

Article

L1 Japanese Perceptual Drift in Late Learners of L2 English

Chikako Takahashi

Department of East Asian Languages and Cultures, Columbia University, New York, NY 10025, USA; ct2423@columbia.edu

Abstract: This study presents evidence of second language (L2) influence on first language (L1) perception of alveolar stops. Sixty-one L1 Japanese late learners of L2 English (onset ~12 years old) in Japan (N = 31) and in the US (N = 30) participated. We examined late L2 learners' L2 perceptual ability and L1 perception drift by administering three perception tasks (AX discrimination, forced categorization, and goodness rating) on word-initial stop consonants. The L2 learners' L1 Japanese and L2 English data were compared to those of Japanese and English monolinguals, respectively (N = 21, N = 16). All participants' production data were also gathered to examine potential perception-production relationships. Late learners' sensitivity patterns along a synthesized /da-ta/ continuum differed significantly from those of monolingual speakers, with a sensitivity peak location between the monolingual Japanese and English groups. This suggests that late learners' voicing category boundaries may have been influenced by L2 English learning. The L2 learners' goodness rating patterns of L1 Japanese stimuli also showed evidence of L1 perceptual drift: L2 learners tended to be more accepting of Japanese stimuli with longer VOTs compared to Japanese monolinguals.

Keywords: L1 drift; perception; phonetic interaction; L2 acquisition

1. Introduction

In the past few decades, research on speech communication and language learning has revealed that phonetic changes in speech do not only occur in one's second language (L2) but also in the first language (L1) upon the onset of L2 learning (e.g., Chang 2012, 2013; Flege 1987; Flege et al. 1999; Guion 2003; Harada 2003; Kartushina et al. 2016a; Lang and Davidson 2019; MacKay et al. 2001; Major 1992; Mora et al. 2015; Mora and Nadeu 2012; Sancier and Fowler 1997). The changes that occur to one's native dialect are especially interesting in the context of second language (L2) learning because they go against the long-held belief that phonetic development does not occur beyond the Critical/Sensitive Period (e.g., Long 1990; Ruben 1999). This change in L1 production and perception is often referred to as phonetic drift (see Chang 2019 for a recent review).

The speaker's perception is often thought of as the key to determining the relationship between L1 and L2 phones. This is one of the factors assumed to influence later learners' L1:L2 interaction, along with the nature of the L1:L2 phonological and phonetic differences (e.g., Flege 1995, 2007). These factors, especially perceptual ability and L1:L2 differences, form the core parts of the Speech Learning Model (SLM; Flege 1995, 2007), which has influenced much of the work on second language phonetic learning. The SLM proposes that L1 and L2 phonetic categories share the same phonological space and, therefore, interact (Flege 1995, 2007). Key to understanding the predictions of SLM is the notion of phonetic differentiability, which depends on perception. If learners perceive enough dissimilarities between a new L2 sound and the closest L1 counterpart, a new phonemic category can be created in the shared system. In this case, close L1 and L2 sounds are expected to dissimilate from one another to keep L1 and L2 sounds distinct in the shared space. However, if learners do not perceive the L2 sounds as different enough from their L1 counterparts, an L2 sound becomes associated/equated with the L1 sound it is most similar to. Until now, the direction of the L1 drift (assimilation vs. dissimilation) has generally



Citation: Takahashi, Chikako. 2024. L1 Japanese Perceptual Drift in Late Learners of L2 English. *Languages* 9: 23. <https://doi.org/10.3390/languages9010023>

Academic Editors: Olga Dmitrieva, Chiara Celata, Esther de Leeuw and Natalia Kartushina

Received: 31 July 2022

Revised: 15 December 2023

Accepted: 23 December 2023

Published: 10 January 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

been determined based solely on the production data, without investigating L2 learners' L1 or L2 perceptions. Assuming that phonetic category formation has some perceptual basis, we can expect to see a similar change in L1 perception (i.e., perceptual assimilation vs. dissimilation).

Another speech learning model we must consider in the discussion of L1:L2 perceptual interaction is the Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2; Best and Tyler 2007), whose postulates extend from the original PAM to "predict success at L2 perceptual learning" (p. 25). While PAM-L2 focuses on how L1 influences the perception of L2 contrasts, it also provides us with possible scenarios as to how L1 and L2 sounds might be associated by an L2 learner. According to PAM-L2, an L2 learner might equate an L1 sound and a new L2 sound, establishing an equivalence not only at the phonetic level but potentially also at the phonological level. This L1-L2 association at different levels (i.e., phonetic vs. phonological level) might mean that L2 influence on L1 perception happens at those varying levels. In the context of the current study, since the AX discrimination task prompts listeners to discern differences between pairs of stimuli without having them categorize them into phonemes, it might be reasonable to say that it taps into listeners' phonetic processing. Conversely, in the categorization and goodness rating tasks, since listeners are asked to associate the stimuli with real words, we consider them to be tasks tapping into listeners' phonological processing. If there are, in fact, varying levels of L1-L2 association and different perception tasks tap into different levels of those associations, then it is possible that we will see differing results (e.g., evidence of L1 drift vs. lack of evidence of L1 drift), depending on the given perception task.

Previous studies have identified several ways that L1 production and perception can change in relation to the learned L2, with a comparison often being made between an L1 sound and an acoustically/perceptually related L2 sound. Some studies have shown an assimilatory effect of L2 on L1 such that L1 sounds are acoustically less nativelike as a result of becoming similar to a phonetically related L2 sound (e.g., Barlow 2014; Chang 2012, 2013; Harada 2003; Kartushina et al. 2016a; Lang and Davidson 2019). However, the L1 sounds of some learners have shown to be less nativelike in the opposite direction (dissimilation), with their acoustic correlates being even further away from their related L2 sounds than with nativelike production, as if learners are attempting to create a larger distance between the related L1 and L2 sounds (e.g., Flege and Eefting 1987a, 1987b; Guion 2003; Takahashi 2020).

L1 phonetic drift has been reported for both early bilinguals and late L2 learners: those who began learning their L2 after the so-called Sensitive Period. Despite the growing number of studies that show L2 learning effects on L1 sound, there is much uncertainty about what factors cause such effects and why there are differences in the nature of L1 change (e.g., dissimilation and assimilation) across studies. Studies of L1:L2 phonetic interaction in early bilinguals indicate that very early learners generally establish two separate phonetic categories for an L1 sound and a similar L2 sound (Kartushina et al. 2016b). These studies generally show no L2 influence on L1 production. In contrast, previous studies suggest that late learners (onset of learning: 12 or later) tend to start with a merged (or overlapped) phonetic category for similar L1 and L2 sounds and that this learner population has shown evidence for L2 influence on L1 (Flege et al. 2003; Flege and Eefting 1987b; Guion 2003).

Late L2 learners have been the focus of many L1 drift studies (e.g., Chang 2012, 2013; Flege 1987; Flege et al. 1999; Guion 2003; Harada 2003; Kartushina et al. 2016a; Lang and Davidson 2019; MacKay et al. 2001; Major 1992; Mora et al. 2015; Mora and Nadeu 2012; Sancier and Fowler 1997). Some studies dealing with the effects of L2 phonetic learning on L1 suggest that the shared L1:L2 category may eventually change as learners make progress (e.g., Casillas 2020; Flege and Eefting 1987b). As is often discussed in the context of Flege's Speech Learning Model (SLM; Flege 1995, 2007), it is possible that as a learner becomes more proficient in L2, they may be able to establish separate categories for similar L1 and L2 sounds, in which case the learner will be inclined to modify these sounds to

keep them distinct from other contrasting sounds in the combined L1:L2 system. This might lead to dissimilation, in which an L1 sound becomes different from that of the monolingual native norm because it is moving away from an associated L2 sound, rather than becoming more like it (e.g., [Flege and Eefting 1987b](#)). While the above line of reasoning predicts that proficient L2 learners will exhibit L1 dissimilatory drift, some studies instead report the opposite—that proficient L2 learners exhibit assimilatory drift (e.g., [Flege 1987](#); [Major 1992](#)). In fact, assimilatory L1 drift has been observed both in novice learners and advanced L2 learners. These contradictory results suggest that in order to determine how post-adolescence L1:L2 phonetic learning interactions evolve, other potential influencing factors must be evaluated in a systematic way.

One of those potential influencing factors is speakers' perceptions of L1 and L2. Perception is thought to either lead to production ([Best and Tyler 2007](#); [Flege 1995, 2007](#)) or co-develop with production ([Flege and Bohn 2021](#)) in L2 phonetic learning. The co-development view indicates that there is no precedence between perception and production development (i.e., perceptual development is not a prerequisite for production development). That is, some learners might exhibit more developed L2 production before L2 perception and vice versa. The specific nature of the relationship aside, some connection between perception and production is expected during the reorganization of the L1 phonetic system. That is, if the L2 learning experience causes a change in L1 production, as has been shown in previous studies, it naturally follows that a similar change would be observed in the speakers' L1 perceptions. For instance, given that assimilatory L1 drift on VOT production has been attested in numerous studies involving various types of learners (e.g., [Chang 2012, 2013](#); [Major 1992](#); [Sancier and Fowler 1997](#)), we would expect to see an assimilatory perceptual voicing boundary shift for those learners.

Despite the seeming importance of L1 and L2 perception data in understanding L1:L2 phonetic interaction, considerably fewer studies have been conducted on L1 perceptual changes compared to those of changes in L1 production. This is especially true for those studies focused on late L2 learners, a few of which have shown either L1 perceptual drift or L1 category shift (e.g., [Dmitrieva 2019](#); [Flege and Eefting 1987a](#); [Lev-Ari and Peperkamp 2013](#); [Tice and Woodley 2012](#); [Williams 1977](#)). Of the few perceptual studies on late L2 learners, [Williams \(1977\)](#) reported an assimilatory L1 category shift showing that five out of eight Spanish–English bilingual participants' perceptual boundaries (between /b/ and /p/) were at a “compromise point lying close to either the Spanish or English monolingual” (p. 295) VOT boundaries. [Tice and Woodley \(2012\)](#) reported that L1 English learners of L2 French studying in an immersion context exhibited a subtle perceptual boundary shift between voiced and voiceless English stops. This change was interpreted as an assimilatory drift toward the L2 French voicing boundary, although after a few weeks, it shifted back to the value observed in week 1. [Tice and Woodley](#) argued that this suggests, “some amount of L1 category confusion as a result of their two weeks of L2 input.” (p. 74). In a more recent study by [Dmitrieva \(2019\)](#) on the perception of final stop voicing, English–Russian bilinguals exhibited L2 English influence on their L1 Russian perception—they relied more on the vowel duration cues and less on glottal pulsing compared to monolingual Russian speakers.

In the current study, late L2 learners' perceptions of L1 Japanese and L2 English stop onsets were investigated via three different perceptual measures (discrimination, categorization, and goodness rating), with the results then compared to those of monolingual groups (Japanese and English). We also included a production task in order to examine the potential relationship between L1 perceptual and production drift. The current study focuses on the voiceless stop /t/ in L1 Japanese and L2 English. The place of articulation of /t/ in the two languages differs: English /t/ are considered alveolar stops, while Japanese ones are considered laminal denti-alveolar. Such differences in the place of articulation could possibly influence the perceptual judgment of the voicing boundary, since there have been studies reporting a relationship between VOTs and the place of articulation ([Benkí 2001](#)). However, it is not clear from previous studies that L2 learners are able to specifically notice

the differences between alveolar and denti-alveolar places of articulation. The voiceless stop consonants were chosen for the current study largely for the comparability of the results to previous studies since many studies conducted on the topic of L1 pronunciation change have focused on this measure of voiceless consonants (e.g., [Chang 2012, 2013](#); [Flege 1987](#); [Harada 2003](#); [Major 1992](#); [Sancier and Fowler 1997](#)). In previous studies examining learners' VOTs, comparisons have been typically made between languages with voiceless stops exhibiting longer VOTs and languages exhibiting significantly shorter VOTs. This makes it possible to see the L1 or L2 perceptual change on a VOT continuum, in relation to the other language, when the equivalent phonemes of the L1 and L2 are thought to occupy two fairly well-separated points on the continuum.

Japanese and English word-initial stops are both described as having two-way contrasts: voiced stops and voiceless stops. While the voiceless consonants in the two languages belong to the same phonemic category, Japanese voiceless stops have noticeably shorter VOTs than their English counterparts ([Harada 2003, 2007](#)). English voiceless stop VOTs have always been described as long-lag VOTs. However, the Japanese voiceless stop VOTs reportedly occupy an intermediate length, ranging between short- and long-lag categories ([Riney et al. 2007](#)). Given the intermediate duration of Japanese VOTs for voiceless stops, Japanese speakers might more easily learn to produce and perceive English voiceless stops; similarly, it is possible that the association between Japanese and English voiceless consonants becomes stronger, it might be more difficult for the two languages to be separated.

English and Japanese voiced stops have been reported both with and without prevoicing. For Japanese voiced stops, while some studies report relatively long voicing leads (e.g., [Shimizu 1990](#)), others report a lack of prevoicing, especially in certain regions and with younger generations (Takada 2011, as cited in [Takada et al. 2015](#)). Similarly, English voiced stops have been reported with both short-lag VOTs and prevoicing ([Flege 1982](#); [Lisker and Abramson 1964, 1967a](#)). Since the use of prevoicing seems to vary among speakers in both languages, there may be more variable judgments of stimuli involving prevoicing.

In terms of L2 influence on L1, whether L1 Japanese VOT production and perception would be modified in some way would depend on the type and strength of the association L1-L2 VOTs hold for the voiceless stops. If the association between the L1 Japanese and L2 English VOTs is strong, learning L2 English (with more exposure and use) might influence L1 Japanese VOTs to be more English-like, both in production and perception. However, if L2 learners try to separate them, the distance between L1-L2 VOTs might become greater (e.g., exhibiting a perceptual voicing boundary at shorter VOT values than monolingual Japanese speakers or producing shorter VOTs in L1 Japanese). While several studies have investigated Japanese VOT drift in early bilinguals (e.g., [Harada 2003, 2007](#); [Yusa et al. 2010](#)), little research has been conducted on how late L2 learning influences L1 Japanese VOT drift in production or perception.

Following the assumption maintained in both the SLM and PAM-L2 that there is a link between production and perception, we predicted that we would observe an L2 learning influence on L1 perception. This L2 influence on perception might be exhibited as a category boundary shift of /d-t/ in categorization tasks given that a similar pattern was observed by [Tice and Woodley \(2012\)](#). Nevertheless, the L1 perceptual drift might be exhibited in various ways, as PAM-L2 suggests that L2 learners might be able to associate L1-L2 sounds at multiple levels. Thus, the L2 influence on L1 perception might only be observable with AX discrimination or categorization tasks (or vice versa), depending on the nature of the L1-L2 association.

The consideration of production and perception data of the same L2 learner participants should inform the nature of the production–perception relationship. Following the assumption maintained in SLM-R ([Flege and Bohn 2021](#)) that production and perception co-develop, we expect to observe a link between VOT production and perception. However, it is possible for the production–perception correlation not to be neatly aligned given that there might not be a precedence relationship between production and perception. That

is, some participants might show nativelikeness in L2 production (and subsequent L1 production drift) without showing perceptual drift and vice versa.

2. Materials and Methods

Adult Japanese late learners of English (onset of learning after about the age of 12) were the target participants for this project. The Japanese learners of English were recruited both in Japan and in the US in order to acquire a sample embodying a range of L2 use and exposure. Monolingual speakers of Japanese and English were also recruited in order to compare the learners' data with that of monolingual native speakers (L1 Japanese, as well as L2 English). While the focus of this paper was the influence of L2 learning on L1 perception, the data from the production experiment was also included so as to examine the potential relationship between L1 perception and production in the context of L2 learning.

2.1. Participants

Sixty-five late L1 Japanese speakers of L2 English (henceforth identified as "L2 learners") were recruited in Japan and in the US (32 in Japan, 33 in the US) for this experiment. Participation criteria for L2 learners were (1) L1 Japanese learners of L2 English who commenced learning English at or about the age of 12, and (2) who self-identified themselves as either actively learning or using English in their daily lives. Four were excluded after data collection because they either had early extensive exposure to English or were trilingual. Data from the remaining 61 L2 learners were analyzed.

Two monolingual control groups consisting of 21 Japanese monolingual speakers and 18 English monolingual speakers were recruited in Japan and the US, respectively. The participants were functional monolinguals who were "linguistically naïve" (Best and Tyler 2007) to languages other than their native languages and not able to communicate in an L2 despite possibly having had some exposure to foreign languages in a school setting. English monolinguals were recruited in New York State, while Japanese monolinguals were recruited in Niigata Prefecture in Japan. Of the 21 Japanese monolinguals who were recruited in Japan, one participant's data was excluded since they later reported having hearing issues, leaving 20 total Japanese monolingual participants. Of the 18 English monolinguals, two exhibited unexpectedly short VOT for /p/ and were, thus, excluded, leaving a total of 16 monolingual English participants.

2.2. Materials

For the AX discrimination task, an 11-step /da-ta/ VOT continuum was synthetically created using a Klatt Grit speech synthesizer in Praat (Boersma and Weenink 2020) with a script that automatically generates a CV sequence modeling a male person's voice.¹ In order to examine the L2 learners' sensitivity to differences in VOTs in a /d-t/ VOT continuum, we chose a VOT that ranged from -20 ms to +80 ms VOT since studies of Japanese and English VOT production have suggested that Japanese production VOT of /t/ is around 25–30 ms and English production VOT of /t/ is around 70–90 ms (e.g., Harada 2003; Lisker and Abramson 1964; Riney et al. 2007). While Japanese voiced stop production has been reported to vary greatly in VOTs (e.g., -200~short lag by Takada et al. 2015), since our focus is on Japanese–English voiceless stops, we did not include many tokens with lead voicing. We included VOTs that reflect obviously voiced (-20 ms) and voiceless (+80 ms) values, with a resulting 11 stimuli from -20 ms to +80 ms VOT with 10 ms steps. As shown in Table 1, for the sake of comparability, the endpoint parameters were taken from Mack (1989), which examined French bilinguals' perceptual boundaries of /da-ta/. A study by Mack (1989) was used as a model for stimuli construction since the VOT contrasts examined in their study (French vs. English) were similar to the current study: French /t/ VOT values (23 ms, Caramazza and Yeni-Komshian 1974) are similar to those of Japanese /t/ VOT values (~25 ms, Harada 2003).

Table 1. Endpoint parameters for da–ta synthesis from Mack (1989): B—formant bandwidth, AF—frication amplitude, AV—voicing amplitude.

ms	Prevoicing (–20)	End (+80)
F1	250	720
F2	500	1240
F3	1500	2500
F4	3600	3600
F5	3850	3850
B1	90	90
B2	500	70
B3	500	110
B4	250	250
B5	200	200
AF	0	0
AV	45	55

From the 11-step stimuli, 9 pairs differing by 20 ms VOT were created for the *different* trials, and 11 pairs with the same VOT length were created for the *same* trials. Three blocks, each containing 40 pairs (22 same, 18 different) were presented to the participants. That is, participants listened to each stimulus set 6 times over the course of the experiment. Half of the different pairs started with a stimulus with a longer VOT (stimulus A), and the other half started with a stimulus that had a shorter VOT (stimulus B). That is, half were presented as AB and the other half as BA. This yielded a total of 120 trials.

For the Category Goodness Rating task, two continua were created using real Japanese and English words (by manipulating VOT values, ranging from 0 or –20 ms to +80 ms, in 10 ms increments): an 11-member synthetic Japanese “tan-dan” and a 9-step synthetic English “den-ten”. For both cases, stimulus VOT was manipulated based on real /d/ words produced by native speakers. Unlike the AX discrimination task, the prevoicing was only included in Japanese stimuli since the native speaker who recorded the Japanese sample exhibited prevoicing.² Since the natural English token did not include prevoicing, the synthesized VOT range for English /den-/–/ten/ started at 0 ms. The modified Praat script mentioned above was used to automatically generate the synthesized stimuli, with the intensities of all stimuli being scaled at 70 dB. For the English task, a total of 36 trials were administered, in two blocks of 18. For the Japanese task, a total of 44 trials were administered, in two blocks of 22.

For the production task, Japanese and English word sets were prepared that focused on eliciting data on voiceless stop VOTs. The Japanese set included 21 target words (Table 2) and 60 filler words. The Japanese target consonants were word-initial /p, t, k/ before mid-low vowel /a/ and high front vowel /i/. Since Japanese does not have the /ti/ sequence natively, although it has become commonly accepted (Crawford 2009), we included the target consonants followed by /e/ as well. The target words were selected so that the first syllable was produced with an H tone whenever possible. Words that included word-initial /p/ and /ti/ were all loan words. Japanese filler items included words with other vowels and other consonants including voiced stops. Care was taken to include words that are semantically very Japanese (e.g., *omotenashi*, “hospitality”) to help speakers be in “Japanese mode.” All of the words were elicited in a carrier sentence “*sorewa — desu*” (equivalent to English “That’s —.”)

The 19 English target words contained word-initial /p, t, k/ before /aɪ/ (or /ɑ/) and /i/ in the stressed syllable (see Table 2); there were 94 filler words. These vowels were chosen to keep the vowel context following the consonants as similar as possible in the two languages. Words were also selected based on difficulty level so that not-so-proficient L2 learner participants would be able to read them without trouble. Participants produced all of the visually presented words in a carrier sentence, “That’s —”.

Table 2. Target words for Japanese and English production tasks.

	Japanese	English
P	Pisutoru (pistol), piza (pizza), pikaso (picaso), pasu (pass), papa (papa), pagu (pug)	Peak, peace, peach, papa, pipe, pie, pile
t	Tiffanii (Tiffany), tisshu (tissue), tiikappu (tea cup), tafu (tough), tasuku (task), tate (vertical), tesuto (test), tefuron (teflon), tetorisu (tetris)	Tease, teach, tea, type, tile, tight
k	Kigu (instrument), kiba (fang), kisu (a type of fish), kasu (sediment), kaba (hippopotamus)	Keep, kiwi, key, kite, Cairo, king

2.3. Procedure

For all participants, the perception tasks (AX discrimination, categorization, and category goodness rating tasks) were conducted before the production tasks. The categorization and category goodness rating tasks used the same stimuli and were part of one experiment. L2 learners completed both Japanese and English tasks in the same experiment session; they were asked to complete both Japanese production and perception tasks before moving on to English tasks. Since the goal was to see how their Japanese might have been influenced by their L2 English, they were asked to complete the Japanese tasks first to minimize the immediate effect of using English on their production and perception of Japanese. The interval between English and Japanese sessions was 2–5 min.³ All of the instructions were given in the language of the current task: Japanese during the Japanese tasks, and English during the English ones.

In the AX discrimination task, the participants were presented with the written prompt, “Same or Different?” while listening to each stimulus pair. After hearing a set of two stimuli on each trial, the participants were asked to judge if the two stimuli were the same or different and were prompted to indicate their decision by pressing a key. The inter-stimulus interval (ISI) was set to 1 s, while the inter-trial interval (ITI) was set to 1.5 s. The participants first completed a practice block with non-test stimuli (12 trials: 6 same, 6 different), which were created from a /pa/-/ba/ continuum. In the practice block, the participants received automatic written feedback on the screen, indicating whether their answers were correct or incorrect. The purpose of the practice block was to have the participants be familiar with the task. After the practice set, each participant completed 3 blocks, each containing 40 fully randomized trials (22 same, 18 different). In the main trials, feedback was not given to the participants. The participants were instructed to take a short break between each block and resume when ready.

In the category goodness rating tasks, the participants listened to a real-word stimulus produced by a native speaker and were asked to identify what they heard by choosing one of two possible words (e.g., *den* or *ten* for English trials). Then, the participants listened to the same token again and were asked to rate how good the given token was as a member of the category they had chosen, using a scale from 1 (bad) to 7 (good). The participants were allowed to listen to each stimulus as many times as needed (up to a maximum of 10). They first completed a practice block of either 9 or 11 tokens presented randomly (9 for English and 11 for Japanese), followed by 2 blocks of 18–22 trials. The purpose of the practice session was to familiarize the participants with the task.

For the production task, the participants were asked to read aloud visually presented target words contained in carrier sentences. The target words were presented in PowerPoint slides one by one in a pseudo-randomized order. Each word was presented once in each set. There were a total of three sets, with each set containing the same items but differing in presentation order. For the English task, before it began, L2 learner participants were given the word list that was going to be presented to familiarize themselves with the words. While they were not allowed to practice word pronunciation with the experimenter during the task, they were encouraged to read the word however they wanted if they did not know it. The recordings took place in quiet rooms easily accessible to the participants (e.g.,

on Stony Brook University campus, in a meeting room in a library, or in rental meeting rooms in Japan). The recording device was a Zoom H4 digital recorder with an SM10A-CN dynamic head-mounted microphone. The sampling rate was 44.1 kHz.

3. Results

3.1. VOT Perception

The perception tasks were conducted to examine L2 learners' perception patterns of L1 Japanese and L2 English voiceless stops to determine whether or not there is an L2 learning influence on L1 perception. The goals of the discrimination task were twofold. One was to determine an individual L2 learner's ability (sensitivity) to detect subtle differences in the da-ta continuum, ranging from -20 ms to 80 ms in VOTs, as indicated by the sensitivity measure (d'). The other, more important, purpose of the discrimination task was to investigate sensitivity peaks along the VOT continuum and compare these among the three language groups (Japanese monolinguals, L2 learners, and English monolinguals).

While the discrimination peak is expected to be around 20–30 ms for both monolingual groups (Shimizu 1977, for Japanese; Mack 1989, for English), Lisker and Abramson (1967b), as cited in Abramson and Lisker (1973), reported a higher VOT boundary for English speakers at around 35 ms. Therefore, it is possible that monolingual English speakers show higher sensitivity in the discrimination of stimuli with longer VOTs, *on average*, compared to monolingual Japanese speakers. More importantly, if L2 English learning influences the learner's perception via exposure to longer VOTs, it is possible that the sensitivity pattern of the L2 group would be more similar to that of English speakers in contrast to the monolingual Japanese group.

Each participant's overall discrimination ability was measured by calculating the d' (d' -prime) score (Macmillan and Creelman 2005), following Mack (1989) and Iverson and Kuhl (1995). We first determined the hit rate (the proportion of correctly identified different stimulus pairs against the total number of different stimulus pairs) and false alarm rate (the proportion of same pairs that were incorrectly reported as "different" against the total number of actual different stimulus pairs) per participant. Since our experiment employed a roving design in which each stimulus pair differed across trials, we used the differencing method discussed in Macmillan and Creelman (2005) to calculate d' -prime, using an R script developed by Paillet (2003). The mean d' -prime per language group was 1.70 (SD: 0.41) for monolingual English speakers, 1.72 (SD: 0.38) for monolingual Japanese speakers, and 1.78 (SD: 0.65) for the L2 learner group.⁴ Welch's one-way ANOVA showed no significant differences between the three groups.

We considered the possibility that the sensitivity peak along the VOT continuum, indicating the voicing boundary, was different between monolingual Japanese and monolingual English groups since the languages utilize different VOT ranges in producing the voiceless /t/, with English /t/ having longer VOTs than Japanese ones.

In order to compare the sensitivity patterns, we calculated d' -prime per stimulus pair. Since calculating d' -prime per stimulus pair required both the hit rate (accurately identified different stimulus pairs) and false alarm rate (same stimulus pairs incorrectly identified as different stimulus pairs) in its calculation, we grouped same stimulus and different stimulus pairs together based on their VOTs. For instance, the 10–10 (ms) same stimulus pair was grouped together with the 0–20 (ms) different stimulus pair in order to obtain the d' -prime score for the stimulus around 10 ms VOT. The two sets of same stimulus pairs at opposite ends of the VOT continuum (-20 ms and 80 ms) were excluded for this reason. Figure 1 shows d' -prime per stimulus VOT (ms) comparing language groups. While both monolingual Japanese and monolingual English speakers exhibit the same d' -prime peak at 10 ms VOT, Japanese monolinguals had a secondary peak at 0 ms, and English monolinguals had their secondary peak at 30 ms VOT.

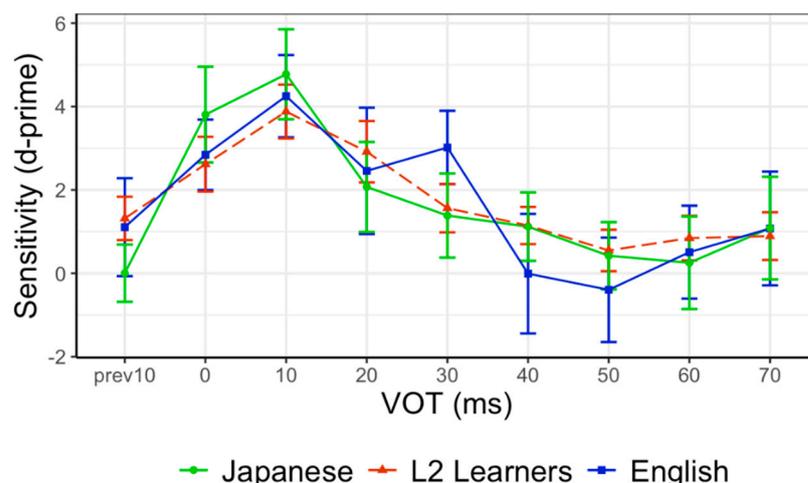


Figure 1. d-prime per stimulus VOT (ms) comparing language groups. Error bars indicate 95% CI.

To further investigate the differences in sensitivity peaks among the three participant groups, we identified the stimulus VOT (ms) that corresponded with the sensitivity peak (d-prime) per participant and compared the means of these values among language groups (Table 3).

Table 3. Summary statistics on sensitivity peak per language group.

Language Groups	Mean	Median	SD
English monolinguals	24.21	10	23.87
L2 learners	18.55	10	19.87
Japanese monolinguals	9.0	10	12.93

We used a one-way Kruskal–Wallis test and post hoc pairwise Wilcoxon tests to examine the group differences in the mean VOT (ms) for the peak d-prime. These methods were used as alternatives to one-way ANOVA and Tukey’s HSD.⁵ The effect of the language group on the d-prime peak VOT was significant ($F(2, 108) = 8.69, p < 0.05$). The post hoc pairwise Wilcoxon test showed a significant difference between Japanese monolingual and English monolingual groups (adj. $p < 0.05$), as well as between Japanese monolinguals and L2 learner groups (adj. $p < 0.05$). The significant difference between monolingual Japanese and English groups suggests that the two languages have different sensitivity patterns. Crucially, a comparison of monolingual Japanese speakers and L2 learners indicates that the L2 learners’ sensitivity patterns are also different from monolingual Japanese speakers, and the L2 learner pattern is similar to that of the monolingual English speakers.

The categorization and category goodness rating tasks were conducted to examine whether L2 learners’ Japanese /d-t/ perceptual boundaries and rating patterns differ from those of Japanese monolinguals, and whether or not L2 learners’ Japanese and English categorization and goodness rating patterns indicate L2 influence on L1 perception. If L2 learning influences L1 perception, we expect to observe L2 learners’ perceptual boundaries shift for L1 Japanese in a way that is consistent with their categorization patterns in L2 English. We included a goodness rating task in order to capture the differences in perception between monolingual and L2 learner groups that might otherwise not be detected via the forced two-choice categorization task. We considered L2 learners’ English task results to be important in interpreting their Japanese perception results.

First, the categorization data (“den” vs. “ten” for English and “dan” vs. “tan” for Japanese) were separately aggregated per stimulus and per language group. Figure 2 shows the categorization function for English “den” vs. “ten”. The categorization function lines show little difference between L2 learners and monolingual English speakers, indicating that L2 English learners, in general, exhibit nativelike perceptual boundaries between

/d-t/ when categorizing L2 English “den” vs. “ten”. In order to examine the potential group differences further, boundary VOT (ms) was estimated by treating the discrete stimulus steps as continuous variables (Xu and Francis 2006) and fitting a local regression line per participant using the loess function in R. This function extracted the estimated VOT (ms) values that corresponded to a 50% categorization rate. The estimated /d-t/ category boundary VOT (ms) for L1 English monolinguals was 23.11 ms (SD: 5.73) and 24.19 ms (SD: 5.93) for L2 learners, with no significant difference between the groups. A generalized linear mixed model with a logit link function was fitted to examine the effect of language group (monolingual vs. L2 learners) on participants’ categorization patterns between English “den” and “ten” (Table 4). The dependent variable was the odds ratio of the binomial response (“den” = 1; “ten” = 0). The model included the fixed effects of the language group and stimulus steps, as well as their interaction terms. The categorical variable of the language group was effect-coded.

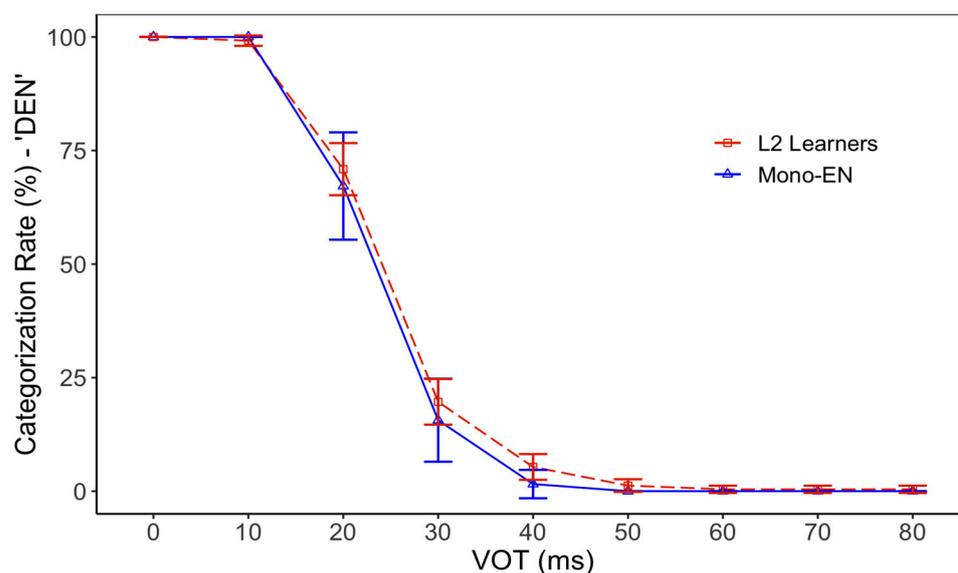


Figure 2. Categorization function of English “den”–“ten”. Error bars indicate 95% CI.

Table 4. Results of the generalized mixed-effects model from English den–ten categorization response.

	β	SE	z	p (> z)
(Intercept)	8.45	0.78	10.79	
Language group	1.04	0.64	1.6	0.01
Steps	−0.36	0.03	−10.36	<0.001
Language group *steps	−0.05	0.02	−1.81	0.06

The effect of steps was significant, which was expected. There was a main effect of the language group, indicating that L2 learners categorized the stimuli as /d/ more often than monolingual English speakers. However, the interaction of stimulus VOTs and the language group did not reach statistical significance ($\beta = -0.05, z = -1.81, p = 0.06$).

For the Japanese contrast “dan” vs. “tan,” monolingual Japanese speakers and L2 learners showed almost identical categorization function patterns (Figure 3), with an estimated voicing boundary of 25.10 ms (SD: 4.01) for monolinguals and 23.84 ms (SD: 6.20) for L2 learner speakers ($p > 0.05$). A generalized linear mixed-effects model was fitted with the comparable structure we used for the English categorization of “den–ten” above, but neither the effect of the language group (monolingual Japanese group vs. L2 learners) nor the interaction between the steps and language group was significant (Table 5).

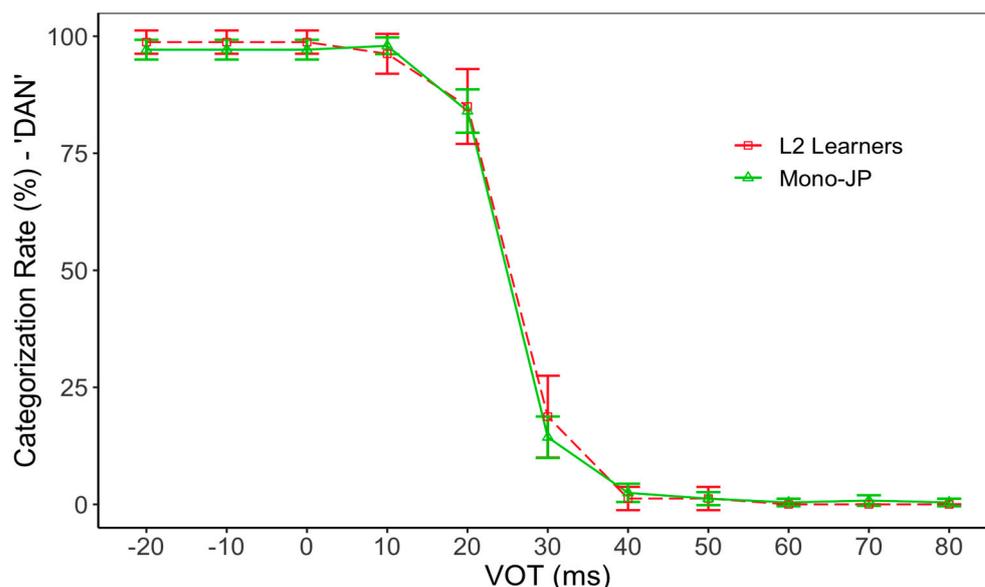


Figure 3. Categorization function of Japanese “dan”–“tan.” Error bars indicate 95% CI.

Table 5. Results of the generalized mixed-effects model from Japanese dan–tan categorization response.

	β	SE	z	p (> z)
(Intercept)	8.62	0.9	9.55	
Language group	−0.14	0.71	−0.2	0.8
Steps	−0.33	0.03	−10.34	<0.001
Language group * steps	0.009	0.02	0.3	0.7

The goodness rating part of the task was included to capture the potential subtle differences in perception between monolingual Japanese speakers and L2 learners for Japanese words, as well as between English monolingual speakers and L2 learners for English words.

In order to analyze the English goodness rating results, we first focused on stimuli whose onset VOTs were longer than 30 ms so as to examine potential group differences in rating /t/-like onsets. Then we adopted the same method to examine group differences in rating /d/-like onsets. This method was used because each participant only rated consonants that they categorized. In other words, for each trial, participants were first asked if the given stimulus sounded like /den/ or /ten/ and were then asked to rate how closely it sounded to the consonant they chose (i.e., /d/ or /t/). Because of this design, fewer stimuli with short VOTs (ms) were rated as /ten/. We chose stimuli with 30 ms to 80 ms as stimuli with /t/-like onsets. This range (30 ms to 80 ms) was chosen because the categorization task result indicated a perceptual category boundary between 20–30 ms. We focused on stimuli with longer than 30 ms VOTs so that we could examine the rating pattern differences in stimuli that were most consistently judged as /t/, as well as the difference in stimuli that were consistently judged as /d/ (Figure 4).

Two separate culminative link models were fitted on the /d/ rating and /t/ rating data sets (Tables 6 and 7).⁶ The dependent variable was the rating response. The models included the fixed effects of both the language group and the stimulus steps, as well as their interaction terms. The models’ random structures included the random intercept of participants. There was a significant effect of the language group on rating for both /d/-like and /t/-like stimuli. However, only the interaction between the stimulus steps and language group for /t/ responses was found to be statistically significant (Table 6). The rating results show that monolingual English speakers tend to give higher ratings to stimuli with longer VOTs than L2 learners of English do, which is not surprising, given that

the L2 learners' L1 Japanese voiceless consonants were generally accompanied by much shorter VOTs than those of English.

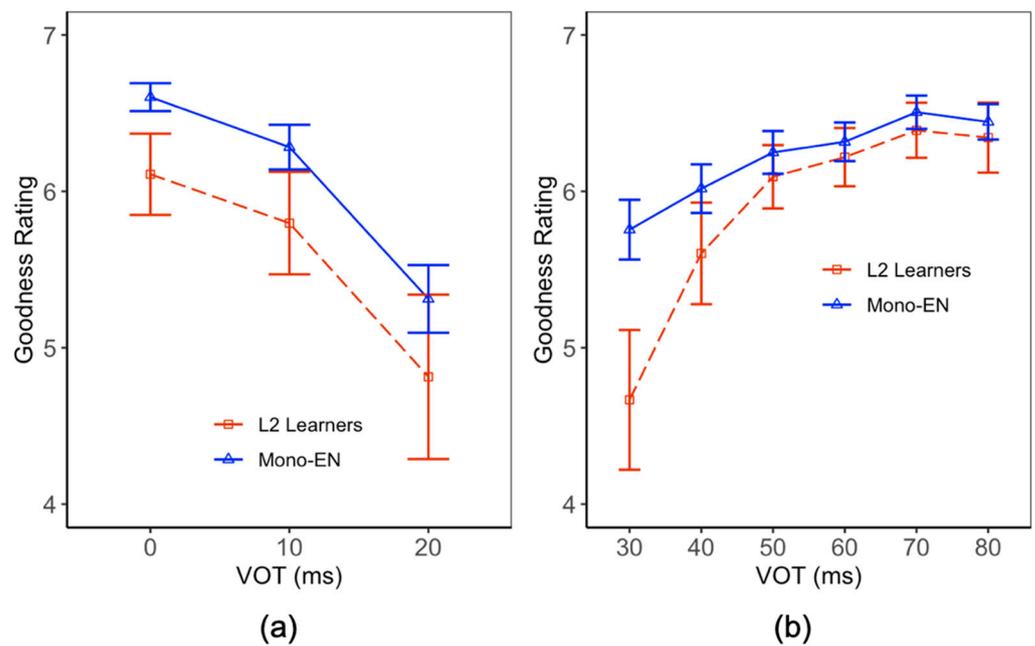


Figure 4. Rating of English stimuli. (a) Stimuli in which the first consonant had been categorized as /d/. (b) Stimuli that were categorized as /t/. The greater value indicates a more favorable goodness rating.

Table 6. Results of the ordinal mixed-effects model from English rating score (/d/ responses).

Fixed Effect	Odds Ratio	SE	z	p (> z)
Language group	1.97	0.67	2.91	<0.005
Steps	−1.38	0.021	−6.59	<0.001
Language group * steps	−0.44	0.23	−1.88	0.05

Table 7. Results of the ordinal mixed-effects model from English rating score (/t/ responses).

Fixed Effect	Odds Ratio	SE	z	P (> z)
Language group	2.16	0.78	2.77	0.005
Steps	0.6	0.06	9.06	<0.001
Language group * steps	−0.14	0.07	−1.99	<0.04

The Japanese rating data were analyzed in a similar manner (Tables 8 and 9). However, since the ordinal model with a random structure for /t/ responses did not converge, a simpler cumulative model with only fixed effects was used (Table 9). Figure 5 shows a rating peak of around 10 ms for both monolingual Japanese speakers and L2 learners, indicating that a stop with a short VOT is perceived to be a good exemplar of /d/ by both groups. In general, L2 learners tended to give lower ratings to stimuli with prevoicing and higher ratings to stimuli (that had been categorized as /tan/) with longer VOT values compared to monolingual Japanese speakers. However, the effect of language groups was significant for only the /d/ ($p < 0.001$) ratings. Additionally, there was a significant interaction of language groups with steps for /d/ ratings ($p < 0.001$). The longer the prevoicing duration was, the lower the L2 learners' rating was for /d/ stimuli.

Table 8. Results of the ordinal mixed-effects model from Japanese rating scores (/d/ responses).

Fixed Effect	Odds Ratio	SE	z	P (> z)
Language group	−1.65	0.63	−2.61	<0.005
Steps	0.20	0.07	2.76	0.005
Language group * steps	0.28	0.08	3.24	<0.005

Table 9. Results of the ordinal model on Japanese rating scores (/t/ responses).

Fixed Effects	Odds Ratio	SE	z	p (> z)
Language group	0.008	0.55	0.01	0.98
Steps	0.22	0.05	4.06	<0.001
Language group * steps	−0.004	0.06	−0.07	0.94

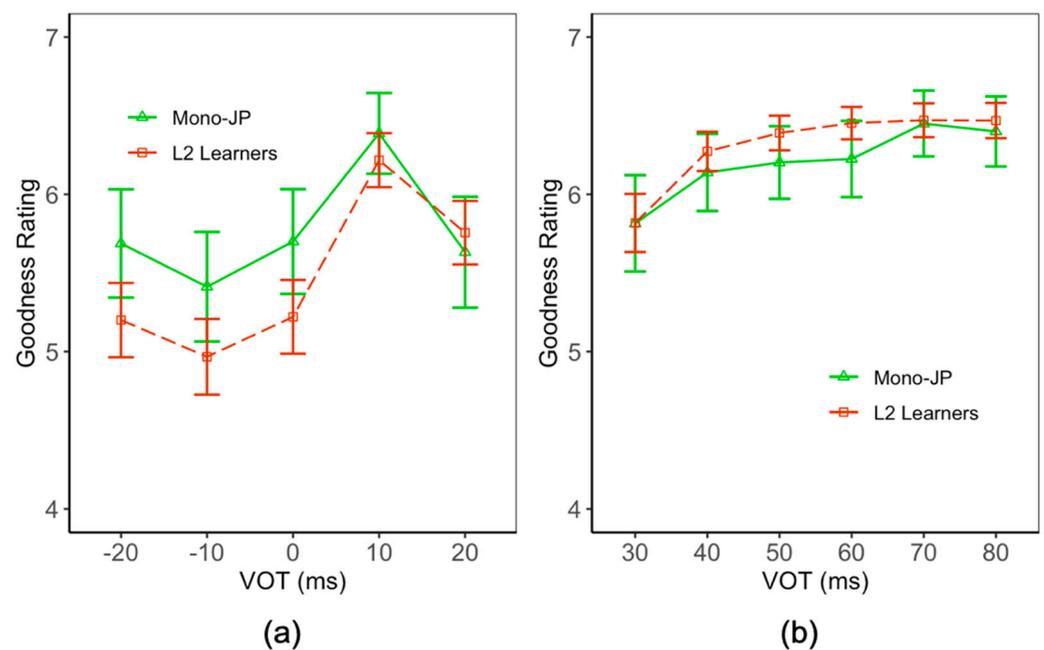


Figure 5. Rating of Japanese stimuli. (a) Stimuli in which the first consonant was categorized as /d/. (b) Stimuli that were categorized as /t/. The greater value indicates a more favorable goodness rating.

3.2. VOT Production

The L2 learner participants completed production tasks in which they were asked to read aloud a list of Japanese/English words ensconced in carrier sentences. Japanese monolinguals completed the Japanese production task, while English monolinguals carried out the English production task. L2 learners completed both production tasks. While the production of word-initial /p, t, k/ was collected in the experiment, in order to compare the results with those of perception tasks, we analyzed the production result of /t/ only. Figure 6 shows the VOT (ms) production of /t/ in each language group. While L2 learners' Japanese VOTs show a similar range to those produced by Japanese monolinguals, their English VOTs ranged from short (closer to 0 ms) to long (longer than 150 ms). This result demonstrates that there is wide variability among the L2 English VOTs produced by the 61 L2 learner participants.

A total of 18 L2 learner participants (out of 61) produced L2 English VOTs within two standard deviations of the monolingual English speakers' group mean. Given that less than half of the L2 learner participants were able to produce relatively more nativelike English VOTs (long VOTs), it is possible that the overall L2 learners' tendencies to give higher ratings for longer Japanese VOTs were driven predominantly by the subgroup of learners who can produce long VOTs in L2 English. As a post hoc analysis, we grouped

L2 learner participants based on their L2 English /t/ production into high-accuracy (in production) and low-accuracy groups.⁷ The high-accuracy group (N = 30) consisted of L2 learners whose individual mean VOTs for /t/ were within 2SD of the English monolingual mean⁸. The remainder of the L2 learners were added to the low-accuracy group (N = 31).

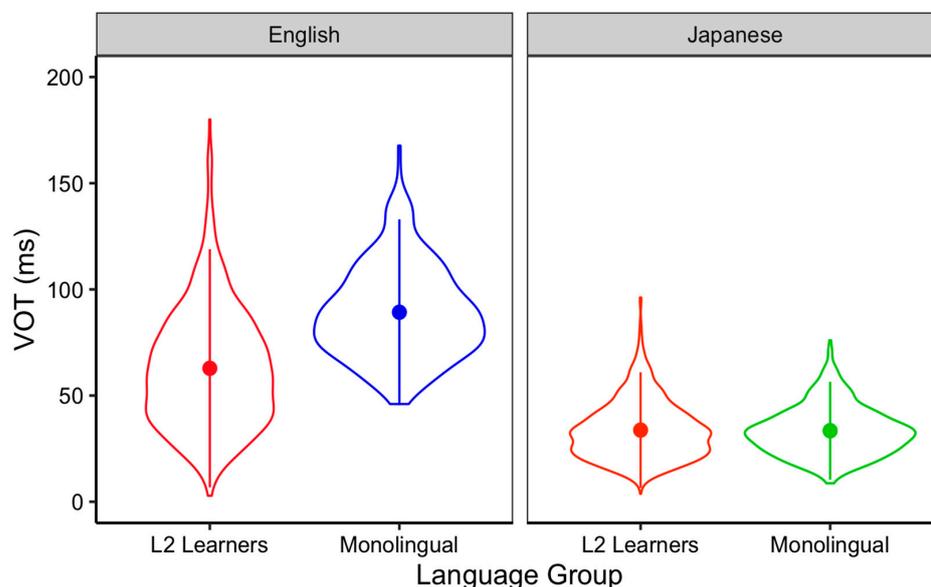


Figure 6. VOT (ms) of /t/ produced by participants in each language group. The dots indicate the group means.

Figure 7 shows the ratings of Japanese stimuli whose first consonant had been categorized as /t/ above 30 ms, comparing the two groups of L2 learners (high accuracy and low accuracy) with monolingual Japanese speakers. It shows that the high ratings on longer VOTs seen among L2 learners are primarily from those in the high-accuracy group. That is, the high-accuracy group consistently gave higher ratings on stimuli with higher VOTs (ones that were judged to be /t/) compared to the remainder of the L1 Japanese speakers (the rest of the L2 learners and monolinguals).

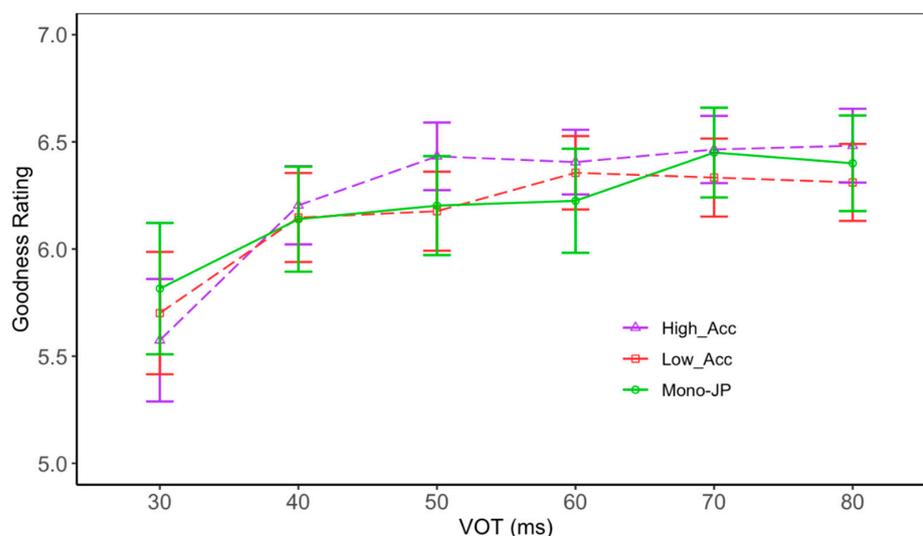


Figure 7. Rating of Japanese stimuli whose first consonant had been categorized as /t/ by two groups of L2 learners (high and low accuracy) and monolingual Japanese speakers. The greater value indicates a more favorable goodness rating.

4. Discussion

The current study employed multiple measures to assess L2 influence on L1 perception with the goal of teasing out the more nuanced aspects of potential L1 perceptual change that might exist among L2 learners.

L2 learners in this study showed evidence of L1 perceptual drift in relation to several aspects of L1 Japanese perception. For one, compared to Japanese monolinguals, the L2 learner discrimination sensitivity peak was higher along the VOT continuum. Since this “shift” is in the direction of the monolingual English speaker sensitivity peak, it could be considered evidence of L1 perceptual boundary drift caused by the L2 learning experience. It is possible that exposure to L2 English (whose voiceless stop onset is usually accompanied by a longer VOT than Japanese voiceless onsets) influenced the L2 learners’ sensitivity patterns in discriminating stimuli on a VOT continuum.

In the Japanese /tan–dan/ categorization task, there were no group differences between Japanese monolinguals and L2 learners. However, the goodness rating pattern revealed that L2 learners, overall, tended to give higher ratings to stimuli with longer VOTs—VOTs that might have been considered too long for Japanese voiceless onset. It is possible that L2 learners were more accepting of longer VOTs as good examples of Japanese voiceless stops, not because of the participant’s L2 learning experiences but because the monolingual Japanese speaker sample size was smaller, introducing more variation in L2 learners. However, the post hoc grouping based on L2 learners’ VOT production accuracies revealed that higher ratings on stimuli with longer VOTs are largely driven by those L2 learners who produced more nativelike English VOTs (within 2SD of the monolingual English speaker mean). This suggests that the rating pattern differences on longer VOTs are likely due to the participants’ L2 learning experiences. It is reasonable to assume that high-accuracy (in their VOT production) L2 learners had sufficient prior L2 English input that helped them be able to produce nativelike L2 English VOTs. The exposure to longer VOTs might have influenced the perception of L1 Japanese, making the longer VOTs more acceptable to the L2 learners.

In addition to the rating pattern difference in /t/-like stimuli, there was an unexpected rating difference in /d/-like stimuli; L2 learners tended to give lower ratings to stimuli with prevoicing. In English, voiced stop onset has been reported to generally have short-lag VOTs and much less often to involve prevoicing (e.g., [Lisker and Abramson 1964, 1967a](#)), although there is variability ([Herd 2020](#)). However, Japanese /b d g/ are generally thought of as accompanied by prevoicing, ([Nasukawa 2008](#); [Ogasawara 2011](#); [Shimizu 1989](#)), although the use of it does not seem to be mandatory and varies depending on region and generation ([Takada et al. 2015](#)). It is possible that L2 learners may have been exposed to L2 onset /d/ without prevoicing, which influenced their rating on prevoiced onset /d/ in their L1 Japanese.

The lack of L1 perceptual drift in the voicing category boundary is not surprising given that Japanese and English monolinguals exhibited similar voicing boundaries in the categorization part of the category goodness rating task. We expected that the two languages would show some difference in terms of /d-t/ voicing boundaries given that the two languages utilize different areas on the VOT continuum for the voiceless /t/. While the mean boundary VOT (ms) reported in this study for monolingual English speakers is shorter than some of the previously reported values (e.g., [Flege and Eefting 1987a](#); [Hitchcock and Koenig 2021](#); [Keating et al. 1981](#)), it is still within the range of the boundary values exhibited in other studies (e.g., [Mack 1989](#)). While the fact that the English monolinguals’ /d-t/ perceptual boundaries reported unexpectedly shorter VOTs might be an artifact of the task included in this study, it is also possible that the result reflects some regional or generational trend and warrants further investigation.

It is worth noting, however, that we did observe L1 drift in the sensitivity peak but did not observe a similar drift pattern in the voicing category boundary. The discrepancy in the result might come from the nature of the task. Specifically, because the stimuli in the AX discrimination task were not presented in real words in English or Japanese, it is

possible that both Japanese and English languages were activated in L2 learners' minds, causing the sensitivity peak to drift away from that of the Japanese monolingual group. L2 learners might also have picked up cues other than VOTs in the AX discrimination task that caused them to be in English mode. As it is difficult to determine which language(s) is(are) activated in learners' minds, the cause of the drift requires further study.

The similarity in perceptual voicing boundaries observed in the categorization task between the two monolingual groups indicates that there is, at most, a very weak link between the voicing boundary location in perception and English VOT production for monolingual English speakers. That is, in English, the VOT values for voiceless stop onset are generally much higher than the VOT values around the perceptual voicing boundary. Nevertheless, it would be hard to explain L2 learners' rating patterns without some perception–production link; learners who were able to produce more nativelike L2 English VOTs tended to give higher ratings on longer VOTs when perceiving Japanese stimuli with onset /t/. We have to note, however, that no direct correlation was found between the L2 learners' English VOT production and Japanese VOT rating on /t/-like stimuli. The lack of direct correlation is consistent with the idea introduced in the Revised Speech Learning Model (SLM-R; Flege and Bohn 2021) that perception and production co-develop without precedence. Since learners differ in their developmental paths, some learners' L2 production might be more developed than their L2 perception and vice versa. It is possible that some learners' ability to notice subtle differences in perception developed before their ability to accurately produce L2 sounds (i.e., long VOTs).

One thing to note is that the aspirated stops in English are allophonic, and their occurrence is determined by context. It is not clear if L1 Japanese learners of English would generalize the use of VOT cues of word-initial stop consonants for the production and perception of stop consonants in other prosodic contexts. In fact, there is a debate in the literature as to the extent to which learners generalize what they have learned about contextual L2 allophones to other variants of the same phoneme, with some reporting that L2 learners do not generalize the learning of word-initial voicing cues to voicing contrasts in other contexts (e.g., de Jong et al. 2009). Further research is required to study how learning to produce nativelike VOTs in the word-initial position in L2 English influences the perception and production of English stop consonants in other prosodic contexts, and how subsequent influence on L1 Japanese stops might take place in various prosodic contexts.

The present study demonstrated several pieces of evidence in favor of the L1 perceptual drift of stop onset. While each piece of evidence alone may not be strong enough to say conclusively that L2 learners exhibited perceptual drift, and further studies are required to confirm each one, taken together, these data suggest an L2 learning influence on L1 (i.e., perceptual drift in terms of the sensitivity pattern and goodness rating pattern). In some L2 phonetic learning studies, phonetic development is observed only in production and not in perception–perceptual development, which is sometimes considered slower or separate from achieving production accuracy (e.g., de Leeuw et al. 2021; Sheldon and Strange 1982). That the change observed in L1 perception is smaller than what we would have expected might be because perception is less susceptible to change compared to production.⁹ Although further research is required to confirm that speakers' perceptions are more resistant to change, this view offers a possible explanation as to why certain mismatches between L2 perception and L2 production have been observed in the literature.

Funding: This research was funded by ETS: Small Grands for Doctoral Research in Second or Foreign Language Assessment, Stony Brook University Faculty-Staff Dissertation Fellowship, as well as Stony Brook University AHLSS Summer Research Fellowship.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Stony Brook University (31 October 2018).

Informed Consent Statement: Written informed consent was obtained from all participants involved in the study.

Data Availability Statement: The data files can be accessed via https://osf.io/f9gxr/?view_only=f0cd581646e54d698352eaf9fef5c47d.

Acknowledgments: I would first like to thank Marie Huffman for her helpful feedback and consultation. I would also like to thank Ellen Browselow, Michael Becker, and Fredericka Bell-Berti, as well for their questions and comments. I thank Kristen Pajkowski for their help in segmentation and Peter Cuce for his editorial help. Finally, I thank the reviewers and the associate editor for their valuable comments and questions.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- 1 A Praat script made by Schertz (2014) was modified for this task.
- 2 Takada et al. (2015) show an ongoing generational change in word-initial voiced stop VOTs from heavily prevoiced variants to less voiced ones.
- 3 It is possible that the L2 participants were in bilingual mode during the entire session as it is difficult to make sure a participant is completely in a “unilingual mode”. However, by giving all the written and spoken instructions in the language of the current task, we aimed to help participants stay in the mode of the language of the given task.
- 4 The d-prime score of 0 indicates hit rate = false alarm rate. A positive score indicates better sensitivity, while a negative d-prime score reflects a higher false alarm rate compared to the hit rate.
- 5 Normality and homogeneity of variance were checked using the Shapiro–Wilk test and Levene’s test using an R package, rstatix (version 0.7.2; Kassambara 2023).
- 6 The data were analyzed using the original package (Christensen 2019).
- 7 The relationship between L2 learners’ production VOTs and goodness rating can also be examined by looking at their correlation. However, no significant correlation was found between them.
- 8 Two standard deviations of the native speaker group mean were chosen as the criterion for determining nativelikeness based on several previous studies (e.g., Flege et al. 1995; Munro et al. 1996).
- 9 bAs it was pointed out by the anonymous reviewer, it is also possible that the strength of the perceptual measure might depend on the task.

References

- Abramson, Arthur S., and Leigh Lisker. 1973. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics* 1: 1–8. [CrossRef]
- Barlow, Jessica A. 2014. Age of Acquisition and Allophony in Spanish-English Bilinguals. *Frontiers in Psychology* 5: 288. [CrossRef] [PubMed]
- Benkí, José R. 2001. Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics* 29: 1–22. [CrossRef]
- Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and Second-Language Speech Perception: Commonalities and Complementarities. In *Language Learning & Language Teaching*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam: John Benjamins Publishing Company, vol. 17, pp. 13–34. [CrossRef]
- Boersma, Paul, and David Weenink. 2020. Praat: Doing Phonetics by Computer [Computer Program]. Available online: <http://www.praat.org/> (accessed on 31 October 2018).
- Caramazza, A., and G. H. Yeni-Komshian. 1974. Voice Onset Time in Two French Dialects. *Journal of Phonetics* 2: 239–45. [CrossRef]
- Casillas, Joseph V. 2020. Phonetic Category Formation Is Perceptually Driven During the Early Stages of Adult L2 Development. *Language and Speech* 63: 550–81. [CrossRef] [PubMed]
- Chang, Charles B. 2012. Rapid and Multifaceted Effects of Second-Language Learning on First-Language Speech Production. *Journal of Phonetics* 40: 249–68. [CrossRef]
- Chang, Charles B. 2013. A Novelty Effect in Phonetic Drift of the Native Language. *Journal of Phonetics* 41: 520–33. [CrossRef]
- Chang, Charles B. 2019. Phonetic Drift. In *The Oxford Handbook of Language Attrition*. Edited by Monika S. Schmid and Barbara Köpcke. Oxford: Oxford University Press. [CrossRef]
- Christensen, Rune H. B. 2019. Ordinal: Regression Models for Ordinal Data. R Package. Available online: <https://CRAN.R-project.org/package=ordinal> (accessed on 20 November 2023).
- Crawford, Clifford James. 2009. Adaptation and Transmission in Japanese Loanword Phonology. Ph.D. dissertation, Cornell University, Ithaca, NY, USA.
- de Leeuw, Esther, Linnaea Stockall, Dimitra Lazaridou-Chatzigoga, and Celia Gorba Masip. 2021. Illusory Vowels in Spanish–English Sequential Bilinguals: Evidence That Accurate L2 Perception Is Neither Necessary nor Sufficient for Accurate L2 Production. *Second Language Research* 37: 587–618. [CrossRef]

- Dmitrieva, Olga. 2019. Transferring Perceptual Cue-Weighting from Second Language into First Language: Cues to Voicing in Russian Speakers of English. *Journal of Phonetics* 73: 128–43. [CrossRef]
- Flege, James Emil. 1982. Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics* 10: 177–92. [CrossRef]
- Flege, James Emil. 1987. The Production of ‘New’ and ‘Similar’ Phones in a Foreign Language: Evidence for the Effect of Equivalence Classification. *Journal of Phonetics* 15: 47–65. [CrossRef]
- Flege, James Emil. 1995. Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 233–72.
- Flege, James Emil. 2007. Language Contact in Bilingualism: Phonetic System Interactions. In *Laboratory Phonology*. Edited by Jennifer Cole and José Ignacio Hualde. Berlin and New York: Mouton de Gruyter, vol. 9, pp. 353–81.
- Flege, James Emil, and Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Rtree Wayland. Cambridge: Cambridge University Press, pp. 3–83.
- Flege, James Emil, and Wieke Eefting. 1987a. Production and Perception of English Stops by Native Spanish Speakers. *Journal of Phonetics* 15: 67–83. [CrossRef]
- Flege, James Emil, and Wieke Eefting. 1987b. Cross-Language Switching in Stop Consonant Perception and Production by Dutch Speakers of English. *Speech Communication* 6: 185–202. [CrossRef]
- Flege, James Emil, Carlo Schirru, and Ian R. A. MacKay. 2003. Interaction between the Native and Second Language Phonetic Subsystems. *Speech Communication* 40: 467–91. [CrossRef]
- Flege, James Emil, Ian R. A. MacKay, and Diane Meador. 1999. Native Italian Speakers’ Perception and Production of English Vowels. *The Journal of the Acoustical Society of America* 106: 2973–87. [CrossRef]
- Flege, James Emil, Murray J. Munro, and Ian R. A. MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* 97: 3125–34. [CrossRef] [PubMed]
- Guion, Susan G. 2003. The Vowel Systems of Quichua-Spanish Bilinguals. *Phonetica* 60: 98–128. [CrossRef] [PubMed]
- Harada, Tetsuo. 2003. L2 Influence on L1 Speech in the Production of VOT. Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9; pp. 1085–88.
- Harada, Tetsuo. 2007. The Production of Voice Onset Time (VOT) by English-Speaking Children in a Japanese Immersion Program. *International Review of Applied Linguistics in Language Teaching* 45: 353–78. [CrossRef]
- Herd, Wendy. 2020. Sociophonetic Voice Onset Time Variation in Mississippi English. *The Journal of the Acoustical Society of America* 147: 596–605. [CrossRef]
- Hitchcock, Elaine, and Laura L. Koenig. 2021. Adult perception of stop consonant voicing in American-English-learning toddlers: Voice onset time and secondary cues. *Journal of the Acoustical Society of America* 150: 460–77. [CrossRef]
- Iverson, Paul, and Patricia K. Kuhl. 1995. Mapping the Perceptual Magnet Effect for Speech Using Signal Detection Theory and Multidimensional Scaling. *The Journal of the Acoustical Society of America* 97: 553–62. [CrossRef] [PubMed]
- Jong, Kenneth J. de, Noah H. Silbert, and Hanyong Park. 2009. Generalization Across Segments in Second Language Consonant Identification. *Language Learning* 59: 1–31. [CrossRef]
- Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich Hans Frauenfelder, and Narly Golestani. 2016a. Mutual Influences between Native and Non-Native Vowels in Production: Evidence from Short-Term Visual Articulatory Feedback Training. *Journal of Phonetics* 57: 21–39. [CrossRef]
- Kartushina, Natalia, Ulrich H. Frauenfelder, and Narly Golestani. 2016b. How and When Does the Second Language Influence the Production of Native Speech Sounds: A Literature Review: L2 Influences on L1: A Literature Review. *Language Learning* 66: 155–86. [CrossRef]
- Kassambara, Alboukadel. 2023. rstatix: Pipe-Friendly Framework for Basic Statistical Texts. R Package Version 0.7.2. Available online: <https://rpkgs.datanovia.com/rstatix/> (accessed on 1 July 2023).
- Keating, Patricia, Michael J. Mikoś, and William F. Ganong. 1981. A cross-language study of range of voice onset time in the perception of initial stop voicing. *The Journal of the Acoustical Society of America* 70: 1261. [CrossRef]
- Lang, Benjamin, and Lisa Davidson. 2019. Effects of Exposure and Vowel Space Distribution on Phonetic Drift: Evidence from American English Learners of French. *Language and Speech* 62: 30–60. [CrossRef] [PubMed]
- Lev-Ari, Shiri, and Sharon Peperkamp. 2013. Low Inhibitory Skill Leads to Non-Native Perception and Production in Bilinguals’ Native Language. *Journal of Phonetics* 41: 320–31. [CrossRef]
- Lisker, Leigh, and Arthur S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20: 384–422. [CrossRef]
- Lisker, Leigh, and Arthur S. Abramson. 1967a. Some Effects of Context on Voice Onset Time in English Stops. *Language and Speech* 10: 1–28. [CrossRef]
- Lisker, Leigh, and Arthur S. Abramson. 1967b. The voicing dimension: Some experiments in comparative phonetics. Paper presented at the 6th International Congress of Phonetic Sciences, Prague, Czech Republic, September 7–13; Edited by Bohuslav Hála, Milan Romportl and Přmysl Janota. pp. 563–67.
- Long, Michael H. 1990. Maturation Constraints on Language Development. *Studies in Second Language Acquisition* 12: 251–85. [CrossRef]

- Mack, Molly. 1989. Consonant and Vowel Perception and Production: Early English-French Bilinguals and English Monolinguals. *Perception & Psychophysics* 46: 187–200. [CrossRef]
- MacKay, Ian R. A., James Emil Flege, Thorsten Piske, and Carlo Schirru. 2001. Category Restructuring during Second-Language Speech Acquisition. *The Journal of the Acoustical Society of America* 110: 516–28. [CrossRef] [PubMed]
- Macmillan, Neil A., and C. Douglas Creelman. 2005. *Detection Theory: A User's Guide*, 2nd ed. Mahwah: Lawrence Erlbaum Associates.
- Major, Roy C. 1992. Losing English as a First Language. *The Modern Language Journal* 76: 190–208. [CrossRef]
- Mora, Joan C., and Marianna Nadeu. 2012. L2 Effects on the Perception and Production of a Native Vowel Contrast in Early Bilinguals. *International Journal of Bilingualism* 16: 484–500. [CrossRef]
- Mora, Joan C., James L. Keidel, and James E. Flege. 2015. Effects of Spanish Use on the Production of Catalan Vowels by Early Spanish-Catalan Bilinguals. In *Current Issues in Linguistic Theory*. Edited by Joaquín Romero and María Riera. Amsterdam: John Benjamins Publishing Company, vol. 335, pp. 33–54. [CrossRef]
- Munro, Murray, James Emil Flege, and Ian R. A. Mackay. 1996. The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics* 17: 313–34. [CrossRef]
- Nasukawa, Kuniya. 2008. Place-Dependent VOT in L2 Acquisition. In *Selected Proceedings of the 2008 Second Language Research Forum*. Edited by Matthew T. Prior, Yukiko Watanabe and Sang-Ki Lee. Somerville: Cascadilla Proceedings Project, pp. 197–210.
- Ogasawara, Naomi. 2011. Acoustic Analysis of Voice-Onset Time in Taiwan Mandarin and Japanese. *Concentric: Studies in Linguistics* 37: 155–78.
- Pailler, Christophe. 2003. Computation of the Sensitivity Parameter (d') of SDT [R Script]. Available online: <http://www.pallier.org/computing-discriminability-a-d-and-bias-with-r.html> (accessed on 1 July 2019).
- Riney, Timothy James, Naoyuki Takagi, Kaori Ota, and Yoko Uchida. 2007. The Intermediate Degree of VOT in Japanese Initial Voiceless Stops. *Journal of Phonetics* 35: 439–43. [CrossRef]
- Ruben, Robert J. 1999. A Time Frame of Critical/Sensitive Periods of Language Development. *Acta Oto-Laryngologica* 117: 202–5. [CrossRef]
- Sancier, Michele L., and Carol A. Fowler. 1997. Gestural Drift in a Bilingual Speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25: 421–36. [CrossRef]
- Sheldon, Amy, and Winifred Strange. 1982. The Acquisition of /r/ and /l/ by Japanese Learners of English: Evidence That Speech Production Can Precede Speech Perception. *Applied Psycholinguistics* 3: 243–61. [CrossRef]
- Schertz, Jessamyn. 2014. klatt_synthesize_vot_f0_series.praat. Available online: <http://individual.utoronto.ca/jschertz/scripts.shtml> (accessed on 25 July 2018).
- Shimizu, Katsumasa. 1977. Voicing features in the perception and production of stop consonants by Japanese speakers. *Studia Phonologica* 11: 25–34.
- Shimizu, Katsumasa. 1989. A Cross-Language Study of Voicing Contrasts of Stops. *Studia Phonologica* 23: 1–12.
- Shimizu, Katsumasa. 1990. A Cross-Language Study of Voicing Contrasts of Stop Consonants in Asian Languages. Ph.D. dissertation, University of Edinburgh, Edinburgh, UK.
- Takada, Mieko, Eun Jong Kong, Kiyoko Yoneyama, and Mary E. Beckman. 2015. Loss of Prevoicing in Modern Japanese /g, d, b/. Paper presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14; pp. 1–5.
- Takahashi, Chikako. 2020. The Interaction between L1 and L2 Phonetic Learning. Ph.D. dissertation, Stony Brook University, New York, NY, USA.
- Tice, Marisa, and Melinda Woodley. 2012. Pagnettes & Bastries: Novice French Learners Show Shifts in Native Phoneme Boundaries. UC Berkeley Phonology Lab Annual Report. Available online: <http://journals.linguisticsociety.org/proceedings/index.php/ExtendedAbs/article/view/589> (accessed on 1 July 2019).
- Williams, Lee. 1977. The Perception of Stop Consonant Voicing by Spanish-English Bilinguals. *Perception & Psychophysics* 21: 289–97. [CrossRef]
- Xu, Yisheng, and Alexander L. Francis. 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America* 120: 1063–74. [CrossRef]
- Yusa, Noriaki, Kuniya Nasukawa, Masatoshi Koizumi, Kim Jungho, Naoki Kimura, and Kensuke Emura. 2010. Unexpected effects of the second language on the first. In Paper presented at the New Sounds 2010: Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, Poznan, Poland, May 1–3; pp. 580–84.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.