The Role of Temporal and Spectral Cues in Non-native Speech Production: Bangla Speakers' L2 English Tense and Lax Vowels

Md. Jahurul Islam
Lecturer, Department of Linguistics, University of British Columbia, Canada
jahurul.islam741@gmail.com | ORCID: 0000-0001-9147-7610

Abdulla Al Masum
Assistant Professor, Center for Language and Cultural Studies, Green University of Bangladesh
masum.gulc@green.edu.bd | ORCID: 0009-0004-1632-8294

Md. Sayeed Anwar
Assistant Professor, Department of Humanities, Rajshahi University of Engineering and Technology, Bangladesh
anwar@hum.ruet.ac.bd | ORCID: 0000-0003-0913-2766

Abstract

This study investigated the role of durational and spectral cues in second language tense and lax vowel contrasts produced by non-native speakers. To test previous claims that speakers primarily rely on durational cues over spectral cues to distinguish L2 tense and lax vowel pairs, citation style speech data were collected from 16 native speakers of Bangla; participants were all undergraduate students. The data were collected via a shadowing task where participants listened to a carefully constructed list of English words in random order and repeated each word immediately after they heard them. The utterances were recorded via a Zoom H1n voice recorder. Collected speech data were annotated and processed using the phonetic analysis software Praat and the semi-automatic annotation toolkit DARLA; statistical analyses were performed using R statistical computing software. Results indicate that Bangla speakers do not emphasize on durational cues to differentiate English tense-lax vowel pairs, contrary to the general patterns reported from other languages; rather, they prefer the spectral cues over the durational cues.

Keywords: non-native vowel production, non-native vowel perception, temporal cue, spectral cue, tense-lax distinctions

Introduction

The production of non-native tense and lax vowels distinctions has been studied from different perspectives, including the temporal and spectral properties of the vowel categories; more specifically, there has been a lot of interest in studying non-native English tense and lax vowels produced by native speakers of other languages (Cebrian, 2007; Leung et al., 2016, Ćavar et al., 2022). Native English speakers can effortlessly differentiate, both in perception and production, between tense and lax vowel sounds in minimal pairs like /bin/ "bean" and /bɪn/ "bin" in terms of their temporal and spectral cues. However, non-native or second language (L2) speakers often struggle to produce these distinctions consistently

and also reported that L2 speakers rely more on durational cues than spectral cues to distinguish between tense and lax vowel sounds (Cebrian, 2007; Chen, 2006; Gao et al., 2020; Rojczyk, 2010).

Research indicates that speakers do not place equal emphasis on the temporal and spectral characteristics of vowel sounds when producing non-native tense and lax distinctions. Studies reported that temporal cues often received more attention from non-native speakers when they try to produce English tense and lax vowel distinctions (Rojczyk, 2010; Mora & Fullana, 2007). In other words, L2 English speakers distinguish English tense and lax vowel pairs primarily via vowel duration and not so via the differences based on their formant frequencies. It has even been claimed that L2 English speakers "over-rely" on duration, and not spectral cues, to distinguish tense vowels from their laxed counterparts (Cebrian, 2006).

That L2 speakers are generally more likely to use durational cues over spectral cues, as claimed in previous studies, can be problematic. Whether speakers put more relative importance on durational or spectral cues for differentiating vowels of similar quality in a second language can depend on the vowel inventory in their first language (L1). For example, if the L1 has a tense-lax contrast that is phonetically implemented via spectral cues, speakers should be able to perceive and produce spectral differences for similar vowel contracts in their L2 as well. If, however, their L1 does not have a tense-lax contrast, speakers are open to using either durational or spectral or both cues; and the durational cue may be the easier choice since it is less complex than spectral cues (because duration is one-dimensional while F1 x F2 plane is two-dimensional).

To widen our understanding of the cues used in second language vowel perception and production, this study investigates whether native Bangla speakers are able to perceive and produce the durational and spectral cues in English tense and lax vowel pairs. While there have been previous attempts at studying the cues important in Bangla consonant production and perception (e.g., Islam, 2019; 2022), there is a general lack of studies attempting vowel production and perception. Previous studies on L2 English speech by Bangla speakers have claimed that Bangla speakers struggle to differentiate English tense-lax vowel contrasts (Rahman, 2018). This phenomenon has actually been reported for Indian English, in general (Payne et al., 2019).

Bangla provides an interesting opportunity to test the "universality" hypothesis of duration as a primary cue over spectral cues for the perception and production of L2 tense-lax vowel contrasts for two reasons. First, Bangla mid vowels have been phonologically described to have tense and lax contrasts though it lacks any tense-lax contrasts among high vowels (Shamim, 2011). Second, Bangla tense and lax vowel contrasts (available for mid vowels) have been reported to have spectral differences but no durational differences. This is a completely different phonetic implementation of the similar vowel categories in the English vowel system.

Besides, Bangla could be an interesting case to study regarding the L2 English tense/lax vowel productions since there is a lack of formal studies on Bangla speakers' ability to distinguish

tense/lax contrast vowel sounds in L2 English from temporal and spectral perspectives. Therefore, we conducted a study to address this research gap where we hypothesized that native Bangla speakers would put more emphasis on spectral cues (F1 and F2 values) than on durational cues since L1 Bangla tense-lax contrasts are not implemented by durational cues, even though English speakers use both durational and spectral cues predominantly

Background

The temporal and spectral characteristics of vowels have been investigated by many studies, including Cebrian (2007), Chen (2006), Gao et al. (2020), Rojczyk (2010), Leung et al. (2016), Fox and Maeda (1999), and Ćavar et al. (2022). Leung et al.(2016) specifically examined how clear-speech and tensity effects interact by comparing clear and plain productions of three English tense-lax vowel pairs (/i-ɪ/, /ɑ-ʌ/, /u-ʊ/). Clearly produced vowels exhibit longer duration and more extreme spectral properties than plain, conversational vowels. These features also characterize tense relative to lax vowels. This study analyzed both the spectral and temporal features of the vowels and found that from plain to clear vowel lengthening was greater for tense vowels compared to lax vowels. Furthermore, clear-speech modifications in spectral change were comparatively larger for lax vowels than tense vowels. The results also indicated that peripheral tense vowels showed more consistency for clear-speech modifications in the temporal domain than in the spectral domain.

L2 speakers of English have been found to have a stronger ability to distinguish tense/lax vowel pairs based on temporal cues rather than spectral cues. Chen (2006) conducted a study on Mandarin native speakers who use English as an L2 to investigate the temporal and spectral features of three English tense-lax vowel pairs (/i/-/ɪ/, /e/-/ɛ/, /u/-/ʊ/). The study compared the production of speech between 40 American English speakers and 40 Mandarin L2 English speakers, analyzing factors such as Euclidean distance based on F1 and F2 frequencies, durational differences, and perceptual judgment. The results indicate a statistically significant difference between the two groups in their ability to distinguish English tense-lax vowel contrasts, with Mandarin speakers showing a stronger reliance on temporal features than spectral features.

Other studies, including Rojczyk (2010), Cebrian (2007), and Fox and Maeda (1999), have reported similar findings. Rojczyk (2010) conducted a study on Polish speakers learning English as an L2 and found that learners tended to rely more on durational cues than on spectral cues when distinguishing between English vowels /æ/, /e/, and /ʌ/. The study showed that learners reweight the cue hierarchy to rely on increased durations differentiating /æ/ and /ʌ/, and listeners demonstrated a strong bias to identify stimuli with longer durational values as /æ/ and with shorter values as /ʌ/. Similarly, Cebrian (2007) found that native Catalan speakers learning English as an L2 relied on the duration of English vowels in their perception of speech, while mostly unable to produce a spectral contrast in their L2 production.

Fox and Maeda (1999) also found evidence of the pattern of relying more on temporal rather than spectral cues for vowel contrast in a study on L2 Japanese speakers. The study

aimed to investigate whether Japanese speakers could differentiate between the two high front vowels (/i/ and /ɪ/) of American English in terms of both production and perception. The results showed that the participants consistently distinguished between the two sounds based on vowel duration, but showed variation in vowel quality where the distribution of the formant values overlapped.

To investigate the pattern of L2 production beyond the English language, Gao et al. (2020) conducted a study on Mandarin learners of German, a language that also has tense-lax pairs. They examined durational and spectral differences of seven German tense-lax vowel pairs in comparison to native German speakers. The results showed that Mandarin learners had a different production pattern for German tense-lax vowel distinction compared to German speakers. The Mandarin learners relied more heavily on temporal features than on spectral features, in contrast to German speakers.

More recently, Ćavar et al. (2022) investigated the reliance on quality and duration cues in the perception of English tense-lax distinction by L2 Polish and Croatian learners. They found that Croatian participants relied mainly on duration as a cue in categorization, whereas Polish learners relied predominantly on quality, even in back vowels where there is no quality contrast in Polish that would be parallel to the tenseness contrast in English. This provides an indication that L2 learners' preference between temporal and spectral cues is not a universal phenomenon; rather, the preference can vary cross-linguistically.

With regards to Bangla speakers, they have been reported to face difficulty in distinguishing the tense/lax contrast in English vowel sounds in their L2 speech production, except for high vowels (Rahman, 2018). This is because the Bangla vowel system does not have tense/lax sounds, and L1 Bangla speakers are not accustomed to distinguishing between vowel length and quality to create meaning contrast. Similar results have been reported for other Indic languages too; for example, Payne et al. (2019) reported that Telugu speakers also face difficulty in producing the tense/lax vowel contrast in their L2 English production due to the absence of such contrast in their L1 vowel system.

Bangla vowel system

Bangla can be an interesting case to study with respect to the L2 English tense/lax vowel productions since Bangla has been described by some researchers (Shamim, 2011) to have some tense/lax contrasts for mid-vowels, at least in terms of phonology, while the high vowels lack any phonological contrasts in tenseness or laxness. Bangla has a seven-vowel system (Islam & Ahmed, 2020; Alam, Habib & Khan, 2008; Barman, 2009; Shamim, 2011) with nasal counterparts for some or all of the oral vowels (cf. Islam, 2018, for this debate).

Figure 1 presents a vowel chart presented in Barman (2009). As the figure indicates, /i/ and /u/ are represented as high vowels while /ɑ/ is shown as a low vowel. The vowels /o/ and /ɔ/ are mid-back vowels while /e/ is a mid front vowel. The vowel /æ/ is an interesting case; it is represented as a low vowel (the position nearly the same as in the core IPA chart). One noticeable feature of this vowel diagram is the height difference between vowels /ɔ/

and /æ/; these two vowels are quite distant from each other in terms of their height. This assumption was, however, proved otherwise, specifically from the acoustic point of view, in some later studies.

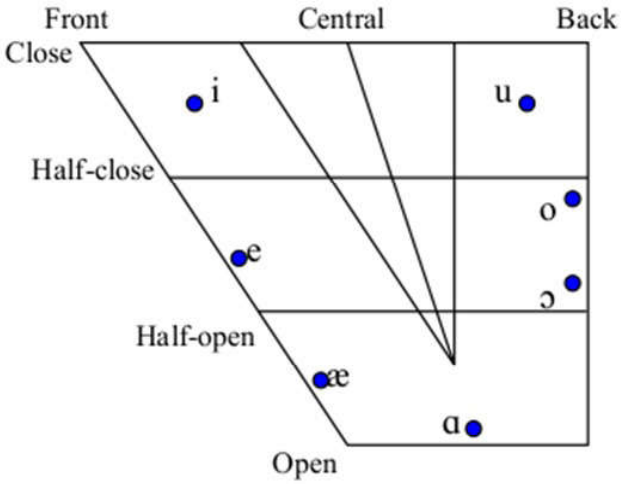


Figure 1: Bangla vowel system (Barman, 2009, p. 27)

Figures 2 and 3 present Bangla vowel charts based on acoustic measurements, as performed in Alam, Habib, and Khan (2008), and Islam and Ahmed (2018), respectively. Figure 2 presents the raw formant values while Figure 3 is based on normalized formant values.

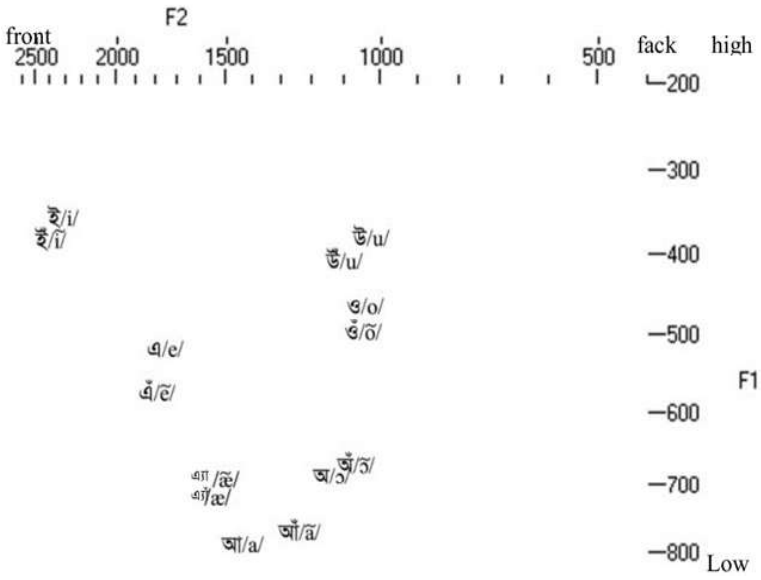


Figure 2: Bangla vowel formants (Alam, Habib, & Khan, 2008, p. 9)

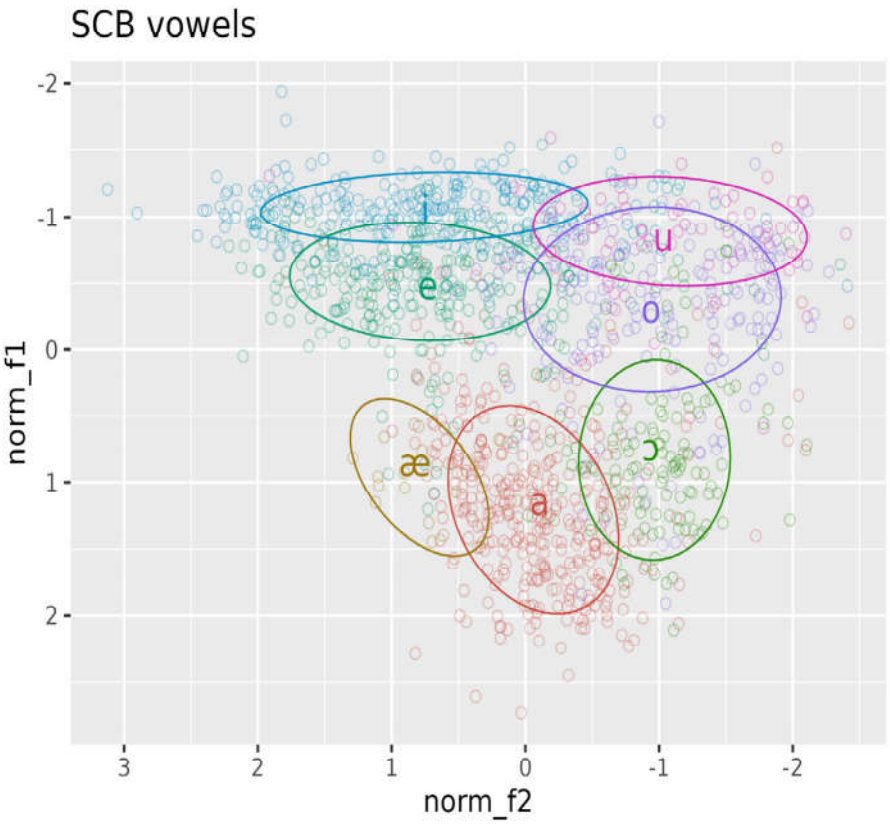
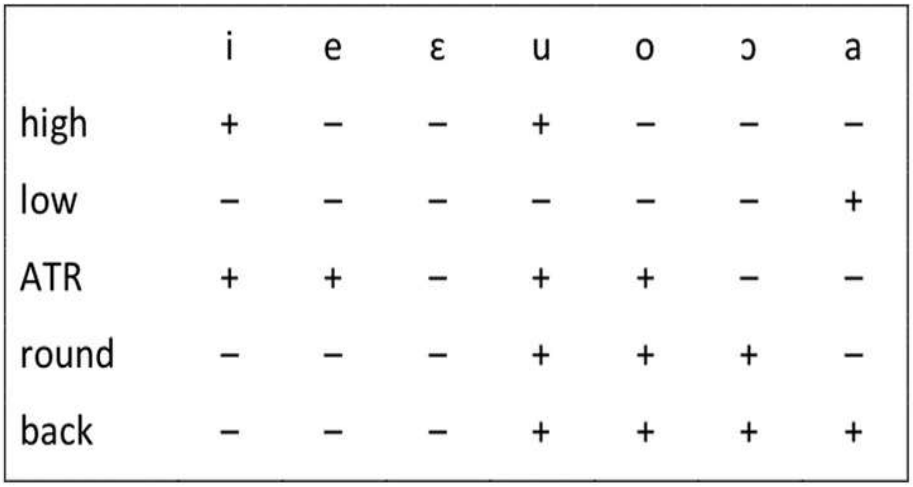


Figure 3: SCB vowels (normalized formants, Islam & Ahmed, 2018, p. 215)

As the figures indicate, the two vowels in question (/ɔ/ and /æ/) have very similar height in the vowel space. This information can have interesting implications when we turn to the phonological aspects of the vowels.

In terms of phonological representations, the features that have been used to describe Bangla vowels include the binary features [±high], [±low], [±back], [±ATR], and [±round]. While the representations of the high vowels and the low vowel /a/ have been unproblematic, the representation of the mid vowels, especially the front ones, is a debated issue among researchers. For example, the /æ/ vowel has traditionally been described as a "low" vowel (Morshed, 1972; Alam et al., 2008; Thompson, 2012) in terms of its phonetic characteristics; thus, it has been represented with a [+low] feature. However, Shamim (2011) preferred the symbol /ɛ/ instead of /æ/ to refer to the low front vowel and claimed that, phonologically speaking, /ɛ/ acts as the lax counterpart of /e/ in Bangla. This analysis of /æ/ as /ɛ/ derives from the phonological process called metaphony in Bangla where /ɛ/ and /e/ pattern together while /ɔ/ and /o/ pattern together. Thus, according to Shamim (2011), phonological features that produce contrastive distinctions among the four mid vowels in Bangla are [±ATR] (representing a tense/lax distinction), and [±round] where /e/

and /o/ are [+ATR] (tense) and /ɔ/ and /ɛ/ are their [-ATR] (lax) counterparts. Figure 4 provides the full featural representations for Bangla vowels.



	i	e	ɛ	u	o	ɔ	a
high	+	−	−	+	−	−	−
low	−	−	−	−	−	−	+
ATR	+	+	−	+	+	−	−
round	−	−	−	+	+	+	−
back	−	−	−	+	+	+	+

Figure 4: Phonological features of Bangla vowels (Shamim, 2011, p. 8)

Now, going back to the issue of temporal vs. spectral cues in vowels, we hypothesize that Bangla speakers will be more sensitive to and successful in following the spectral cues than the temporal cues in their L2 English tense/lax vowel productions. Our hypothesis is based on two assumptions. First, vowel duration in Bangla is not contrastive; in other words, vowel duration differences do not produce any phonological contrasts. (There can be systematic durational differences in the phonetics though, conditioned by discourse functions (cf. Ghosh, 2018)). And, second, the tense/lax contrasts in Bangla are phonetically implemented primarily in the form of spectral differences. Tense vowels have been reported to have similar duration values compared to their lax counterparts; and in some cases, the lax vowel has been reported to be longer than its tense counterpart (cf. Alam et al., 2008). Table 1 presents the average duration of Bangla tense and lax vowels in Alam et al. (2008, p. 7).

Table 1: Duration of tense and lax vowels in Bangla (based on Alam et al. 2008, p. 7)

Vowel	Tenseness	Avg. duration (in ms)
/e/	tense	75.45
/ɛ/	lax	90.89
/o/	tense	75.74
/ɔ/	lax	74.65

The above discussion indicates that Bangla tense and lax vowels are implemented in a completely different way than it is done in English and many other languages. Previous studies have predominantly reported that tense vowels are generally longer in duration and more peripheral in the vowel space (Roesler & Song, 2018; Lee, 2008; Zsiga, 2013; Botma

et al., 2012; Harrington et al., 2011), which is evidenced by their relative position in the vowel space. In contrast to the generally reported patterns, Bangla tense and lax vowel counterparts do not differ in terms of duration. Now, when it comes to the question of L2 English tense and lax vowel production (and perception), we hypothesize the following:

- Bangla speakers will be able to successfully produce the distinctions between English tense and lax vowel pairs in terms of their spectral properties (i.e., the formant values), unlike what has been reported from speakers of other languages.
- Bangla speakers will not be able to produce consistent durational differences between English tense and lax vowel pairs.

We will pursue these two hypotheses in the remainder of this paper.

Methods

The study was conducted in two distinct phases with comparable methodology. In the first phase, data were collected for only the high vowels in the English vowel inventory whereas the second phase involved the English mid vowels. The two phases included distinct sets of participants of similar demographics; more details on the participant demographics and experimental setup are provided below.

Participants

Participants in this study were recruited based on the following criteria:

A. *Demographic information*: The sample consisted of 16 participants, ranging in age from 18 to 25 years old (M=19.53, SD=1.50), with 8 female speakers and 8 male speakers. All participants were students enrolled in different undergraduate programs, including English, Business Studies, and Engineering. All participants were native speakers of Bangla with English as their second language, and all of them learned English primarily in the academic (in schools, for example) settings with no significant exposure to native speakers of English.

B. *Eligibility criteria*: Participants were eligible to participate if they were fluent in English, had access to the internet, and had not previously participated in a study on the same topic.

C. *Recruitment procedures*: Participants were recruited through various means, including online postings on social media platforms, flyers posted in the university notice boards, and word of mouth. Participants did not receive any direct compensations for their participation in the project.

D. *Informed consent*: Before participating in the study, participants were given a detailed explanation of the study's purpose, procedures, and risks and benefits. Participants were informed that their participation was voluntary and that they could withdraw at any time. Participants were required to give their informed consent before beginning the study.

As mentioned above, half of these participants took part in the first phase of data collection which focused on the high tense/lax vowels only while the other half provided data for the

second phase where data were collected on the mid tense/lax vowels.

Materials

The entire word selection process was meticulously designed. This study avoided the words carrying nasal and rhotic sounds as they might affect phonetic environment, though few appeared due to consonant joining.

During the first phase of the study, we constructed a list of 30 isolated monosyllabic English words containing the high front and high back vowels (/i/, /ɪ/, /u/, and /ʊ/). 20 of them contained high front vowels (10 tense and 10 lax vowels) while the remaining 10 contained high back vowels (5 tense and 5 lax). The number of words containing lax vowels was smaller because of the dearth of tense/lax contrasts in monosyllabic words for high back vowels. The wordlist for the second phase included 20 monosyllabic English words containing the front and back mid vowels (/e/, /ɛ/, /o/, and /ɔ/). Out of the 20 words, 10 contained mid front vowels (5 tense and 5 lax) while the remaining 10 contained mid back vowels (5 tense and 5 lax). Tables 2 and 3 provide the list of words constructed for both phases of data collection. It can be noted that many mid vowels are diphthongized in quality; we chose them primarily based on their quality in the initial steady part of the vowel (often the first 30-40% of the vowel) and not based on the off-glides.

Table 2: List of words for phase-1 data collection

Front Vowel				**Back Vowel**	
tense		**lax**		**tense**	**lax**
/i/		**/ɪ/**		**/u/**	**/ʊ/**
bean	keel	bin	kill	boot	book
feat	leak	fit	lick	food	foot
deal	peak	dill	pick	kook	cook
heat	seat	hit	sit	pool	pull
green	teen	grin	tin	suit	soot

Table 3: List of words for phase-2 data collection

Front Vowel		**Back Vowel**	
tense	**lax**	**tense**	**lax**
/e/	**/ɛ/**	**/o/**	**/ɔ/**
bait	bed	boat	boy
cape	dev	goat	coy
lake	fed	toad	joy
pave	get	vote	toy
sage	peg	dope	fall

Stimuli

The stimuli used in this study consisted of isolated words spoken by a female native speaker of American English. The speaker was chosen based on her fluency in American English, and the fact that she was a native speaker of the language and had formal training as a linguist. The audio was recorded using a Zoom H4n voice recorder; the recording took place in a quiet room with minimal background noise. The speaker was given a systematically constructed list of real English words to read out loud. The list was designed to cover the contrasts between tense and lax vowels in the high front and high back region of the vowel space. Once the recording was completed, the audio files were transferred to a computer and were then reviewed for sound quality; any extraneous noise or errors were removed. Sound clips for individual words were extracted from the main audio file and were used as stimuli to be played to the participants in the study.

Overall, the recording process was designed to ensure that the stimuli used in the study were of high quality and accurately represented the phonetic characteristics of American English. The equipment used and recording conditions were carefully chosen to minimize noise and ensure clarity of sound. The resulting audio files were stored securely and organized in a way that facilitated ease of use during the experiment.

For the second phase (where we collected data on the mid vowels), the stimuli were collected from online Oxford Advanced Learner's Dictionary (OLD, 2023) as sound clips. For each word, the entry for the word was searched in the dictionary and then the American English pronunciation of the word was downloaded as an audio media file.

Experimental set up and procedure

For data collection, this study used "elicited imitation (EI)" method which McDonough (2017) defines as "a testing technique in which a speaker is asked to repeat a series of sentences verbatim" (p. 562). Tomita, Suzuki and Jessop (2009) stated that EI claims the subjects to hear a stimulus and imitate it in their default linguistic competency following reconstructive nature. As a method, EI has been widely used to study oral proficiency in second language. For example, Kwon (2021) used EI to study contrasting phonetic properties of voiceless stop sounds in South Korean English. In this study, participants were asked to shadow 50 words (each repeated three times) in random order. That is, a set of audio stimuli was played to the participants whose imitated productions were then audio-recorded.

As mentioned above, the data in this study were collected in two separate phases that were identical in design. In the first phase, we collected data for four high vowels (IY, IH, UW, and UH); participants were exposed to an imitation task where they shadowed 30 stimuli presented to them (with 3 non-consecutive repetitions of each). Sound clips of the stimuli words were presented to the participants using Praat's (Boersma & Weenink, 2023) ExperimentMFC; the order of the words was completely randomized. Participants were allowed to listen to each stimulus two times at most. Participants listened to the stimuli via an A4TECH Bloody G525 Virtual 7.1 Surround Sound Gaming Headphone, and then immediately repeated the words loudly. The second phase of data collection followed

the same procedure to collect data on mid vowels from a different set of demographically similar participants.

The recording sessions were done in a sound-attenuated media lab at Green University, Bangladesh; a ZoomH1N Professional Voice Recorder (Version 1.19) was used to record the imitated productions. All the recorded files were stored in WAV format. Participants spent about 5-7 minutes in the first phase and about 8-10 minutes in the second phase to complete the task.

The recorded files were transcribed at word-level in Praat (Boersma & Weenink, 2023), an open-source software. Transcribed files were then processed and forced-aligned to phoneme-level annotations using DARLA (Reddy & Stanford, 2015). All the annotated phonemes were manually checked and cleaned. A custom-developed Praat script was used to extract the first two formants of the target vowels; results were extracted in a CSV file. Statistical analysis including visual inspections of the data was performed in the R Programming language (R Core Team, 2023).

In general, the formants (F1 and F2) were measured at the midpoint of the vowels. For the diphthong(-ized) vowels (/e/, /o/, and /ɔɪ/), the measurements were taken at 30% into the vowel after the vowel onset so that the measurements reflect the steady portions (and not the more dynamic gliding portion) of the vowels.

Results

This section presents the results of the study. We present the results organized as two separate phases: first for high vowels and then for the mid vowels.

High vowels

L1 vowels

Before diving into the results of the L2 productions, we need to confirm whether the tense/lax distinctions were consistently present in the L1 stimuli. Figure 5 presents the distributions of the four high vowels (front and back) on the F1 and F2 planes. The x-axis represents F2 which corresponds to the frontness of a vowel while the y-axis represents F1 which corresponds to the height of a vowel. Both axes are reversed since F1 and F2 are inversely correlated with vowel height and vowel frontness, respectively. The labels indicate the centers of the vowel distributions (in the form of mean F1 and F2); the ellipses show the confidence ellipses at a 95% confidence level (based on 2 standard deviations) for each vowel category. The degree of overlap between two ellipses indicates how overlapped the two categories are; a complete overlap would indicate no difference between the two vowels while no overlap would indicate that the two categories are completely different.

The symbols used for vowels in Figure 5 and the remaining plots follow the ARPAbet convention (Klautau, 2001), primarily because the DARLA forced aligner returned the vowel transcriptions in this convention. ARPAbet is the alphabetical presentation of phonetic transcription. It was developed by the Advanced Research Projects Agency (ARPA). The list of ARPAbet corresponded to IPA used in this study is presented in Table 4.

Table 4: List of ARPAbet with IPA

ARPAbet	IPA	ARPAbet	IPA
IY	/i/	EY	/e/
IH	/ɪ/	EH	/ɛ/
UW	/u/	OW	/o/
UH	/ʊ/	OY	/ɔɪ/

As Figure 5 shows, all four vowel categories have non-overlapping distributions. This confirms that the distinction between the tense vowels and their lax counterparts is clearly available in the L1 audio stimuli. In terms of the durational characteristics, the tense vowels were longer than the lax ones, in general; Figure 6 confirms this trend. As Figure 6 shows, we checked for the durational differences between tense and lax vowels separately for the contexts where the vowel preceded a voiced vs. a voiceless consonant since vowels before a voiced consonant are typically longer than vowels before a voiceless consonant (Durvasula & Luo, 2012; Jacewicz et al., 2007). The figure clearly indicates that the tense vowels were longer than the lax vowels in the L1 stimuli presented.

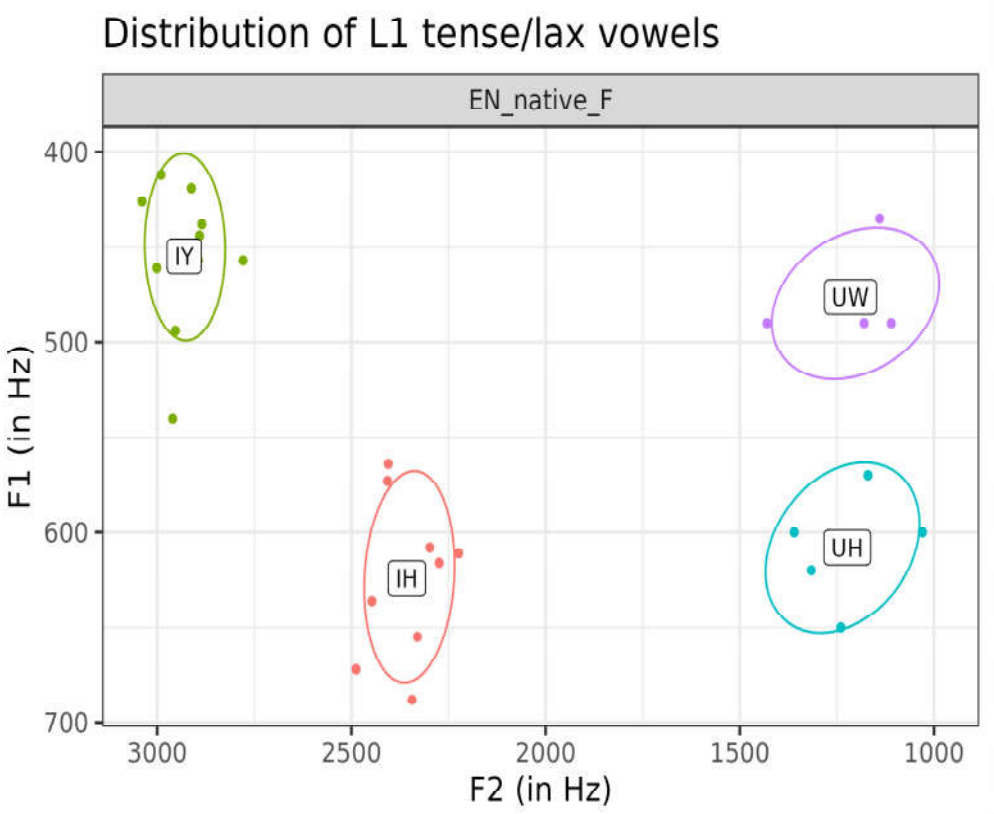


Figure 5: Distribution of vowels in L1 speech on the F1 x F2 plane (high vowels)

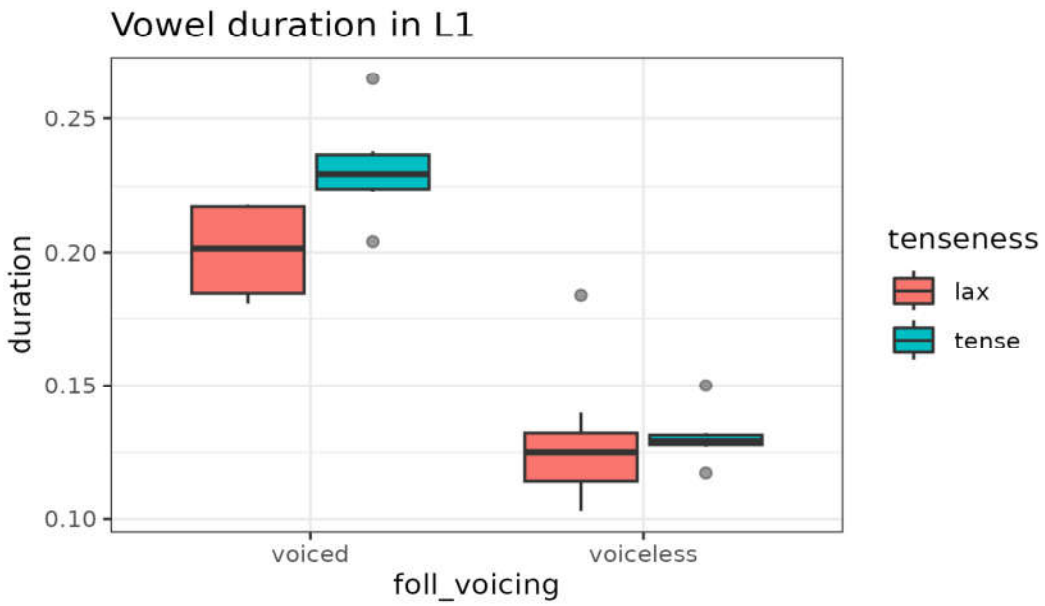


Figure 6: Duration of vowels in L1 English stimuli (Higher vowels)

L2 vowels: Spectral features

Figure 7 shows the F1 and F2 space for the L2 productions of the high vowels. Again, the x-axis shows the F2 values (in Hz) and the y-axis represents F1 values (in Hz); the ellipses represent 95% confidence ellipses for each vowel. Each speaker's data have been plotted in separate panels. As the figure indicates, most speakers tend to distinguish between the two front vowels more successfully than the two back vowels. We did not normalize the formant values since we are comparing vowel categories only within speakers (i.e., each speaker acts as their own baseline).

To focus on the overlaps between the pairs of confidence ellipses, Figure 8 shows only the ellipses without the individual tokens for the vowels. As the figure indicates, the two front vowels (IY and IH) had a considerably smaller degree of overlap, except Speaker "22M," compared to the two back vowels (UW and UH). The front lax vowel IH was less front (with lower F2) and lower (with higher F1) than its tense IY counterpart. The two back vowels, on the other hand, were less distinguishable for most speakers; only Speaker "03F" appeared to have produced the distinction between UW and UH. For other speakers, the two categories had considerable overlap or were often completely merged.

Thus, the results for the high vowels show that speakers were noticeably successful in producing the tense/lax contrasts with the front vowels (IY vs. IH) while the same speakers were not much successful in producing the distinctions for the two back vowel categories (UW vs. UH).

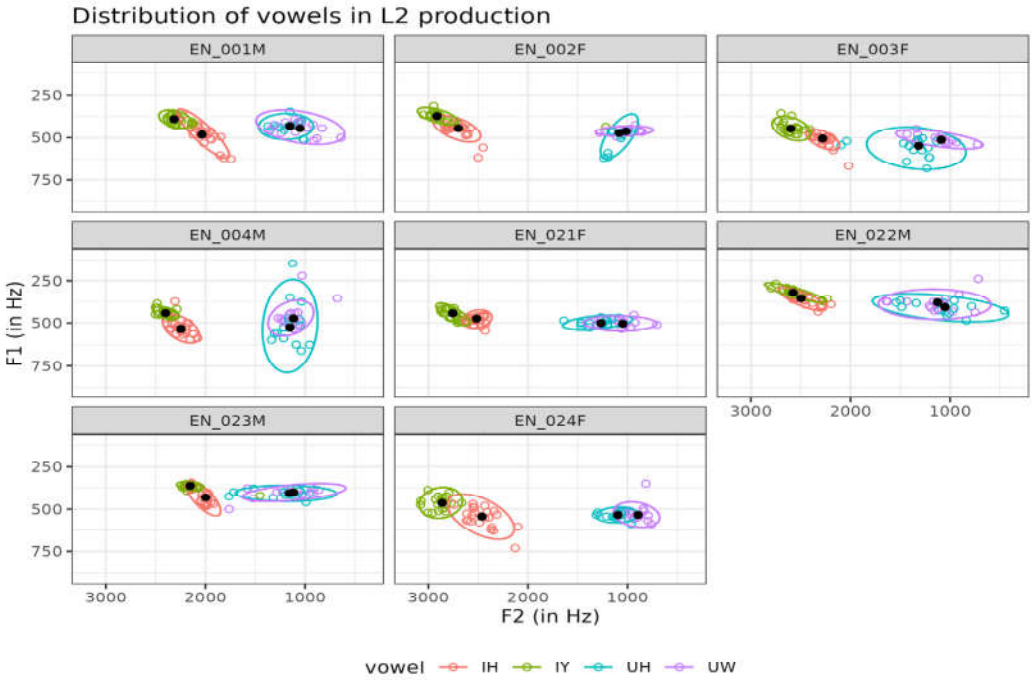


Figure 7: Formant distribution in tense vs. lax vowel in L2 speech (high vowels)

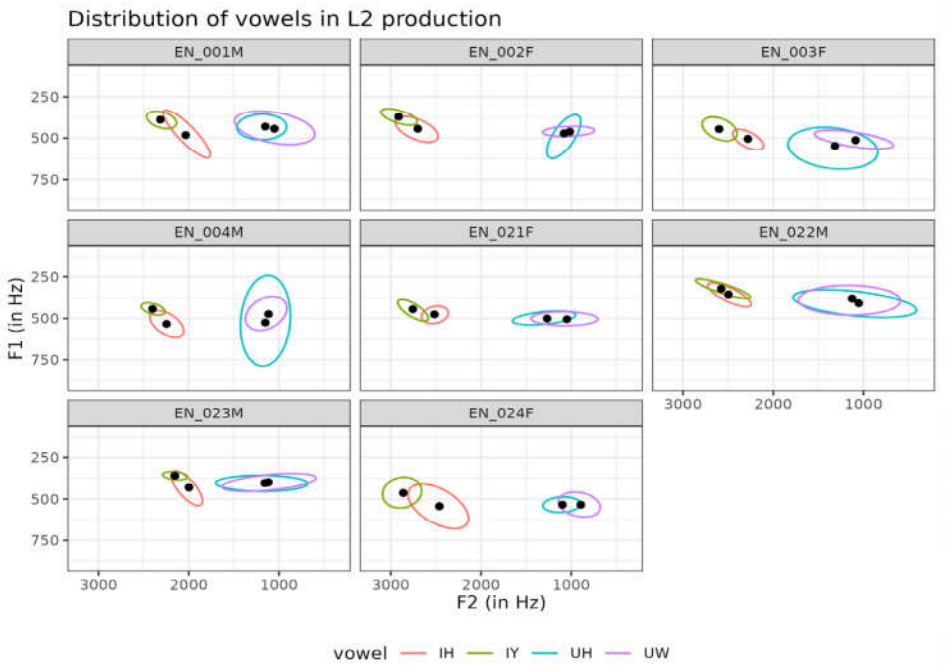


Figure 8: Formant distribution in tense vs. lax vowel in L2 speech (only ellipses)

Pillai-Bartlett Traces comparison

The Pillai-Barlett Trace has been a popular and effective method among phoneticians for determining the degree of overlap between distributions of two neighboring vowel categories (Nycz & Hall-Lew, 2013). The Pillai-Barlett Trace, informally known as the Pillai score, refers to a statistic that is produced by a Multivariate Analysis of Variance (MANOVA) model which is a particular case of regular ANOVA "that models variation with respect to more than one dependent variable simultaneously" (Nycz & Hall-Lew, 2014, pg. 5). Vowels are typically described by two dependent variables (F1 and F2), and that is why a MANOVA model can fit very well for the purpose of comparing two vowels in a bi-dimensional space. In addition, MANOVA output provides a p-value that indicates whether the Pillai score produced is statistically significant or not.

Pillai score is a value bounded between 0 and 1; a higher score indicates non-overlapping distributions between the two categories being compared. That is, a Pillai score of 1 indicates that the two vowels are completely distinct, while a score of 0 indicates that the two categories are completely merged. Pillai statistic for vowel comparison was introduced by Hay et al. (2006) and it has been widely used in linguistics to compare two mergers and splits. For example, Hay et al. (2006) used this method to compare distributions of [ɪə] and [ɛə] vowels in New Zealand English. Pre-lateral (before /l/) [ɔ] and [ʊ] vowels in New Zealand English were analyzed using this method by Kennedy (2006). Pillai trace was also used by Hall-Lew (2009) and Wong and Hall-Lew (2014) to study the differences between [ɑ] and [ɔ] vowels in San Francisco and New York City. In addition, Islam and Ahmed (2020) used this method to compare the mergers of the mid and low front vowels in the Mymensingh dialect of Bangla.

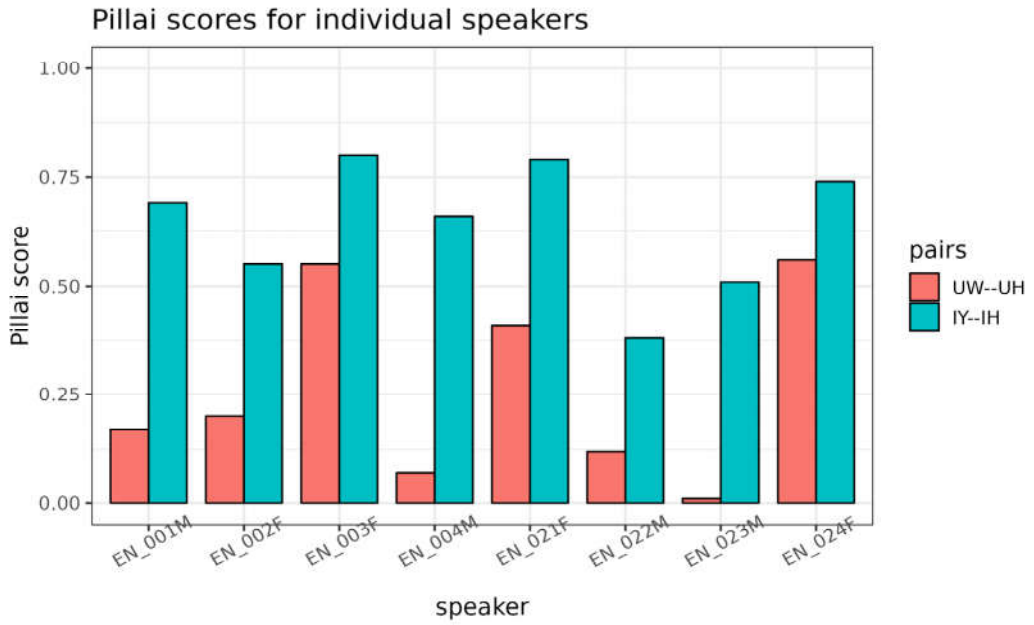


Figure 9: Pillai scores for tense vs. lax vowels distributions by individual speakers (high vowels)

As Figure 9 shows, the Pillai score for the tense/lax comparisons for the front vowels was considerably higher than for the back vowels for each speaker. MANOVA models returned statistically significant p-values for the front vowel comparisons for all speakers ($p < .001$ in each case). For the back vowel comparison, Pillai scores for only speakers "EN_003F" and "EN_024F" were statistically significant ($p < .001$); others were not statistically significant. These results provide a clear indication that speakers were consistently better at producing the tense/lax contrast for front vowels than the back vowels. We ran a Wilcoxon signed rank exact test to check if the difference of performance between the front vs. the back vowels was statistically significant; the results of the test confirmed that the difference is indeed statistically significant ($V < 0.001$, $p = .008$) with a large Wilcoxon effect size of 0.89.

L2 vowels: Duration

In terms of temporal characteristics, Figure 10 presents the temporal differences between the tense and lax vowels, separately for the voiced and voiceless contexts.

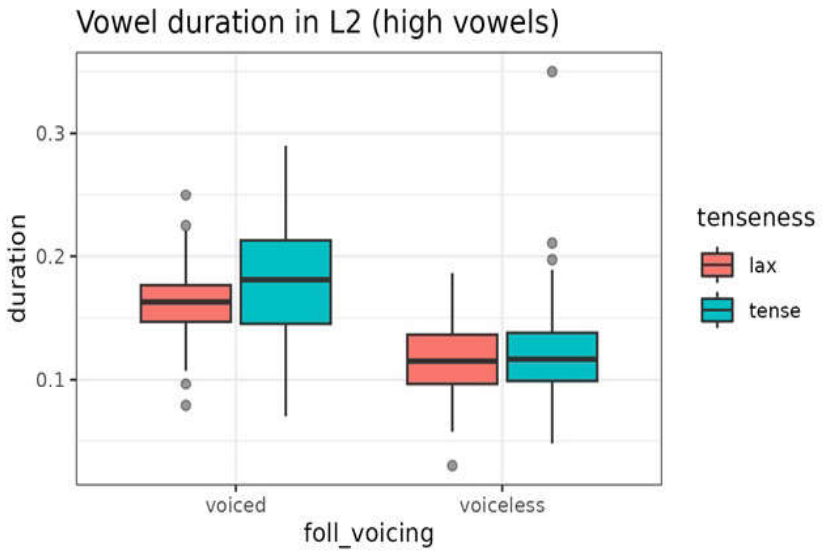


Figure 10: Duration of vowels in L2 tense vs. lax vowels (high vowels)

As the figure indicates, there are no obvious durational differences between the tense and lax vowels in the voiceless context. However, in the voiced context, the trend is not obvious. To investigate whether the tense vowels were significantly different from the lax vowels in terms of their length, we performed two separate two samples t-tests (two-tailed): one for the voiceless and another for the voiced context. Results revealed that the difference in the voiceless context was not statistically significant ($t = -1.49$, $df = 374$, $p = .136$). Contrarily, the difference between tense and lax vowels in the voiced context was found statistically significant ($t = -2.73$, $df = 164$, p-value = 0.007); it can be noted, however, that the effect

size of this difference, obtained via the R package rstatix (Kassambara, 2022), was found to be "small" (Cohen's D = 0.4).

Mid vowels

L1 English vowels (stimuli)

Next we come to the tense/lax contrasts among the mid vowels (the data collected in the second phase). Similar to what was done for the high vowels, we first aimed to confirm whether the tense/lax distinctions were consistently present in the L1 stimuli. Figure 11 presents the distributions of the four mid vowels (front and back) on the F1 and F2 planes. The x-axis represents F2 which corresponds to the frontness of a vowel while the y-axis represents F1 which corresponds to the height of a vowel. As before, the labels indicate the centers of the vowel distributions (in the form of mean F1 and F2); the ellipses show the confidence ellipses at a 95% confidence level. As Figure 11 shows, all four vowel categories have non-overlapping distributions. The two front vowels are considerably far apart from each other, indicating their complete distinctness in the vowel space. The two back vowels too are distinct from each other since the two ellipses do not overlap; however, the two categories appear to be very close to each other in the vowel space (compared to the two front vowels).

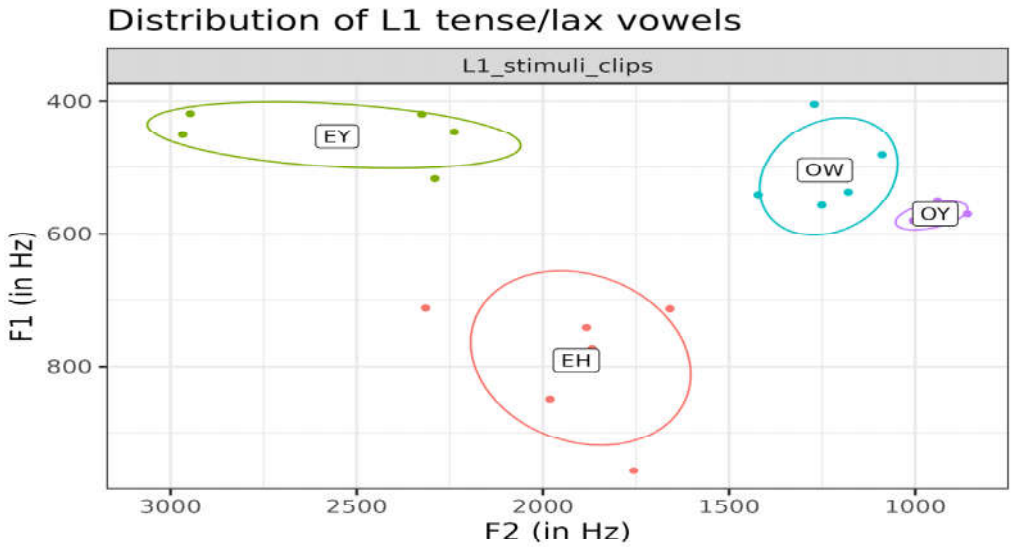


Figure 11: Distribution of vowels in L1 speech on the F1 x F2 plane (mid vowels)

L2 vowels

Analogous to high vowels, Figure 12 shows the F1 and F2 space for the L2 productions of the high vowels, with F2 values on the x-axis, F1 values on the y-axis and ellipses representing the 95% confidence ellipses for each vowel. Individual panels represent separate speakers. A quick visual inspection of the plot indicates that the tense vs. lax vowels among the front vowels had less overlap between them compared to those among the back vowels.

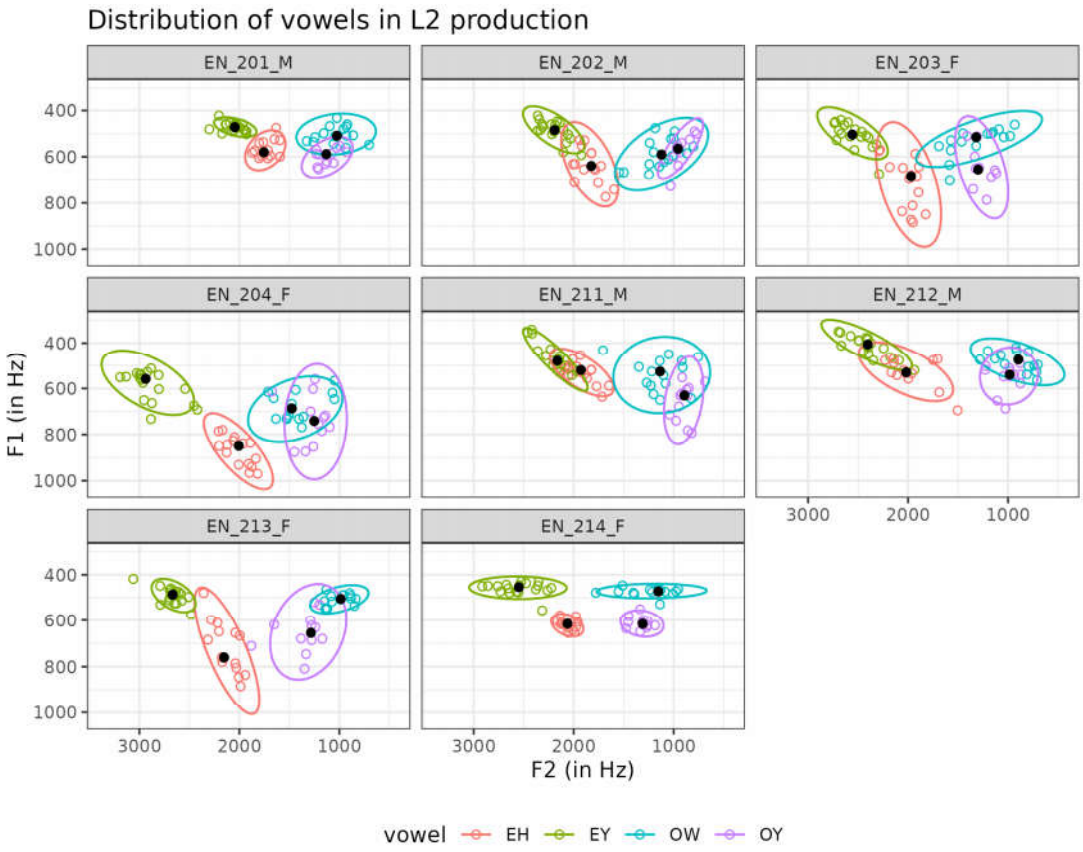


Figure 12: Formant distribution in tense vs. lax vowel in L2 speech (mid vowels)

Figure 13 shows only the ellipses without the individual tokens for the vowels; this plot makes it easier to see that 6 out of the 8 speakers were very successful in producing the distinction between EY and EH. The scenario was completely the opposite for the back vowels where six out of the eight speakers had a large degree of overlap between OW and OY vowels. Thus, the results look pretty similar to what we saw for the high vowels: speakers were more successful in distinguishing the tense/lax contrasts in the front of the vowel space than in the back of the vocal tract.

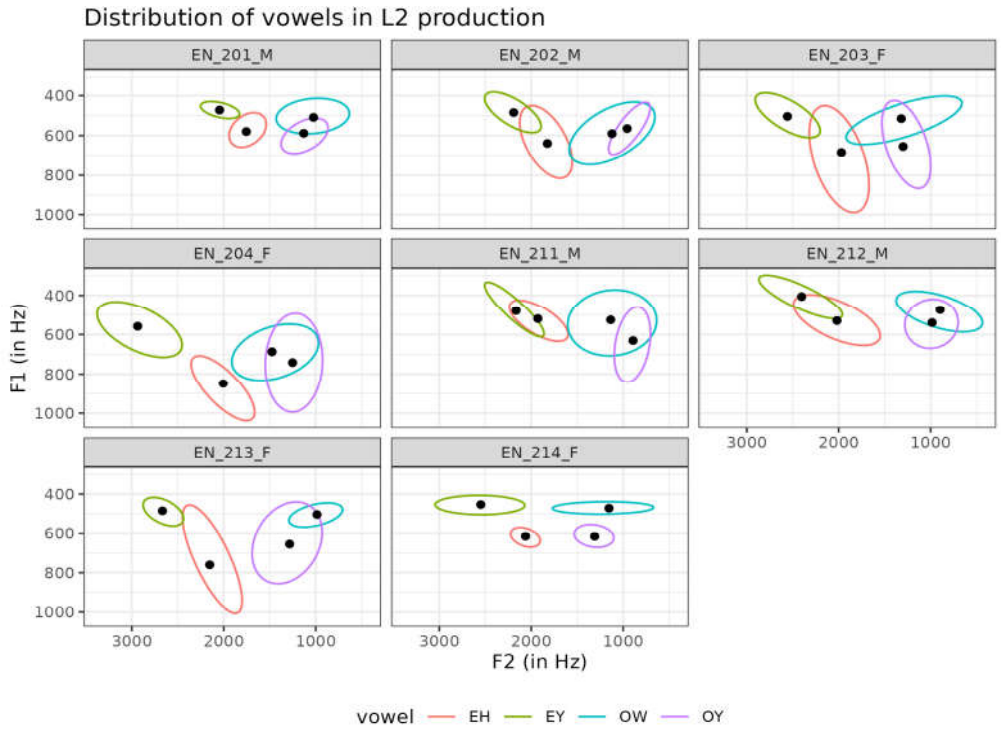


Figure 13: Formant distribution in tense vs. lax vowel in L2 speech (mid vowels, only ellipses)

Pillai scores

To statistically assess the tense-lax distinctions, Figure 14 presents the Pillai scores for comparisons between distributions of concerned vowel pairs in the front and back of the vocal tract.

As Figure 14 shows, the Pillai score for the tense/lax comparisons for the front vowels was consistently higher, with two clear exceptions (speakers EN_211_M and EN_214F). MANOVA models returned statistically significant p-values for the front vowel comparisons for all speakers ($p < .001$ in each case). For the back vowel comparisons, Pillai scores were statistically significant for speakers EN_201_M, EN_213_F, and EN_214_F ($p < .001$ in each case); others were not statistically significant. These results provide an indication that speakers were consistently better at producing the tense/lax contrast for front vowels than the back vowels. A. Wilcoxon signed rank exact test revealed that the difference of performance between the front vs. the back vowels in the form of Pillai scores was statistically significant ($V < 0.001$, $p = .008$, effect size = 0.89).

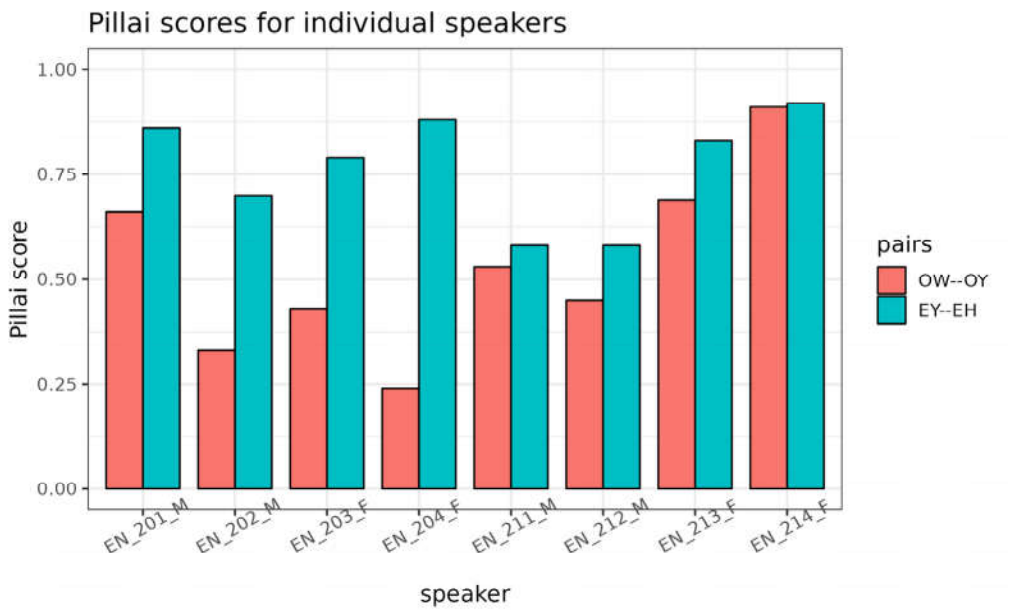


Figure 14: Pillai scores for tense vs. lax vowel distributions by individual speakers (mid vowels)

Discussion

While most previous studies reported that second language learners primarily rely on temporal cues (compared to spectral cues) while listening to tense/lax vowels, our study indicates that this may not be a universal phenomenon. We had two hypotheses in this study: 1) Bangla speakers will be able to successfully produce the distinctions between English tense and lax vowel pairs in terms of their spectral properties (i.e., the formant values), unlike what have been reported from speakers of other languages; and, 2) Bangla speakers will not be able to produce consistent durational differences between English tense and lax vowels pairs. And, generally speaking, we have found sufficient evidence in support of both our hypotheses.

Participants in this study did not consistently produce the durational differences between English tense and lax vowels even though the differences were salient in the L1 stimuli the participants shadowed. This indicates that Bangla speakers are not sensitive to the durational differences or they do not prioritize duration as a cue to differentiate between tense and lax vowels; this is consistent with L1 Bangla tense and lax vowels where duration does not differ systematically (as seen in Alam et al., 2008).

Contrarily, Bangla speakers were able to produce the spectral differences in terms of the F1 and F2 values in English tense and lax vowels more consistently than durational differences. Some noticeable patterns were observed for the spectral properties with respect to the vowel height and backness. First, speakers were able to distinguish the high front tense vowel (/i/) from its lax counterpart (/ɪ/) significantly better than the tense/lax pair in the high back position (/u/ vs. /ʊ/); in fact, the two high back vowels were nearly the same for

most speakers. One reason behind this discrepancy between the degree of success between the high front and high back vowels could be diagonal differences between the vowels, as presented in the L1 English stimuli. For example, /i/ is different from /ɪ/ in two ways: 1) it has a lower F1 value and 2) it has a higher F2 value (see Figure 5). Therefore /i/ and /ɪ/ are different on two dimensions. The two high back vowels, however, differ primarily on F1 (/u/ has a lower F1 value than /ʊ/); they have very similar F2 values. This reduced differences in dimensions may have contributed to the decreased success for differentiating the tense and lax vowels in the high back region. It remains to be seen whether Bangla speakers would be able to perform better if /u/ and /ʊ/ differed in F2 values as well (which is very much possible to observe in different dialects of English).

Secondly, even though Bangla speakers were not able to distinguish the English tense and lax in the high back region, they were much better for the mid back tense/lax vowels (less than 50% overlap between the two categories, compared to 70-80% overlap for the high vowels). A potential explanation could be that Bangla vowel inventory already has a tense/lax contrast in this region which is why they possess the perceptual sensitivity to detect the differences here (especially given that the Euclidean distance between /o/ and /ɔ/ was very small in the L1 stimuli (see Figure 9). This also indicates that Bangla speakers are less sensitive to F1 differences in lower frequency regions.

The results in this study contradict the claims from the previous studies from two perspectives. First, the claim that L2 English learners predominantly over-rely on durational features to distinguish tense and lax vowels (compared to spectral features) could not be corroborated; rather, this suggests that it is completely possible for L2 English speakers to emphasize spectral cues over durational cues. Second, unlike Rahman's (2018) study, our study indicates that Bangla speakers are indeed capable of distinguishing English tense vowels (though not like the native English speakers), even without being significantly exposed to live interactions with native English speakers, when they are forced to pay attention to the necessary details. It needs to be noted that the data used in this study are inherently different from Rahman's study. In Rahman (2018), participants read passages and sentences while this study collected isolated words through a shadowing task; therefore, the data used in this study is inherently less natural. In a shadowing task, participants have to listen to the auditory or perceptual cues very carefully before they can produce the sounds; contrarily, in a reading task, there is no direct perception involved.

Furthermore, most L2 English learners in Bangladesh have never had considerable exposure to native English speech in real life; therefore, it is very unlikely that they would be able to recognize and then learn those categorical differences between the tense and lax vowels. While it is possible to be exposed to native pronunciations through the media (e.g., movies, shows), very few people have access to them; more importantly, research has shown that language learning does not happen effectively in the media form without any live interaction with human beings (Kuhl et al., 2003). Therefore, Bangla speakers have possibly never had the necessary exposure to native English tense and lax vowels, and have not developed the perceptual categories in the first place. But this does not mean that

Bangla speakers are completely incapable of perceiving the spectral differences between English tense and lax vowels; rather, with proper exposure to L1 English, Bangla speakers can successfully distinguish between the concerned vowel categories.

Conclusion

In this study, we investigated whether temporal cues are more important to L2 English speakers than spectral cues when distinguishing English tense vowels from their lax counterparts. We also tested whether Bangla speakers are incapable of distinguishing English tense and lax vowels, as proposed in previous studies. Our results indicate that the over-reliance of temporal cues is not a predominant or universal phenomenon; rather, whether the L2 speaker prioritize temporal or spectral cues are primarily determined by the nature of the phonetic implementation of tense and lax vowels in the first language of the speakers. Our results also confirm that Bangla speakers are indeed capable of perceiving and producing spectral differences between the English tense and lax vowels, even without significant exposure or live interaction with native speakers of English. Furthermore, Bangla speakers put more emphasis on spectral cues, unlike many other languages, over temporal cues when perceiving and producing tense and lax vowels.

References

Alam, F., Habib, S.M., & Khan, M. (2008). "Acoustic analysis of Bangla vowel inventory." Center for Research on Bangla Language Processing, BRAC University. https://dspace.bracu.ac.bd/xmlui/handle/10361/643

Barman, B. (2009). A contrastive analysis of English and Bangla phonemics. *Dhaka University Journal of Linguistics*, *2*(4), 19-42.

Boersma, P. & Weenink, D. (2023). Praat: Doing phonetics by computer [Computer program]. Version 6.3.09. Retrieved 2 March 2023 from http://www.praat.org/

Botma, B., Sebregts, K., & Smakman, D. (2012). The phonetics and phonology of Dutch mid vowels before /l/. *Laboratory Phonology*, *3*(2), 273-297.

Ćavar, M.E., Rudman, E.M., & Oštarić. A. (2022). Temporal versus spectral cues in L2 perception of vowels: A study with Polish and Croatian learners of English. *Journal of Slavic Linguistics 30*(1): 85-107.

Cebrian, J. (2007). Old sounds in new contrasts: L2 production of the English tense-lax vowel distinction. *ICPhS,* 6-10 August 2007, ID 1576.

Chen, Y. (2006). Production of tense-lax contrast by Mandarin speakers of English. *Folia phoniatrica et logopaedica*, *58*(4), 240-249.

Durvasula, K., & Luo, Q. (2012). Voicing, aspiration, and vowel duration in Hindi. In *Proceedings of Meetings on Acoustics 164ASA* (Vol. 18, No. 1, p. 060009). Acoustical Society of America.

Fox, M.M. & Maeda, K. (1999). Perception and production of American English tense and lax vowels by Japanese speakers. *University of Pennsylvania Working Papers in Linguistics 6*(1), Article 21.

Gao, Y., Ding, H., & Birkholz, P. (2020). An acoustic comparison of German tense and lax vowels produced by German native speakers and Mandarin Chinese learners. *The Journal of the Acoustical Society of America*, *148*(1), EL112-EL118.

Ghosh, B. (2018). Effect of contrastive focus on vowel duration in Bangla. *The Journal of the Acoustical Society of America*, *143*(3), 1757-1757.

Hall-Lew, L. (2009). *Ethnicity and phonetic variation in a San Francisco neighborhood*. PhD thesis, Stanford University.

Harrington, J., Hoole, P., Kleber, F., & Reubold, U. (2011). The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *Journal of Phonetics*, *39*(2), 121-131.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458-484.

Islam, M.J. (2018). Phonemic status of Bangla nasal vowels: A corpus study. *Acta Linguistica Asiatica*, *8*(2), 51-62.

Islam, M.J. (2019). *Phonetics and Phonology of 'Voiced-Aspirated' Stops: Evidence from Production, Perception, Alternation and Learnability*. PhD Thesis. Georgetown University.

Islam, M.J. (2022). The role of prevoicing, breathy-voicing and aspiration in the perception of breathy-voiced stops in Bangla. *Poznan Studies in Contemporary Linguistics*, *58*(1), 29-58.

Islam, M.J., & Ahmed, I. (2020). Mid-front and back vowel mergers in Mymensingh Bangla: An acoustic investigation. *Linguistics Journal*, *14*(1), 206-232.

Jacewicz, E., Fox, R.A., & Salmons, J. (2007). Vowel duration in three American English dialects. *American Speech*, *82*(4), 367-385.

Kassambara A (2022). rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.7.1. https://CRAN.R-project.org/package=rstatix

Klautau, A. (2001). ARPABET and the TIMIT alphabet. An archived file. https://web.archive.org/web/20160603180727/http://www. laps. ufpa. br/aldebaro/papers/ak_arpabet01.pdf

Kuhl, P.K., Tsao, F.M., & Liu, H.M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096-9101.

Kwon, H. (2021). A non-contrastive cue in spontaneous imitation: Comparing mono-and bilingual imitators. *Journal of Phonetics*, *88*, 101083.

Lee, J.Y. (2008). Perception of English high vowels: Duration as a cue by Korean speakers of English. *Kansas Working Papers in Linguistics 30*, 195.

Leung, K.K., Jongman, A., Wang, Y., & Sereno, J.A. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *The Journal of the Acoustical Society of America*, *140*(1), 45-58.

McDonough, K. (2017). Experimental research methods. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 562-576). Routledge.

Mora, J.C., & Fullana, N. (2007). Production and perception of English/i:/-/I/and/æ/-/^/in a formal setting: Investigating the effects of experience and starting age. In *Proceedings of the 16th international congress of phonetic sciences 3*, 1613-1616.

Morshed, A.K.M. (1972). *The phonological, morphological and syntactical patterns of standard colloquial Bengali and the Noakhali dialect*. PhD Thesis. University of British Columbia.

Nycz, J., & Hall-Lew, L. (2013, December). Best practices in measuring vowel merger. In *Proceedings of Meetings on Acoustics 166 ASA* (Vol. 20, No. 1). Acoustical Society of America.

OLD. (2023, February). *Oxford Learners' Dictionary*. https://www.oxfordlearnersdictionaries.com

Payne, E., Maxwell, O., & Volchok, B. (2019, August). Tense-lax contrasts in Indian English vowels: Transfer effects from l1 Telugu at the phonetics-phonology interface. In

Proceedings of the 19th International Congress of Phonetic Sciences. Melbourne, Australia. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Rahman, A.M. (2018). Tense-lax merger: Bangla as a first language speakers' pronunciation of English monophthongs. *Asian Englishes*, *20*(3), 220-241.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reddy, S. & Stanford, J. (2015). Toward completely automated vowel extraction: Introducing DARLA . *Linguistics Vanguard*, *1*(1), 15-28. https://doi.org/10.1515/lingvan-2015-0002

Roesler, L., & Song, J.Y. (2018). Acoustic characteristics of tense and lax vowels across sentence position in clear speech. *The Journal of the Acoustical Society of America*, *144*(6), EL535-EL540.

Rojczyk, A. (2010). Overreliance on duration in nonnative vowel production and perception: The within lax vowel category contrast. *Achievements and Perspectives in SLA of Speech: New Sounds*, *2*, 239-249.

Shamim, A. (2011). *A reanalysis of Bengali vowel assimilation with special attention to metaphony*. MA Thesis. The City University of New York.

Thompson, H. (2012). *Bengali*. John Benjamins Publishing Company.

Tomita, Y., Suzuki, W., & Jessop, L. (2009). Elicited imitation: Toward valid procedures to measure implicit second language grammatical knowledge. *TESOL Quarterly*, 345-350.

Wong, A.W.M., & Hall-Lew, L. (2014). Regional variability and ethnic identity: Chinese Americans in New York City and San Francisco. *Language & Communication*, *35*, 27-42.

Zsiga. E. C. (2013). *The sounds of language: An introduction to phonetics and phonology*. Wiley-Blackwell.