Word count (29.7.2022): 13388 – remove 4388

Short title: Phonetic transfer

Full title: Mixed language processing increases cross-language phonetic transfer in Bengali-English bilinguals

Authors:

Auromita Mitra
EFL University,
Hyderabad,
India

Indranil Dutta
Jadavpur University,
Kolkata,
India

Address for correspondence:
Auromita Mitra,
EFL University,
Hyderabad- 500007,
auromita.mitra@gmail.com

Abstract:

This study investigates the phonetic outcome of mixed-language speech in Bengali and Indian English, towards understanding cross-language interaction in highly proficient bilinguals. We compare spectral properties of L2 vowels [æ] (common to L1, L2) and [ʌ] (absent in L1) in code-switched (mixed) vs. non-switched productions. Results reveal asymmetrical shifts in both vowels during mixed productions, towards related L1 categories. We interpret this as a temporary increase in cross-language phonetic interaction during mixed-language use, leading to a shift towards L1 norms, evidencing transfer effects on L2 vowels. We elicit mixed productions through two tasks: cued picture-naming and code-switching, to assess if experimental paradigm independently influences the behavior under study. Results reveal parallel patterns, but small magnitude differences, across paradigms. We discuss these findings in light of recent proposals about asymmetries in short-term phonetic interaction, postulated discursive factors in code-switching, and the issue of comparability between paradigms in transfer studies.

1 Introduction: Phonetic transfer

Bi/multilingual speakers can distinguish between the phonetic norms of their languages and maintain separate sound categories for each language (Bosch & Sebastián-Gallés, 2003; Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; MacLeod & Stoel-Gammon, 2010). However, these categories are not autonomous— they influence each other across languages in both perception and production (e.g. Flege, 1995; Flege, MacKay, & Piske, 2002; Fowler, Sramko, Ostry, Rowland, & Hallé, 2008), and the nature of such interaction provides crucial insights into how language 'systems' are cognitively represented and processed. Based on the productions of proficient bilingual speakers of Bengali and English, this study reports that cross-language phonetic influence temporarily increases during mixed-language use.

Cross-language influence at the level of sounds can be studied in broadly two kinds of conditions: (i) While a bilingual speaker is operating in any one of their languages, by comparing bilingual speech to monolingual norms (e.g. Caramazza et al., 1973; Flege, 1987; Guion, 2003); (ii) When both languages of a bilingual speaker are co-activated, by comparing speakers' mixed-language speech to their own norms while using a single language (e.g. Bullock & Toribio, 2009; Elias, McKinnon, & Milla-Muñoz, 2017; Grosjean & Miller, 1994; Simonet, 2014). The present study is concerned with the latter. In production, this kind of influence has been variously termed transfer, drift, accommodation, and interference. We use the term "transfer" here, to indicate any interaction between two sets of phonetic norms.

Existing research on the phonetic effects of mixed-language production has largely focused on a limited set of phonologically related languages. A majority of these studies use temporal properties of consonants (in particular, voice onset time or VOT) to measure transfer. However, reported results in the literature vary greatly across studies, as discussed in the following subsections, and appear to be contingent upon both language-specific features and language experience of the participants. This emphasizes the importance of considering

data from a wider variety of populations, language pairs, and phonetic features in order to make meaningful generalizations. Widespread multilingualism in the Indian subcontinent suggests that phonetic behavior in these populations can be particularly valuable towards understanding the nature of such cross-language interactions, as they are likely to reflect real-world experience with mixed-language processing. However, there is no work yet on short-term phonetic transfer in the Indian subcontinent, or in any Indo-Aryan language.

The present study examines phonetic transfer between Bengali and English in a group of highly proficient bilingual speakers in India. We measure spectral properties, F1 and F2, of two English vowels to ascertain if L1 influence on L2 increases during mixed-language use, relative to a participant's baseline production of L2. Mixed-language data is elicited in two switching paradigms: cued picture-naming and code-switching. We compare these results to assess if differences between the paradigms independently influence the outcome of phonetic interaction. The results demonstrate a time-invariant effect of transfer on L2 vowels during mixed-language production, and parallel pattern but different magnitude of transfer between the two paradigms. We discuss these findings in light of recent proposals about asymmetries in short-term phonetic interaction and the role of connected speech in introducing discursive factors to transfer studies.

The rest of this section is structured as follows: 1.1 provides background information on Bengali and Indian English focusing on vowel systems. 1.2 discusses two potential sources of transfer during language processing: global co-activation of multiple languages vs. the local act of switching between languages. Given the highly multilingual setting that characterizes this population, we note that the former is expected to be a constant feature and therefore any observed phonetic interaction is better understood as resulting from the latter. The design of the present study is motivated on the basis of this discussion. A logical consequence of this model is that such phonetic interaction must be highly localized. 1.3 elaborates on this by reviewing literature on the duration and 'reversibility' of transfer effects. A recurring

pattern that emerges from existing research on short-term phonetic transfer is that both languages of the bilingual speaker are not equally affected, although the nature of these reported differences varies greatly across studies. In 1.4, we review some of these findings, focusing on proposed sources of asymmetries, in order to highlight the factors that mediate such phonetic interaction (or fail to do so). We argue that these results point to an urgent need for expanding the range of language pairs and sound categories examined in order to make meaningful generalizations about the nature of short-term phonetic transfer. Following from this, 1.5 discusses reported asymmetries between different sound categories of a single language, and develops the hypotheses of the present study on the basis of these. In 1.6, we motivate a secondary aim of this study: to verify the postulated differences between two paradigms for eliciting short-term phonetic interaction in an experimental setup, namely cued picture-naming and code-switching. Finally, we summarize the research questions and hypotheses.

Existing studies have studied phonetic transfer along different acoustic features. While discussing the literature in subsequent sections, we will often indicate the acoustic feature in parentheses following a citation, for clarity.

1.1 Bengali and English in India

Demography

Bengali (also, Bangla) is an Indo-Aryan language primarily spoken in India and Bangladesh. In India, more than 97 million people speak Bengali as a first language (Census of India, 2011), mostly in the state of West Bengal. A majority of this population also speaks additional languages.

Indian English (IE) refers to the variety of English that has developed in the Indian subcontinent. In India, it is spoken as an L2 by 129 million people (Census of India, 2011).

English is one of the two official languages, used in education, law, media, as a lingua franca mainly for an educated elite in most metropolitan regions, and carries a high prestige value (B. Kachru & Smith, 1981; B. B. Kachru, 1983; Pandey, 2015; Tollefson & Tsui, 2014).

Note that "Indian English" is not a monolithic entity; there are large regional variations in the English speech of Indian speakers, and the literature on IE phonology is thus neither "standardized" nor uncontested. For the present study, we focus on a variety of IE spoken by an educated elite, used in news channels, often taught as a target in classrooms, and generally considered to be free of any particular regional affiliation (c.f. Masica, 1972, who reports this variety to be "a de-facto norm", and thus calls it General(ized) Indian English). This sound system is described in section 1.1. However, recent research comparing the segmental and suprasegmental properties of IE spoken in different parts of the country also report many commonalities (see Sirsa and Redford (2013) for a review). It has thus been suggested that in spite of the varied L1s of its speakers, IE has a target phonology that is distinct from any of these, as well as from other native varieties of English. Thus, it is fruitful to think of regional variations as resulting from L1-influence on a common underlying target. Excepting a small minority of specific communities (Coelho, 1997; Pandey, 2015; Wells & Wells, 1982), there are very few L1 speakers of IE. Thus, English is largely learned and used as an L2.

Considering the language usage patterns suggested by this demography, the present study focuses specifically on short-term phonetic interaction during mixed-language use, because:

(i) In multilingual populations, long-term representations of sound categories are expected to be affected by multiple languages.

(ii) Given that examples of an IE phonology without L1 'influence' are very rare, it is more meaningful to think of cross-language transfer in L2 as relative to a speaker's own production in a given baseline condition.

The socio-linguistic facts about English in India also suggest that, for the present pop-

ulation, mixed-language processing of English and an L1 is an ecologically valid paradigm that is likely to be a part of everyday language experience. Thus, data from this population is valuable to the understanding of short-term phonetic interaction.

Vowel systems

The vowel inventory of Western Bangla (the variety spoken by the participants in this study) consists of [i, e, æ, a, ɔ, o, u], and their nasalized counterparts (Rubino & Garry, 2001). The vowel chart in figure 1 shows their distribution in the F1×F2 formant space. Note that there is no mid-central vowel category.

The vowel system of IE contains the monophthongs [ɪ, i, ɛ, e, æ, ə/ʌ/ɜː, aː, ɔ, o, ʊ, u] (Masica, 1972; Wells & Wells, 1982). The vowel chart in figure 2 shows their distribution in the F1×F2 formant space. A single mid-central vowel corresponds to the categories [ʌ,ə,ɜː], which are treated as distinct in many native varieties of English (Bansal, 1969; Hickey, 2005; Nihalani, Tongue, Hosali, & Crowther, 1979; Wells & Wells, 1982). Since the English items used in this study are traditionally transcribed with [ʌ], we use this symbol to indicate the mid-central vowel throughout.

<Insert figure 1 about here>

<Insert figure 2 about here>

1.2 What causes transfer and what does it affect?

Existing research has distinguished changes in category representations due to the acquisition of multiple sound systems (cf. Speech Learning Model (Flege, 1995, 2007); Perceptual Assimilation Model-L2 (Best & Tyler, 2007)), from interaction during accessing, processing, or articulation of these categories (cf. transfer vs. interference (Grosjean, 2012); competence vs. performance interference (Paradis, 1993)). Given their transient nature, dynamic

changes in production during mixed-language use are generally attributed to the latter, e.g. online processing costs (Olson, 2013; Šimáčková & Podlipskỳ, 2015; Tsui, Tong, & Chan, 2019, VOT), language mode (Simonet, 2014, vowel quality), context-awareness (Khattab, 2013, phonological variables). What triggers this interaction? Olson (2016) argues that while cross-language phonetic effects are largely measured at the point of language switch, it could have two potential sources:

1. The local point of switch itself

2. Global co-activation of two languages: bilingual language mode (Grosjean, 1998)

A number of studies have specifically manipulated language mode, both in the presence and absence of switching. Overall, results suggest that: (i) In the absence of other manipulations, productions in a bilingual language mode show increased cross-language influence compared to a monolingual mode (Simonet, 2014; Simonet & Amengual, 2020, vowel quality); (ii) However, language mode is not the sole source of influence during mixed language use—studies comparing switched and non-switched tokens produced in the same test block (identical language mode) (Olson, 2016; Tsui et al., 2019, VOT) or spontaneous conversation (Piccinini & Arvaniti, 2015, VOT), have still reported a difference, suggesting that independently of mode, switching between languages triggers a local increase in cross-language transfer. (iii) How the two sources interact to influence the final outcome of transfer is not fully understood: (Olson, 2016, VOT) found no additive effects of language mode, (Olson, 2013, VOT) found a balanced language context to inhibit transfer compared to unbalanced contexts. Other studies have not analyzed the two separately, eliciting switched tokens in a bilingual test block and non-switched tokens in a separate monolingual test block, separated by a few hours to days (Elias et al. (2017); vowel quality), (Antoniou, Best, Tyler, & Kroos, 2011; Bullock & Toribio, 2009; Schwartz, Balas, & Rojczyk, 2015; Šimáčková & Podlipskỳ, 2015, 2018, VOT).

Grosjean (1998) suggests that various aspects of the communicative setting, including exposure to (spoken or written) stimuli in multiple languages and awareness of the interlocutor's being bilingual, could trigger a bilingual language mode. The participants in the present study are in an environment which largely contains mixed-language input, multilingual interlocutors, and no stable ambient language. Therefore we expect that phonetic transfer during everyday language use, if any, takes place in a bilingual mode. To preserve ecological validity, we elicited both switched and non-switched utterances in a bilingual language mode. Any observed differences in this paradigm would arguably result from interaction during online processing of sounds while switching between languages.

In a bilingual mode, both language systems are expected to be nearly equally accessible throughout the test block. Thus, a consistent difference between switched and non-switched tokens in such a paradigm would be possible only if the effects were highly localized— if not, we should expect a gradual convergence over the course of the experiment. This is discussed in the next section.

1.3   Duration of transfer effects

Research comparing bilingual speakers' phonological systems with monolinguals treats them as relatively stable over time. The difference from monolingual norms is interpreted as the cumulative result of cross-language influence over long periods (e.g. Caramazza et al., 1973; Guion, 2003). However, longitudinal studies and between-subject comparisons of bilingual speakers with different durations of L2 exposure suggest that interaction between the sound systems is dynamic: over time, increasing exposure to an L2 can reduce effects of L1 influence, leading to a more fine-grained separation and native-like production of foreign contrasts. For example, (Bohn & Flege, 1992, vowel quality) observed that approximately 7 years' difference in L2 exposure between groups having the same age of acquisition led to a

significant difference in the amount of cross-language influence. These have been interpreted as changes to long-term phonological representations.

Moreover, changes due to transfer are not necessarily unidirectional or irreversible. (Sancier, Fowler, et al., 1997, VOT) first demonstrated that spending 2-5 months in an L1 or L2 environment causes productions in both languages to "drift" towards the ambient language. This not only evidenced that cross-language interaction can be triggered in an order of months, but also that the effects can be reversed within a similar time range. A more recent study by (Tobin, Nam, & Fowler, 2017, VOT) reported comparable effects in an even shorter duration (2-4 weeks). Based on a short-term longitudinal study, (Chang, 2012, VOT) demonstrates that over the course of the first five weeks of learning an L2, there is a gradual convergence of L1 towards L2. For some sounds and features (though not others, cf. sec.1.5), transfer was additive over time. How long these effects last in the absence of regular L2 input was not tested.

The research discussed above concerns situations where a speaker is operating in any one of their languages. When it comes to the phonetics of mixed-language use, the majority of existing studies only analyze the switched (target) token. Thus, there are few direct measurements of the duration of short-term transfer effects. One study which measured this in a code-switching paradigm (Bullock & Toribio, 2009, VOT) did not find any residual effects on the matrix language following a switch, suggesting that transfer during code-switching is localized. Indirect evidence in support of this comes from experiments which have elicited switched and non-switched tokens in the same test block and still reported differences between the two (e.g. Olson, 2013; Tsui et al., 2019, VOT), suggesting that changes due to transfer are quickly 'reset'— in an order of seconds.

Note that all the studies of mixed-language production discussed above measure VOT, which is a temporal feature. We have no apriori reason to assume that these durations generalize to vowel quality. However, findings from sub-categorical phonetic shifts triggered by other

factors (such as convergence towards an interlocutor) do evidence rapid shifts in vowel quality within comparable time-frames (e.g. Babel, 2010, 2012; Pardo, 2010). Thus in the current study, we present switched and non-switched tokens randomly within the same test block to induce a bilingual mode. We expect the intervening words between two subsequent targets to undo any residual effects of transfer.

1.4   Asymmetries between languages in extent and direction of transfer

A recurring pattern that emerges from existing research on short-term phonetic transfer is that the extent and patterns of phonetic shift are not equivalent across the two languages of the bilingual speaker. Studies vary greatly in the nature of differences that they report. The potential causes for such asymmetry are of interest because they point towards the factors that mediate cross-language phonetic interaction. In this section, we discuss two proposed sources of such asymmetry that are relevant to the present study:

L1 vs. L2 status

Flege's Speech Learning Model (SLM) (1995; 2007) posits that the sound categories of a bilingual speaker exist in a common phonological space, and therefore in principle, both L1 and L2 categories can influence one another. Thus, the patterns of cross-language transfer depend on how L2 phonemes are mapped in relation to the existing L1 categories. In non-switched production, this is evidenced through a pervasive L1 influence on non-native contrasts that are perceptually linked to an existing L1 contrast (cf. equivalence classification: Flege (1987); Flege and Hillenbrand (1984)), and the observation that both L1 and L2 sound systems of bilinguals differ from those of corresponding monolingual speakers (e.g. Guion, 2003, vowel quality).

In mixed-language production, the role of language status is less clear— studies have

reported unidirectional influence of L1 on L2 (Antoniou et al., 2011; Balukas & Koops, 2015; Goldrick, Runnqvist, & Costa, 2014; Šimáčková & Podlipskỳ, 2015, VOT), L2 on L1 (Olson, 2013; Tsui et al., 2019, VOT), (Elias et al. (2017), vowel quality) bidirectional convergence (Bullock & Toribio, 2009; Olson, 2016, VOT), divergence (Bullock & Toribio, 2009; Šimáčková & Podlipskỳ, 2018, VOT), and no influence (Muldner, Hoiting, Sanger, Blumenfeld, & Toivonen, 2019, vowel quality), (Schwartz et al., 2015, phonological process). Since most existing studies measure shifts in VOT, the observed asymmetries between languages have been variously explained either in terms of their L1 vs L2 status, or language-specific differences between long- and short-lag VOT languages.

Olson (2013) first reported a unidirectional VOT shift of L1 towards L2 in two different groups: native speakers of Spanish (short-lag) and English (long-lag), matched for proficiency and age of L2 acquisition. This established that beyond language-specific differences, the L1 vs L2 status of the language does mediate transfer. They interpret this asymmetry in terms of the Inhibitory Control Model (ICM) (Green, 1998): to select a phonetic realization from one language, the other must be inhibited. Since L1 is the 'stronger' language, the amount of inhibition required on L1 while using an L2 is greater than the inhibition required on an L2 while using the L1. This greater initial inhibition on L1 means that switching into L1 incurs a greater switch cost (more cross-language influence) than switching into L2. Lower switch cost results in a lack of visible cross-language transfer effects on L2. Tsui et al. (2019) report comparable results, but only in participants who were not equally dominant in both languages. Balanced bilinguals did not demonstrate any transfer effects in VOT. They propose that this is because balanced bilinguals have better inhibitory control (low switch cost in both languages) due to greater experience with language switching.

Since differences between long-lag and short-lag languages independently affect the phonetic realization of VOT, focusing on other phonetic features as sites for transfer can avoid this conflation, and clarify the precise effect of language status. However, given the impli-

cations of the ICM, it is necessary to first establish that these L2 categories can indeed be affected by dynamic interference, particularly in bilinguals with extensive language-switching experience. The few existing studies on vowel quality (Elias et al., 2017; Muldner et al., 2019; Simonet, 2014) or phonological processes (Schwartz et al., 2015; Simonet & Amengual, 2020) have examined transfer effects on L1 categories. As far as we are aware, there is no work yet on short-term transfer in L2 vowels. Thus, this study builds on the existing work by examining whether the L2 vowel quality of proficient bilinguals can be affected by mixed-language processing.

Differences between sound systems of the languages

In addition to the cognitive factors discussed above, many researchers have attributed the observed asymmetries to language-internal factors. Specifically, a pattern that emerges from the body of work on VOT is that a shift in long-lag VOT towards the short-lag norms of the other language is much more consistent and systematic than the converse (Antoniou et al., 2011; Bullock & Toribio, 2009; Chang, 2012; Olson, 2016; Tobin et al., 2017). In general, short-lag VOTs seem to resist accommodative shift. Bullock and Toribio (2009) suggest that because long-lag languages offer a greater range of acceptable VOTs, there is more 'room' for movement. In contrast, shift in short-lag languages could risk the loss of a phonological contrast (VOT being the primary cue for voicing contrast in these languages). This is a language-specific phonological constraint on transfer. In a similar vein, Antoniou et al. (2011) observe that while the VOT of English stops shifted towards Greek (a short-lag language), the reverse was seen only in a specific subset of sounds: word-medial non-nasalized stops. They attribute this to the low frequency of these sounds in Greek, making them less "stable", and thus more amenable to cross-language influence. This suggests that the distribution of sounds in a language could affect the kind of transfer patterns observed.

These results highlight that beyond interaction during the cognitive processing of sounds,

phonetic transfer is ultimately a linguistic phenomenon, and thus subject to language-specific phonological constraints. Therefore, to make meaningful generalizations, it is imperative to consider data from a variety of language pairs. Existing research on phonetic transfer largely revolves around a limited set of phonologically related languages. Thus, the present study extends the scope of this research to a new pair of languages— Indian English and Bengali.

Since Bengali has a four-way laryngeal contrast, VOT does not encode a binary voiced-voiceless distinction. Moreover, VOT is not the only relevant cue for the corresponding phonological categories in the language (Dmitrieva & Dutta, 2020). This means that any effect of L1 Bengali on L2 English VOT is likely to be affected by additional phonological factors. Thus, avoiding any potential confounds was another reason we chose to focus on vowel quality as a target for transfer in the present study.

1.5   Asymmetries between sounds

Asymmetries in the extent and patterns of transfer have not only been observed between languages, but also between different sounds/features of a language, in both long-term and transient interactions. Studies that have examined multiple sound categories have found that interactions between individual sound pairs do not necessarily reflect the overall pattern of global (system-wide) shift (e.g. Chang, 2012; Elias et al., 2017, vowel quality), suggesting that 'extent of transfer' cannot treated as an atomic measure. In light of the discussion about VOT in section 1.4, this is not surprising — if a general linguistic principle of 'room for movement' and contrast constrains transfer, then we should expect it to apply to individual sounds too. Once again, this emphasizes the importance of examining a wider range of sound contrasts.

In the present study, we focus on two vowel categories in Indian English: the mid-central vowel [ʌ] and the low front vowel [æ]. There are two plausible sources of asymmetry between

these:

1. Position in the IE vowel space: compared to [æ], [ʌ] exists in a part of the vowel space that has a lower vowel density. This affords a greater latitude for movement without risking the loss of a contrast, particularly in the vowel height (F1) dimension. Thus, considering purely phonological constraints on IE, we expect that a greater degree of shift is possible in [ʌ] compared to [æ]. Any movement in [æ] is expected to be primarily in the backness (F2) dimension.

2. Target of transfer: findings from existing research on cross-language phonetic interaction suggest that changes in production during mixed-language use are not random, but rather targeted with respect to categories in the other language. Thus, another source of asymmetry is the fact that the category [ʌ] is absent in Bengali, whereas [æ] is a common category across both languages (cf. sec.1.1).

Flege's Speech Learning Model (SLM) (1995; 2007) posits that sound categories which are common across languages influence each other because they share a common acoustic-phonetic space. Thus, if Bengali and English differ in their canonical realizations of [æ], then we should expect the English [æ] to shift towards the corresponding Bengali category in the mixed condition.

There is no obvious competing L1 category during the production of [ʌ]. However, it is unlikely that this should altogether preclude a shift in [ʌ], since at least one existing study has reported transfer effects on a non-common vowel category in a comparable paradigm (Simonet, 2014). Here, participants' Catalan [ɔ] shifted towards an acoustically close category [o], which is common to Catalan and Spanish, suggesting that movement might be towards a set of phonetic norms rather than a specific corresponding category.

We expect the vowel [ʌ] to shift towards an acoustically/perceptually close category in IE— a probable candidate being the low vowel [aː], as anecdotal observations of Bengali-

accented English suggest that the sounds are perceived as close by Bengali speakers.

In the present study, our research question concerns whether L2 vowels are produced differently in nonswitch vs switch conditions. Previous literature on phonetic transfer suggests that such shifts are not random, but rather directional and targeted. Although our primary research question merely concerns whether productions in the two conditions differ, we wanted to understand how the direction of shift compares to corresponding Bengali categories. To estimate average baseline differences in vowel quality between the relevant English and Bengali categories, we processed and used existing speech data from a freely available Bengali corpus, SHRUTI (Das, Mandal, & Mitra, 2011). The procedure is described below:

Baseline Bengali data from the SHRUTI corpus

The SHRUTI corpus is the largest free corpus of Western Bengali speech that we are aware of. It comprises of 21.64 hours of read speech, containing 7383 unique sentences (22012 unique words) from Bengali newspapers in different domains, covering the most frequently used words of the language. The speakers are 34 adults (age range 20-40, 75% male, 25% female) who have grown up in the Indian state of West Bengal, and speak standard colloquial Bengali. All of the speakers have university-level education or higher. Given these parallels in linguistic, regional, and socio-economic background, we believe that Bengali speech from this corpus is representative of the average target Bengali vowel categories that are of interest in this study. To estimate average vowel positions, we extracted formant values from syllable-initial productions of [æ] and [aː], at 55% into the vowel, to get estimates from the steady state, which should most closely approximate the production target. Next, we compared these to our participants' productions in the non-switched English condition to understand the relative positions of these categories. This is visualized in figure 3.

Note that we refrained from using corpus data for average Indian English vowel positions, opting instead to use the participant data from this study, because IE has high regional

variability (c.f. discussion in section 1.1). To the best of our knowledge, there is no freely available corpus of IE speech from Bengali speakers in India.

The F1×F2 plot indicates that the Bengali category [æ] is more fronted, and [aː] lower and more fronted, compared to the English categories [æ] and [ʌ] respectively.

1.6   A note on paradigms: connected speech or not?

This section motivates a secondary aim of this study: to verify if the extent and patterns of phonetic interaction in bilingual participants are independently mediated by the choice of experimental paradigm. Studies of phonetic transfer during mixed-language use have largely employed spontaneous or scripted code-switching (CS) to elicit mixed productions. CS is defined as the use of multiple languages within a single utterance, i.e. in connected speech (Myers-Scotton, 1993). Other paradigms which have been used include cued picture-naming, delayed repetition, interpreting across languages, reading word-lists etc. Olson (2013) argues that code-switching introduces discursive factors such as discourse planning and pragmatic considerations, which ultimately reflect patterns of language practice, rather than cognitive behavior. This finds support from studies which report transfer effects before the switch point (Bullock & Toribio, 2009), or other evidence for planning such as hyperarticulation (Muldner et al., 2019) and interlocutor-awareness in spontaneous CS (Khattab, 2013). To avoid such confounds, Olson (2013) endorses cued picture-naming (where pictures are named in rapid succession in either language, based on the given language cue) as an alternative paradigm to isolate purely phonetic effects in controlled setups. This observation has important implications for how the existing literature on short-term phonetic interaction is interpreted: if the discursive properties of a CS paradigm independently affect the phonetic outcome of mixed-language use, this calls to question the comparability of results across studies using these different paradigms. Because the reported outcomes of short-term pho-

netic interaction vary so greatly across studies (cf. section 1.4), generalizations about the mechanisms that underlie these patterns often rely crucially on the assumption of such comparability. Therefore, identifying paradigm-specific differences is particularly relevant for this line of research.

Note that while there is evidence for planning in CS, precisely how this affects phonetic interaction cannot be predicted from the existing literature: while Bullock and Toribio (2009) found tokens immediately preceding the switch point to be phonetically distinct from the rest, one participant group converged towards the switch language (explained in terms of proficiency) while the other showed divergence (explained in terms of extra-linguistic factors). The participants in Muldner et al.'s (2019) study hyperarticulated switched tokens (indicating some degree of planning), and did not evidence a significant shift during mixed-language use. Moreover, recall that the findings of Olson (2016) suggest that there is a limit on the extent of transfer— multiple factors do not necessarily result in additive effects. Thus, the presence of additional factors in CS does not necessitate that this should alter the extent of observable transfer compared to any other paradigm. It is possible that the outcomes in the studies discussed above are better attributed to some other aspect of the experimental setup, rather than to the act of code-switching in particular.

The precise contribution of various (linguistic and extra-linguistic) factors and their interactions to the phonetics mixed-language use merits detailed investigation in its own right. In the present study, however, our focus is limited to verifying whether, in an experimental setting, the postulated discursive elements of a CS paradigm alter the extent of observable transfer compared to a paradigm that arguably lacks these: cued picture-naming. We are less concerned with the exact nature of differences than with their comparability. While researchers have successfully employed each of these paradigms to induce phonetic interaction, comparison across studies is difficult since they also differ in other factors. We provide a starting point for such comparison by eliciting productions of the same tokens from the

same set of participants, in both paradigms. A difference in results will suggest that the planning involved in connected speech and code-switching affects transfer above and beyond the phonetic outcome of switching itself, which has implications that future work should consider.

Research questions and hypotheses

- RQ1—L2 vowels: Are L2 vowel categories of proficient bilingual speakers affected by dynamic phonetic transfer during mixed-language production?

- H1—L2 vowels: We expect to see differences in the spectral properties of L2 vowels in switched vs non-switched tokens

- RQ2—Asymmetry: Do patterns of transfer differ between the two vowel categories?

- H2—Asymmetry: Since existing literature demonstrates targeted and phonologically-constrained shifts, we expect transfer effects to be category-specific. Based on the position of IE vowels compared to those of Bengali vowels from Das et al. (2011), we expect [ʌ] to be lowered and fronted in the mixed condition. [æ] is expected to be fronted. We hypothesize that the extent of shift in [æ] to be lesser than [ʌ], as it is in a part of the IE vowel space that has greater vowel density.

- RQ3—Paradigm: Is there an observable difference between the extent of transfer in a cued picture-naming vs. a code-switching paradigm?

- H3—Paradigm: While researchers have discussed code-switching as a source of discursive factors likely to affect transfer behavior, this has not been systematically studied. Existing literature does not allow us to predict precisely what form this effect might take (c.f. discussion in Section 1.6). Thus, this question is exploratory, and meant to serve as a starting point for more within-study comparisons.

2 Methodology

The experiment had the following factorial design: 10 participants* 20 target words* 2 conditions* 2 tasks* 4 iterations = 3200 target tokens.

2.1 Participants

Participants were recruited through an open call at the EFL University campus in Hyderabad. Interested volunteers completed a brief language background questionnaire which was used to screen participants as described below.

As noted in section 1.1, the population is multilingual and linguistically diverse. The university campus comprises of students from different parts of the country and world, with varying L1s and different varieties of English. Our aim in screening participants was to minimize between-participant random variability, while maintaining ecological validity considering the actual day-to-day settings in which transfer behaviors likely take place. We chose two metrics to do this: (i) where the participant grew up; and (ii) amount of formal education in Bengali. The first ensures that the variety of Bengali that individuals are exposed to in their surroundings is similar across participants. This would mean that the target representations for Bengali vowels would be expected to be similar. This was especially important because we were not eliciting unilingual L1 Bengali data in this study, and were using an existing data source to estimate the position of Bengali vowels in the formant space. Based on previous literature (c.f. section 1.5), we expect that L2 vowels shift towards related L1 categories during mixed language utterances. Thus, we wanted to ensure that any between-participant asymmetries in the pattern of shift reflected individual differences in transfer, rather than different acoustic targets. Another reason to control for location was that it is a fairly reliable predictor for language(s) of education(see Supplementary materials for details).

The second metric is Medium of Instruction (MoI) in school. Sustained formal education in a language is a reliable predictor of LSRW (listening, speaking, reading, writing) skills. Since this study uses written stimuli, we wanted to ensure that all participants were sufficiently comfortable with reading and speaking both Bengali and English, to avoid variability due to reading/task effort.

Given these considerations, we only recruited participants who had grown up in the state of West Bengal and lived there for the majority of their lives, had parents who both spoke Bengali, and had received at least 5 years of formal education in Bengali (all respondents had received 5+ years of education in English and were pursuing university degrees that are taught in English), and self-reported being proficient in speaking, understanding, and reading Bengali.

A total of 28 bilingual speakers of Bengali and English (18 female) responded to the recruitment call. Using the inclusion criteria reported above, 10 volunteers (5 female, 5 male, age range 19 to 28) were invited to participate in a speech production study, and were compensated for their time. After completion of the study, we followed up with a detailed language background survey at a later date to learn more about the language profiles of the participants. This survey was a modified and consolidated version of three popular bilingualism profiling tools: the Bilingual Language Profile (BLP) (Birdsong, Gertken, & Amengual, 2012), the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld, & Kaushanskaya, 2007) which incorporates language history and self-assessed proficiency, and the Bilingual Switching Questionnaire (BSWQ) (Rodriguez-Fornells, Kramer, Lorenzo-Seva, Festman, & Münte, 2012). Together, these incorporate questions about language history, usage, attitudes, and switching experience. The complete questionnaire, along with responses, is available in the "Data" component of the OSF repository for this project (c.f. section 5). –footnotes?

Given our inclusion criteria, we were able to recruit 10 participants for this study. Al-

though existing studies in the field have used comparable participant numbers, we acknowledge that this is a small sample size. As detailed above, even after a recruitment call that already controlled for language background and education level (recruiting Bengali-English bilinguals in a university campus), we were able to include only 10 out of 28 volunteers (35.7%) in the final study. Note that the criteria used here establish only a minimal level of between-participant similarity in exposure, dominance, habits, and proficiency, such as is extremely common in most comparable studies. This demonstrates the degree of linguistic variability in the present population. Norms for experimental research on bilingualism have largely been established in a Western context. In settings where linguistic variability at the population level is much greater than these canonically studied groups, establishing comparable levels of experimental control requires significantly greater resources. Given that small sample sizes are likely to have low statistical power, replication with larger sample sizes are important to confirm the observed effects, and provide much-needed data diversity (c.f. discussion in section 1). A more detailed discussion of power analyses in bilingualism research is included as supplementary material on OSF (c.f. section 5). We hope that this can facilitate more conversation around establishing norms for power and sample size in our field.

2.2 Stimuli

Twenty monosyllabic English words: ten containing the vowel [ʌ] and ten containing [æ], were selected as target words. To minimize coarticulatory effects on the vowel, the pre-vocalic consonants were all voiced plosives ([b] or [d]). 10 monosyllabic English words that do not contain the target vowel served as filler items. We avoided English words that are already lexicalized loanwords in Bengali. The complete list of target words can be found in the "Materials" component of the OSF repository for this project (c.f. section 5).

2.3   Procedure

Participants were recorded individually in a sound attenuated room. Productions were recorded in Praat (Boersma & Weenink, 2016) at a 44.1K sampling rate. Each participant completed four iterations each of two tasks— cued picture-naming and code-switching. Utterances in the two conditions (nonswitch and switch/mixed) were randomly interspersed within each iteration. In the nonswitch condition, the target word was preceded by another English word in the utterance. In the mixed condition, the target was preceded by a Bengali word and thus produced as a switch from Bengali. Every target word appeared twice during a test block; once in a nonswitch and once in a switch condition. Participants alternated between the two tasks, and were allowed to take a short break after each test block. The tasks are described below:

1. Cued picture-naming: Participants were presented with slides in the following sequence:

   - Language cue: a word in either English or Bengali orthography
   - A picture displayed for 50 ms; named by the participant in either English or Bengali, depending on the language cue. The picture of a fish was used in all trials, prompting the participant to say either [fɪʃ] or [maː tʃʰ], depending on the language cue
   - Target: an English word; read out by the participant
   - Distracter math problem

   Since all the target words were in English, the language cue was used to manipulate the preceding word, giving target words in non-switched and switched (mixed) utterances. The orthographies of English and Bengali are visually distinct, and were therefore used to cue language.

2. Code-switching: Each slide contained one English target word embedded in either an English (nonswitch) or a Bengali (mixed; target produced as a switch from Bengali) carrier sentence which was read out by the participant. The carrier sentences are given below:

- Unilingual sentence: That is a yellow [target word].

- Mixed-language sentence: o-ʈa æk-ʈa kalo [target word].

Gloss and translation for Bengali sentence:

o-ʈa   æk-ʈa   kalo [target word]

that-clf one-clf black [target word]

"that is a black [target word]"

Carrier sentences were chosen such that the sound preceding the target word was uniform across all sentences (the final sound in yellow is produced as [o] in Indian English), to minimize any possible differences due to coarticulatory effects on the target vowel. To ensure identical word order in both languages, we chose a specifier-noun construction, as this has a parallel structure in Bengali and English. This gave a uniform prosodic context for the target word across conditions. While presenting the mixed-language utterances, we used mixed orthography.

2.4 Acoustic analysis

The first iteration of each task was treated as a trial, and excluded from further analyses. An additional 67 tokens (2.79%) were excluded due to errors in presentation/production, giving 2333 target word tokens ([ʌ]: 1137, [æ]: 1196) for analysis. Recordings were manually

segmented and annotated in Praat (Boersma & Weenink, 2016), and the first and second formant frequencies (F1 and F2) of the target vowels [ʌ, æ] were measured at five points: in 10% increments starting at 5% into the vowel. This gave measures of F1 and F2 at 5%, 15%, 25%, 35% and 45% into the vowel, allowing us to observe any dynamic changes in the extent of transfer over the course of producing the vowel. Individual vocal tract differences affect formant frequencies across speakers. To reduce between-speaker differences, we used z-score normalization (using the PhonR package (McCloy, 2016) in R (R Core Team, 2013), representing each speaker's formant frequencies as the scaled distance from their own mean). Plotting these vowels shows a difference between male and female speakers, such that male speakers have overall lower formant frequencies (figure 5a). Speaker normalization using the z-score method does not get rid of this difference. Since gender-related differences in formant frequencies are not of interest in this study, we normalized individual values by gender (representing each value as a scaled distance from the group mean; figure 5b). Statistical analyses were performed on these normalized values.

< Insert figure 5a,5b about here >

3   Results

All statistical analyses were performed in R (R Core Team, 2013). We used linear mixed effects models to test the effect of experimental conditions on vowel height (F1) and backness (F2) respectively. Since formants were extracted from five points over the course of articulating the vowel, time was coded numerically (1-5), and treated as an independent variable. This was scaled and centered prior to analysis and treated as a continuous predictor. We fitted models using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015), and used lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) to obtain estimates of p values for individual variables. For all mixed models, the alpha criterion was set at $|t| > 2$. The de-

pendent variables were normalized F1 and F2, and we fitted separate models for these. The independent variables or fixed factors, along with their levels, are given in Table 2. Subject and Item were treated as random factors.

< Insert Table 2 about here >

3.1 Models

Fixed effects

Our study involved two IE vowels. Vowel category is expected to systematically affect formant frequencies (FFs). Thus, Vowel is a fixed effect in the models. The experimental treatment of interest is Context, i.e. whether the English target item was presented in an English (nonswitch) or Bengali (switch/mixed) context. Our main research question (L2 vowels) asks whether L2 vowel quality (formant frequencies) of proficient bilinguals is affected by the Context of utterance. We hypothesize that the direction of such shift in the mixed condition, if any, is towards a corresponding L1 category. Therefore, we expect the effect of Context on formant frequencies, i.e. the direction of shift, to be moderated by the vowel category (c.f. the discussion in Section 1.5, leading to our second research question Asymmetry). Thus, the main effect of interest is the interaction term Context*Vowel. This corresponds to the transfer pattern. Our third research question (Paradigm) asks whether the language-mixing paradigm, or Task, affects the extent and direction of transfer. If this is the case, we expect Task to interact with the transfer term, giving a three way Task*Context*Vowel interaction. Since we measured FFs at five points in the vowel, Time is expected to affect the FF. However, the precise effect of formant dynamics is expected to depend on the vowel category. Thus, we expect the interaction of Vowel*Time to predict

formant frequency. If transfer patterns change across time, then we should expect it to interact with the transfer term, giving a Context*Vowel*Time interaction.

Random effects structure

Subject and Item were treated as random factors. Given our hypotheses, we included random intercepts for Subject and Item, by-subject random slopes for Vowel (formant targets for vowel categories differ across individuals) and Context (individuals differ in their response to the experimental condition), and a by-item random slope for Context (the effect of experimental conditions differ across words).

We tested the significance of our hypothesized explanatory variables using t-values and p-values for individual variables in the models, and tested overall model fit by comparing the full model to a corresponding null model that lacks the variable of interest using ANOVAs. The following subsections summarize and discuss the model outputs for F1 and F2 respectively.

3.2   Vowel Height – F1

Model output summary

First, to test our first and second research questions (L2 vowels and Asymmetry), we fitted a model with vowel*time, and vowel*context as fixed effects. The transfer term (vowel*context) is significant in this model (model coefficient/ unstandardized effect size (β)= -2.330e-01, standard error (SE)= 5.244e-02, t= -4.44, p=0.0002). Comparing this to a null model that lacks the transfer term shows that the full model is significantly better ($\chi^2(2) = 14.42, p = 0.0007$). To confirm whether the effect of context is moderated by vowel identity, we compared a second null model that contained Context as a fixed effect, but lacked the Context*Vowel interaction. The full model was better ($\chi^2(1) = 12.51, p = 0.0004$).

To examine the third research question, Paradigm, we fitted a model with vowel*time and task*vowel*context as fixed effects. This three-way interaction term is significant (β= -1.955e-01, SE= 5.453e-02, t= -3.59, p= 0.0003). We also find a significant effect of task on FFs (β=3.300e-01, SE=2.668e-02, t=12.371, p=2e-16). Comparing this model to a null model that lacks the fixed effect Task ($\chi^2(4) = 572.61, p = 2.2e - 16$) shows that the full model is better. Because Task has an independent effect on F1, and adding interactions increases model parameters, we fitted a second null model with Task as a separate fixed effect without the interaction term to test whether the interaction term is needed. However, the full model was still a better fit for the data ($\chi^2(3) = 23.82, p = 2.722e - 05$). Next, we tested two other null models to verify that the full model with the three-way interaction term is indeed the optimal model for the data, by replacing the three-way interaction term with: (i) vowel*context + task*context (to test if task only interacts with context, rather than the transfer term vowel*context). The full model is better ($\chi^2(2) = 23.28, p = 8.78e - 06$); (ii) vowel*context + task*vowel (to test if task only interacts with vowel, rather than the transfer term vowel*context). The full model is better ($\chi^2(2) = 13.36, p = 0.001$). This shows that the three-way interaction is the best fit for the data. Because an independent effect of task on FFs is unexpected (c.f. model interpretation below), we wanted to examine if the effect might be local, i.e. decrease over time. We added a task*time interaction term to the full model, which is significant (β= -1.216e-01, SE= 1.347e-02, t= -9.028, p= 2e-16) and improves model fit ($\chi^2(1) = 81.22, p = 2.2e - 16$).

Next, to examine the dynamics of transfer, we added a three-way interaction between the transfer term (vowel*context) and time. This term has a marginally significant effect (β= -4.604e-02, SE= 2.692e-02, t= -1.71, p= 0.08), but does not significantly improve model fit compared to a null model lacking the vowel*context*time term ($\chi^2(2) = 3.41, p = 0.18$).

Thus, the optimal model was: f1 $\sim$ time*vowel + vowel*context*task + time*task + (1 + context|word) + (1 + context + vowel|subject). Table 3 summarizes the model coefficient

estimates (β), standard errors (S.E.), t-values, and p-values for this model. p-vales smaller than 0.05 are starred.

Model interpretation

We extracted formant estimates from five uniformly spaced time points in the vowel: from the time at 5% into the vowel (at the start of articulation) to 45% (at the mid-part, or steady state, which is expected to correspond most closely to the articulatory target for the category). As expected, we see that F1 values are significantly predicted by Time in the model, reflecting the dynamic nature of articulation which causes acoustic variation over the course of time. The positive slope of the estimate for Time indicates that F1 increases, i.e. the two categories in the study show lowering, over the course of articulation. Moreover, the significant Time*Vowel interaction shows that the F1 trajectory is different for the two vowels. This is expected, as the categories are expected to present different articulatory targets.

Vowel is not a significant predictor of F1. This is initially surprising, given that the phonological category is expected to systematically affect vowel quality. However, a boxplot visualization (Figure 6) shows that the participants in this study vary greatly both in the relative heights of the two IE categories **[ʌ, æ]** and the extent of height difference between them. Note that these measures are only from the nonswitch condition, suggesting that this variability is a feature of the participants' IE vowel categories, independent of the effects of switching. Given that estimates for fixed effects are pooled across participants, this explains the lack of reliable differences between **[ʌ]** and **[æ]**.

The negative slope of the estimate for Context shows that F1 is lower, i.e. vowels are higher, in the nonswitch (e) condition compared to the mixed (b) condition. Put more intuitively, vowels are lowered in the mixed condition. However, the lack significance indicates that using English words in mixed utterances does not lead to a uniform lowering/raising

across vowels. Instead, the significant interaction between Vowel and Context signals that the extent and direction of shift is moderated by vowel category. There is an asymmetrical shift in the categories **[ʌ]** and **[æ]** in the mixed-language condition. Specifically, the Vowel*Context interaction has a negative slope, showing that the lowering is greater in **[ʌ]**. The lack of a significant Vowel*Context*Time interaction shows that transfer patterns do not change across time. This indicates a uniform scaling of formant trajectories in the mixed condition, suggesting that the underlying acoustic targets for these vowels are different during non-switched vs mixed utterances.

The results also reveal a significant effect of Task, such that the sentence task (code-switching) leads to higher F1 values in both vowels relative to the cued picture-naming task as: vowels are lowered in the code-switching task, independent of experimental condition. This is surprising, as we have no aproiri reason to expect a uniform difference in vowel quality across tasks. A significant interaction with Time with a positive slope of the estimate shows that this effect is local: it is strongest at the beginning of the vowel, and reduces over the course of the vowel. This is consistent with the interpretation that the unexpected Task effect is an artifact of the paradigm, possibly resulting from coarticulatory effects owing to the specific stimuli used in this study. Of greater interest to us is the significant Vowel*Context*Task interaction, which indicates a difference in transfer patterns across paradigms. Specifically, the negative slope of the estimate suggests that transfer in the F1 (height) dimension are greater in the code-switching (s) task compared to the picture-naming (p) task.

< Insert Figure 6 about here >

3.3   Vowel backness – F2

Model output summary

First, to test RQ1 (L2 transfer) and RQ2(Asymmetry), we fitted a model with vowel*time, and vowel*context as fixed effects. The transfer term (vowel*context) is significant in this model (standardized effect size (β)= 1.234e-01, standardized error (SE)= 2.055e-02, t= 6.00, p=5.46e-06). Comparing this to a null model that lacks the transfer term shows that the full model is significantly better ($\chi^2(2) = 16.20, p = 0.0003$). To confirm whether the effect of context is moderated by vowel identity, we compared a second null model that contained Context as a fixed effect, but lacked the Context*Vowel interaction. The full model was better ($\chi^2(1) = 16.02, p = 6.245e - 05$).

To test RQ3 (Paradigm), we fitted a model with vowel*time and task*vowel*context as fixed effects. While task has an independent effect on F2 (β= -8.758e-02, SE= 1.756e-02, t= -4.98, p= 6.18e-07), the three-way interaction term is not significant (β= 3.349e-02, SE= 3.565e-02, t= 0.94, p= 0.34). However, there are significant two-way interactions between task*context (β= -1.099e-01, SE= 2.483e-02, t= -4.427, p= 9.66e-06) and task*vowel (β= -8.983e-02, SE= 2.535e-02, t= -3.544, p= 0.000396). Comparing this to a null model that lacks the fixed effect Task confirms that the full model is still a better fit for the data ($\chi^2(4) = 572.61, p = 2.2e - 16$), indicating that an optimal model should contain the term Task. To confirm whether Task interacts with any of the other terms, we fitted a model with Task as a separate fixed effect without the interaction term. Comparing this to the full (interaction) model shows that the latter is still a better fit ($\chi^2(3) = 44.8, p = 1.02e - 09$). Next, we tested two other null models to verify that the three-way interaction term is indeed optimal for the data, by replacing the three-way interaction term with: (i) vowel*context + task*context (to test if task only interacts with context, rather than the transfer term vowel*context). The full model is better ($\chi^2(2) = 17.28, p = 0.0001$); (ii) vowel*context + task*vowel (to

test if task only interacts with vowel, rather than the transfer term vowel*context). The full model is better ($\chi^2(2) = 28.45, p = 6.633e - 07$). This shows that the three-way interaction is the best fit for the data.

Because an independent effect of task on FFs is unexpected (c.f. model interpretation below), we wanted to see if the effect might be local, i.e. decrease over time. We added a task*time interaction term to the full model, which is significant (β= 6.237e-02, SE= 8.876e-03, t= 7.027, p= 2.23e-12) and improves model fit ($\chi^2(1) = 49.271, p = 2.23e - 12$).

Next, to examine the dynamics of transfer, we added a three-way interaction between the transfer term (vowel*context) and time. This term is not significant (β= -6.274e-03, SE= 1.774e-02, t= -0.35, p= 0.72), and the addition does not significantly improve model fit compared to a null model lacking the vowel*context*time term ($\chi^2(2) = 2.2443, p = 0.32$).

Thus, the optimal model is: f2 ~ vowel*time + vowel*context*task + task*time + (1+context|word) + (1+context+vowel|subject). Table 4 summarizes the model coefficient estimates (β), standard errors (S.E.), t-values, and p-values for this model. p-values smaller than 0.05 are in starred.

Model interpretation

As with F1, we find that Time significantly influences F2 values in the model, reflecting the dynamic nature of speech which causes acoustic variation over the course of articulation. Moreover, the Time*Vowel interaction shows that this trajectory differs across the two vowels: while the positive slope for Time indicates that F2 increases i.e. vowels are fronted with time, the negative slope of the Time*Vowel interaction shows that this effect is smaller for **[ʌ]**. This is expected, as figure 3 confirms that the category **[æ]** is overall more fronted than **[ʌ]**.

Unlike in the F1 dimension, Vowel is a significant predictor of F2, with **[ʌ]** having significantly lower F2, i.e. being produced as more backed, than **[æ]**. As with vowel height, there is

no significant independent effect of Context, showing that mixed language use does not lead to uniform fronting or backing across vowels. This is not surprising, as we expect movement to be towards the corresponding L1 category, and thus differ across vowels. The negative slope of the estimate indicates that on average, vowels have lower F2 i.e. are backed, in the nonswitch (e) condition. In other words, vowels are fronted in the mixed condition. This effect is greater in **[æ]**, as indicated by the negative slope of the Vowel*Context interaction. In the F2 dimension, too, the lack of a significant Vowel*Context*Time interaction shows that transfer patterns do not change across time, suggesting that the underlying acoustic targets are different in the two conditions.

We see a significant effect of Task on F2, such that vowels in the code-switching (s) task have lower F2, i.e. are more backed than those in the picture-naming (p) task. As with F1, this is an unexpected result; we have no principled reason to expect code-switching to induce an across-board backing of vowels. Vowels are backed in the code-switching task, independent of experimental condition. A significant interaction with Time, with a positive slope for the estimate, shows that this effect is local: it is strongest at the beginning of the vowel, and reduces over the course of the vowel. This aligns with the possibility that the unexpected Task effect is an artifact of the paradigm and the particular stimuli used. The Vowel*Context*Task interaction is of interest, as it relates to our research question of how the test paradigm affects transfer. Unlike vowel height, we do not find patterns of shift in vowel backness (F2) to differ across paradigms, as indicated by the lack of a significant effect for the fixed effect.

< Insert Table 4 about here >

.

Statistical analyses reveal that the vowel **[ʌ]** is lowered and fronted, while **[æ]** is fronted, in the mixed condition. Based on previous literature, we expect that such shifts in vowel categories result from increased cross-language influence during mixed-language use, i.e.

shifts are targeted with respect to L1 categories. As discussed in section 1.5, although our primary research question is about whether productions in switch vs nonswitch conditions differ, we wanted to understand how the direction of shift compares to corresponding Bengali categories. Figure 3 shows the non-switched IE productions of the participants in relation to average positions of related vowel categories in Bengali. Figure 4 overlays productions in the mixed condition over this, showing F1×F2 vowel plots for IE vowel productions in the nonswitch and mixed conditions, in comparison to the related Bengali categories from the SHRUTI corpus Das et al. (2011). Given that the Bengali corpus data represent average vowel positions, and are not from the same set of speakers as the participants of this study, we refrained from including these in statistical analyses. However, figure 4 provides a heuristic confirmation for our informal hypotheses in section 1.5) that the IE **[æ]** moves towards the corresponding **[æ]** category in Bengali, whereas the shift in **[ʌ]** is towards the related Bengali category **[aː]**. The following sections discuss the implications of these results.

< Insert Figure 4 about here >

4   Discussion

Taken as a whole, these results bear on our original research questions in the following ways: regarding our first question of whether L2 vowels of proficient bilinguals can be affected by short-term transfer from L1 (L2 vowels), our results indicate that this indeed happens during mixed-language use. Both the common category **[æ]** and the L2-exclusive category **[ʌ]** show spectral shifts during mixed-language productions compared to non-switched productions. As hypothesized (c.f. RQ2 and H2, Asymmetry), the results reveal asymmetries in the extent and direction of shift between the two vowels. Specifically, **[ʌ]** is lowered and fronted, while **[æ]** is fronted, with the extent of shift being greater in **[ʌ]**. Comparison with canonical Bengali vowel productions shows that the shift in **[æ]** is towards the corresponding Bengali

category **[æ]**, while that in **[ʌ]** is towards the Bengali category **[aː]**. Regarding our third research question of whether the paradigm used to elicit language switching independently affects patterns of transfer (RQ3: Paradigm), analyzing the role of Task in the results confirms this. We find a greater degree of transfer during code-switching compared to the cued picture-naming task. However, this difference is limited to the F1 dimension, and needs to be replicated. These results and their implications are discussed in the following subsections.

4.1 Phonetic transfer in L2 vowels

The main aim of this study was to examine short-term phonetic transfer in L2, towards understanding language processing in proficient bilinguals. Two interrelated questions, discussed in section 1, bear on this: (i) what causes transfer in mixed-language speech; and (ii) what causes observed asymmetries between transfer in L1 vs L2. Together, these serve as the basis for our theories about the language systems of multilingual individuals exist in relation to one another, and consequently, what cognitive models of language should look like.

What causes transfer: language mode vs switching   An often discussed issue in transfer research is whether cross-language interaction at the phonetic level during language mixing is triggered by the co-activation of languages that makes representations from both languages available, or by the local act of switching during online processing and articulation of those sounds. In this study, we presented nonswitched English and code-switched Bengali-English stimuli in the same session, such that the language mode (participants' global language "setting") remained the same throughout. In spite of this, results demonstrate consistent differences between switched and non-switched tokens. This confirms that the local act of switching does trigger transfer, even in a bilingual language mode, paralleling results from Olson (2016); Piccinini and Arvaniti (2015); Tsui et al. (2019). These studies all examined

the effect of switching on a temporal feature of speech, namely voice onset time or VOT. The results of the present study indicate that the same happens in vowels. Moreover, the exact effect of transfer is interesting: VOT is a one-dimensional measure. Thus, observing an increase or decrease in VOT in a mixed language condition makes it difficult to infer whether a consonant has moved towards a specific L1/L2 counterpart, or reflects a more general system-wide shift in articulatory norms. Since vowel quality measures show directionality as well as magnitude, we can observe finer details in the patterns of shift. In the present study, we find that productions of English vowels do not merely shift towards a Bengali counterpart, but rather towards a set of Bengali-like norms: a vowel category that doesn't exist in Bengali nonetheless moves towards a related (perceptually close) Bengali category. This result parallels Simonet (2014), who found the Catalan **[ɔ]** of bilingual speakers shifts towards a related Spanish category, **[o]**, in a bilingual language mode (note that they did not examine code-switched speech, and focused on vowel height as a target of transfer). The study showed that co-activating multiple languages within a discourse causes a transient increase in cross-language phonetic interaction, leading to changes in both L1 and L2 (moderated by language dominance). This is interpreted as a follows: a bilingual mode temporarily co-activates both language systems, making both L1 and L2 representations available during processing. These representations interact while producing speech in such a language mode, leading to cross-language phonetic transfer. But what does this interaction look like, and how are these abstract representations processed during the act of producing speech? Examining speech during switching can shed light on this.

The time course of such shifts provides an insight into this. Vowels in the mixed condition do not start out the same, eventually deviating in acoustics from their nonswitch counterparts. Instead, we see a difference between the conditions right from the start of articulation. This suggests that such interaction involves the speech planning stage. The pattern of transfer observed in this study, particularly for **[ʌ]**, also demonstrates that be-

yond a simple L2-L1 mapping, transfer during online processing there makes reference to entities beyond the level of the individual sound. Specifically, such transfer demonstrates a system-level interaction: a sound that does not exist in the L1 repertoire nonetheless acquires a different target, and the target is related to the L1 system in a meaningful way. This cannot be reasonably explained as articulatory inertia carried over from uttering L1 sounds. It is also difficult to account for in a purely exemplar framework, as the discourse does not contain any instances of "L1 [ʌ]" that could affect the L2 category. If representations are combined in creative ways during online speech processing within an utterance, then can we think of intermediate representations not just for individual sounds, but rather for sound systems? While more detailed empirical data will eventually enable us to clarify such questions, this approach aligns with sociophonetic understandings of bilingual phonetic variation, for example as discussed by Khattab (2009, 2013) based on the use of phonological variables by Arabic-English bilingual children. This visualizes the bilingual phonetic system as "a rich store of phonological representations", comprising of both language-internal and cross-language variation, from which the bilingual speaker is constantly choosing, guided by a variety of pressures. We propose that vowels are a good feature to examine the plausibility of such systems as cognitive models for language processing: in addition to continuous measures (unlike phonological variables), they also provide two-dimensional data (unlike VOT) that is suited to further test and refine predictions from such a model.

Differences between L1 vs L2 transfer: Studies of short-term phonetic transfer consistently find differences between transfer patterns in L1 vs L2. The bulk of this literature is based on the feature VOT. While vowels have been shown to induce L2-to-L1 influence, the present study demonstrates L1-to-L2 transfer in vowels.

Previous studies of language-switching have often reported phonetic transfer from L2 to L1, but a lack of parallel transfer from L1 to L2. One postulated model that is consistent

with this finding is the Inhibitory Control Model (ICM), first proposed by Green (1986) at the lexical level, but extended to sound systems by Olson (2013); Tsui et al. (2019). This proposes that the lack of transfer effects on L2 reflects the lower inhibition required on L2 representations in the first place, which makes switching into L2 easier, compared to the converse situation, when the high initial inhibition on L1 induces phonetic transfer while switching back into L1. As discussed in section 1.2, vowels provide a fertile testing ground for this theory in sound systems, as they avoid the confound of independent language-specific differences between long- and short-lag VOT languages. Both vowels in this study showed systematic shifts in the mixed-language condition compared to baseline non-switched productions. In the context of an ICM-based analysis of between-language asymmetries in phonetic transfer as proposed by Olson (2013), this is a significant finding: it demonstrates that the postulated lower switch cost on an L2 does not entirely mask transfer effects on L2 vowels. There are a couple of possible explanations for this. To begin with, the present study focuses solely on L1 (Bengali) to L2 (English) transfer, and not the converse. It is theoretically possible that these participant are L2 English dominant. That would mean that switching from L1 into L2 induces a greater switch cost, leading to transfer effects on switched L2 productions. This would lead us to expect that if the same participants are asked to switch from English to Bengali, we would not see any evidence of transfer. While more empirical data could clarify this, the language background of the present participant pool makes this scenario unlikely. All participants report regularly using both English and Bengali in a variety of settings, and feeling equally comfortable using bot languages. Listening and speaking skills are expected to be most relevant to phonetic transfer in daily language use scenarios, and participants' self-ratings for these skills are at ceiling for both languages. One point to note is that participants rated their reading proficiency in Bengali to be lower than that in English. Given that the stimuli in this task were presented visually, differences in reading proficiency are a potential source of explanation. However, anecdotal observation

suggests that phonetic transfer in L2 productions is common in daily language use too, even in the absence of reading. This raises questions about an ICM-based analysis of cross-language asymmetries in phonetic transfer that future research should investigate.

Moreover, our findings differ from Tsui et al. (2019), who report on the basis of VOT measurements that balanced bilinguals do not evidence an effect of language-switching. Tsui et al. attribute the absence of transfer effects to greater experience with language switching in these participants, leading to lower switch-costs in both languages. As noted in 2.1, the participants in this study are highly proficient bilinguals with extensive experience of language switching. Moreover, they live in an environment where they are constantly exposed to mixed-language stimuli, and are thus expected to be habituated to drawing on multiple language systems during online language processing. They nonetheless demonstrate cross-language transfer effects in L2 vowel production. One way to think about these differences is that vowels are generally more prone phonetic transfer than temporal features such as VOT. Thus, following from the discussion in sec. 1.4, it might be fruitful for future studies to focus on vowels as a potential site for transfer. This would both expand the empirical scope of existing research (cf. sec. 1.5) and avoid confounding factors arising from independent differences between long- and short-lag VOT languages. More generally, these findings point to a need for expanding the range of features in which transfer effects are studied, in order to make robust generalizations about the mechanisms underlying observed patterns. This is an example where studying different features of speech (e.g. VOT vs vowel quality) allows us to compare the predictions of different underlying process models. Transfer studies usually focus on a single feature due to ease of creating materials. Measuring multiple features within a single design, with the same group of participants, promises to be a powerful tool to compare different models at little added cost to researchers. This would sidestep the unavoidable confounds that come with comparing results across studies, or even replication.

Beyond methodological considerations, however, we argue that these differences ulti-

mately raise a broader question about a purely inhibition-based understanding of transfer. In light of a compelling body of literature on the sociophonetic and communicative implications of phonetic variation along the L1-L2 continuum (e.g. Agnihotri, 1979; Khattab, 2013; Stuart-Smith et al., 2011), is it wise to think of transfer as resulting from an inability to suppress and separate language systems? Given that young children are able to use minute phonetic variation in sophisticated ways, might it not be more intuitive to think of transfer as a communication-enabling feature of language, even at a purely cognitive level? Research on the cognitive mechanisms underlying cross-language interaction and variation in bilingual speech more generally is still an emerging and exciting line of inquiry. We believe that a communication-centered approach to bilingual variation is essential to understand how bilingual language processing functions in the real world. To this end, native multilingual populations serve as an important source of data on language behavior that has been largely ignored, but should be harnessed.

4.2   Asymmetries in Phonetic Transfer

The literature on short-term phonetic transfer has frequently noted differences in transfer patterns between different sound categories in a language. This has been attributed to both general linguistic principles such as contrast preservation, and language-specific factors such as frequency and distribution of categories. Examining these asymmetries more closely, and understanding their sources, helps us figure out what factors affect/moderate transfer and thus, by extension, what level of representation and processing this interaction happens at, and the nature of short-term intra-speaker variation. The present study examined transfer in two L2 categories, **[ʌ]** and **[æ]**. We hypothesized an asymmetrical shift between the two vowels (Asymmetry), owing to differences in:

- their position in the Indian English vowel space: **[æ]** is in a more crowded part of the

vowel space

- the presence/absence of a corresponding category in the L1 Bengali inventory: **[æ]** is a common phonological category in Bengali and English, while **[ʌ]** does not exist as a category in Bengali

The results of the study indicate that the effect of the experimental condition differed across the two vowels, showing differences in both the magnitude and direction of shift. Overall, the magnitude of shift is greater in **[ʌ]**. Recall (figure 2) that the category **[æ]** is in a denser part of the IE vowel space than **[ʌ]**. Specifically: **[æ]** is already at the lowermost part of the vowel space. Moreover, large movements upwards in the F1 (vowel height) dimension would risk a conflation of the category with the neighboring **[ɛ]**. In contrast, **[ʌ]** (indicated as **[ə]** in the chart) is in the central part of the vowel space and is not flanked by any immediate neighboring categories. Thus, the difference in magnitude of transfer aligns with our hypothesis Asymmetry that the extent of shift in individual phonological categories is constrained by the linguistic principle of contrast preservation. An alternative way to interpret this result is that the initial distance between IE **[ʌ]** and its target L1 category **[aː]** was greater than that between the IE and Bengali **[æ]** categories, leading to a greater visible shift in the former. However, our results also parallel previous studies with VOT that have consistently observed greater shifts in the long-lag VOT language (which has a greater acceptable range of VOT, i.e. more room for movement) compared to the short-lag VOT language, and attribute these differences to the language-specific distribution of categories (Antoniou et al., 2011; Bullock & Toribio, 2009). This suggests that a general principle of "room for movement" does play a role in moderating transfer. More broadly, these results collectively suggest that phonetic transfer (i) makes reference to system-level representations that go beyond individual sounds; and (ii) respects linguistic principles. This has implications both for our understanding of phonetic transfer as a phenomenon (e.g. on

which mental representations and at what level of processing and is this occurring?) and language as a system (e.g. how have transfer phenomena shaped phonological systems in multilingual settings where languages have co-existed for long periods of time? What are the linguistic implications of increasingly cosmopolitan urban spaces and identities?).

Our results also indicate asymmetry in the direction of shift between the two categories: the category **[ʌ]** is lowered and fronted in the mixed language condition, while **[æ]** is fronted. On the basis of previous literature indicating that both long-term and short-term shifts in phonological categories of multilingual speakers is targeted towards specific categories in the other language, we hypothesized (Section 1.5) that participants' IE **[æ]** would shift towards the Bengali **[æ]**, and **[ʌ]** towards the Bengali **[aː]**. As figure 4 indicates, comparing transfer patterns with Bengali categories from Das et al. (2011) suggests that this is borne out. The results of **[ʌ]** align with Simonet's (2014) finding that a category which is not common to L1 and L2 nonetheless shifts towards an acoustically/perceptually close category. Once again, we see that transfer phenomena make reference to system-level representations.

4.3   Switching paradigm

Our third research question, Paradigm, concerned the effect of experimental paradigm on the phonetic outcome of mixed-language use. Olson (2013) proposes that in both scripted and spontaneous code-switching paradigms, "discursive properties may partially drive the production of code-switches, masking underlying effects of interaction", and thus proposes cued picture-naming as a preferable alternative. If such discursive properties independently affect the phonetic outcome of language mixing in an experimental setup, this calls to question the comparability of results across studies using these different paradigms. We verified this by comparing productions of the same items from the same group of participants in a scripted CS paradigm and a cued picture-naming paradigm. We found that the direction of

transfer did not change across the two paradigms, with **[ʌ]** getting lowered and fronted, and **[æ]** getting fronted, in the mixed condition in both cases. However, there was a difference in the magnitude of transfer. Specifically, in the F1 dimension, the extent of phonetic transfer is greater in the code-switching task compared to the cued picture-naming task. This asymmetry is not seen in the F2 dimension. One way to think about this is that during code-switching, there is a greater degree of language co-activation because switches are being made within an utterance, where planning is required: Since participants see the entire utterance and then read it aloud, there is room for utterance planning. Compared to this, in the cued picture-naming paradigm participants see the language cue immediately prior the target word, inducing extremely "local" switches and thus a smaller degree of language co-activation. While both of these are local switches and take place in a bilingual language mode where both languages are already co-activated, it is possible that this minute difference in time-frame induces a difference of degrees in the extent of cross-language phonetic interaction. This aligns with the findings of Bullock and Toribio (2009), who found evidence of transfer in pre-switch tokens, suggesting planning effects. In our study specifically, another factor that could have contributed to this is orthography. In both tasks, Bengali and English words were presented in their respective orthographies. In the picture-naming task, this was used to cue language— as described in section 2.3, participants saw one slide at a time. In the CS task, mixed utterances were presented in mixed orthography. It is possible that seeing mixed orthography, particularly for participants who are used to processing mixed-language visual input in their surroundings, triggered increased language co-activation, leading to greater cross-language interaction.

We do not see any difference in the pattern of transfer, or the relative behaviors of the two categories. Our study used a frame sentence to induce code switching, and not all the items led to pragmatically meaningful sentences. Thus, one possible explanation for the overall lack of asymmetry is that the discursive effects discussed by Olson (2013) are

triggered by more naturalistic paradigms where attention to meaning and interlocutor are important. More within-study comparisons with less structured paradigms, and spontaneous code-switching data, can clarify this. The results of the present study indicate that for the purposes of observing dynamic phonetic transfer in an experimental setup, elicitation through cued-picture-naming vs. (controlled) code-switching does not significantly alter the behavior under study. However, elicitation through code-switching offers greater planning, potentially increasing the magnitude of observable effects.

An unexpected result in the present study is that the paradigm directly affected vowel quality, across experimental conditions: all vowels were more backed and lowered in the code-switching task. There is no principled reason to expect this. A closer look at the dynamics of this effect indicate that it is greatest at the beginning of articulation, and decreases steadily with time. Given that it is localized at the beginning of the target vowel, it seems likely that this lowering is an artifact of the particular stimuli used here— resulting from some coarticulatory effect with the preceding word in the carrier phrase that was absent in the picture-naming task. While intriguing, a more detailed analysis of this effect is beyond the scope of the present study.

Although the findings of this study are interesting, we acknowledge that it has a number of limitations, including (i) a small sample size; (ii) a frame sentence to elicit code-switched productions such that not all the sentences were pragmatically well-formed. This could have blocked some of the discursive effects that might result from speech planning in ordinary connected speech, thus obscuring differences between the two paradigms under study; (iii) focus on a single acoustic feature; (iv) the absence of unilingual Bengali data from the same set of speakers as the participants of the study, which would have allowed us to compare each individual's transfer patterns to their own Bengali productions, rather than to average Bengali vowel formants. Future work should address these. Our results also raise a number of open questions for future research, particularly concerning the paradigm differences and

between-sound asymmetries in phonetic transfer.

5   Conclusion

This study investigated how the phonetic systems of highly proficient bilingual speakers interact during online speech processing, by examining the phonetic outcome of mixed-language production in speakers of Bengali and Indian English. Specifically, we examined the spectral quality of two L2 vowel categories **[ʌ]** and **[æ]** produced in two experimental conditions: nonswitch (English) and switch/mixed (switching from Bengali to English). Comparisons of F1 and F2 across conditions reveal that mixed language processing temporarily increases phonetic interaction between the languages, causing the English **[æ]** of participants to converge towards the Bengali category **[æ]**, and **[ʌ]** to converge towards a related L1 category **[aː]**. The extent and direction of shift is asymmetrical across the two vowel categories, which parallels findings in prior studies. This asymmetry appears to be mediated by the language-specific distribution of vowel categories. These findings evidence an effect of phonetic transfer on L2 vowels for the first time, and raise questions for proposed inhibition-based accounts of cross-language phonetic interaction. Mixed productions were elicited in two switching paradigms: code-switching and cued picture-naming, to verify whether the postulated discursive elements in code-switching independently influence the outcome of phonetic transfer. Results suggest that in an experimental setup, elicitation through cued-picture-naming vs. (controlled) code-switching does not significantly alter the behavior under study. However, the magnitude of observed transfer is greater in the code-switching paradigm. These findings add to empirical data, highlight the need for diversifying the set of language pairs and populations in transfer studies, and encourage more within-study comparisons of paradigms and features, towards establishing better methodological norms. Ultimately, these results contribute to ongoing discussion about the implications of short-term phonetic interaction

as a window to understanding language as a mental and social system.

Abbreviations

clf = classifier

Data availability

The materials, data files, and analysis code for this study are openly available in OSF at: https://doi.org/10.17605/OSF.IO/DSB2X.

Competing interests

The authors declare none.

References

Agnihotri, R. K. (1979). Processes of assimilation: A sociolinguistic study of sikh children in leeds. (Unpublished doctoral dissertation). University of York.

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. Journal of Phonetics, 39(4), 558–570.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. Language in Society, 437–456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. Journal of Phonetics, 40(1), 177–189.

Balukas, C., & Koops, C. (2015). Spanish-english bilingual voice onset time in spontaneous code-switching. International Journal of Bilingualism, 19(4), 423–443.

Bansal, R. K. (1969). The intelligibility of indian english: Measurements of the intelligibility of connected speech, and sentence and word material, presented to listeners of different nationalities. Central Institute of English;[available from Orient Longmans, Madras].

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. doi: 10.18637/jss.v067.i01

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & M. J. Munro (Eds.), Language experience in second language speech learning: In honour of James Emil Flege (pp. 13–34).

Birdsong, D., Gertken, L., & Amengual, M. (2012). Bilingual language profile: An easy-to-use instrument to assess bilingualism. coerll, university of texas at austin. web. 20 jan. 2012.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (version 5.4. 15)[computer program]. retrieved april 7, 2016.

Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. Studies in Second Language Acquisition, 131–158.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. Language and speech, 46(2-3), 217–243.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. Journal of cognition, 1(1).

Bullock, B. E., & Toribio, A. J. (2009). Trying to hit a moving target: On the sociophonetics of code-switching. Multidisciplinary approaches to code switching, 41, 189–206.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals.

The Journal of the Acoustical Society of America, 54(2), 421–428.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. Journal of phonetics, 40(2), 249–268.

Coelho, G. M. (1997). Anglo-Indian English: A nativized variety of indian english. Language in Society, 561–589.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Routledge.

Das, B., Mandal, S., & Mitra, P. (2011). Bengali speech corpus for continuous auutomatic speech recognition system. In 2011 international conference on speech database and assessments (oriental cocosda) (pp. 51–55).

Dmitrieva, O., & Dutta, I. (2020). Acoustic correlates of the four-way laryngeal contrast in Marathi. Phonetica, 77(3), 209–237.

Dutta, I. (2021, Jun). Bengali Vowel Formant Data extracted from the Shruti Corpus. OSF. Retrieved from osf.io/2vpdm

Elias, V., McKinnon, S., & Milla-Muñoz, Á. (2017). The effects of code-switching and lexical stress on vowel quality and duration of heritage speakers of Spanish. Languages, 2(4), 29.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. Journal of phonetics, 15(1), 47–65.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. Speech perception and linguistic experience: Issues in cross-language research, 92, 233–277.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. Laboratory phonology, 9, 353–382.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. The Journal of the Acoustical Society of America, 76(3), 708–721.

Flege, J. E., MacKay, I. R., & Piske, T. (2002). Assessing bilingual dominance. Applied Psycholinguistics, 23(4), 567–598.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Hallé, P. (2008). Cross language phonetic influences on the speech of French–English bilinguals. Journal of Phonetics, 36(4), 649–663.

Gelman, A. (2019). Don't calculate post-hoc power using observed estimate of effect size. Annals of surgery, 269(1), e9–e10.

Ghosh, S. (2016, 10). Vowel sound change in progress: A cross-generational study of navaron dialect. International Journal of English: Literature, Language Skills IJELLS, 5, 2278-742.

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less nativelike. Psychological Science, 25(4), 1031–1036.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. Brain and language, 27(2), 210–223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. Bilingualism: Language and cognition, 1(2), 67–81.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. Bilingualism: Language and cognition, 1(2), 131–149.

Grosjean, F. (2012). An attempt to isolate, and then differentiate, transfer and interference. International Journal of Bilingualism, 16(1), 11–21.

Grosjean, F., & Miller, J. L. (1994). Going in and out of languages: An example of bilingual flexibility. Psychological science, 5(4), 201–206.

Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. Phonetica, 60(2), 98–128.

Hickey, R. (2005). Legacies of colonial English: Studies in transported dialects. Cambridge University Press.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. The American Statistician, 55(1), 19–24.

Jhingran, D. (2009). Hundreds of home languages in the country and many in most classrooms: Coping with diversity in primary education in india. In Social justice through multilingual education (pp. 263–282). Multilingual Matters.

Kachru, B., & Smith, L. (1981). English for cross-cultural communication. London: Macmillan Press.

Kachru, B. B. (1983). The indianization of English: the English language in India. Oxford University Press.

Khattab, G. (2009). Phonetic accommodation in children's code-switching. na.

Khattab, G. (2013). Phonetic convergence and divergence strategies in English-Arabic bilingual children. Linguistics, 51(2), 439–472.

Khubchandani, L. M. (1997). Language policy and education in the indian subcontinent. In Encyclopedia of language and education (pp. 179–187). Springer.

Khubchandani, L. M. (2021). Plurál languages, plurál cultures. In Plural languages, plural cultures. University of Hawaii Press.

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in r. Behavior research methods, 53(6), 2528–2543.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software, 82(13), 1–26. doi: 10.18637/jss.v082.i13

MacLeod, A. A., & Stoel-Gammon, C. (2010). What is the impact of age of second language acquisition on the production of consonants and vowels among childhood bilinguals? International Journal of Bilingualism, 14(4), 400–421.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (leap-q): Assessing language profiles in bilinguals and multilinguals.

Masica, C. (1972). The Sound System of Indian English. Monograph No. 7.

McCloy, D. R. (2016). phonr: tools for phoneticians and phonologists [Computer software manual]. (R package version 1.0-7)

Mohanty, A. K. (2009). Multilingual education: A bridge too far? In Social justice through multilingual education (pp. 3–16). Multilingual matters.

Muldner, K., Hoiting, L., Sanger, L., Blumenfeld, L., & Toivonen, I. (2019). The phonetics of code-switched vowels. International Journal of Bilingualism, 23(1), 37–52.

Myers-Scotton, C. (1993). Dueling languages: Grammatical structure in code-switching. Clarendon Press.

Nihalani, P., Tongue, R. K., Hosali, P., & Crowther, J. (1979). Indian and British English: A handbook of usage and pronunciation. Oxford University Press Delhi.

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. Journal of Phonetics, 41(6), 407–420.

Olson, D. J. (2016). The role of code-switching and language context in bilingual phonetic transfer. Journal of the International Phonetic Association, 46(3), 263–285.

Pandey, P. (2015). Indian English Pronunciation. In The handbook of English pronunciation (pp. 301–319). Wiley Online Library.

Paradis, M. (1993). Linguistic, psycholinguistic, and neurolinguistic aspects of "interference" in bilingual speakers: The activation threshold hypothesis. International Journal of Psycholinguistics, 9, 133–145.

Pardo, J. S. (2010). Expressing oneself in conversational interaction. In E. Morsella (Ed.), Expressing oneself/expressing one's self: Communication, cognition, language, and identity (pp. 183–196). London: Taylor & Francis.

Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish–English spontaneous code-switching. Journal of Phonetics, 52, 121–137.

R Core Team. (2013). R: A language and environment for statistical computing [Computer

software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Ramanathan, V. (2005). Rethinking language planning and policy from the ground up: Refashioning institutional realities and human lives. Current issues in language planning, 6(2), 89–101.

Rodriguez-Fornells, A., Kramer, U., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. Frontiers in Psychology, 2, 388.

Rubino, C. R. G., & Garry, J. (2001). Facts about the world's languages : an encyclopedia of the world's major languages, past and present / edited by Jane Garry and Carl Rubino; contributing linguistics editors, Adams B. Dodomo, Alice Faber, Robert French [Book]. H.W. Wilson New York.

Sancier, M. L., Fowler, C. A., et al. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. Journal of phonetics, 25(4), 421–436.

Schwartz, G., Balas, A., & Rojczyk, A. (2015). Language mode vs. L2 interference: Evidence from L1 Polish. In Proceedings of the International Congress of Phonetic Sciences (ICPhS). Glasgow, UK.

Šimáčková, Š., & Podlipskỳ, V. J. (2015). Immediate phonetic interference in code-switching and interpreting. In Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS). Glasgow, UK.

Šimáčková, Š., & Podlipskỳ, V. J. (2018). Patterns of short-term phonetic interference in bilingual speech. Languages, 3(3), 34.

Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. Journal of Phonetics, 43, 26–37.

Simonet, M., & Amengual, M. (2020). Increased language co-activation leads to enhanced cross-linguistic phonetic convergence. International Journal of Bilingualism, 24(2), 208–221.

Sirsa, H., & Redford, M. A. (2013). The effects of native language on Indian English sounds and timing patterns. Journal of phonetics, 41(6), 393–406.

Stuart-Smith, J., Timmins, C., Alam, F., Gregersen, F., Parrot, J., & Quist, P. (2011). Hybridity and ethnic accents. Selected Papers ICLaVE Amsterdam: Benjamins, 5, 43–57.

Tobin, S. J., Nam, H., & Fowler, C. A. (2017). Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. Journal of phonetics, 65, 45–59.

Tollefson, J. W., & Tsui, A. B. (2014). Language diversity and language policy in educational access and equity. Review of Research in Education, 38(1), 189–214.

Tsui, R. K.-Y., Tong, X., & Chan, C. S. K. (2019). Impact of language dominance on phonetic transfer in Cantonese–English bilingual language switching. Applied Psycholinguistics, 40(1), 29–58.

Wells, J. C., & Wells, J. C. (1982). Accents of english: Volume 1 (Vol. 1). Cambridge University Press.

Tables

Table 1: Key biographical information summarized across participants. Scales: age 0(since birth)—5(since age 5) etc.; likert scale: 0(not proficient at all)—6(highly proficient)

		Bengali			English		
		Mean	SD	Range	Mean	SD	Range
Age of acquisition		0.25	0.71	0 - 2	2.6	2.13	0 - 6
Years of formal education		11.8	4.92	8 - 15	16+	4.15	7 - 20+
Self-rated proficiency	speaking	6	0	6 - 6	6	0	6 - 6
	listening	5.87	0.35	5 - 6	6	0	6 - 6
	reading	4.87	1.24	3 - 6	6	0	6 - 6

Table 2: Fixed effects

Variable name	Levels
Vowel	[ʌ], [æ]
Task (corresponding to the two paradigms)	p (Picture: cued picture-naming), s (Sentence: code-switching)
Context (corresponding to the experimental conditions)	e (English: nonswitch), b (Bengali: switch/mixed)
Time	continuous variable, range -1.6 to 1.6, scaled and centered

Table 3: Optimal model: F1

	Estimate	Std. Error	t value	p value
(Intercept)	-0.053	0.179	-0.300	0.769
time	0.139	0.012	12.040	< 0.001 ***
vowelʌ	-0.147	0.223	-0.661	0.519
contexte	-0.013	0.068	-0.196	0.846
tasks	0.330	0.027	12.414	< 0.001 ***
time:vowelʌ	-0.171	0.014	-12.688	< 0.001 ***
vowelʌ:contexte	-0.137	0.060	-2.251	0.031 *
vowelʌ:tasks	0.009	0.038	0.234	0.815
contexte:tasks	0.073	0.038	1.951	0.051 .
time:tasks	-0.122	0.013	-9.028	< 0.001 ***
vowelʌ:contexte:tasks	-0.196	0.054	-3.600	< 0.001 ***

Table 4: Optimal model: F2

	Estimate	Std. Error	t value	p value
(Intercept)	0.232	0.221	1.05	0.304
vowelʌ	-0.386	0.148	-2.60	0.015 *
time	0.043	0.008	5.63	< 0.001 ***
contexte	-0.017	0.028	-0.61	0.548
tasks	-0.087	0.017	-5.00	< 0.001 ***
vowelʌ:time	-0.163	0.009	-18.35	< 0.001 ***
vowelʌ:contexte	0.110	0.028	4.01	< 0.001 ***
vowelʌ:tasks	-0.090	0.025	-3.55	< 0.001 ***
contexte:tasks	-0.110	0.025	-4.44	< 0.001 ***
time:tasks	0.062	0.009	7.03	< 0.001 ***
vowelʌ:contexte:tasks	0.033	0.036	0.94	0.34652

Figures
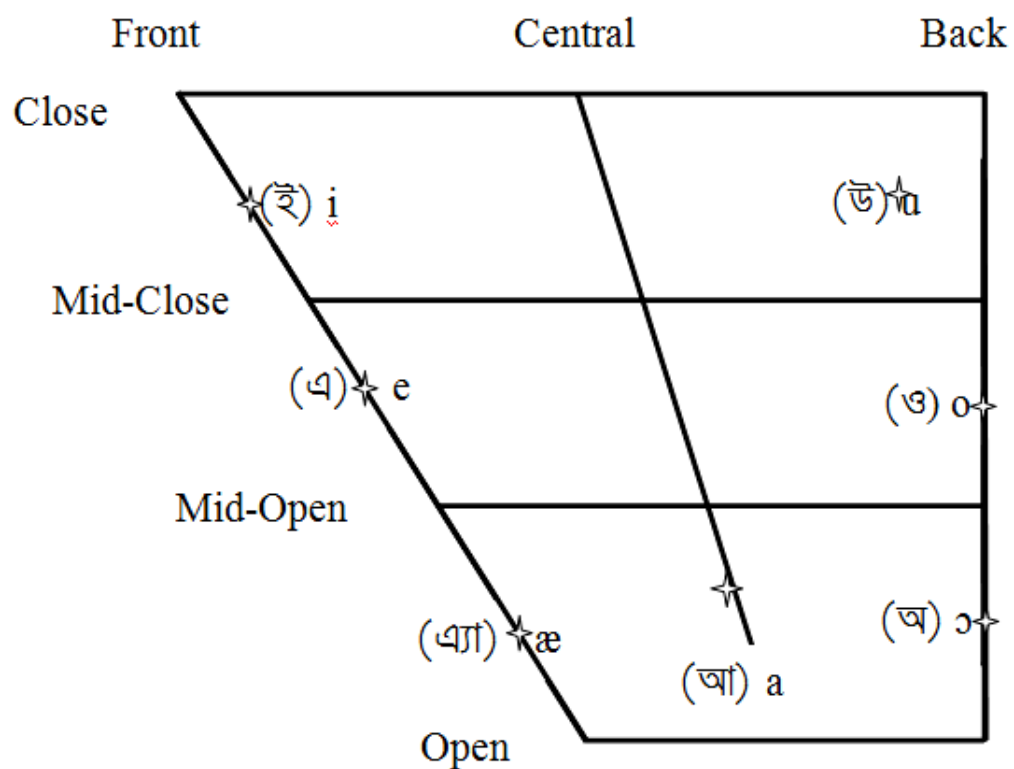
Figure 1: Vowel chart for oral monophthongs in Bengali, from Ghosh (2016)
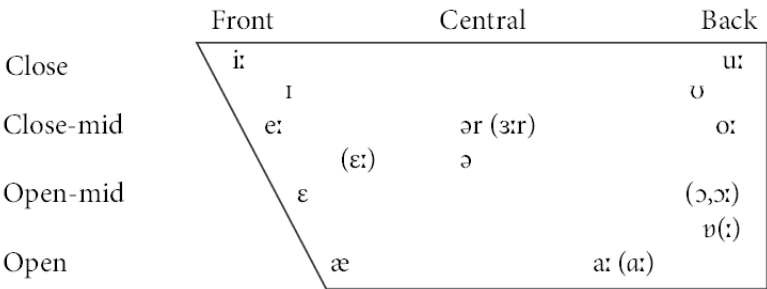
Figure 2: Vowel chart for monophthongs in General(ized) Indian English, from Masica (1972)
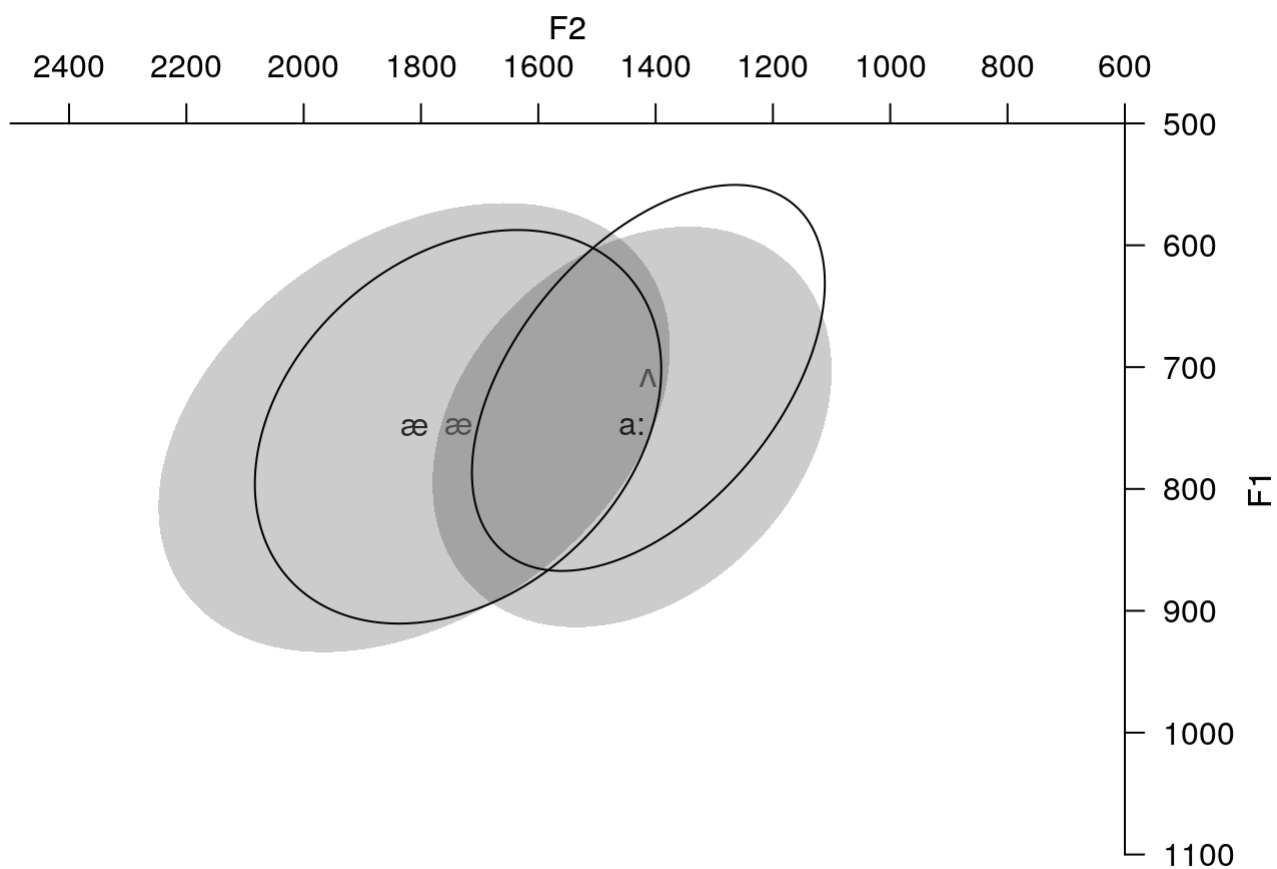
Figure 3: Participants' unilingual English vowels (solid ellipse) compared to Bengali categories from the SHRUTI corpus (Das et al., 2011)
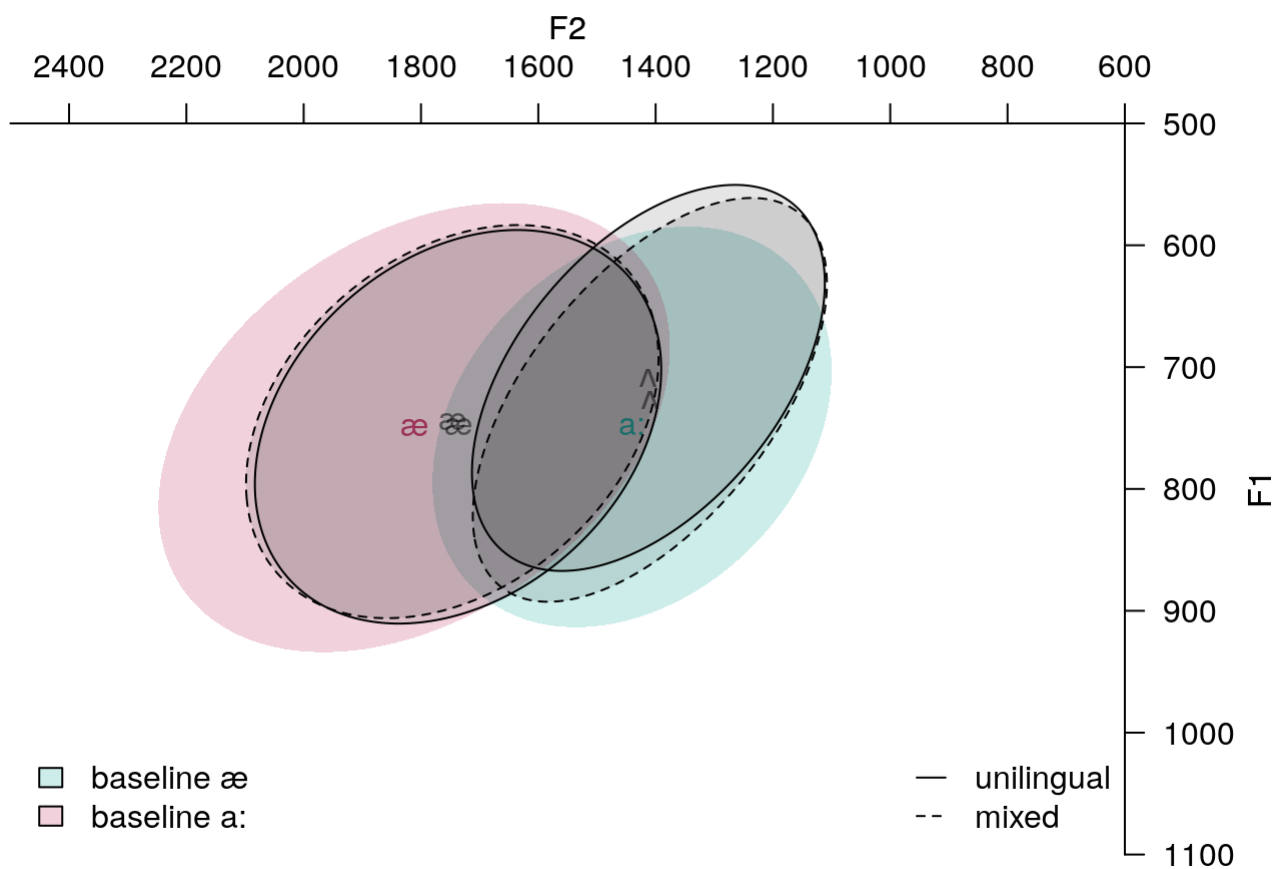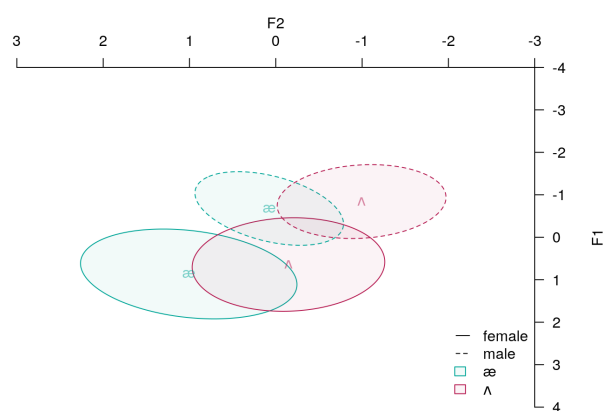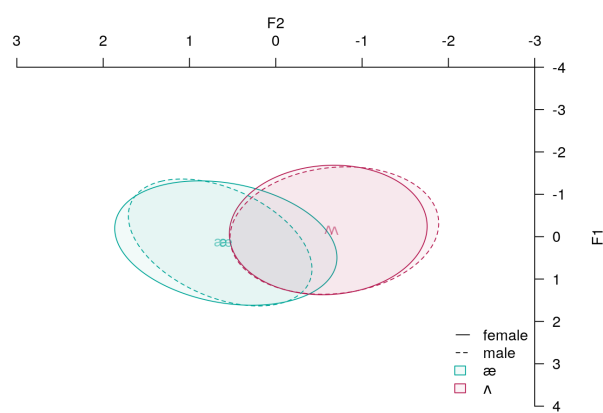
Figure 4: Vowel quality in unilingual English, mixed, and baseline Bengali productions

(a) before gender normalization

(b) after gender normalization

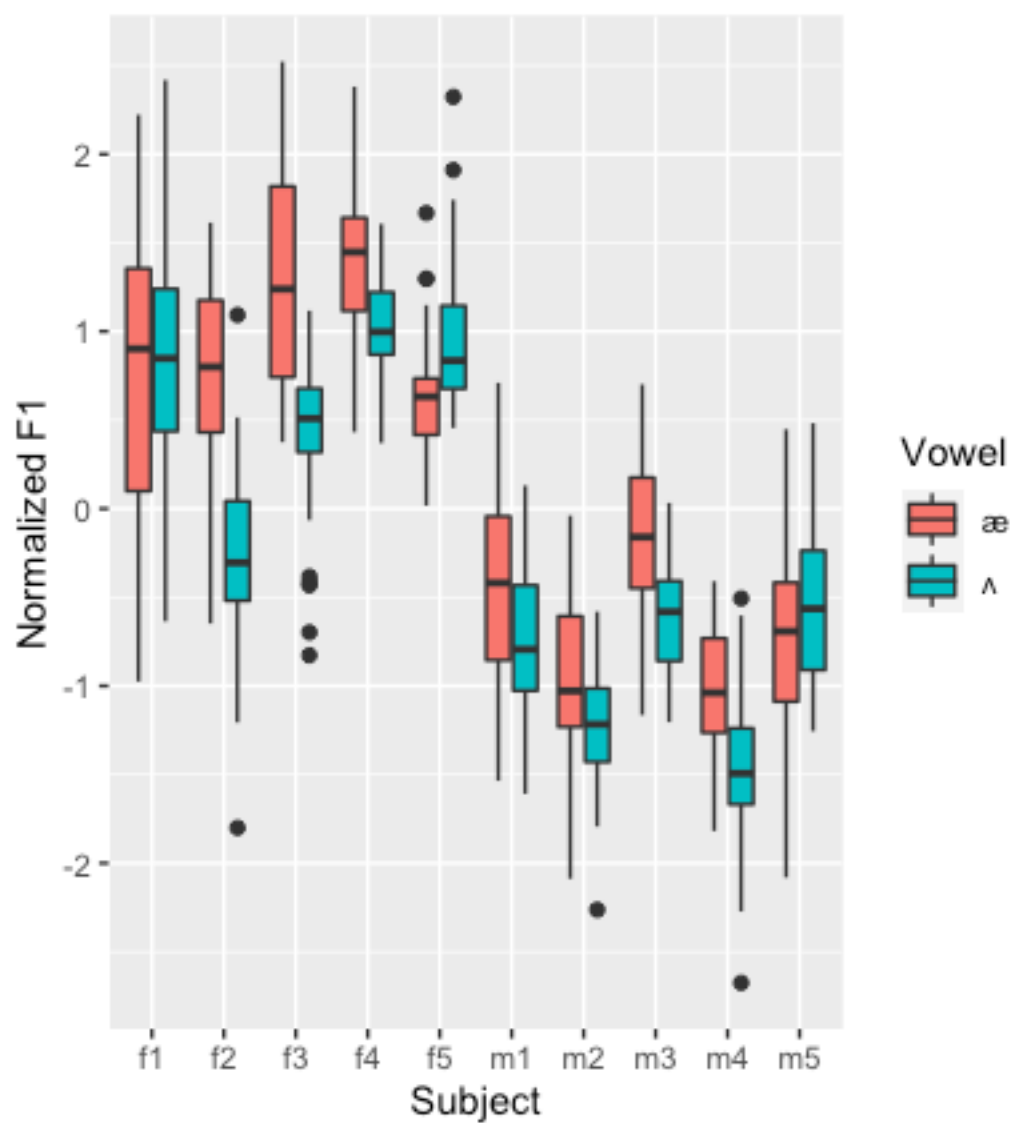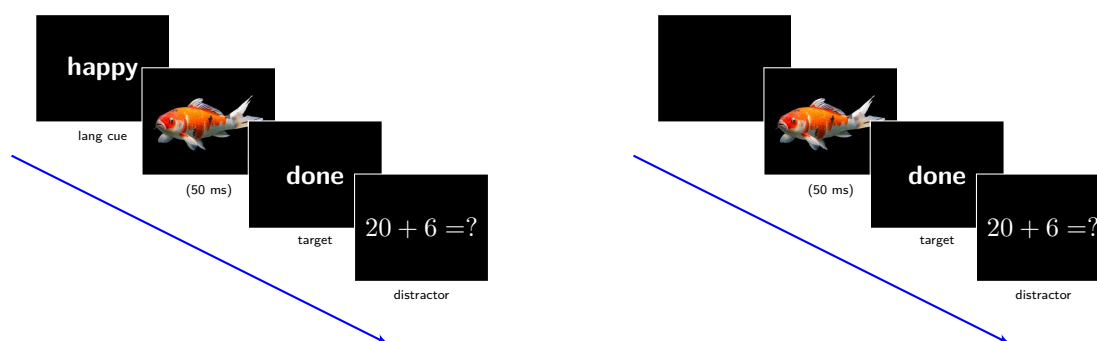Figure 5: Gender differences in Lobanov-normalized vowels

Figure 6: Subject-wise mean F1 in unilingual English productions

(a) English (nonswitch) trial   (b) Bengali-English (switch) trial

Figure 7: Picture-naming task sample sequence: language cue (not read out), picture (named as per lang cue), target (read out), distractor