intrusions from the phone in the unattended language/ear, with significantly more intrusions when the attended language was in the left ear (ear of attended language main effect), $F(1, 12) = 8.58, p < 0.05$; and significantly more intrusions for voiced targets (target voicing main effect), $F(1, 12) = 52.38, p < 0.01$; but no significant effect of attended language, $F(1, 12) = 2.69, p > 0.05$. We found no significant interactions (all $p > 0.05$).

In sum, our results indicate that participants are able to maintain attention and successfully track the target phone in the attended language/ear the majority of the time, independently of which language was being attended. There were, however, a significant number of intrusions, indicating that the stimulus in the unattended ear influenced responses, and this occurred more often for voiced stimuli than for voiceless stimuli.

5 Discussion

The patterns of intrusion attested in our results are as follows. First of all, in line with other work on dichotic listening, our results showed a right ear advantage. There were fewer intrusions from the unattended ear when participants focused on target stimuli presented to the right ear than when they focused on the left ear. This likely reflects a bias towards processing phonology in the left hemisphere, according to the widely-accepted interpretation that stronger contralateral than ipsilateral ear to cortex connections reflect a right-ear advantage and hence left hemisphere superiority in speech perception, at least for right-handed listeners (e.g., Kimura 1961; 1967; Liberman et al. 1967). Second, we found that there were no voicing errors in this dichotic perception task. When listeners selected incorrect responses, which was not rare, the incorrect response was always one that matched the correct response in the voicing feature, e.g., /p/ was sometime selected for [fata] in English listening mode but /b/ was never selected. Third, we found that there

were more intrusions in manner for voiced stimuli than for voiceless stimuli. Fourth, we found that the attended language had no significant effect on the rate of intrusions. That is, the right ear advantage as well as the effect of voicing on manner intrusions (more intrusions for voiced stimuli than for voiceless stimuli) was the same regardless of the attended language.

With respect to our feature-based analysis of Malayalam-English bilinguals (Table 2), the effect of voicing on manner intrusions is of particular interest. Voiced stimuli in both languages/ears made it more difficult for the participants to maintain attention to the target language/ear (mean intrusions 34.04%) than did voiceless stimuli in each ear/language (mean intrusions 20.55%). On our account, voiceless phones are the unmarked specification, which is consistent with the tendency for languages to develop voiceless obstruents before voiced ones. However, our results indicate that markedness cannot be equated with the psycholinguistic notion of salience, since in our study it is the marked segments (voiced) that presented greater perceptual salience by permitting intrusions upon attentional mechanisms more often than the unmarked (voiceless) segments do. That is, participants found it harder to tune out unattended voiced segments than voiceless ones.

This result speaks to the possibility raised in Section 2 (Introduction) that the perceptual salience of a segment is related to the number of phonological features needed to specify it. Consider, for instance, a trial in the current design where a listener is presented with two dichotic stimuli of the pattern [vata] – [bata], and thus faces one of two possible tasks – to attend to the English item [vata] (if the trial is English-Attend), or to attend to the Malayalam item [bata] (if the trial is Malayalam-Attend). In an English-Attend trial, the listener's unattended ear is receiving a signal (Malayalam /b/) that, on our analysis, can only trigger the feature [voice], while the attended signal (English /v/) triggers [voice] and [continuant]. The listener's rate of correct detection of [voice] should be high, given that evidence for this feature is present in both ears. This is attested in our results. However, intrusions do also occur to a substantial extent, as our results also attest, and this also demands an explanation. When listeners have to recognize both [voice] and an additional feature [continuant], intrusions occur to a greater degree than when the listener must recognize the absence of [voice] and the presence/absence of [continuant]. Given our feature-based account, it is possible that there is a general principle at play in these results—selective perception is more difficult when there are a greater number of features to recognize.

The voiced stimuli, which involved a greater number of features, on our account, also showed larger acoustic differences from each other in harmonic-to-noise ratio (HNR) than did the voiceless stimulus pairs (see Table 4). English /v/ and Malayalam /b/ have a large HNR difference (19.859 dB); the difference between English /f/ and Malayalam /p/ is notably smaller (3.555 dB). This large difference appears to make it less likely that listeners will be able to completely tune out the unattended ear. Recall that Cherry's (1953) original selective attention experiment showed that a large spectral change in the unattended ear was always noticed. The attended English /v/, with the highest HNR value of all the phones in English, is influenced by the Malayalam /b/, with the lowest HNR value of all the Malayalam phones. The perceived phoneme is the one with the second highest HNR value in English, a /b/.

In the Malayalam-Attend condition, however, the listener faces the opposite challenge. It is now necessary to tune out a signal that has greater triggering capacities (unattended English /v/ has two features, [voice] and [continuant]), while attending to the Malayalam target /b/ that triggers only one feature ([voice]). The more marked signal in the unattended ear in this case causes intrusions into the attended ear. When

our participants' reports indicated a misperceived phoneme, it was always an aspirated voiced stop, Malayalam /b^h/. This is easily accounted for under our current set of assumptions. The signal phonetics will trigger features that help maintain contrast between competing sounds within a given language's phonological system. The competing signal, in this Malayalam-Attend case, differs from the attended signal in terms of the feature [continuant]. Given the higher feature specification of the competing signal and large periodicity difference, it will intrude, but the listener's lexical access will be faced with a gap in Malayalam, in that the intrusion would now trigger a feature that the lexicon does not utilize at this place of articulation (labial): [continuant]. The consequence is that the listener selects /b^h/, which has an increased HNR higher than Malayalam /b/.

We argue that this situation is best explained by referring to the notion of perceptual assimilation of closest resembling phonological category as elucidated in the Perceptual Assimilation Model, or PAM (Best 1995). In this framework, the greater the phonetic-articulatory similarity between two unrelated consonants, the more likely they are to be perceptually assimilated to one another. With respect to determining what counts as similar, for the present paradigm we use the articulatory phonetic correlates of features. The feature [continuant] in fricatives correlates (roughly) to a narrow constriction that produces continuous turbulent airflow at some location in the vocal tract. This is captured, as described in Section 3.1 (Stimuli), by the Harmonics-to-Noise Ratio (HNR) of the signal, which relates its periodic and aperiodic components. Similarly, the feature [spread glottis] in stops indicates aspiration – the opening gesture of spreading the vocal cords to allow turbulent airflow through the glottis. This also increases aperiodicity in the signal. Given that the glottal vs. oral turbulence distinction, in featural terms a [spread glottis] versus [continuant] distinction, is not present in the Malayalam phonology for labial consonants, the intruding phonetic information, in this case the [continuant] feature of the English fricative, triggers the closest available featural correlate in Malayalam, which is [spread glottis]. Since the feature [voice] is already triggered by the attended signal (and is not contradicted by the competitor), the combination of this with the intrusion from the unattended signal leads to the detection of a voiced, aspirated segment. A (highly harmonic) /v/ in the unattended ear affects the perceived noisiness of the attended phone /b/, and the resultant percept is thus a more highly harmonic Malayalam phoneme, namely /b^h/.

When target and competitor are both voiceless segments, fewer features are involved in the contrast (on our account, the feature [voice] is absent from voiceless stops and voiceless fricatives). We suggested above that this feature sparsity accounts for the relatively lower intrusion rates. However, the nature of intrusions found in voiced and voiceless trials has a unified explanation. Both can be explained by how variation in periodicity (as measured by HNR) across ears can either trigger the feature [continuant] when that feature is relevant for the attended phonology or be assimilated to the closest phonological match in the case when the feature [continuant] is absent. We can identify the closest phonological match with reference to the HNR hierarchy we can extract from Table 4 for English and Malayalam, as indicated below:

English HNRs: /v/ > /b/ > /p/ > /f/

Malayalam HNRs: /b^h/ > /p/ > /p^h/ > /b/

For the voiceless target-competitor pair, effects of harmonic pull follow the HNR hierarchy, as they do for voiceless pairs. Thus, for English /f/ and Malayalam /p/, intrusions effect a shift to the adjacent step in the harmonicity hierarchy. In the English-attend

mode, this results in /p/ responses to [fata]; in the Malayalam-attend mode, this results in /p^h/ responses to [pata], by which [fata] in the unattended ear triggers activation of [spread glottis] as the closest matching feature in the attended language phonological system. In this way, the phonological inventory constrains participants' available options in choosing a matching phoneme, limiting them to a particular set of grammatical representations.

If an unattended phoneme with a higher harmonic ratio pulls a less harmonic attended phone up the scale of harmonicity, the larger HNR distinction between the voiced pairs (/v/-/b/) will increase the perceived harmonicity of less harmonic Malayalam /b/ to the more highly harmonic /b^h/ when an English /v/ is presented to the unattended ear. On the other hand, the very high harmonicity of the target /v/ is only slightly affected by low HNR of /b/, and thus the perceptual HNR is reduced to the second most harmonic phoneme in English, also a /b/. In this way, intrusion patterns for both voiced and voiceless stimuli can be understood phonetically in terms of harmonic pull.

On the premise that harmonic properties of the signal trigger featural distinctions in the mind, the featural account provided above complements the phonetic analyses. At least for the current case of voicing, a greater number of features in the representation corresponds to larger acoustic differences and greater harmonic pull. There is a greater harmonic difference for voiced stimulus pairs than for the voiceless stimulus pairs. Thus, the sparsity in feature specifications for the voiceless stimulus pairs corresponds to smaller differences in HNR. As can be seen in Table 4, the competing sounds in a voiced target vs. voiced competitor pairing differ greatly in HNR (nearly 20 dB); but in the voiceless trials the HNR difference is much smaller, about 3dB. Thus, while an unattended English /f/, with the lowest harmonicity of the English consonants, causes the highly harmonic Malayalam /p/ to drop to a slightly less harmonic /p^h/, the unattended Malayalam /p/, with a high HNR, increases the perceptual harmonicity of the English /f/ to that of the English stop /p/. These intrusions are in the expected direction but occurred less often than for voiced stimuli.

The feature-based approach to representing voicing and manner contrasts that we have pursued (see Table 2) assumes privative features and underspecification. Additionally, in the Introduction, we took the "standard approach" (e.g. Lisker & Abramson 1964; Keating 1984; Lombardi 2018) to voicing features as opposed to "laryngeal realism" (Jakobson 1949; Honeybone 2005). Thus, we represent the laryngeal distinctions between stops in English with the feature [voice], the same feature that distinguishes /f/ and /v/ in English fricatives and /p^h/ and /b^h/ in Malayalam. As reviewed in Section 2, there is an ongoing debate over the proper characterization of laryngeal features in English, centered on whether the contrast between stops is better captured by the feature [spread glottis], the "laryngeal realism approach" or by the feature [voice], the "standard approach" (Hall 2001). For Malayalam, it is clear that both features are required. The "standard approach" to feature representation was crucial to our interpretation of the intrusion patterns in Malayalam-English bilinguals. Specifically, it allowed us to maintain that phonological features are triggered in perception when the phonetic signal is consistent with those features. When properties of the phonetic signal do not map directly to features in the listener's attended phonology, they map instead to the nearest phonological category. This mechanism of perceptual assimilation to the nearest phonological category is well-established from research on cross-language speech perception (e.g., Best 1995; Faris et al. 2016), L2 speech perception (Best & Tyler 2007; Bundgaard-Nielsen et al. 2011), bilingual speech perception (Antoniou et al. 2012; Antoniou et al. 2013), and cross-accent speech perception (Best et al. 2015; Best et al. 2015; Shaw et al. 2018). Here, we observed its effects in bilinguals during dichotic listening.

While additional research is needed to evaluate the generality of our account, we have presented here an explicit phonological proposal for a population of speakers, Malayalam-English bilinguals, and explained how that proposal dictates behavior in a controlled perception experiment. The dichotic listening paradigm was motivated in part by the connection between this experiment and the real life task faced by bilinguals of balancing selective attention and two rather different sets of phonological contrasts. We are optimistic about the prospect of future work linking phonological theory to a broader range of speech behaviors, both for the potential of theory to guide our understanding of speech behavior and for experimental results to constrain the development of phonological theory.

6 Conclusion

In this study we tested adult L2 dominant bilinguals using a task that combines the involvement of selective-attention mechanisms alongside normal signal processing, in a manner that listeners – especially bilingual listeners – are often faced with in situations akin to cocktail-party paradigms. Our tasks presented stimuli from listeners' L1 and L2, dichotically, and required the participants to undertake a phonemic categorization task for one of the two languages. Consistent with past work using this paradigm, we found that listeners showed a right ear advantage, suggesting left hemisphere superiority in phonological processing. When the stimulus presented in the unattended ear contained a phonological feature absent in the language of the attended ear, that feature was perceptually assimilated to the closest matching feature. Intrusions were more likely for segments containing more distinctive features. These accounts of the results hinge crucially on our particular feature-based analysis of the languages involved, which underscores the potential for experiments such as this to provide new lines of evidence for phonological representations in bilinguals.

Abbreviations

HNR = Harmonic to Noise Ratio, dB = decibel, CVCV = consonant vowel consonant vowel, ANOVA = Analysis of Variance, Hz = Hertz.

Acknowledgements

We thank participants at the Australian Linguistics Society meeting where parts of this work were presented, for their comments, as well as the *Glossa* Associate Editor and three anonymous reviewers, whose comments greatly improved the paper. This research was supported by a MARCS post-graduate scholarship to the first author.

Competing Interests

The authors have no competing interests to declare.

References

- Antoniou, Mark, Catherine T. Best & Michael D. Tyler. 2013. Focusing the lense of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *The Journal of the Acoustical Society of America* 133(4). 2397–2411. DOI: <https://doi.org/10.1121/1.4792358>
- Antoniou, Mark, Michael D. Tyler & Catherine T. Best. 2012. Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics* 40(4). 582–594. DOI: <https://doi.org/10.1016/j.wocn.2012.05.005>
- Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5. 183–208. <https://www.jstor.org/stable/4419977>. DOI: <https://doi.org/10.1017/S095267570002268>

- Asher, Ronald E. 1985. *Tamil*. London: Croom Helm.
- Beckman, Jill, Michael Jessen & Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics* (49)2. 259–284. DOI: <https://doi.org/10.1017/S0022226712000424>
- Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: issues in cross-language research*, 171–204. Maryland: York Press.
- Best, Catherine T., Jason A. Shaw, Gerard Docherty, Bronwen G. Evans, Paul Foulkes, Jen Hay, Jalal Al-Tamimi, Katharine Mair, Karen E. Mulak & Sophie Wood. 2015. From Newcastle MOUTH to Aussie ears: Australians' perceptual assimilation and adaptation for Newcastle UK vowels. *Paper presented at the Interspeech 2015*, Dresden, Germany.
- Best, Catherine T., Jason A. Shaw, Karen E. Mulak, Gerard Docherty, Bronwen G. Evans, Paul Foulkes, Jennifer Hay, Jalal Al-Tamimi, Katharine Mair, Sophie Wood. 2015. Perceiving and adapting to regional accent differences among vowel subsystems. *Paper presented at the 18th International Congress of Phonetic Sciences*, Glasgow, UK.
- Best, Catherine T. & Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Murray J. Munro & Ocke-Schwen Bohn (eds.), *Second language speech learning: The role of language experience in speech perception and production*, 13–34. Amsterdam: Johns Benjamins. DOI: <https://doi.org/10.1075/llt.17.07bes>
- Bundgaard-Nielsen, Rikke L., Catherine T. Best & Michael D. Tyler. 2011. Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition* 33(3). 433–461. DOI: <https://doi.org/10.1017/S0272263111000040>
- Cherry, Edward Colin. 1953. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America* 25(5). 975–979. DOI: <https://doi.org/10.1121/1.1907229>
- Chomsky, Noam, & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Faris, Mona M., Catherine T. Best & Michael D. Tyler. 2016. An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America* 139(1). EL1–EL5. DOI: <https://doi.org/10.1121/1.4939608>
- Golston, Chris. 1996. Direct OT: Representation as pure markedness. *Language* 72(4). 713–748. <https://www.jstor.org/stable/416100>. DOI: <https://doi.org/10.2307/416100>
- Grosjean, François. 1988. Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes* 3(3). 233–274. DOI: <https://doi.org/10.1080/01690968808402089>
- Hall, Alan T. 2001. Introduction: Phonological representations and phonetic implementation of distinctive features. *Distinctive Feature Theory*, edited by Tracy Hall, 1–40. DOI: <https://doi.org/10.1515/9783110886672.1>
- Honeybone, Patrick. 2005. Diachronic evidence in segmental phonology: the case of laryngeal specifications. *The Internal Organization of Phonological Segments*, 319–354.
- Iverson, Gregory K. & Joseph C. Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12(3). 369–396. <https://www.jstor.org/stable/4420084>. DOI: <https://doi.org/10.1017/S0952675700002566>
- Iverson, Gregory K. & Joseph C. Salmons. 1999. Glottal spreading bias in Germanic. *Linguistische Berichte*, 135–151.
- Iverson, Gregory K. & Joseph C. Salmons. 2003. Laryngeal enhancement in early Germanic. *Phonology* 20(1). 43–74. DOI: <https://doi.org/10.1017/S0952675703004469>

- Iverson, Gregory K. & Sang-Cheol Ahn. 2007. English voicing in dimensional theory. *Language Sciences* 29(2–3). 247–269. DOI: <https://doi.org/10.1016/j.langsci.2006.12.012>
- Jakobson, Roman. 1949. *On the identification of phonemic entities*. The Hague: Mouton. DOI: <https://doi.org/10.1080/01050206.1949.10416304>
- Jessen, Michael & Catherine Ringen. 2002. Laryngeal features in German. *Phonology*, 19(2). 189–218. DOI: <https://doi.org/10.1017/S0952675702004311>
- Kager, Rene, Suzanne V. H. van der Feest, Paula Fikkert, Annemarie Kerkhoff & Tania S. Zamuner. 2007. Representations of [voice]: evidence from acquisition. In Erik Jan van der Torre & Jeroen van de Weijer (eds.), *Voicing in Dutch: (De)voicing – phonology, phonetics, and psycholinguistics*, 41–80. Amsterdam & Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/cilt.286.03kag>
- Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language*, 286–319. <https://www.jstor.org/stable/413642>. DOI: <https://doi.org/10.2307/413642>
- Kehrein, Wolfgang & Chris Golston. 2004. A prosodic theory of laryngeal contrasts. *Phonology* 21(3). 325–357. <https://www.jstor.org/stable/4615515>. DOI: <https://doi.org/10.1017/S0952675704000302>
- Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.
- Kimura, Doreen. 1961a. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology/Revue canadienne de psychologie* 15(3). 166–171. DOI: <https://doi.org/10.1037/h0083219>
- Kimura, Doreen. 1961b. Some effects of temporal-lobe damage on auditory perception. *Canadian Journal of Psychology/Revue canadienne de psychologie* 15(3). 156–165. DOI: <https://doi.org/10.1037/h0083218>
- Kimura, Doreen. 1967. Functional asymmetry of the brain in dichotic listening. *Cortex* 3(2). 163–178. DOI: [https://doi.org/10.1016/S0010-9452\(67\)80010-8](https://doi.org/10.1016/S0010-9452(67)80010-8)
- Lahiri, Aditi & Henning Reetz. 2010. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38(1). 44–59. DOI: <https://doi.org/10.1016/j.wocn.2010.01.002>
- Lieberman, Alvin M., Franklin S. Cooper, Donald Shankweiler & Michael Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review* 74(6). 431–461. DOI: <https://doi.org/10.1037/h0020279>
- Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20(3). 384–422. DOI: <https://doi.org/10.1080/00437956.1964.11659830>
- Lombardi, Linda. 2018. *Laryngeal features and laryngeal neutralization*: Routledge. DOI: <https://doi.org/10.4324/9780429454929>
- Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511753459>
- Petrova, Olga, Rosemary Plapp, Catherine Ringen & Szilárd Szentgyörgi. 2006. Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review* 23(1). 1–35. DOI: <https://doi.org/10.1515/TLR.2006.001>
- Shaw, Jason A., Catherine T. Best, Gerard Docherty, Bronwen G. Evans, Paul Foulkes, Jennifer Hay & Karen E. Mulak. 2018. Resilience of English vowel perception across regional accent variation. *Laboratory Phonology* 9(1). 1–36. DOI: <https://doi.org/10.5334/labphon.87>
- Spivey, Michael J. & Viorica Marian. 1999. Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science* 10(3). 281–284. DOI: <https://doi.org/10.1111/1467-9280.00151>

- Steriade, Donca. 1995. Underspecification and markedness. In John A. Goldsmith (ed.), *Handbook of Phonological Theory*, 114–174. Cambridge, MA: Blackwell.
- Weber, Andrea & Anne Cutler. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50(1). 1–25. DOI: [https://doi.org/10.1016/S0749-596X\(03\)00105-0](https://doi.org/10.1016/S0749-596X(03)00105-0)

How to cite this article: Mandal, Sayantan, Catherine T. Best, Jason Shaw and Anne Cutler. 2020. Bilingual phonology in dichotic perception: A case study of Malayalam and English voicing. *Glossa: a journal of general linguistics* 5(1): 73.1–17. DOI: <https://doi.org/10.5334/gjgl.853>

Submitted: 09 November 2018 **Accepted:** 05 May 2020 **Published:** 22 July 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Glossa: a journal of general linguistics is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS The Open Access logo, which is a stylized 'O' with a person icon inside, representing accessibility.