

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

Acknowledgments:

We are grateful to Maumita Bhowmik and Anannya Mondal for their help with data collection and processing, and to all the participants of the study.

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 unilingual vs. mixed productions. Results reveal dynamic and asymmetrical changes in both vowels; [ʌ] shifting towards a related L1 category [a:]. 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 for the first time. 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 no difference 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.

Keywords: phonetic transfer; vowel quality; bilingual speech production; L1-L2 interaction; Indian English; Bengali

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, the reported results vary greatly across studies, 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 unilingual 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 dynamic effect of transfer on L2 vowels during mixed-language production, *but no difference in the extent of transfer between the two paradigms –and slightly different patterns of shift among the 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. 1.7 summarizes 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 other 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). Recent research comparing the segmental and suprasegmental properties of IE spoken in different parts of the country report many commonalities (see Sirsa and Redford (2013) for a review). This suggests 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.

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 population, 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). Note

that there is no mid-central vowel category.

The vowel system of IE contains the monophthongs [ɪ, i, ɛ, e, æ, ə/ʌ, ɑ, ɔ, ʊ, u], represented by the lexical set KIT, FLEECE, DRESS, FACE, and TRAP, STRUT, PALM, LOT/CLOTH, GOAT, FOOT, GOOSE (Masica, 1972; Wells & Wells, 1982). A single mid-central vowel corresponds to the categories [ʌ, ə, ɜ:], 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.

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 & Podlipský, 2015; Tsui, Tong, & Chan, 2019) (VOT), language mode (Simonet, 2014) (vowel quality), context-awareness (Khattab, 2013). 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 manip-

ulations, 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) 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 & Podlipský, 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 unilingual and 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 and Flege \(1992\)](#) 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 transfer. 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, and Fowler \(2017\)](#)(VOT) reported comparable effects in an even shorter duration (2-4 weeks). Based on a short-term longitudinal study, ([Chang, 2012](#), VOT)(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 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 & Podlipský, 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 & Podlipský, 2018), and no influence (Muldner, Hoiting, Sanger, Blumenfeld, & Toivonen, 2019; Schwartz et al., 2015). 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 profi-

ciency 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 implications 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 vowels in L2. 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](#)), suggesting that ‘extent of transfer’ cannot be 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’ 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 [a:], as anecdotal observations of Bengali-accented English suggest that the sounds are perceived as close by Bengali speakers.

In the present study, we do not elicit unilingual Bengali productions. To estimate average baseline differences in vowel quality between the relevant English and Bengali categories, we used Bengali production data from five female speakers from Dutta (2021). Formant values were extracted from syllable-initial productions of [æ] and [a:], at 55% into the vowel. Next, we compared these to our participants’ productions in the unilingual English condition. Comparisons of Lobanov-normalized F1 and F2 suggest that the Bengali category [æ] is lower, and [a:] lower and more fronted, than 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 phonetic 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.

1.7 Research questions and hypotheses

- L2 vowels: Are L2 vowel categories of proficient bilingual speakers affected by dynamic phonetic transfer during mixed-language production, as evidenced by differences in the spectral properties of L2 vowels in switched and non-switched tokens?

- Asymmetry: If so, we expect an asymmetry between the vowels in the extent and direction of shift. Based on comparisons with Bengali vowels from [Dutta \(2021\)](#) we expect [ʌ] to be lowered and fronted in the mixed condition. [æ] is expected to show lowering, but to a lesser extent than [ʌ], as it is in a part of the IE vowel space that has greater vowel density.
- Paradigm: Is there an observable difference between the extent of transfer in a cued picture-naming vs. a code-switching paradigm?

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

10 Bengali-English bilingual speakers (male=5, age range 19 to 28), completed the tasks. All the participants speak other languages in addition to Bengali and English. At the time of recording, they were living in or around the EFL University campus in Hyderabad, which has a linguistically diverse population. Speakers were selected on the basis of responses to a Language Background Questionnaire to ensure comparable LSRW (listening, speaking, reading, writing) skills in L1 and L2 across participants, and were compensated for their time.

2.2 Stimuli

Twenty monosyllabic English words containing the vowels [ʌ] or [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 lexicalized loanwords in Bengali.

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. Every participant performed four iterations each of two tasks– cued picture-naming and code-switching. Utterances in the two conditions (unilingual and mixed) were randomly interspersed within each iteration. In the unilingual 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 unilingual and once in a mixed 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 50ms; named by the participant in either English or Bengali, 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 (unilingual) 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 target word embedded in either an English (unilingual) 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]”

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. A specifier-noun construction was used to ensure identical word order in both languages, giving a uniform prosodic context for the target word across condition. In 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 avoid this confound, we used Lobanov 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 (citation for lobanov normalization?). Plotting the vowels (figure) shows a difference between male and female speakers, such that male speakers have lower formant frequencies. Normalizing by speaker 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 speaker’s values as a scaled distance from the group mean). Statistical analyses were performed on these normalized values.

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. Models were fit using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015), and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) was used to estimate p values for individual variables. For all mixed models, the alpha criterion was set at $|t| > 2$. The dependent variables were Lobanov 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

1. Subject and Item were treated as random factors.

< Insert Table 1 about here >

3.1 Models

Fixed effects

Our study involved two IE vowels. Vowel category is expected to systematically affect formant frequencies (FFs).

The experimental treatment of interest is Context, i.e. whether the target item was presented in a unilingual (English) or mixed (Bengali) 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 shift in the mixed condition is towards the corresponding L1 category. Thus, we expect the effect of Context on formant frequencies, i.e. the direction of shift, to be affected by the vowel category (c.f. the second hypothesis Asymmetry). Thus, our main effect of interest is the interaction term Context*Vowel. This corresponds to the transfer pattern.

Our third hypothesis (Paradigm) was that the language-mixing paradigm, or Task, will affect 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, formant dynamics are expected to depend on the vowel category, and so 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 Con-

text*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). (convergence problems with by-word random slope for context– changing the optimizer to bobyqa solved the problem)

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 Power analysis

We were able to recruit 10 participants. To confirm whether the experiment design was powerful enough to detect the effects we are investigating, we carried out post-hoc power analyses using the `simr` package (?) in R. The target of the analysis was to calculate the power for detecting an effect for our main variables of interest: vowel*context (to test the hypotheses L2-vowels and Asymmetry), and vowel*context*task (to test the hypothesis Paradigm).

Power is a function of sample size and effect size. Thus, a key issue in power analyses is to estimate the expected effect size for the treatment/factor of interest (??). This may be expressed as a standardized effect size or estimate (β), Cohen's d, or partial η^2 . Effect sizes in `simr` are expressed as unstandardized effect sizes (model estimates). In the absence

(to the best of our knowledge) of literature on typical effect sizes for spectral differences in cross-language vowel production, we made a conservative estimate of effect sizes as $|\beta| > 0.1$. This amounts to the assumption that the smallest effect size that is meaningful to us (smallest effect size of interest; SESOI (?)) is 0.1, and we want to confirm whether the study has enough power to detect such an effect, if it were to exist. Note that this is an arbitrary assumption; only more discussion among researchers in this field will lead to a consensus on what is considered a typical cross-language effect. In the Calculations of d values from our models using the package EMAtools (?) show that they range from 0.01 to 1.06 (no need).

While differences between paradigms have been discussed in previous literature and theoretically motivated, there is no work yet which has measured differences in a controlled experimental setup, and thus no reported effect sizes to act as a guideline. Since we are assuming $\beta = 0.1$ to be the smallest difference across conditions that we are interested in, we will assume that a difference of similar magnitude across paradigms will be of theoretical interest to us. Thus, the power analysis fixes effect sizes of both variables at $\beta = 0.1$, and report the power.

-> in cover letter: power analyses for complex mixed models are not trivial and is an active pursuit. While we have done our best to follow available literature, we would love to have more feedback on this.

transfer term: vowel*context

To calculate the power for the interaction term using simr, we set the estimate for the term to 0.05, and compare the full model against a model that lacks the interaction term. On the basis of 100 simulations, the power for

paradigm differences: vowel*context*task

3.3 Vowel Height – F1

Model output summary

First, to test hypothesis 1 (L2 transfer), 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)= -2.330e-01, standardized 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 test the hypothesis 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 null models that lack 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. **later, during interpreting part? -> Want to see: is the effect of task local (decreases with time)? Since the effect of just task on F1 is unexpected, want to see if it might be an artifact of the paradigm/stimuli** – Because an independent effect of task on FFs is unexpected (c.f.), 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 ($\beta = -1.216e-01$, $SE = 1.347e-02$, $t = -9.028$, $p = 2e-16$) 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 ($\beta = -4.604e-02$, $SE = 2.692e-02$, $t = -1.71$, $p = 0.08$), but does not significantly improve model fit compared to a null model lacking the vowel*context*time term ($\chi^2(2) = 3.41, p = 0.18$).

Table summary:

Interpretation

—older stuff: Table 2 summarizes the results of the fixed effects from the model for F1. As expected, we observe a significant effect of gender, such that male speakers have a lower mean F1 than female speakers. Vowel is not a significant predictor of F1. This is surprising, given that the phonological category is expected to systematically affect vowel quality. However, a boxplot visualization (Figure 2) shows that the participants in this study vary greatly both in the relative heights of the two categories [ʌ, æ] and the extent of height difference between them. Note that these measures are only from the English (unilingual) condition, suggesting that this variability is a feature of the participants' IE vowel categories, independent of the effects of switching. The lack of any independent effect of Context on F1 suggests that using English words in mixed utterances does not lead to a uniform lowering/raising across vowels. Instead, the significant interaction between Context and Vowel indicates a

lowering in [ʌ], but not in [æ], in the mixed-language condition. 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. This effect is strongest at the beginning of the vowel, and disappears by 45%. This is an unexpected result, since we have no principled reason to expect code-switching to induce an across-board lowering of vowels. As noted earlier, there is no significant interaction of Task with any of the other fixed factors. We used a series of ANOVAs to compare the model elaborated above with a null model that differs in the absence of the main factor which concerns our hypothesis, Vowel*Context, at each point in the vowel. We find that the difference is not significant at 5%, and increases over the course of articulating the vowel.

< Insert Table 2 about here >

< Insert Figure 2 about here >

—end older stuff

3.4 Vowel backness – F2

title

First, to test hypothesis 1 (L2 transfer), 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 the hypothesis Paradigm, we fitted a model with vowel*time and task*vowel*context as fixed effects. While task has an independent effect on F2 ($\beta = -8.758e-02$, SE= 1.756e-02, t= -4.98, p= 6.18e-07), the three-way interaction term is not significant ($\beta = 3.349e-02$, SE= 3.565e-02, t= 0.94, p= 0.34). However, there are significant two-way interactions between task*context ($\beta = -1.099e-01$, SE= 2.483e-02, t= -4.427, p= 9.66e-06) and task*vowel ($\beta = -8.983e-02$, 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.

later, during interpreting part? -> Want to see: is the effect of task local (decreases with time)? Since the effect of just task on F1 is unexpected, want to see if it might be an artifact of the paradigm/stimuli – Because an independent effect of task on FFs is unexpected (c.f.), 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 ($\beta = 6.237e-02$, SE= 8.876e-03, t= 7.027, p= 2.23e-12) and improves model fit ($\chi^2(1) = 49.271, p = 2.23e - 12$). **interpretation:**

direction of effects: tasks has negative estimate (vowels are backed in the s task). task*time has a positive estimate— means that the effect of task is mitigated/reduced with increase in time

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 ($\beta = -6.274e-03$, $SE = 1.774e-02$, $t = -0.35$, $p = 0.72$), 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: vowel*time + vowel*context*task + task*time + (1+context|word) + (1+context+vowel|subject)

Table .. summarizes the fixed effects for the optimal model for F2. Model coefficient estimates (β), standard errors, and p-values. p-values smaller than 0.05 are in bold. (do the same for F1)

[put at end of subsection: after summary of model results, say— want to see how the direction of shift compares to corresponding Bengali categories] Figure 1 summarizes the findings of the experiment ($F1 \times F2$ plots for vowel productions in the unilingual and mixed conditions, in comparison to the related Bengali categories from Dutta (2021)). The following sections discuss the results for vowel height and backness respectively.

< Insert Figure 1 about here >

—older stuff:

Table 3 summarizes the results of fixed effects from the model for F2 at each point. Once again, we see that male speakers have a lower mean F2 than female speakers, as expected. The estimates for Vowel show that excepting the beginning of the vowel, the category [ɒ] is consistently produced as more backed than [æ]. Unlike vowel height, Context has an independent effect on vowel backness, such that both the vowels in this study have higher F2 values in the Bengali (mixed) condition, i.e. they are fronted in the mixed condition. A significant interaction between Vowel and Context reveals an asymmetry in the extent of

fronting between the two vowels – the shift is greater in the case of [æ]. We see a significant effect of Task on F2, such that vowels in the Sentence task (code-switching) were more backed than those in the picture-naming task. As with F1, this is an unexpected result, as we have no principled reason to expect code-switching to induce an across-board backing of vowels. A lack of significant interaction with Context suggests that there is no difference in the extent of shift between the two paradigms. The results of ANOVAs comparing the full model to a null model that lacks the Vowel*Context interaction factor reveal that as with vowel height, the change in vowel backness between the two conditions is not significant at 5%, and increases over the course of the vowel.

< Insert Table 3 about here >

4 Discussion

With regard to our first research question L2 vowels, the results confirm that the L2 vowels of proficient bilinguals can be affected by mixed-language processing. Both vowels in this study showed systematic shifts in the mixed-language condition compared to baseline unilingual 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. 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. We find that even proficient bilinguals with extensive experience of language switching demonstrate cross-language transfer effects in vowel production. 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.

We hypothesized an asymmetrical shift between the two vowels (Asymmetry), owing to differences in (i) their positions in the Indian English vowel space, and (ii) the presence/absence of a corresponding category in the L1 Bengali inventory. The significant interaction between Vowel and Context in both F1 and F2 confirm that the experimental condition affected both height and backness differently for each vowel. As expected, we find a greater degree of shift in [ʌ], possibly owing to its position in a less dense part of the IE vowel space. The category [ʌ] is lowered and fronted in the mixed language condition as expected (Figure 1). The direction of shift in [æ] is surprising— it is fronted in the mixed-language condition. While phonological constraints on IE suggested that a shift in the F1 dimension could potentially risk the loss of contrast and is thus expected to be minimal, Figure 1 shows that [æ] in Bengali is not fronted with respect to the English category. Thus, fronting in the mixed condition cannot be explained in relation the former, at least from the present data. Recall that the Bengali vowel productions used here are from a different set of speakers than our participants. Assuming some amount of individual differences in the precise phonetic realization of vowel categories in both languages, this could be a potential source of explanation for this result. Data from a larger number of speakers for both languages, as well as within-speaker comparisons of English and Bengali vowels, are required to clarify this.

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 no significant interaction of Task either with Context, or with the Vowel*Context interaction, suggesting that the extent of phonetic transfer was not affected by the paradigm. There are at least two plausible reasons for this– (i) it is possible that the planning effects of CS do not translate to an increase/decrease in observable transfer at all. This would mean that the pre-switch effects observed, for example, in [Muldner et al. \(2019\)](#) and [Bullock and Toribio \(2009\)](#) are better interpreted as resulting from some other aspect of the experimental setup, than the act of code-switching itself. While planning could additionally affect the words preceding the switch, the effects on the switched token itself do not change. (ii) Our study used a frame sentence to induce code switching, and not all the items led to pragmatically meaningful sentences. Thus, another possible explanation 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. Thus, the choice of paradigm can be based on other considerations. Moreover, this suggests that (all else being equal) the results from such studies can be compared directly.

The paradigm did affect vowel quality, but this applied uniformly across experimental

conditions: all vowels were more backed and lowered in the code-switching task. Given that at least for F1, the effect is limited to the early stages of articulating the vowel, it is plausible that this lowering is an artifact of the particular stimuli used here—resulting from some corarticulatory effect that was absent in the picture-naming task. While this is intriguing, it is beyond the scope of the present study.

5 Conclusion

This study examined the phonetic outcome of mixed-language production in a group of highly proficient bilingual 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: unilingual (English) and mixed (switching from Bengali to English). Comparisons of F1 and F2 reveal that mixed language processing temporarily increases phonetic interaction between the languages, causing **[ʌ]** to converge towards a related L1 category **[a]**. 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.

This study has a number of limitations: (i) the ‘baseline’ values for Bengali vowel categories were estimated from a different set of speakers than the participants of this study,

potentially obscuring the role of individual differences in the realization of vowel categories; within-speaker comparisons of L1 and L2 vowels is likely to offer a more fine-grained view of the direction of shift in relation to L1 categories. (ii) we used 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. Future work should address this by considering data from more naturalistic code-switching paradigms.

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

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).
- 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.
- Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, 23(3), 389.
- 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.
- 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.
- Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological Science*, 25(4), 1031–1036.
- 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.
- 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. (2013). Phonetic convergence and divergence strategies in English-Arabic bilingual children. *Linguistics*, 51(2), 439–472.
- 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.
- 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)
- 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/>
- 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., & Podlipský, 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., & Podlipský, 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.
- 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: Fixed factors

| Factor | Levels |
|--|---|
| Gender | male, female |
| 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: unilingual), b (Bengali: mixed) |
| Time | continuous variable, -1.6 - 1.6 |

Table 2: Model outputs for F1 at each of the five points in the vowel

| | Dependent variable: | | | | |
|-------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | F1_5 (Model 1) | F1_15 (Model 2) | F1_25 (Model 3) | F1_35 (Model 4) | F1_45 (Model 5) |
| (Intercept) | 0.447** (0.211) | 0.718*** (0.176) | 0.750*** (0.153) | 0.863*** (0.152) | 0.921*** (0.150) |
| Gender(male) | -1.171*** (0.237) | -1.526*** (0.184) | -1.586*** (0.168) | -1.668*** (0.177) | -1.681*** (0.175) |
| Context(e) | -0.010 (0.058) | -0.015 (0.057) | 0.033 (0.042) | 0.025 (0.040) | 0.037 (0.035) |
| Vowel(Λ) | 0.022 (0.170) | -0.070 (0.175) | -0.034 (0.170) | -0.131 (0.174) | -0.216 (0.168) |
| Task(s) | 0.319*** (0.091) | 0.235*** (0.083) | 0.172** (0.081) | 0.133* (0.077) | 0.102 (0.073) |
| Context(e):Vowel(Λ) | -0.126** (0.056) | -0.138** (0.060) | -0.185*** (0.055) | -0.191*** (0.064) | -0.192*** (0.066) |
| Observations | 2,333 | 2,333 | 2,333 | 2,333 | 2,333 |
| Log Likelihood | -1,771.558 | -1,589.316 | -1,278.310 | -1,313.795 | -1,380.083 |
| Akaike Inf. Crit. | 3,599.116 | 3,234.631 | 2,612.621 | 2,683.589 | 2,816.165 |
| Bayesian Inf. Crit. | 3,760.253 | 3,395.769 | 2,773.758 | 2,844.727 | 2,977.303 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 3: Model outputs for F2 at each of the five points in the vowel

| | Dependent variable: | | | | |
|-------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | F2_5 (Model 1) | F2_15 (Model 2) | F2_25 (Model 3) | F2_35 (Model 4) | F2_45 (Model 5) |
| (Intercept) | 0.607*** (0.222) | 0.838*** (0.213) | 0.778*** (0.211) | 0.827*** (0.203) | 0.826*** (0.204) |
| Gender(male) | -0.839*** (0.124) | -0.891*** (0.106) | -0.807*** (0.107) | -0.750*** (0.116) | -0.748*** (0.109) |
| Context(e) | 0.003 (0.036) | -0.088** (0.036) | -0.065** (0.032) | -0.063 (0.044) | -0.103*** (0.033) |
| Vowel(Λ) | -0.217 (0.175) | -0.615*** (0.183) | -0.651*** (0.175) | -0.801*** (0.168) | -0.836*** (0.166) |
| Task(s) | -0.234*** (0.029) | -0.194*** (0.035) | -0.152*** (0.050) | -0.141*** (0.047) | -0.092** (0.045) |
| Context(e):Vowel(Λ) | 0.031 (0.050) | 0.110*** (0.041) | 0.128*** (0.040) | 0.091** (0.046) | 0.147*** (0.043) |
| Observations | 2,333 | 2,333 | 2,333 | 2,333 | 2,333 |
| Log Likelihood | -1,283.531 | -1,478.494 | -1,174.087 | -1,222.682 | -1,019.033 |
| Akaike Inf. Crit. | 2,623.062 | 3,012.988 | 2,404.175 | 2,501.363 | 2,094.066 |
| Bayesian Inf. Crit. | 2,784.199 | 3,174.125 | 2,565.312 | 2,662.501 | 2,255.204 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Figures

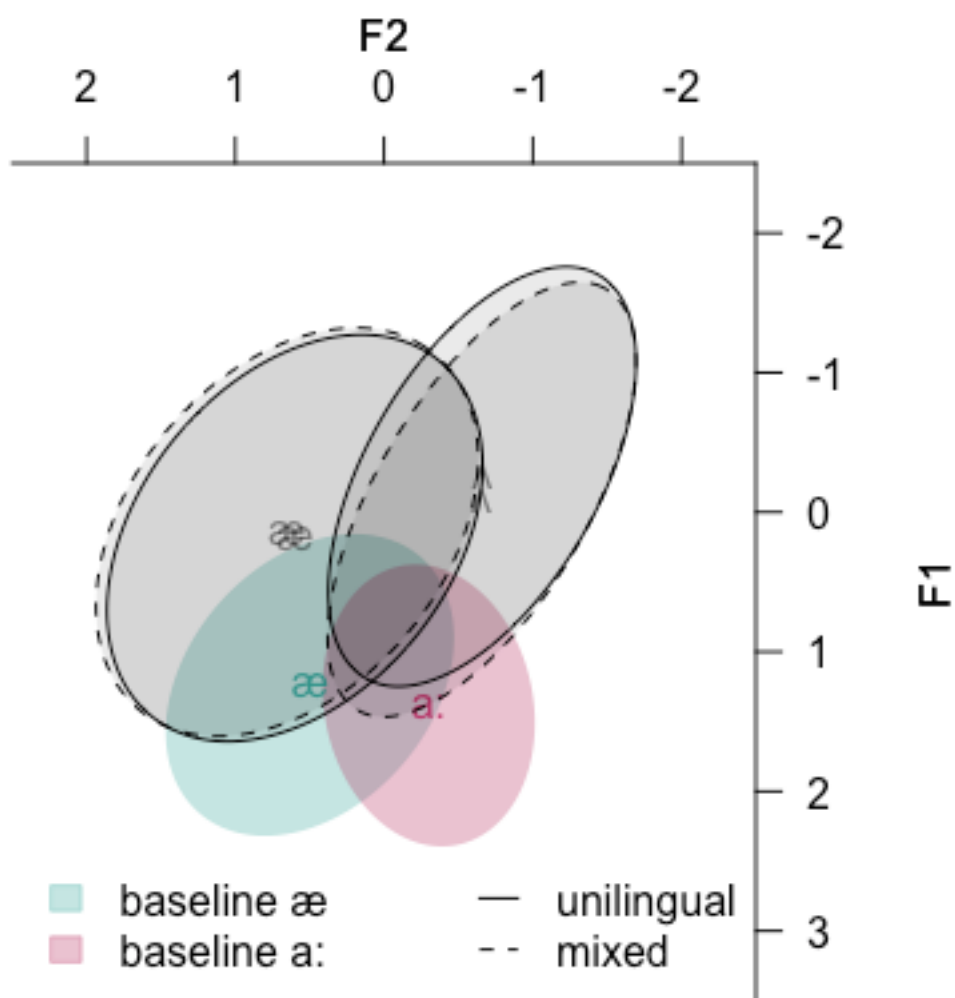


Figure 1: Vowel quality in unilingual English, mixed, and baseline Bengali productions

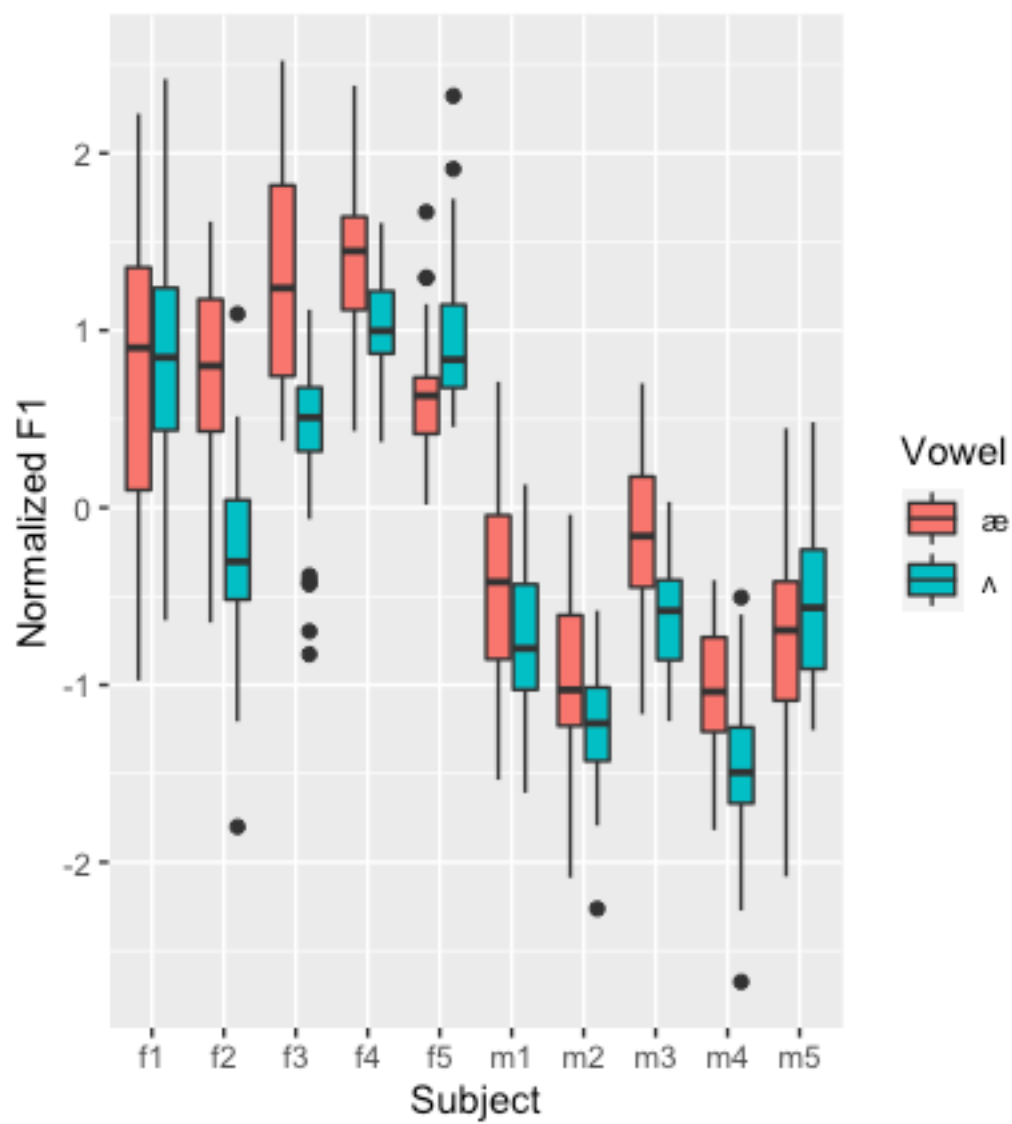


Figure 2: Subject-wise mean F1 in unilingual English productions

Table 4

| | Dependent variable: |
|-----------------------|----------------------|
| | f1 |
| time | 0.139*** (0.012) |
| vowel | −0.147 (0.223) |
| contexte | −0.013 (0.068) |
| tasks | 0.330*** (0.027) |
| time:vowel | −0.171*** (0.013) |
| vowel :contexte | −0.137** (0.061) |
| vowel :tasks | 0.009 (0.038) |
| contexte:tasks | 0.073* (0.038) |
| time:tasks | −0.122*** (0.013) |
| vowel :contexte:tasks | −0.196*** (0.054) |
| Constant | −0.053 (0.178) |
| Observations | 11,665 |
| Log Likelihood | −12,954.910 |
| Akaike Inf. Crit. | 25,951.830 |
| Bayesian Inf. Crit. | 26,106.480 |

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5: optimal model: F1

| | Estimate | Std. Error | t value | Pr(> t) |
|-----------------------|------------|------------|---------|-------------|
| (Intercept) | -5.347e-02 | 1.784e-01 | -0.300 | 0.76863 |
| time | 1.391e-01 | 1.156e-02 | 12.040 | < 2e-16 *** |
| vowel | -1.475e-01 | 2.230e-01 | -0.661 | 0.51870 |
| contexte | -1.342e-02 | 6.836e-02 | -0.196 | 0.84627 |
| tasks | 3.300e-01 | 2.659e-02 | 12.414 | < 2e-16 *** |
| time:vowel | -1.709e-01 | 1.347e-02 | -12.688 | < 2e-16 *** |
| vowel :contexte | -1.371e-01 | 6.088e-02 | -2.251 | 0.03130 * |
| vowel :tasks | 9.002e-03 | 3.850e-02 | 0.234 | 0.81512 . |
| contexte:tasks | 7.334e-02 | 3.760e-02 | 1.951 | 0.05112 . |
| time:tasks | -1.216e-01 | 1.347e-02 | -9.028 | < 2e-16 *** |
| vowel :contexte:tasks | | | | |