Accepted for publication in *Journal of Second Language Pronunciation*, January 14, 2021
https://doi.org/10.1075/jslp.20033.gro

Perception of American English Consonants /v/ and /w/ by Hindi Speakers of English

Vikas Grover[a], Valerie L. Shafer[b], Luca Campanelli[c,d], D. H. Whalen[b,c,e], and Erika S. Levy[f]

[a]Department of Speech-Language Pathology, School of Health Sciences and Practice, New York Medical College, Valhalla, NY, USA; [b]Program in Speech-Language-Hearing Sciences, The Graduate Center, CUNY, New York, NY, USA; [c]Haskins Laboratories, New Haven, CT, USA; [d]Department of Linguistics, University of Georgia, Athens, GA, USA; [e]Department of Linguistics, Yale University, New Haven, CT, USA; [f]Department of Biobehavioral Sciences, Teachers College, Columbia University, New York, NY, USA

Abstract

This study examined perception of the American English (AE) /v/-/w/ consonant contrast by Hindi speakers of English as a second language (L2). A second aim was to determine whether residence in the US modulated perception of this difficult contrast for proficient bilingual Hindi-English listeners. Two groups of Hindi-English bilinguals (the first resided in the US for more than five years, the second lived in India) and a group of AE-speaking listeners participated in the study. Listeners' identification and discrimination of nonsense words (e.g., "vagag" vs. "wagag") were examined. Hindi listeners performed significantly less accurately than AE controls. Accuracy by Hindi listeners was near chance for identification and higher-than-chance for discrimination. Exposure to AE in the US did not improve performance. These results are consistent with previous studies of late L2 learners and extend findings to a population that was proficient in an L2 before arriving in the L2 environment.

Keywords: Speech perception, Hindi, Bilingualism, Length of residence

Address correspondence to: Vikas Grover, Department of Speech-Language Pathology, School of Health Sciences and Practice, New York Medical College, 40 Sunshine Cottage Road, Valhalla, NY 10595, USA. E-mail: vgrover2@nymc.edu

1. Introduction

In the United States (US) alone, approximately 650,000 individuals report speaking Hindi at home (United States Census Bureau, 2015). Many speakers of Hindi first learn English in India, where English serves as an important second language (L2). Thus, US immigrants from the Hindi-speaking urban regions of India often arrive in the US with fairly proficient English skills. Even so, Hindi speakers' pronunciation of certain English speech sounds, in particular /v/ and /w/, exhibits differences from native English speakers. These differences suggest that the Hindi speakers may have difficulty perceiving these speech sounds contrastively. Poor speech intelligibility and speakers being stereotyped are some common problems related to this /v/-/w/ contrast (Chand, 2009).

The goal of the current study was to investigate whether proficient Hindi-English bilinguals who learned English as an L2 in India exhibit perceptual difficulty with the English /v/ versus /w/ contrast. A second goal was to examine whether extended residence in the US improves perception of /v/ and /w/ in these bilinguals. This knowledge will contribute to our understanding of the nature of L2 speech learning and inform pedagogical and clinical practice.

1.1. Status of /v/ and /w/ in English versus Hindi

The consonants /v/ and /w/ (e.g., in "vine" versus "wine") are phonemically contrastive in AE and in other varieties of English. Hindi, however, does not contrast these speech sounds. It is unclear to what extent Hindi speakers of Indian English produce or perceive speech categories that resemble /v/ or /w/ found in English varieties spoken by speakers of American or British English (Chand, 2009; Iverson, Wagner, Pinet, & Rosen, 2011). The lack of contrast between /v/ and /w/ in Hindi is reflected in Hindi orthography, in which the grapheme "व" represents a single Hindi speech sound /ʋ/, which is the closest match to English /v/ or /w/ (Ohala, 1999; Pierrehumbert & Nair, 1996; Sahgal & Agnihotri, 1988; Whitney, 1896). Production of this labiodental approximant /ʋ/ involves a place of articulation similar to that of English /v/, with the upper teeth and lower lip coming in contact, and the manner of articulation similar to that of an English /w/, with rounded lips. Hindi speakers use the labiodental place of articulation and the approximant manner in a single Hindi speech sound (Ohala, 1999). Because Hindi speakers produce this labiodental approximant /ʋ/ in Hindi, it is important to explore how they perceive English /v/ and /w/. Previous studies showed that Hindi speakers assimilated English /v/ and /w/ to the single Hindi labiodental approximant /ʋ/ (Iverson, Ekanayake, Hamann, Sennema, & Evans, 2008; Iverson et al., 2011). In line with the previous findings, our pilot data with eight Hindi speakers also showed that Hindi speakers assimilated 100 % of English /v/ and /w/ tokens to the single labiodental approximant /ʋ/ in Hindi.

Surprisingly, only one study has used a rigorous experimental approach to test perception of English /v/ and /w/ by Hindi first language (L1) speakers for whom English was an L2 (Iverson et al., 2011). Iverson et al. (2011) reported poorer perception of this contrast by Hindi listeners compared to native British-English listeners. A second study examined /v/-/w/ perception in Sinhala L2 learners of English, for whom the L1 also does not have a /v/-/w/ contrast (Iverson et al., 2008). These L2 learners also showed significantly lower accuracy than British English speakers on /v/ and /w/ perception. Thus, these two studies suggest that /v/-/w/ is a difficult contrast for L2 learners for whom this contrast is not present in the L1.

However, the conclusion that /v/-/w/ is a difficult contrast cannot be broadly generalized for several reasons. First, both studies by Iverson and colleagues (2008, 2011) examined perception using synthetic speech. It is possible that the richer phonetic information found in natural speech would allow for better perception by Hindi-English bilinguals. Second, the goal of Iverson et al. (2011) was to examine cross-language specialization of phonetic processing and little information about participants' language background or exposure to English was provided. The current study aimed to address these limitations by using natural speech stimuli and by comparing two groups of Hindi speakers with different degrees of exposure to AE.

In Fuchs's (2019) comparison of British-English and Indian English speakers' production of the /v/-/w/ contrast, significant differences in the spectral centroids were found between the two groups. Fuchs hypothesized a possible merger between /v/ and /w/ in Indian English speakers and recommended a perception study on this contrast. The Indian English participants in his study, however, included speakers from more than one Indian language (i.e., Hindi, Bengali, Malayalam, and Telugu). Fuchs found slight, but non-significant, differences in the spectral centroid and the normalized second formant (F2) for /v/ and /w/ in the different Indian English groups. The differences in the L1 Hindi and Bengali groups were slightly greater than those in the L1 Malayalam and Telugu groups. However, it would be difficult to draw conclusions regarding influence from other phonologies because the *n* was only five in each group. The labiodental approximant or fricative /v/, for example, might not be present in Bengali phonology (Alam, Habib, & Khan, 2008). Anecdotally, some Bengali speakers also use /b^h/ instead of /v/ or /w/ (e.g., 'bhery' for 'very' or 'bhater' for 'water'). The goal of the present study was to examine perception of the /v/-/w/ contrast in Hindi speakers; we included only standard Hindi speakers for whom Hindi was the L1.

1.2. Bilingual experience

Studies have shown a strong relationship between the age of arrival in a new country (where an L2 is based) and speech perception and production skills in the L2 (Flege, 1991; Flege & Fletcher, 1992). These studies, however, have largely focused on learners who have had little experience with the L2 in the home country, beyond classroom learning (e.g., Italian learners of English). English in India, however, is a special case because English is used in a greater variety of contexts. English was introduced to India during British colonization. Eventually, the Indian population adopted English for administrative tasks (Schneider, 2007; Sharma, 2006; Fuchs, 2016).

The status of English in India, as well as other countries that were ex-British colonies, is different from that found in other nations in which English is viewed as a foreign language (Buschfeld et al., 2014; Melchers et al., 2019; Schneider, 2007). Since the independence of India, the knowledge and use of English in India have increasingly been viewed as prerequisites for success in Indian society (Sharma, 2006; Schneider, 2007). In contemporary India, English is considered an integral part of culture (Sahgal, 1991). It is now a fundamental part of the education system in India. English is increasingly the medium of teaching in schools in India starting in elementary schools (Piller & Skillings, 2005), although more so in urban areas, private schools, as well as certain regions of India (Fuchs, 2016). Fuchs reported that English is increasingly serving as the medium of instruction in schools in India. In Hindi-speaking metropolitan regions such as New Delhi, the medium of instruction in schools can range from only Hindi (i.e., negligible English exposure) to only English (i.e., courses taught using English

as a medium). In regions where a language other than Hindi is the L1 (e.g., Malayalam, Assamese, Bengali, and Tamil), a three-language formula is often used in education, in which English is one of the three (Aggarwal, 1988). In the current study, we focus on participants with Hindi as L1, who received an education in English and Hindi from an early age in school in New Delhi. These are speakers of an educated variety of English and it is this variety that is developing into a standardized Indian English (Fuchs, 2016).

However, the nature of the input varies across schools. With regard to the /v-w/ contrast, English teachers in some schools are bilingual Hindi-English speakers who may not make a distinction between /v/ and /w/ in their own productions. Even so, American and British English media may influence perception. This may be less the case for Hindi speakers born before approximately 1985 because internet access was not consistently available until the 1990s. Thus, it is not known whether English experience in India has allowed native Hindi speakers to develop distinct /v/ and /w/ phonological categories.

Investigations of how the age of arrival in a country modulates speech perception and production of an L2 have focused almost exclusively on L2 learners who have relatively weak skills, if any, in the language at the time of arrival. Hindi-English bilinguals, however, often have strong skills in English when they arrive in a predominantly English-speaking country. Thus, on some measures of language, Hindi speakers might be expected to perform similarly to early learners of an L2. For example, Hindi speakers might be expected to resemble Spanish-Catalan bilinguals, who have high proficiency in both languages, rather than adult L2 learners, such as the Italian-English late bilinguals studied by Flege et al., (1999), who learned their L2 starting after the age of 13 years. The few studies that have carefully examined speech perception in early bilinguals suggest that even in this population, speech perception and processing can differ from those of monolinguals (e.g., Baigorri, Campanelli, & Levy, 2018; Sebastian-Galles & Begoña, 2012; Hisagi et al., 2014). Given that Iverson et al. (2011) is, to our best knowledge, the only study of Hindi-English listeners' perception of /v/ and /w/, and that the bilinguals in the study had resided in London for less than five years, it is an open question whether a longer period of English experience would allow for improvement in perception of English /v/ and /w/.

1.3. Automatic Selective Perception Model

Strange's (2011) Automatic Selective Perception Model (ASP) can serve as a framework for explaining how Hindi-English bilinguals might perceive English L2 speech sounds. The underlying premise of the ASP model is that L1 speech perception is rapid and automatic. L1 learners develop selective perception routines (SPRs). These SPRs allow them to quickly and automatically identify L1 speech sounds and recover lexical meanings. According to this model, L2 learners are able to learn to discriminate and identify novel phoneme categories of the language; however, this process is attention-dependent and effortful. As a result, under difficult task conditions, L2 learners often show poor performance because they fall back on L1 SPRs. This is problematic when the L1 SPRs do not result in the correct phoneme identification. In the case of English /v/ and /w/, the claim is that monolingual Hindi listeners would assimilate these phonemes into a single Hindi category. As a consequence, identification of /v/ and /w/ would be particularly poor because this difficult task relies on long-term memory to match tokens to stored phonological representations. Discrimination of the two stimuli would show better performance than identification if the interstimulus interval is fairly short (less than approximately 1 sec) because the stimuli can be compared in short-term memory (Strange, 2011).

Support for the ASP model comes from studies showing more accurate L2 speech perception in simpler tasks (e.g., Strange, 2011) and in tasks allowing attention to be directed towards discrimination (e.g., Hisagi, et al., 2010; Hisagi, et al., 2015). Studies in which task difficulty is manipulated reveal that naïve listeners can discriminate difficult contrasts more accurately when working memory load is minimized (Strange, 2011; Hisagi & Strange, 2011). For example, Hisagi, et al. (2010) showed that when attention was directed to a Japanese vowel duration contrast (in an auditory oddball counting task, where the target was the vowel duration change), naïve English listeners showed good behavioral discrimination (although poorer than the Japanese group). They also showed comparable neural discrimination (using the Mismatch Negativity -MMN), to that of native Japanese listeners. However, when attention was directed away from the vowel duration change (in a visual oddball counting task), the non-native listeners showed an attenuated neural response, whereas the native group showed no change with attention. A number of studies have shown an attenuated MMN in passive tasks (attention focused on the visual modality) to non-native speech contrasts in naïve listeners or adult learners of an L2 (Hisagi et al., 2015; Näätänen, et al., 2007). Thus, studies using the MMN design support the ASP model's main claim, that native language speech perception is largely automatic.

Studies of early bilinguals, who have become proficient in an L2 before puberty, indicate considerable variability in speech perception skills (Flege et al., 1987). For example, Hisagi et al. (2015) showed that early Spanish-English bilinguals, who began learning English before five years of age, were as accurate as English monolinguals at identifying and discriminating a difficult English vowel contrast. However, when attention was focused away from the stimuli, many of these bilingual listeners showed attenuated MMN to the contrast. In a follow-up study, which was designed to examine more directly how attention to the speech modality modulated neural processing in early Spanish-English bilinguals, no difference in MMN was observed between monolinguals and bilinguals, although other attention-related responses did differ between groups (Datta et al., 2020).

Hindi L2 learners of English, however, might show a different pattern of results compared to other late L2 learners because of the special status of English in India. In particular, higher L2 proficiency in grammar and lexical knowledge would allow for more resources to be directed towards speech perception. It is also possible that Hindi-English bilinguals would have had more experience with native English varieties early in their English learning, which might allow for more accurate perception of /v/ and /w/.

Strange's ASP provides a theoretical bases for the listeners' ability to detect various acoustic parameters and their interaction with the selective perception routines of L1 while perceiving L2 speech sounds. Two other influential models of cross-language speech perception are the Perceptual Assimilation Model (Best, 1995; Best and Tyler, 2007) and the Speech Learning Model (SLM) (Flege, 1995). The PAM and PAM-L2 models examine the relationship between perceptual assimilation and discrimination. In part, because perceptual assimilation was not investigated here, we did not refer to this model. The main objective of SLM was to investigate L2 pronunciation, although there are a few studies examining perception (Flege, 1995). The current study only focuses on the perception of the /v/-/w/ contrast.

1.4. The current study

The current study examined Hindi-English bilinguals' perception of AE consonants /v/ and /w/. This is the first study to examine Hindi speakers' perception of this contrast with the use of natural speech stimuli and comparison of two groups of Hindi speakers with different degrees of exposure to English. We further investigated whether long-term experience with the L2 (> five years) in an Anglophone country (the US), would modulate L2 perception of /v/ and /w/. Hindi listeners were asked to identify and discriminate tokens of /v/ and /w/ in AE nonsense word forms. The identification task required reliance on long-term memory. In contrast, an AXB discrimination task was designed to allow participants to match the target (X) to the A or B token. The identification task was expected to yield less accurate outcomes than the discrimination task because it required access to long-term memory representations, as well as requiring the participant to hold the target stimulus in working memory to allow comparison (Strange, 2011). The /v/ and /w/ targets were presented both in word-initial and word-medial positions, and with different stress patterns to allow generalization of the results across a wider range of word shapes.

There is some controversy about whether Hindi has lexically contrastive stress (Abbasi, Pathan, & Channa, 2018). The current paper does not directly examine this factor; however, given that AE prosody includes stress, realized as increased duration, greater intensity and higher pitch (Bolinger, 1962), stress was included as a factor in the design. L2 experience was included as a group factor as we compared performance of Hindi listeners who lived in India (Hindi IND) and Hindi listeners who lived in the US (Hindi US) with a length of Residence (LOR) of more than 5 years. Variables including age of first exposure to English, length or experience (e.g., years of schooling in English), LOR in the US, and self-reported proficiency were collected by means of a language background questionnaire (LBQ).

The criterion of LOR of more than 5 years was selected to explore perceptual performance beyond the time frame examined in a previous study (Iverson et al., 2011). Iverson et al. (2011) included Hindi speakers who resided in London for 1-5 years. Although LOR is not always directly correlated with the language experience, of interest here were the effects of a longer length of residence on /v/ and /w/ perception. All 16 participants in the Hindi US group were exposed to AE in their daily work environment in the New York/New Jersey region. Also, 10 out of the 16 participants in the Hindi US group had school-aged children in the New York/New Jersey region and socialized with their AE-learning children in AE environments. The participants' immersion in AE professional and sometimes social environments likely exposed them to the AE /v/ and /w/ contrast.

We hypothesized that the Hindi-English bilinguals would show less distinct categorization of /v/ and /w/ in the identification task than AE monolingual listeners. In contrast, we predicted higher accuracy in discrimination than identification for Hindi-English bilinguals, although accuracy still might be lower than for the AE control group. We further hypothesized that performance by the Hindi US group, compared to the Hindi Ind group, would more closely resemble that of the AE control group because of their increased exposure to the /v/-/w/ contrast in AE. More specifically, we expected that the Hindi US group would perform above chance levels in categorizing /v/ and /w/ as two distinct phonemes and would show higher identification and discrimination accuracy than the Hindi IND group. We further hypothesized that more accurate discrimination would be observed for stressed than unstressed syllables because the

increased salience, in terms of intensity, duration, and fundamental frequency, would allow the listeners to match the stimuli in the AXB task more easily.

2. Methods

2.1. Participants

Fifty-two adults (25 males, 27 females, age range: 30-45 years) were recruited for this study. Of these participants, 16 (eight females and eight males) were Hindi listeners who had lived in the US for more than five years. Participants also included 20 (10 females and 10 males) Hindi listeners who had lived only in India. Finally, 16 (nine females and seven males) were monolingual AE listeners who spoke a relatively standard (Northeast) variety of AE. The 20 Hindi listeners who had lived only in India were recruited from New Delhi, a region where Standard Hindi is spoken. All spoke Hindi as their L1. Most of the listeners had also been exposed to Punjabi phonology, which includes the same labiodental approximant /ʋ/ as Hindi. All spoke English as an L2. The two Hindi groups included in this study demonstrated significant (qualitative and quantitative) differences in terms of self-reported exposure ($p < .001$) and self-reported proficiency in English ($p = .023$), but the difference between the two groups' self-reported proficiency in Hindi was not significant ($p < .444$). Detailed analysis for the language backgrounds of the two groups is provided in the Appendix. Data from four of these participants (Hindi listeners who lived in India) were removed from the analysis due to excessive ambient noise during testing. In total, data from 48 participants were used, 16 in each group. Mean age in years (*SD*) of the English group was 34.1 (4.74), of the Hindi US group was 38.6 (3.65), and of the Hindi IND group was 39.1 (4.28). The mean (*SD*) for the LOR in the US for the Hindi US group was 12.6 (4.30). The participants in both Hindi groups had learned English while growing up in India. For all participants, the medium of instruction in school was either English or both English and Hindi—none had received their education in only Hindi. Based on the institutionalization of English in India, it is likely that English in India was taught by teachers who spoke Indian English (Sharma, 2006); thus, any exposure to AE for the Hindi US group only occurred when they arrived in the US.

2.2. Stimulus materials

Four monolingual female speakers of AE from the New York State or Baltimore region recorded naturally-produced English nonsense word stimuli for the experimental study. Table 1 shows these nonsense word forms. Consonants included /v, w, b, f/ in initial and medial positions in Consonant (C) Vowel (V) Consonant (C) Vowel (V) Consonant (C) [CVCVC] in **C**VCVC and CV**C**VC combinations (the underlined and bolded consonant indicates the experimental consonant). The phonemes /b/ and /f/ served as control consonants because they are contrastive in Hindi and should be easily categorized and discriminated. The other (non-target) consonant used in the nonsense words was /g/, and the vowel in both syllables was /ɑ/. The target /v/ and /w/ tokens were recorded in both stressed and unstressed syllables and in initial and medial position, with one target per word. Multiple tokens of nonsense words were recorded on a Dell computer with a Turtle Beach, Motego II sound card, using Shure (Model SM 10A) head-mounted microphone in a sound-shielded booth and digitized at 22050 Hz using Sound Forge (version 4.5). From this large set, final stimuli were selected that were similar in fundamental frequency (F0) (range: 189-220 Hz.; Mean: 192 Hz.). Table 1 provides the mean syllable

duration and standard deviation for each word type. The final selected word forms were normalized for intensity by root mean square (RMS) using Adobe Audition (version 6). Stimuli were labeled and verified by three AE listeners who did not participate in the experiment, in order to ensure accuracy in the production of the intended consonant. Any items that were not identified with 100% accuracy were removed and rerecorded.

Table 1. Nonsense word forms and syllable duration in milliseconds by condition. Mean (SD).

	Initial Position		Medial Position	
Consonants	Unstressed	Stressed	Unstressed	Stressed
/v/	vaˈgag 54 (8.5)	ˈvagag 178.5 (16.8)	ˈgavag 170.3 (11)	gaˈvag 267.1 (16.6)
/w/	waˈgag 81 (12)	ˈwagag 198.1 (14.5)	ˈgawag 189 (8.9)	gaˈwag 290.1 (11.2)
/b/	baˈgag 47.3 (4.3)	ˈbagag 152.5 (8)	ˈgabag 191.6 (6.2)	gaˈbag 268.2 (18.8)
/f/	faˈgag 80.5 (12.7)	ˈfagag 180.2 (9.3)	ˈgafag 259.1 (28.5)	gaˈfag 348 (30.4)

2.3. Procedure

The participants were tested in a quiet room. Each participant's hearing was screened before the start of the experiment at 20 dB HL (1000, 2000, and 4000 Hz.) A brief interview was conducted with the Hindi-English speaking participants to assess their conversational skills in Hindi. Questions such as "Where do you work?" and "Tell me about your favorite Hindi movies" were asked in Hindi. In addition, participants were asked to name as many animals as possible in one minute (timed by the researcher). This task was included as a means to evaluate conversational fluency in Hindi and English (Hurks et al., 2006).

For all participants, the stimulus words were presented from a laptop computer (Lenovo, ThinkPad T430) by means of E-Prime software (version 2.0.10.248; Psychology Software Tools, Pittsburgh, PA) and a high-quality Razer Kraken 7.1 Chroma headset at a range of 75 dB SPL to 90 dB SPL. Prior to the study, stimulus intensity was calibrated using a sound level meter (Larson-Davis 800B precision Integrating Sound Level Meter). Participants were allowed to adjust the volume to a comfortable level at the beginning of the experiment within this range. Each task took approximately 30 minutes (M = 27, SD = 4) with breaks, for a total experimental time of 2.5 hours. This includes production tasks that are not reported here.

2.3.1. Task 1-Identification. In the identification task, participants were presented with the series of nonsense words. For each word presented, they were asked to press a button on the keyboard to identify the target sound in that word. The subsequent word was delivered two seconds after the response. There were two blocks of approximately 10 minutes each in the identification task. Each block presented the target in one position only (i.e., initial or medial position). E-Prime software recorded the response accuracy and reaction times.

The first block included the target sounds in the initial position of the nonsense word (for example, /ˈvagag, ˈwagag, vaˈgag, waˈgag/ (/ˈ/ indicates stress). The second block included the target sound in the medial position of the nonsense word (for example, /ˈgavag, ˈgawag, gaˈvag,

gaˈwag/). The order of tokens in each block was randomized using E-Prime software. For each block, each /v/ and /w/ token was presented four times (4 speakers × 4 word forms × 4 presentations = 64) and each /b/ and /f/ token was presented twice (4 speakers × 4 word forms × 2 presentations = 32), resulting in a total of 96 tokens per block. Participants were instructed to press a key on the laptop labeled with "v", "w", "b" and "f" to identify the first consonant (in the first block) or the second consonant (in the second block) of the nonsense word. The blocks were presented in the same order for all participants (block 1, then block 2).

Practice (with 16 tokens) was provided before each block for task familiarization using a different consonant set (/s, t, n, m/) and keys labeled as "S, T, N, M". All participants were expected to identify these consonants easily. Feedback was provided for the practice task. After the participants succeeded on this task (accuracy > 95%), they were given instructions for the experimental stimuli.

2.3.2. Task 2-Categorial AXB discrimination. In this AXB task, participants were presented with a series of three stimuli. The interstimulus interval (ISI) between the tokens was 250 ms (Massaro, 1974). For each triplet sequence (AXB), the participant was required to decide whether the second word (X) sounded like the first word (A) or the third word (B) in the sequence. If X sounded like A, the participant was instructed to press, "1" on the laptop keyboard and if X sounded like B, the participant pressed "3". Two seconds after the response, the participant was presented with the next trial.

There were four 10-minute blocks. In the first block, the target was in initial position with the target syllable stressed (for example, /ˈvagag, ˈwagag/). In the second block, the target was in initial position with the target syllable unstressed (for example, /vaˈgag, waˈgag/). In the third block, the target was in medial position with the target syllable unstressed (for example, /ˈgavag, ˈgawag/). In the fourth block, the target was in medial position with the target syllable stressed (for example, /gaˈvag, gaˈwag/).

In each block, there were 80 total presentations, out of which 16 presentations were for the 'v/w' tokens. For example, eight AXB triplets were presented with the order 'v-v-w' and eight with the order 'w-w-v'. The other 64 presentations used 'b/f' with either 'v' or 'w' and included eight presentations each with the AXB order 'b-b-v, v-v-b, b-b-w, w-w-b, f-f-v, v-v-f, f-f-w, w-w-f'. Sequences including only 'b and f', such as 'b b f or f f b', were not presented in order to maintain the task time under 45 minutes. The triplets were randomly selected (without replacement) by E-Prime from the list of triplet sets from the four different AE speaker stimulus sets, as described below. A total of 64 'v/w' tokens (16 target tokens × 4 blocks) were presented in the discrimination task.

In a triplet (AXB), all three nonsense words were tokens from the same speaker; however, the two "same" nonsense words were different tokens from the same speaker. This was to encourage the participants to attend to the phonemic-level, rather than simply to acoustic similarity. The order of experimental tasks was counterbalanced such that half of the participants began with the identification task and subsequently performed the discrimination task and the other half performed the tasks in the reverse order.

2.4. Data analysis

For both identification and AXB tasks, the independent variables consisted of *Group* (English, Hindi IND, and Hindi US), *Position* (Initial and Medial), *Stress* (target stressed and unstressed),

and *Stimulus Type* (for the identification task, the consonants /b/, /f/, /v/, and /w/; for the discrimination task, the pairs 'b-v, b-w, f-v, f-w, v-w'). The dependent variables were *Response Accuracy* and *Response Time* (RT).

For response accuracy, data were analyzed using mixed-effects logistic regression. The approach allows modeling the nonlinear component of the dependent variable (data bounded by 0 and 1) and has shown advantages over other techniques such as analysis of variance on aggregated and transformed data (Agresti, 2002; Jaeger, 2008). To reduce Type I error rates, all models included the maximal random effects structure justified by the design (Barr, Levy, Scheepers, & Tily, 2013): Random intercepts for subjects, random slopes for the within-subjects predictors, and their interactions. For three analyses, the random slopes for the interactions between predictors were not retained because of convergence failures.

For both identification and AXB tasks, experimental and control consonants or pairs were examined in separate analyses. This facilitated the presentation and interpretation of the results and reduced model complexity and degrees of freedom. In addition, a direct comparison between control and experimental consonants was not relevant to addressing our research questions.

For each task, condition, and dependent variable, subjects with average proportion accuracy or average response time ± 3 *SD* from the mean were excluded. No more than 1.5% of the data were excluded following this procedure. Data were analyzed with R version 3.4.3 (R Core Team, 2017) using the functions `glmer` and `lmer` from the **lme4** package, version 1.1-15 (Bates, Mächler, Bolker, & Walker, 2015). To facilitate the interpretation of the results, we reported type-III ANOVA tables generated using the `joint_tests` function from the **emmeans** package (Lenth, Love, & Herve, 2018). Post-hoc Tukey adjusted comparisons were carried out using the `emmeans` function from the **emmeans** package (Lenth et al., 2018). Effects were considered statistically significant for $p < .05$.

The Response Time (RT) data showed the same pattern of results as the accuracy data. The RT data are reported in the Appendix for reference.

3. Results

3.1. Identification task: Control consonants

Accuracy of responses to the control consonants /b/ and /f/ was high overall and comparable in English, Hindi IND, and Hindi US participants, but there were some significant differences related to consonant identity, position, and stress. In addition, the Hindi groups were slower at responding to /b/ than the English group. Results of the control contrasts are available in the Appendix.

3.2. Identification task: Experimental contrast

3.2.1. Accuracy. Response accuracy to the experimental consonants /v/ and /w/ was overall at ceiling for the English group (97% accuracy) and near chance for the two Hindi groups (Hindi US, 55%; Hindi IND, 58% correct). Descriptive and inferential statistics are reported in Tables 2 and 3, respectively. Post-hoc Tukey tests confirmed that performance was comparable in the two Hindi groups ($p = .883$) and less accurate in the Hindi groups than in the English group ($p < .001$). In addition to the effect of Group, we also found significant main effects of Consonant

and Position. Identification accuracy was higher for /v/ (72% correct) than /w/ (68% correct) and higher in the medial than the initial position (75% and 65% accuracy, respectively).

The interactions of Group × Consonants and Group × Consonants × Position were statistically significant, as well (Table 3 and Figure 1). For both interactions, Tukey post-hoc comparisons replicated the findings described for the main effect of Group: In no condition did performance accuracy of the two Hindi groups differ significantly ($p > .21$), and all comparisons between the Hindi groups and the English group were statistically significant or approached significance ($p < .054$)[1].

The two-way interaction Group × Consonant was driven by the Consonant factor. Identification accuracy was comparable between /v/ and /w/ for the Hindi US group ($p = .785$). Identification accuracy was higher for /w/ than /v/ for the English group ($p = .008$), but lower for /w/ than /v/ for the Hindi IND group ($p = .007$). Similarly, the three-way interaction Group × Consonants × Position pointed to differential effects of Position and Group on response accuracy for /v/ and /w/ (Figure 1). For the English group, identification accuracy in the Initial position was higher for /w/ than /v/ (99% and 90% correct, respectively; $p < .001$). The reverse pattern was observed in the Hindi IND group; they identified /v/ more accurately than /w/ (61% and 40% accuracy, respectively; $p = .004$). For all other comparisons the difference between /v/ and /w/ was not statistically significant ($p > .12$).

Lastly, a significant Consonant × Position × Stress interaction was found, indicating that the effect of Stress was moderated by Consonant and Position. Stress had a negative effect on identification accuracy of /v/ in Initial position (Unstressed, 79%, Stressed 60%; $p < .001$) and a positive effect on the identification of /w/ (Unstressed 48%, Stressed 74%; $p < .001$). The effect of Stress for consonants in Medial position was not statistically significant ($p > .191$).

Table 2. Identification task: Descriptive statistics for response accuracy by group, condition, and consonant. Proportion correct. Mean (SD).

Position, Stress	Group	/b/	/f/	/v/	/w/
Initial, Stressed	English	0.96 (0.06)	0.82 (0.17)	0.90 (0.04)	1.00 (0.02)
	Hindi US	0.81 (0.13)	0.91 (0.08)	0.44 (0.27)	0.62 (0.31)
	Hindi IND	0.80 (0.20)	0.69 (0.35)	0.45 (0.30)	0.60 (0.33)
Initial, Unstressed	English	0.95 (0.07)	0.85 (0.15)	0.90 (0.12)	0.98 (0.07)
	Hindi US	0.76 (0.27)	0.93 (0.06)	0.71 (0.15)	0.26 (0.25)
	Hindi IND	0.75 (0.23)	0.85 (0.20)	0.77 (0.21)	0.21 (0.18)
Medial, Stressed	English	1.00 (0.00)	0.98 (0.03)	0.99 (0.03)	1.00 (0.02)
	Hindi US	0.96 (0.11)	0.99 (0.02)	0.62 (0.27)	0.59 (0.30)
	Hindi IND	0.97 (0.07)	0.99 (0.03)	0.75 (0.21)	0.49 (0.26)
Medial, Unstressed	English	1.00 (0.02)	1.00 (0.02)	0.98 (0.04)	1.00 (0.00)
	Hindi US	0.96 (0.06)	0.99 (0.03)	0.48 (0.25)	0.70 (0.27)
	Hindi IND	0.99 (0.03)	1.00 (0.02)	0.63 (0.29)	0.71 (0.23)

[1]Note that for all AE listeners response accuracy to /w/ in the Medial, Unstressed condition was 100% (Table 2). Because of this lack of variability, Tukey post-hoc tests for the group comparisons need to be interpreted with caution. The p value of .054 for the Eng-Hindi US contrast, for example, does not correctly reflect the large and reliable difference between the two groups (100% vs. 70% accuracy). Wilcoxon rank sum tests performed on aggregated data confirmed this intuition: For both the English-Hindi US and the English-Hindi IND contrasts, $W > 247$ and $p < .001$.

Table 3. Identification task: ANOVA summary table for response accuracy to the critical consonants /v/ and /w/.

Model term	*df*	*F* ratio	*p*-value
Group	2	15.012	**<.001**
Cons	1	4.929	**.026**
Posit	1	7.85	**.005**
Stress	1	1.395	.238
Group × Cons	2	5.552	**.004**
Group × Posit	2	2.317	.099
Group × Stress	2	1.874	.153
Cons × Posit	1	2.239	.135
Cons × Stress	1	0.91	.340
Posit × Stress	1	3.439	.064
Group × Cons × Posit	2	3.124	**.044**
Group × Cons × Stress	2	1.306	.271
Group × Posit × Stress	2	1.432	.239
Cons × Posit × Stress	1	8.553	**.003**
Group × Cons × Posit × Stress	2	1.897	.150

Note. Group (English, Hindi IND, Hindi US); Cons = Consonant (/v/, /w/); Posit = Position (initial, medial); Stress (target stressed and unstressed). Significant effects in bold.

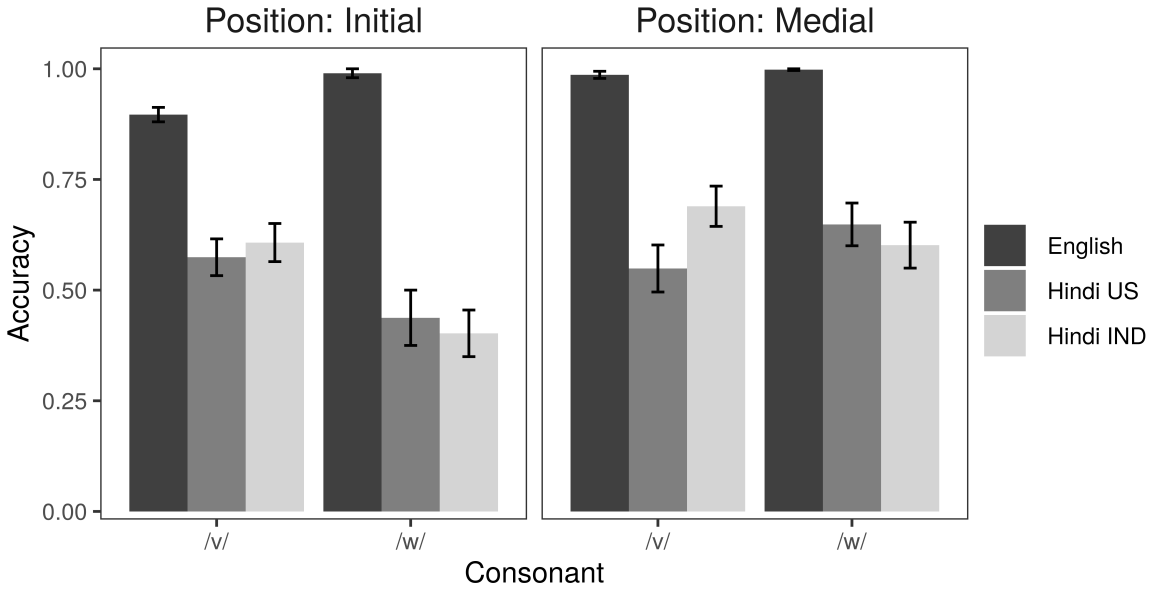

Figure 1. Identification task: Proportion of correct responses by Group, Consonant, and Position. Error bars: ± SE.

3.3. AXB categorial task: Control pairs

Discrimination accuracy of the control pairs was high or near ceiling for all participants, and all groups showed similar RTs. The full description of the analyses for the control pairs is found in the Appendix.

3.4. AXB categorial task: Experimental pair

3.4.1. Accuracy. Categorial accuracy of the experimental pair /v/ and /w/ (referred to as /vw/ in tables) was at ceiling for the English participants (97% correct) and between 75% and 80% for the Hindi groups (Hindi US, 79%; Hindi IND, 77%; See also descriptive statistics in Table 4). Inferential statistics are reported in Table 5. The main effect of Group was followed up with Tukey's post-hoc tests, which showed no significant difference between the two Hindi groups ($p = .596$) and less accurate discrimination accuracy for both Hindi groups compared to the English groups (English-Hindi US, $p < .001$; English-Hindi IND, $p < .001$). There was also a main effect of Position, such that response accuracy was higher for consonants in the Medial than Initial position (Initial, 78%; Medial, 91%).

Table 4. Discrimination task: Descriptive statistics for response accuracy by group, condition, and pair. Proportion correct. Mean (SD).

Position, Stress	Group	/bv/	/bw/	/fv/	/fw/	/vw/
Initial, Stressed	English	0.90 (0.04)	0.97 (0.05)	0.87 (0.08)	0.99 (0.03)	0.96 (0.04)
	Hindi US	0.92 (0.07)	0.96 (0.05)	0.94 (0.08)	0.96 (0.06)	0.71 (0.11)
	Hindi IND	0.92 (0.06)	0.97 (0.05)	0.95 (0.06)	0.93 (0.10)	0.68 (0.12)
Initial, Unstressed	English	0.92 (0.07)	0.98 (0.05)	0.89 (0.08)	0.98 (0.03)	0.96 (0.07)
	Hindi US	0.85 (0.10)	0.91 (0.11)	0.88 (0.08)	0.96 (0.06)	0.70 (0.09)
	Hindi IND	0.87 (0.08)	0.93 (0.10)	0.89 (0.08)	0.98 (0.04)	0.68 (0.08)
Medial, Stressed	English	0.97 (0.06)	0.98 (0.04)	0.96 (0.06)	0.99 (0.03)	0.99 (0.02)
	Hindi US	0.96 (0.06)	0.98 (0.04)	0.98 (0.03)	0.99 (0.02)	0.89 (0.07)
	Hindi IND	0.96 (0.05)	0.96 (0.04)	0.98 (0.03)	0.98 (0.04)	0.84 (0.13)
Medial, Unstressed	English	0.98 (0.05)	0.99 (0.02)	0.98 (0.04)	0.99 (0.02)	0.99 (0.02)
	Hindi US	0.97 (0.06)	0.98 (0.04)	0.99 (0.02)	0.99 (0.03)	0.90 (0.08)
	Hindi IND	0.97 (0.04)	0.97 (0.05)	0.98 (0.03)	0.99 (0.02)	0.87 (0.07)

Table 5. Discrimination task: ANOVA summary table for response accuracy to the critical pair /vw/.

Model term	*df*	*F* ratio	*p*-value
Group	2	56.094	**<.001**
Posit	1	75.715	**<.001**
Stress	1	0.53	.467
Group × Posit	2	0.357	.700
Group × Stress	2	0.403	.669
Posit × Stress	1	0.004	.949
Group × Posit × Stress	2	0.177	.838

Note. Group (English, Hindi IND, Hindi US); Posit = Position (initial, medial); Stress (target stressed and unstressed). Significant effects in bold.

4. Discussion

This study investigated the accuracy with which Hindi listeners perceived AE /v/ and /w/ and whether exposure to AE in the US for at least five years influenced their perception. Hindi listeners were unable to identify /v/ and /w/ consistently, performing at chance level on the identification task. This differed from the AE controls, who identified these speech sounds with high accuracy. As predicted, the Hindi listeners demonstrated better-than-chance performance on the AXB task; but again, this performance was much lower than AE listeners, who showed near-ceiling performance. These findings support our main hypothesis, that Hindi listeners would show less accurate perception of /v/ and /w/ than AE speakers.

Our findings, however, did not support our second hypothesis, that experience with AE in the US would improve performance. Hindi listeners in the US and those in India did not differ significantly. We did observe a Group by Consonant interaction in identification for the Hindi US and Hindi IND group. Specifically, the Hindi IND group favored /v/ over /w/ responses in initial position, whereas the Hindi US group showed no preference.

In addition, we observed some significant effects of position and stress, but these generally did not interact with group. We had no strong evidence for predicting the direction of effects for position and stress, beyond that stress might improve perception because stressed syllables tend to be more salient as they are characterized by greater duration and intensity, and a higher fundamental frequency. Thus, we consider the results on position and stress to be exploratory and in need of replication. Below, we discuss these findings in greater detail in relation to our hypotheses.

4.1. Perception patterns for /v/ and /w/

Researchers of Hindi have expressed uncertainty with regard to the phonemic status of Hindi /ʋ/ (Sahgal & Agnihotri, 1988). The results of this study provided no evidence indicating that one of the AE phones was perceptually closer to the Hindi /ʋ/, since identification was equally poor for both categories. This does not preclude the possibility that Hindi speakers might perceive one of the AE categories as a better exemplar of Hindi /ʋ/. The findings that Hindi speakers performed better on the AXB discrimination task than on the identification task might indicate that while

the two non-native sounds are assimilated into a single native category, they might differ in goodness of fit (i.e., Category Goodness assimilation instead of a Single Category assimilation in the PAM; Best & Tyler, 2007). A future perceptual assimilation study could investigate Hindi listeners' assimilation of AE /v/ and /w/ to the Hindi /ʋ/ category, as well as their goodness ratings of the AE stimuli to more directly assess the listeners' assimilation patterns.

These findings showed a similar pattern to those of the Hindi and Sinhala listeners tested by Iverson et al. (2008, 2011). Specifically, in both studies, the L2 listeners showed equally poor categorization of English /v/ and /w/, performing at near chance levels. Taken together, the results of the present study along with these two studies, suggest that perception of the /v/-/w/ contrast is difficult for late L2 learners of English who do not have this contrast in their L1. It would be of interest to investigate perception of /v/ and /w/ by speakers of languages that also lack this contrast but are typologically different from Hindi and Sinhala, such as German or Dutch. German and Dutch are also similar to English in a number of phonological characteristics; thus, evidence that perception of English /v/ and /w/ is difficult for speakers of these languages, as well, will more strongly support the suggestion that the /v/ versus /w/ contrast is challenging for L2 learners beyond native Hindi speakers.

4.2. Models of second-language learning

Performance by Hindi listeners was less accurate on the identification task than on the categorial AXB task. The Automatic Selective Perception (ASP) Model (Strange, 2011) can explain these task differences. When the Hindi listeners were required to identify L2 /v/ and /w/, they may have had insufficient resources to maintain the detailed phonetic information in memory (Werker & Logan, 1985). As a result, Hindi listeners fell back on their L1 SPR, which did not allow them categorize /v/ and /w/ as different phonemes. In the categorial AXB task, Hindi listeners performed more accurately than on the identification task because task difficulty was decreased and they could use selective attention to identify phonetic differences.

It will be important in future studies to examine which tasks serve to better train an L2 contrast that is assimilated into a single L1 phonological category. Our findings suggest that a first step might be an AXB categorial task, to allow the L2 learner to perceive the difference between /v/ and /w/ (see Bradlow, 2008, for a review of training studies).

4.3. Second language experience

The results of this study did not support our hypothesis that the Hindi US group would show higher accuracy than the Hindi IND group because of their exposure to AE for at least five years. There was, in fact, little difference in behavior between the two Hindi groups. Iverson et al. (2011) also showed no effect of living in an English-speaking country on perception by Hindi L2 speakers of English, although their participants had lived in the country for a shorter time period (1-5 years).

Studies of other later L2 learners of English have previously suggested the length of residence for L2 speakers who have arrived as adults does not strongly correlate with L2 speech production and perception skills (Flege, 1999). However, we could not assume that this would be the case for the Hindi participants in the current study because the sociolinguistic situation for English as an L2 in India is so different from other languages pairs (e.g., Italian-English). The higher level of proficiency for Hindi speakers of English compared to many other groups that come to the US might have allowed for a stronger effect of LOR because Hindi speakers would

be able to use English in more settings. The participants' self-ratings of proficiency did reveal that those who lived in the US consistently rated their English higher (by approximately 1 point on a 7-point scale). However, this self-reported increased proficiency did not extend to perception of the /v/ and /w/ contrast. Future studies will be needed to examine the types of experience that are best able to improve L2 speech perception in participants with this background.

While the differences were not significant, we did observe trends in the Hindi IND and Hindi US groups: On /v/, the Hindi IND group showed slightly higher accuracy than the Hindi US group, and on /w/, the Hindi US group showed slightly higher accuracy than the Hindi IND group. (Only the difference for /v/ approached significance.) This shift, however, cannot be interpreted as resulting from improved perception for the Hindi US group, but rather suggests a shift in response bias away from a slight preference for labeling both AE /v/ and /w/ as "v" to labeling them as "w". It is possible that Hindi listeners who have been in the US are more aware of the contrastive nature of /v/ and /w/, and thus operate under the assumption that many words begin with /w/; however, the current findings indicate that the Hindi listeners' perceptual skills were not sufficient to determine accurately when to label a word as /w/ rather than /v/. More specifically, this finding may indicate the 'v' orthographic character is initially favored, but after coming to the US, Hindi listeners begin to recognize 'w' as an onset and attempt to detect its differences from 'v'.

4.4. Phonemic context effects

An exploratory objective of this study was to determine any effects of syllable position and/or stress on the target consonant performance of Hindi listeners. We did not have a direct hypothesis regarding the possible effects of stress or position, beyond the hypothesis that more accurate perception would be found in stressed syllables. Studies suggest that L2 clear speech is perceived more accurately than L2 conversational speech (Smiljanić & Bradlow, 2005, 2008), supporting the hypothesis that stressed syllables might facilitate perception.

Results of our manipulations suggest that neither stress nor word position led to a significant improvement of perception by the Hindi groups in the identification task, but that these factors did influence performance. Hindi listeners from both groups showed a preference for the 'v' label for /v/ or /w/ in initial unstressed syllables, and for the 'w' label in medial unstressed syllables. This pattern was reversed for stressed syllables, for which listeners showed a slight preference for the 'w' label for initial stressed syllables and the 'v' label for unstressed syllables. This pattern suggests that the Hindi listeners may have been utilizing prosodic cues in their perception of /v/ and /w, but that these cues were not beneficial to their identification. One possibility for the different pattern of performance with the target in medial versus initial position is that the medial consonants are facilitated by the neighboring vowel.

4.5. Limitations and future directions

A limitation of our study is that participants living in India versus those living in the US are likely to have differences other than length of residence in the US. Nonetheless, the two Hindi groups showed similar performance. A second limitation is that we did not gather specific information regarding the nature of the participants' L2 learning (e.g., whether their English teachers spoke English as an L1 or L2 and which social or regional variety of English they spoke). It will be important in future studies to examine how much and what sort of early input

from an L1 English teacher is necessary to allow for higher accuracy in /v/ versus /w/ perception. It will also be important to explore this question in relation to English as an L2 in other countries where English has a special status (e.g., Nigeria, Jamaica) to further our understanding of how the L1 phonology influences the emerging standardized form of English.

Future directions include examinations of the relationship between perception and production of /v/ and /w/ in Hindi listeners and of what types of targeted training of this contrast will lead to significant improvement in its perception and production. A formal perceptual assimilation study with a goodness of fit task could be included with a larger population of Hindi speakers of English to examine the relationship between listeners' perceptual assimilation and discrimination within the framework of the PAM-L2 (Best & Tyler, 2007). It will also be interesting to examine the extent to which age of arrival in the US modulates acquisition of this contrast. Studies of L2 learning suggest a rapid decline in the ability to learn a difficult L2 contrast not found in the L1, particularly for arrival in the L2 country in later childhood and beyond (Flege, 1991; Hisagi et al., 2015; Levy, 2009; Levy & Strange, 2008; Mackay & Flege, 2004). It would be particularly interesting to examine whether Hindi-English participants who arrive in late childhood (between approximately 9 and 14 years) and who are already fairly proficient in English will shift their production and perception to closely resemble AE /v/ and /w/ categories.

4.6. Conclusion

This study revealed that Hindi-English bilingual listeners have difficulty perceiving the AE /v/-/w/ contrast. Findings are extended to a population that was proficient in an L2 before arriving in the L2 environment. It is concluded that their prolonged exposure to this contrast in the US does not lead to improvements in perception. This finding is in line with the small number of studies that have examined /v/ and /w/ in L2 learners of English (e.g., Iverson, et al., 2008; 2011), suggesting that this is a particularly difficult contrast to learn.

References

Abbasi, A. M., Pathan, H., & Channa, M. A. (2018). Experimental phonetics and phonology in Indo-Aryan & European languages. *Journal of Language and Cultural Education, 6*(3), 21–52. https://doi.org/10.2478/jolace-2018-0023

Aggarwal, K. S. (1988). English and India's three-language formula: An empirical perspective. *World Englishes, 7*(3), 289–298.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York, NY: Wiley.

Alam, F., Habib, S.M., & Khan, M. (2008). Acoustic analysis of Bangla consonants. *Spoken Languages Technologies for Under-resourced Languages (SLTU-2008)*, 108-113. Retrieved from https://isca-speech.org/archive/SLTU_2008/papers/su08_108.pdf

Baigorri, M., Campanelli, L., & Levy, E. S. (2018). Perception of American English vowels by early and late Spanish-English bilinguals. *Language and Speech, 62*(4), 681-700. https://doi.org/10.1177/0023830918806933

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross language research* (pp. 171–204). Baltimore: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & M. J. Munro (Eds.), *Second Language speech learning: The role of language experience in speech perception and production* (pp.13–34). Amsterdam, The Netherlands: John Benjamins.

Bolinger, D. (1962). *Forms of English*. Cambridge, MA: Harvard University Press.

Bradlow, A. R. (2008). Training non-native language sound patterns: lessons from training Japanese adults on English /ɹ/-/l/ contrast. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology of second language acquisition* (pp. 287–308). Philadelphia, PA: John Benjamins.

Brown, V. A., Hedayati, M., Zanger, A., Mayn, S., Ray, L., Dillman-Hasso, N., & Strand, J. F. (2018). What accounts for individual differences in susceptibility to the McGurk effect? *PLOS ONE, 13*(11), e0207160. https://doi.org/10.1371/journal.pone.0207160

Buschfeld, S., Hoffman, T., Huber, M., & Kautzsch, A. (Eds.). (2014). *The Evolution of Englishes: The Dynamic Model and beyond.* Amsterdam/Philadelphia: John Benjamins.

Census of India (2011). *Comparative Speaker's strength of Scheduled Languages*. Retrieved from http://www.censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf

Chand, V. (2009). [V]at is going on? Local and global ideologies about Indian English. *Language in Society, 38*, 393-419. https://doi.org/10.1017/S0047404509990200

Chand, V. (2009). *Who Owns English? Political, Social and Linguistic Dimensions of Urban Indian English Language Practices* (Doctoral dissertation). University of California, Davis, CA.

Datta, H., Hestvik, A., Vidal, N., Tessel, C., Hisagi, M., Wróblewski, M., & Shafer, V. L. (2020). Automaticity of speech processing in early bilingual adults and children. *Bilingualism: Language and Cognition, 23*(2), 429–445. https://doi.org/10.1017/S1366728919000099

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America, 89*(1), 395–411. https://doi.org/10.1121/1.400473

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and language experience: Issues in cross-language research* (pp. 233–277). Baltimore, MD: York Press.

Flege, J. E., & Eefting, W. (1987). Cross-Language Switching in Stop Consonant perception and Production by Dutch Speakers of English. *Speech Communication, 6*(3), 185-202. https://doi.org/10.1016/0167-6393(87)90025-2

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of Acoustical Society of America, 91*(1), 370–389. https://doi.org/10.1121/1.402780

Flege, J. E, MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of Acoustical Society of America, 106*(5), 2973–2987. https://doi.org/10.1121/1.428116

Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language, 56*(2), 246–269. https://doi.org/10.1016/j.jml.2006.06.007

Fuchs, R. (2019). Almost [w]anishing: The elusive /v/-/w/ contrast in Educated Indian English. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 1382-1386). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Fuchs, R. (2016). *Speech rhythm in varieties of English: Evidence from educated Indian English and British English.* Singapore: Springer.

Hisagi, M., Shafer, V.L., Strange, W., & Sussman, E.S. (2010). Perception of a Japanese Vowel length contrast by Japanese and American English listeners: behavioral and electrophysiological measures. *Brain Research, 1360,* 89-105. https://doi.org/10.1016/j.brainres.2010.08.092

Hisagi, M., & Strange, W. (2011). Perception of Japanese temporally-cued contrasts by American English listeners. *Language and speech, 54,* 241-64. https://doi.org/ 10.1177/0023830910397499

Hisagi, M., Tajima, K., & Kato, H. (2014). The effect of language experience on the ability of non-native listeners to identify Japanese phonemic length contrasts. *Proceedings of Meetings of Acoustics, 21*, 060003. https://doi.org/10.1121/1.4887491

Hisagi, M., Garrido-Nag, K., Datta, H., & Shafer, V. L. (2015). ERP indices of vowel processing in Spanish–English bilinguals. *Bilingualism: Language & Cognition, 18*(2), 271–289. https://doi.org/10.1017/S1366728914000170

Hurks, P. P. M., Vles, J. S. H., Hendriksen, J. G. M., Kalff, A. C., Feron, F. J. M., Kroes, M., ... Jolles, J. (2006). Semantic category fluency versus initial letter fluency over 60 seconds as a measure of automatic and controlled processing in healthy school-aged children. *Journal of Clinical and Experimental Neuropsychology, 28*(5), 684–695. https://doi.org/10.1080/13803390590954191

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., & Evans, B. G. (2008). Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1305–1316. https://doi.org/10.1037/0096-1523.34.5.1305

Iverson, P., Wagner, A., Pinet, M., & Rosen, S. (2011). Cross-language specialization in phonetic processing: English and Hindi perception of /w/-/v/ speech and nonspeech. *The Journal of the Acoustical Society of America, 130*(5), 297–303. https://doi.org/10.1121/1.3632048

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Lenth, R., Love, J., & Herve, M. (2018). emmeans: Estimated marginal means, aka least-squares means (Version 1.1). Retrieved from https://CRAN.R-project.org/package=emmeans

Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *Journal of the Acoustical Society of America, 125*(2), 1138–1152. https://dx.doi.org/10.1121/1.3050256

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics, 36*(1), 141–157. https://doi.org/10.1016/j.wocn.2007.03.001

Lewandowsky, S., Oberauer, K., Yang, L. X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods, 42*(2), 571–585. https://doi.org/10.3758/brm.42.2.571

Mackay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics, 25*(3), 373–396. https://doi.org/10.1017/S0142716404001171

Massaro, D. W. (1974). Perceptual units in speech recognition. *Journal of Experimental Psychology, 102*(2), 199-208. http://dx.doi.org/10.1037/h0035854

Melchers, G., Shaw, P., & Sundkvist, P. (2019). *World Englishes* (3rd ed.). Routledge. https://doi.org/10.4324/9781351042581

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55*(4), 601–626. https://doi.org/10.1016/j.jml.2006.08.009

Ohala, M. (1999). Hindi. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (pp. 100–103). New York, NY: Cambridge University Press.

Pierrehumbert, J., & Nair, R. (1996). Implications of Hindi prosodic structure. In J. Durand & B. Laks (Eds.), *Current trends in phonology: Models and methods* (pp. 549–584). Salford, UK: University of Salford Press.

Piller, B., & Skillings, M. J. (2005). English language teaching strategies used by primary teachers in one New Delhi, India School. *The Electronic Journal for English as a Second Language, 9*(3). Retrieved from http://www.tesl-ej.org/wordpress/issues/volume9/ej35/ej35cf/

Psychology Software Tools, Inc. [E-Prime 2.0.10.248]. (2015). Retrieved from http://www.pstnet.com.

R Core Team (2017). R: A language and environment for statistical computing (Version 3.4.3). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Sahgal, A. (1991). Patterns of language use in a bilingual setting in India. In J. Cheshire (Ed.), *English around the World: Sociolinguistic Perspectives* (pp. 299-307). Cambridge: Cambridge University Press.

Sahgal, A., & Agnihotri, R. (1988). Indian English phonology: A sociolinguistic perspective. *English World-Wide, 9*(1), 51–64. https://doi.org/10.1075/eww.9.1.04sah

Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Sebastian-Galles, N., & Diaz, Begoña. (2012). First and second language speech perception: Graded learning. *Language Learning, 62*(2), 131-147.

Sharma, M. (2006). Institutionalization of English in India-A Historical Background. *South Asian Review, 27*(2), 175-189. https://doi.org/10.1080/02759527.2006.11932448

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of Acoustical Society of America, 118*(3), 1677–1688. https://doi.org/10.1121/1.2000788

Smiljanić, R., & Bradlow, A. R. (2008). Temporal organization of English clear and plain speech. *Journal of Acoustical Society of America*, *124*(5), 3171–3182. https://doi.org/10:1121/1.2990712

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39*(4), 456–466. https://doi.org/10.1016/j.wocn.2010.09.001

United States Census Bureau (2015). Detailed languages spoken at home and ability to speak English for the population 5 years and over for United States: 2009-2013. [Data File]. Retrieved from https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35–44. https://doi.org/10.3758/BF03207136

Whitney, W. D. (1896). *Sanskrit grammar*. Boston, MA: Ginn & Co.

Perception of American English Consonants /v/ and /w/ by Hindi Speakers of English

Vikas Grover, Valerie L. Shafer, Luca Campanelli, D. H. Whalen, and Erika S. Levy

Appendix

S.1. Language background questionnaire (LBQ)

Table S1. Descriptive statistics for the average of self-reported exposure to and use of Hindi and English. Responses were measured on a 7-point Likert scale with 1 = all Hindi and 7 = all English.

	Average[a]	Parents	Siblings	Children	Spouse	Grandparents	Friends	Community	Colleagues
Hindi IND									
N	16	16	16	15	15	10	16	15	15
Mean	3.19	1.59	2.88	3.83	3	1.4	3.75	3.63	4.83
Median	3.13	1	3	4	3	1	4	4	5
SD	0.77	1.08	1.22	0.82	1.13	1.26	0.78	0.77	1.19
Min-Max	1.86-5	1-5	1-5	2.5-5	1-5	1-5	2-5	2.5-5	3-7
Hindi US									
N	16	16	16	10	13	1	16	16	16
Mean	4.17	2.12	2.75	5.65	3.23	1	4.34	4.66	6.72
Median	4.1	1.5	2.75	6	3.5	1	4	4	7
SD	0.75	1.3	1.15	1.2	1.63	NA	1.19	1.34	0.75
Min-Max	3.12-5.64	1-4.5	1-4.5	4-7	1-6	1-1	2-6.5	3-7	4-7
Difference (Hindi US − Hindi IND)									
Mean	0.98[b]	0.53	-0.13	1.82	0.23	-0.4	0.59	1.03	1.89

Note. [a]Arithmetic mean of the responses to the eight questions (parents, siblings, children, spouse, grandparents, friends, community, colleagues). [b]A two-sample Wilcoxon test indicated that the difference between the two groups on the average variable was statistically significant, $W = 40$, $p < .001$.

Table S2. Descriptive statistics for English proficiency self-ratings. Responses were measured on a 7-point Likert scale with 1 = poor and 7 = good.

	Average[a]	General Proficiency	Grammar	Pronunciation	Reading	Understanding	Writing
Hindi IND							
N	16	16	16	16	16	16	16
Mean	5.6	5.5	5.25	5.44	5.88	5.88	5.69
Median	5.75	5.5	5	5	6	6	5.5
SD	0.74	0.89	0.86	0.89	0.81	1.09	0.79
Min-Max	4.33-6.83	4-7	4-7	4-7	5-7	4-7	5-7
Hindi US							
N	16	16	16	16	16	16	15
Mean	6.23	6.19	6.12	5.94	6.38	6.38	6.33
Median	6.25	6	6	6	6.5	6.5	6
SD	0.63	0.66	0.81	0.77	0.72	0.72	0.72
Min-Max	5-7	5-7	5-7	5-7	5-7	5-7	5-7
Difference (Hindi US − Hindi IND)							
Mean	0.63[b]	0.69	0.87	0.5	0.5	0.5	0.64

Note. [a]Arithmetic mean of the responses to the six questions (general proficiency, grammar, pronunciation, reading, understanding, writing). [b]A two-sample Wilcoxon test indicated that the difference between the two groups on the average variable was statistically significant, $W = 67.5$, $p = .023$.

Table S3. Descriptive statistics for Hindi proficiency self-ratings. Responses were measured on a 7-point Likert scale with 1 = poor and 7 = good.

	Average[a]	General Proficiency	Grammar	Pronunciation	Reading	Understanding	Writing
Hindi IND							
N	16	16	16	16	16	16	16
Mean	6.14	6.19	6	6.12	6.12	6.25	6.12
Median	6.5	6.5	6	6	6	7	6.5
SD	1.07	1.11	1.1	1.09	1.09	1.13	1.15
Min-Max	3-7	3-7	3-7	3-7	3-7	3-7	3-7
Hindi US							
N	16	16	16	16	16	16	15
Mean	6.55	6.75	6.75	6.69	6.5	6.81	5.73
Median	6.58	7	7	7	6.5	7	6
SD	0.43	0.45	0.45	0.48	0.52	0.4	1.03
Min-Max	5.67-7	6-7	6-7	6-7	6-7	6-7	4-7
Difference (Hindi US − Hindi IND)							
Mean	0.41[b]	0.56	0.75	0.57	0.38	0.56	-0.39

Note. [a]Arithmetic mean of the responses to the six questions (general proficiency, grammar, pronunciation, reading, understanding, writing). [b]A two-sample Wilcoxon test indicated that the difference between the two groups on the average variable was not statistically significant, $W = 107.5$, $p = .444$.

S.2. Identification task: Control consonants

S.2.1. Accuracy. Accuracy of responses to the control consonants /b/ and /f/ was high overall and comparable in English, Hindi IND, and Hindi US participants: The main effect of Group and the interactions between Group and the within subject predictors were not statistically significant (Tables 2 and S4). Average identification accuracy for English, Hindi IND, and Hindi US participants was 96%, 95%, and 95% respectively. For the within-subject experimental manipulations, a reliable effect of Position was observed, such that identification accuracy was higher for consonants in medial than initial position (for initial, 84% correct; for medial, 99% correct; $p < .001$).

Table S4. Identification task: ANOVA summary table for response accuracy to the control consonants /b/ and /f/.

Model term	df	*F* ratio	*p*-value
Group	2	1.347	.260
Cons	1	1.009	.315
Posit	1	18.915	**<.001**
Stress	1	0.158	.691
Group × Cons	2	1.968	.140
Group × Posit	2	1.063	.345
Group × Stress	2	0.689	.502
Cons × Posit	1	0.772	.380
Cons × Stress	1	5.212	**.022**
Posit × Stress	1	0.389	.533
Group × Cons × Posit	2	0.572	.565
Group × Cons × Stress	2	0.901	.406
Group × Posit × Stress	2	0.597	.550
Cons × Posit × Stress	1	4.762	**.029**
Group × Cons × Posit × Stress	2	0.808	.446

Note. Group (English, Hindi IND, Hindi US); Cons = Consonant (/b/, /f/); Posit = Position (initial, medial); Stress (target stressed and unstressed). Significant effects in bold.

Two significant interactions, Consonant × Stress and Consonant × Position × Stress (Table S4) were found. The former interaction was driven by a difference in sign between the effect of Stress on the identification of the two consonants: for /b/, accuracy was slightly higher in the Stressed than in the Unstressed condition; for /f/, the reverse pattern was found, identification was poorer in the Stressed than in the Unstressed condition. Those differences, however, were small and none of the post-hoc comparisons resulted in statistically significant differences ($p > .09$. /b/ Unstressed, 90% correct; /b/ Stressed, 92%; /f/ Unstressed, 94%; /f/ Stressed, 90%).

The three-way interaction Consonant × Position × Stress pointed to a differential effect of Stress on different consonants and positions, but provided no new relevant information as compared to the effects described above. The difference between Stressed and Unstressed conditions was not statistically significant for any consonant in any position ($p > .38$) and the only significant differences were driven by the effect of Position (for all comparisons between initial and medial position, $p < .018$.

S.2.2. Response Time. Descriptive statistics for identification response time to the control consonants are reported in Table S5 and inferential statistics in Table S6. Linear mixed-effects regression analysis revealed a statistically significant effect of Consonant, indicating faster responses for /f/ (489 ms) than /b/ (532 ms). There was also a significant Group × Consonants interaction. Post-hoc Tukey tests revealed a statistically significant difference between groups for /b/ but not for /f/. For /b/, response times were comparable in the two Hindi groups ($p = .984$) but slower in the Hindi participants than in the English group ($p < .035$. English, 407 ms; Hindi US, 587 ms; Hindi IND, 601 ms). For /f/, response times were

comparable in all participants (p > .258. English, 439 ms; Hindi US, 540 ms; Hindi IND, 487 ms). No other significant effects were detected. The RT data, in line with the previous studies, were collected using a laptop keyboard (Brown et al., 2018; Fedorenko et al., 2007; Lewandowsky et al., 2010; Oberauer & Kliegl, 2006).

Table S5. Identification task: Descriptive statistics for response time in milliseconds by group, condition, and consonant. Mean (SD).

Position, Stress	Group	/b/	/f/
Initial, Stressed	English	371 (91)	504 (162)
	Hindi US	562 (219)	635 (190)
	Hindi IND	640 (292)	658 (438)
Initial, Unstressed	English	427 (144)	440 (121)
	Hindi US	594 (203)	517 (170)
	Hindi IND	592 (239)	491 (152)
Medial, Stressed	English	387 (116)	405 (164)
	Hindi US	644 (224)	516 (177)
	Hindi IND	626 (362)	481 (152)
Medial, Unstressed	English	449 (98)	436 (164)
	Hindi US	546 (134)	486 (129)
	Hindi IND	534 (169)	453 (139)

Table S6. Identification task: ANOVA summary table for response time to the control consonants /b/ and /f/.

Model term	df1	df2	*F* ratio	*p*-value
Group	2	51.2	3.067	.055
Cons	1	51.16	6.721	**.012**
Posit	1	51.26	0.007	.934
Stress	1	47.87	3.913	.054
Group × Cons	2	51.14	5.173	**.009**
Group × Posit	2	51.24	0.013	.987
Group × Stress	2	47.82	2.073	.137
Cons × Posit	1	49.53	2.819	.100
Cons × Stress	1	47.59	3.244	.078
Posit × Stress	1	51.62	0.003	.956
Group × Cons × Posit	2	49.46	0.053	.948
Group × Cons × Stress	2	47.53	0.27	.764
Group × Posit × Stress	2	51.59	1.4	.256
Cons × Posit × Stress	1	46.98	2.537	.118
Group × Cons × Posit × Stress	2	46.91	0.129	.880

Note. Group (English, Hindi IND, Hindi US); Cons = Consonant (/b/, /f/); Posit = Position (initial, medial); Stress (target stressed and unstressed). Significant effects in bold.

S.3. AXB categorial task: Control pairs

S.3.1. Accuracy. Average correct responses by Group and Pair was greater than 92%. Descriptive statistics by Group, Condition, and Pair are reported in Table 4. Results of mixed-effects logistic regression analysis showed main effects of Pair and Position (Table S7). For the effect of Pair, it emerged that /bv/ and /fv/ tended to be a bit more challenging than /bw/ and /fw/ (/bv/, 93% correct; /bw/, 97%; /fv/, 94%; /fw/, 98%). Discrimination was significantly poorer in /bv/ than /bw/ (p = .01), in /bv/ than /fw/ (p < .001), and in /fv/ than /fw/ (p < .001). The comparison between /bw/ and /fv/ approached significance (p = .06). No significant differences were observed for the remaining comparisons, p > .128. For the effect of Position, we found that discrimination accuracy was significantly higher when consonants were in the Medial than in the Initial position (98% and 93% accuracy, respectively).

There was also a significant Group × Pair interaction which, however, was of limited interest, both because overall accuracy was high (greater than 92%) and because of the 12 Tukey post-hoc comparisons (three group comparison by four pairs) only one was statistically significant (for /bw/, English-Hindi IND, p = .048).

Table S7. Discrimination task: ANOVA summary table for response accuracy to the control pairs /bv/, /bw/, /fv/, /fw/.

Model term	*df*	*F* ratio	*p*-value
Group	2	1.172	.310
Pair	3	8.465	**<.001**
Posit	1	40.012	**<.001**
Stress	1	0.031	.861
Group × Pair	6	2.262	**.035**
Group × Posit	2	1.513	.220
Group × Stress	2	0.223	.800
Pair × Posit	3	2.497	.058
Pair × Stress	3	0.197	.898
Posit × Stress	1	1.014	.314
Group × Pair × Posit	6	0.407	.875
Group × Pair × Stress	6	1.052	.389
Group × Posit × Stress	2	0.57	.566
Pair × Posit × Stress	3	0.179	.911
Group × Pair × Posit × Stress	6	0.176	.983

Note. Group (English, Hindi IND, Hindi US); Pair (/bv/, /bw/, /fv/, /fw/); Posit = Position (initial, medial); Stress (target stressed and unstressed). Significant effects in bold.

S.3.2. Response Time. Descriptive and inferential statistics for RT to the control pairs are reported in Tables S8 and S9, respectively. There were main effects of Pair and Position. The effect of Pair was driven by a slightly slower RT to /fv/ than /bv/, /bw/, and /fw/ (/bv/, 498 ms; /bw/, 500 ms; /fv/, 537 ms; /fw/, 486). Only two comparisons, however, were statistically significant: /bw/-/fv/ (p = .021) and /fv/-/fw/ (p < .001). No significant differences were found

for the remaining comparisons, $p > .11$. For the effect of Position participants showed faster RT for the Medial than the Initial position (Initial, 529 ms; Medial, 484 ms).

There was a significant Pair × Stress interaction, indicating that the Stress effect differed between pairs. The unstressed condition was significantly slower than the stressed condition for the /bv/ pair ($p = .02$). No differences were observed for all other comparisons ($p > .162$). We also found a significant Position × Stress interaction, with Stress showing a positive effect in Initial position (Unstressed, 506 ms; Stressed, 553; $p < .001$), but negatively affecting response time in Medial position (Unstressed, 507 ms; Stressed, 459 ms; $p < .001$). No other significant effects were found.

Table S8. Discrimination task: Descriptive statistics for response time in milliseconds by group, condition, and pair. Mean (SD).

Position, Stress	Group	/bv/	/bw/	/fv/	/fw/
Initial, Stressed	English	444 (147)	489 (230)	567 (234)	560 (365)
	Hindi US	526 (212)	543 (222)	613 (167)	553 (167)
	Hindi IND	540 (205)	562 (190)	599 (203)	571 (231)
Initial, Unstressed	English	500 (292)	421 (216)	560 (329)	413 (185)
	Hindi US	548 (136)	516 (171)	534 (142)	459 (133)
	Hindi IND	572 (227)	500 (292)	603 (343)	474 (176)
Medial, Stressed	English	373 (144)	425 (140)	440 (228)	412 (160)
	Hindi US	522 (157)	473 (151)	489 (157)	431 (146)
	Hindi IND	429 (134)	509 (245)	523 (274)	459 (176)
Medial, Unstressed	English	483 (210)	498 (296)	486 (256)	451 (213)
	Hindi US	492 (154)	510 (175)	497 (130)	513 (136)
	Hindi IND	544 (204)	550 (256)	529 (184)	533 (186)

Table S9. Discrimination task: ANOVA summary table for response time to the control pairs /bv/, /bw/, /fv/, /fw/.

Model term	*df1*	*df2*	*F* ratio	*p*-value
Group	2	51.21	0.977	.383
Pair	3	44.6	6.249	**.001**
Posit	1	51.29	6.012	**.018**
Stress	1	51.1	0.014	.906
Group × Pair	6	58.03	1.328	.260
Group × Posit	2	51.29	0.093	.912
Group × Stress	2	51.1	0.773	.467
Pair × Posit	3	11272.75	2.314	.074
Pair × Stress	3	11277.91	4.314	**.005**
Posit × Stress	1	11286.41	79.056	**<.001**
Group × Pair × Posit	6	11273.74	0.559	.763
Group × Pair × Stress	6	11279.17	0.755	.605
Group × Posit × Stress	2	11285.63	2.156	.116
Pair × Posit × Stress	3	11278.25	1.052	.368
Group × Pair × Posit × Stress	6	11279.94	1.082	.370

Note. Group (English, Hindi IND, Hindi US); Pair (/bv/, /bw/, /fv/, /fw/); Posit = Position (initial, medial); Stress (target stressed and unstressed). Significant effects in bold.

S.4. Identification task: Experimental contrast

S.4.1. Response Time. Descriptive and inferential statistics for RT to the experimental consonants /v/ and /w/ are reported in Tables S10 and S11. Similar to identification accuracy, there was a main effect of Group, indicating comparable performance of the two Hindi groups and significantly slower RT of the Hindi groups as compared to the English group. Average RT was 420 ms for the English group, 808 ms for the Hindi US participants, and 707 ms for the Hindi IND group (for the comparison English-Hindi US, $p < .001$; for English-Hindi IND, $p < .001$; for Hindi US-Hindi IND, $p = .711$).

We also found significant main effects of Position and Stress, such that RT was on average faster for consonants in the Initial than in the Medial position (570 and 711 ms respectively), and in the Stressed than Unstressed condition (639 and 652 ms respectively). Lastly, there was a statistically significant interaction of Consonant × Position × Stress (Table S11 and Figure S1). For /v/, stress had no effect in Initial position (Unstressed, 583 ms; Stressed, 603 ms; $p = .989$), but a positive effect in Medial position (Unstressed, 732 ms; Stressed, 669 ms; $p = .031$). The opposite pattern emerged for the consonant /w/: Responses to the stressed condition were faster than those to the unstressed condition in Initial position (Unstressed, 635 ms; Stressed, 560 ms; $p = .041$) but not in Medial position (Unstressed, 732 ms; Stressed, 781 ms; $p = .997$).

Table S10. Identification task: Descriptive statistics for response time in milliseconds by group, condition, and consonant. Mean (SD).

Position, Stress	Group	/v/	/w/
Initial, Stressed	English	353 (123)	378 (120)
	Hindi US	813 (328)	611 (211)
	Hindi IND	657 (212)	705 (292)
Initial, Unstressed	English	415 (134)	395 (126)
	Hindi US	733 (321)	797 (430)
	Hindi IND	601 (248)	763 (299)
Medial, Stressed	English	460 (229)	430 (161)
	Hindi US	870 (275)	1026 (628)
	Hindi IND	690 (363)	886 (541)
Medial, Unstressed	English	468 (176)	461 (119)
	Hindi US	902 (321)	969 (447)
	Hindi IND	832 (349)	767 (293)

Table S11. Identification task: ANOVA summary table for response time to the critical consonants /v/ and /w/.

Model term	*df1*	*df2*	*F* ratio	*p*-value
Group	2	51.39	16.936	**<.001**
Cons	1	57.45	0.646	.425
Posit	1	52.7	28.192	**<.001**
Stress	1	59.95	8.643	**.005**
Group × Cons	2	55.54	1.531	.225
Group × Posit	2	52.16	0.693	.505
Group × Stress	2	56.75	0.73	.487
Cons × Posit	1	56.95	0	.991
Cons × Stress	1	58.68	0.165	.686
Posit × Stress	1	58.51	0.078	.781
Group × Cons × Posit	2	54.41	1.032	.363
Group × Cons × Stress	2	55.42	0.316	.730
Group × Posit × Stress	2	56.25	0.83	.442
Cons × Posit × Stress	1	58.24	8.408	**.005**
Group × Cons × Posit × Stress	2	56.43	2.765	.072

Note. Group (English, Hindi IND, Hindi US); Cons = Consonant (/v/, /w/); Posit = Position (initial, medial); Stress (target stressed and unstressed). Significant effects in bold.

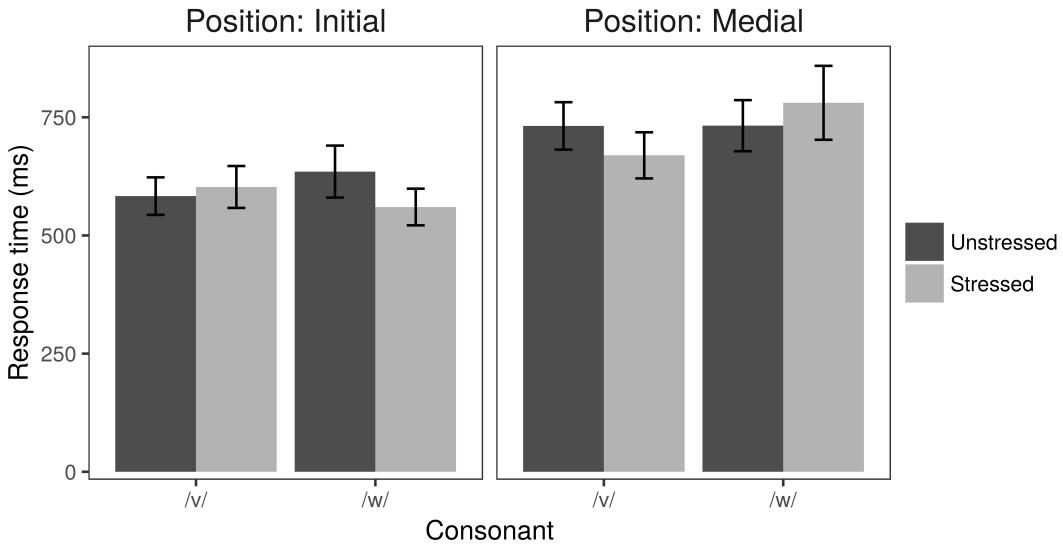

Figure S1. Identification task: Response time by Consonant, Position, and Stress. Error bars: ± SE.

S.5. AXB categorial task: Experimental contrast

S.5.1. Response time. The effects of Group and Position on RT to the /v/ and /w/ pair replicated the pattern found for discrimination accuracy of the same experimental pair. Descriptive and inferential statistics are reported in Table S12 and S13. The English group's RTs were faster than those of the Hindi groups, and performance of the two Hindi groups was comparable (see Table S12 for descriptive statistics). Average RT was 465 ms for the English participants, 701 ms for the Hindi US group, and 718 ms for the Hindi IND participants (post-hoc comparisons: English-Hindi US, p = .013; English-Hindi IND, p = .002; Hindi US-Hindi IND, p = .772). For the main effect of Position, participants' responses were faster when the consonant was in Medial than Initial position (Initial, 682 ms; Medial, 582 ms).

A significant Position × Stress interaction was also found, indicating that the effect of stress was negative when consonants were in Initial position and positive when they were in Medial position (Initial Unstressed, 645 ms; Initial Stressed, 725 ms; Medial Unstressed, 633 ms; Medial Stressed, 540 ms). For the effect of Stress in Initial position, p = .002; for the effect of Stress in Medial position, p = .001.

Table S12. Discrimination task: Descriptive statistics for response time in milliseconds by group, condition, and pair. Mean (SD).

Position, Stress	Group	/vw/
Initial, Stressed	English	505 (247)
	Hindi US	820 (331)
	Hindi IND	849 (342)
Initial, Unstressed	English	452 (213)
	Hindi US	727 (316)
	Hindi IND	751 (318)
Medial, Stressed	English	444 (189)
	Hindi US	562 (187)
	Hindi IND	613 (316)
Medial, Unstressed	English	461 (264)
	Hindi US	734 (266)
	Hindi IND	703 (276)

Table S13. Discrimination task: ANOVA summary table for response time to the critical pair /vw/.

Model term	*df1*	*df2*	*F* ratio	*p*-value
Group	2	51.26	7.473	**.001**
Posit	1	51.56	17.235	**<.001**
Stress	1	51.83	0.099	.754
Group × Posit	2	51.5	1.511	.230
Group × Stress	2	51.72	0.435	.650
Posit × Stress	1	51.59	19.349	**<.001**
Group × Posit × Stress	2	51.53	1.717	.190

Note. Group (English, Hindi IND, Hindi US); Posit = Position (initial, medial); Stress (target stressed and unstressed). Significant effects in bold.

S.6. Pilot Assimilation Task

Before undertaking the current study, we conducted a pilot test with eight native speakers of Hindi (who did not participate in the current study) to examine how Hindi speakers assimilate English /v/ and /w/. The participants (5 males, and 3 females, age range: 28-42 years) listened to nonsense words with /v/ and /w/ in the same stimuli utilized in the current study: vaˈgag, ˈvagag, ˈgavag, gaˈvag, waˈgag, ˈwagag, ˈgawag, and gaˈwag. Participants were presented with various Hindi graphemes that corresponded to labials, stops (aspirated and unaspirated), fricatives, and the labio-dental approximant. They were asked to indicate the grapheme that corresponded with the specified speech sound in the nonsense words they heard. All participants assimilated 100% of the /v/ and /w/ tokens to the single native labio-dental approximant /ʋ/ category. Based on

these highly consistent findings, which were in line with previous studies (Iverson et al., 2011; Iverson et al., 2008), we decided to not include an assimilation task in the current study.