NATIVE SPEAKERS' REDUNDANT ACOUSTIC CUE MAY BE LEARNERS' TREASURE: CAN PITCH BE REPURPOSED?

Keiji Iwamoto, Indiana University Bloomington

Isabelle Darcy, Indiana University Bloomington

Instructors of Japanese commonly notice Mandarin speakers' perceptual difficulty in identifying Japanese voicing categories. The Mandarin phonological system seems to make learners 'deaf' to essential/primary acoustic cues (pre-voicing cues) used by Japanese native speakers, so they may need to rely on other acoustic cues, such as pitch cues on post-stop vowels which are redundant/secondary for category identification. However, since pitch categorically co-varies with voicing cues (high pitch: voiceless, low pitch: voiced), we hypothesized that pitch could play an important role for Mandarin speakers because of their sensitivity to pitch thanks to the L1 tonal system. We used a gating paradigm to test whether learners rely on pitch cues to identify voicing categories. Learners (L1: Mandarin or English) and Japanese native speakers heard fragments of non-words and identified voicing categories in word-initial position based only on pitch cues on following vowels (e.g., /tasɯbiː/ without /t/, i.e., /asɯbiː/). The results showed that Mandarin speakers indeed used pitch cues and outperformed L1 Japanese/English speakers, i.e., the way pitch is used in L1 matters in determining how important or redundant pitch cues can be. Based on the results, we discuss the goal of perception training in the spirit of the Intelligibility Principle.

INTRODUCTION

Among multiple acoustic cues consisting of sound (e.g., loudness, duration, pitch), some cues are more essential or important (primary cues) than others (redundant cues). For example, primary cues of word-initial voicing contrasts of stops in many varieties of English (e.g., /t/ vs. /d/ as in *ten* vs. *den*) are duration of aspiration (i.e., a puff of air; longer aspiration = voiceless, shorter aspiration =voiced) and one of the redundant cues are pitch cues on following vowels (/ɛ/ in *ten*/*den*) which co-vary with voicing cues (voiceless = higher-pitch, voiced = lower-pitch). Redundant cues could play a primary role when primary cues become ambiguous, for example, when duration of aspiration is in between typical voiceless or voiced stops (Whalen et al., 1993).

This study examined the possibility that a redundant acoustic cue for native speakers (NS) can be an important cue for some learners by disentangling multiple cue factors (Gating paradigm)

Iwamoto, K. & Darcy, I. (2024). Native Speakers' Redundant Acoustic Cue may be Learners' Treasure: Can Pitch be Repurposed? In D. J. Olson, J. L. Sturm, O. Dmitrieva, & J. M. Levis (eds), *Proceedings of 14th Pronunciation in Second Language Learning and Teaching Conference*, (pp. 1-12). Purdue University, September 2023. https://doi.org/10.31274/psllt.17555

and by comparing three L1 groups: (Midwest/Indiana) American English, Tokyo-Japanese, and Mandarin Chinese (henceforth, English, Japanese, Mandarin). The rationale is that these three languages differ in the degree of pitch usage for lexical contrasts, which has been shown to affect the ability to discriminate pitch (Schaefer & Darcy, 2014). The target cue is *pitch*, a cue that is often claimed to be a redundant cue for Japanese voicing contrasts of stops, which can be difficult for L1 Mandarin speakers to identify (as often observed by Japanese language instructors).

Knowing which acoustic cues are important for learners could allow teachers/learners to focus on specific prosodic positions (e.g., word-medial positions) where important-for-learners acoustic cues are not available, referring to production studies. Additionally, in the spirit of the Intelligibility Principle (e.g. Levis 2020) for L2 pronunciation (production), we ponder if 'native-like' perception is the goal of L2 perception teaching and suggest instead that listeners can reach contrastive perception with different cues.

Voicing Contrasts

Voicing contrasts are attested in many languages but primary cues for voicing contrasts vary. In so-called 'voicing' languages, the presence or absence of pre-voicing in initial position is attended to as primary cue by some language speakers, such as Japanese. In so-called 'aspirating' languages, the primary cue that speakers attend to is aspiration differences, such as in Mandarin, and in many varieties of English (Cho et al., 2019). If these pre-voicing/aspiration cues are primary, voiceless stops (/t/) realized as [t] in a *voicing* language (with acoustically short or zero aspiration) could be interpreted as voiced stops (/d/) by *aspirating* language speakers who would rely on aspiration as their main cue. The L2 perception models such as the Speech Learning Model (SLM, Flege, 1995) predicts that such perceived similarity results in the merger of voiced and voiceless sounds, hindering new L2 category formation for *pre-voiced* voiced stops. As a result, both Japanese voiceless (/t/) and voiced (/d/) may be interpreted as voiced stops (/d/) in L1 Mandarin speakers' minds, causing perceptual confusion.

Redundant Cues: Pitch

The aforementioned predictions only consider pre-voicing/aspiration cues (primary cues) but redundant cues cannot be ignored since redundant cues may play an important role for L2 learners who have such perceptual confusion. Rather, since primary cues can be misleading cues due to the mismatch in voicing contrasts between L1 and L2 (i.e., 'voicing' vs. 'aspirating'), redundant cues may be promoted to a *primary* cue. Unlike the primary cues, it appears that oft-cited redundant cues, vowel pitch cues created by consonantal perturbation (e.g., Kang & Hirayama, 2023; Kirby, 2018), could be used for voicing category identification more straightforwardly, i.e., in many languages, phonologically voiceless stops induce higher pitch on post-stop vowels than phonologically voiced ones, regardless of its phonetic realizations (Dmitrieva et al., 2015). For example, in Japanese, /a/ in 'ta' has higher pitch than /a/ in 'da' in word-initial position (Gao & Arai, 2019).

Learners may be able to make use of the phonologically specified categorical pitch differences for L2 perception of voicing contrasts, especially for Mandarin speakers whose L1 (a tone language) employs pitch height and contours for lexical contrasts extensively. The *redundant* pitch cues may be a precious cue for L1 Mandarin learners of L2 Japanese thanks to their learned sensitivity to pitch (Schaefer & Darcy, 2014). It is this potential relationship between pitch cues in voicing contrasts and learned sensitivity to pitch cues in L1 that we investigated in this study. We assume that the way pitch is used for lexical contrasts (i.e., lexical function of pitch) in L1 characterizes their sensitivity to pitch even in L2.

Research Questions and Hypotheses

RQ1: Can listeners identify L2 voicing categories only with pitch cues (without primary cues, e.g., pre-voicing and aspiration)?

RQ2: If possible, is there a correlation between the lexical function of pitch in L1 and pitch usage for segmental contrast in L2?

Three possibilities exist to characterize the relationship between pitch cues in L2 voicing contrasts and the lexical function of pitch in L1, resulting in different possibilities for learners to use pitch as a secondary (backup) cue.

The *Extension Hypothesis*

Tone language speakers (e.g., Mandarin) are more sensitive to pitch differences than non-tone language speakers (e.g., English, Japanese) because pitch differences contrast more words in tone languages than non-tone languages. Also, there seems to be a difference even among non-tone language speakers depending on the degree of pitch usage for lexical contrasts in L1 (Schaefer & Darcy, 2014). For example, pitch-accent language speakers (e.g., Japanese) are more sensitive to pitch than lexical stress language speakers (e.g., English). This is presumably because pitch is used exclusively to decode pitch-accent patterns in Japanese whereas pitch is just one of the acoustic cues to signal lexical stress in combination with vowel quality, duration, and loudness. Based on the different sensitivity to pitch due to the lexical function of pitch in L1, we hypothesize that the more pitch differentiates lexical items in L1, the more sensitive to pitch people become even in segmental contrasts, and the better able they may be in using it for identifying voicing (Figure 1A).

The *Attenuation Hypothesis*

Previous production studies showed that tone language speakers seem to attenuate pitch perturbation effects on post-stop vowels (Francis et al., 2006; Kirby, 2018). This is presumably because tone language speakers should preserve pitch cues for tonal contrasts, i.e., lexical contrasts, rather than segmental contrasts. This attenuation may happen in perception too. The more pitch differentiates lexical contrast in L1, the more people attenuate pitch perturbation effects, resulting in a decreased ability to use pitch for voicing contrasts (Figure 1B).

The *Pitch Saliency Hypothesis*

Francis et al. (2006) showed that although Cantonese speakers attenuate the pitch perturbation effect in production, they seem to be able to employ pitch differences to identify voicing categories in perception. In other words, their behavior in production does not reflect their perception (i.e., in perception, they can do what they are not used to doing). They suggested the possibility of pitch being a language-independent cue for voicing contrast, in which case anyone, irrespective of pitch usage in their L1, would be able to rely on pitch if useful (Figure 1C).

Rationale and Predictions

To test our hypotheses, we recruited learners of Japanese with Mandarin and English as L1, and a group of native Japanese listeners. A gating paradigm was implemented to disentangle multiple cue factors: Participants chose 'voiced' or 'voiceless' only with pitch cues on post-stop vowels (e.g., /tasɯbiː/ without /t/, i.e., /asɯbiː/). Each hypothesis predicts the following order (high to low) of the accuracy scores:

The *Extension Hypothesis*: Mandarin > Japanese > English (Fig. 1A)
The *Attenuation Hypothesis*: English > Japanese > Mandarin (Fig. 1B)
The *Pitch Saliency Hypothesis*: English ≈ Japanese ≈ Mandarin (Fig. 1C)

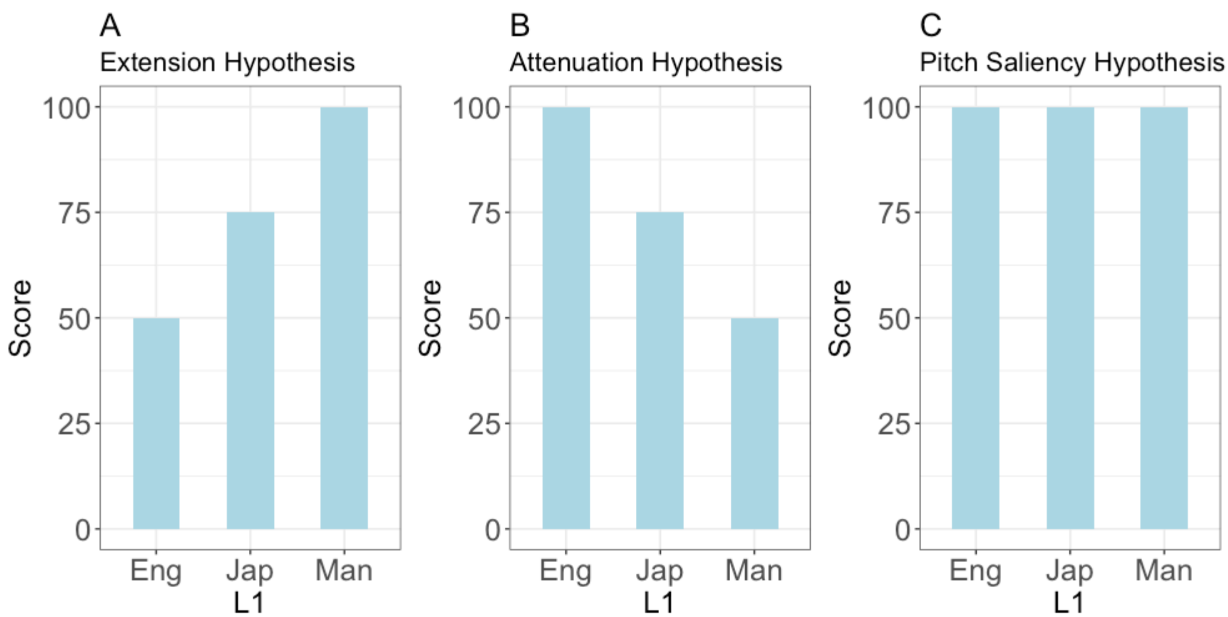

Figure 1. Predictions for each hypothesis (Eng: English, Jap: Japanese, Man: Mandarin)

METHODS
Participants

Thirty-two participants were recruited from three language groups: American English (N = 11), Japanese (N = 10), and Mandarin (N = 11). No Japanese participant spoke an accentless variety. Most learners were beginning learners of Japanese, enrolled in second semester courses. No participants reported any hearing impairment.

Five participants were excluded from the final analysis: (1) four participants had extensive exposure to other tone languages; One English speaker speaks Vietnamese; One Japanese speaker had learned Mandarin; Two Mandarin speakers speak other Chinese varieties as L1s (Shanghainese and Cantonese); (2) One Mandarin speaker could not read Japanese orthography (*katakana*) without the researcher's assistance. Therefore, only datapoints from ten English, nine Japanese, and eight Mandarin speakers were analyzed.

Stimuli

Non-word minimal pairs were used as stimuli to minimize any lexical effect. Non-words consist of four morae (CV.CV.CV.V) in Japanese where the onset of the first mora is a stop which perturbs pitch patterns on post-stop vowels. The nucleus of the first mora is a low-mid vowel, /a/, which is used in all L1s of the three L1 groups. The first author (Japanese NS) assigned the remaining segments without violating Japanese phonotactics. The target stops in word-initial position consist of three places of articulation (POA: bilabial, alveolar, and velar) and two voicing categories, resulted in 6 non-word items (Table 1).

Recording

These non-words were recorded by two Japanese native speakers (Tokyo dialect; a female and a male born after the 1990s) who did not participate in the task. After providing instructions and practicing the pitch-accent patterns, each talker recorded carrier sentences with each item four times. The first author (a Japanese native speaker) checked the pitch-accent pattern. Talkers were recorded individually in a sound-proof room with a Blue Yeti X microphone. The sound software used for the recordings was Praat (Boersma & Weenink, 2022, Version 6.2.04) installed in Mac Book Air (M1 2020, Version 13.3.1).

Table 1

Japanese orthography and IPA for the non-word items

Japanese orthography	IPA	Pitch-accent patterns[1]
タスビー	/tasɯbiː/	HLLL
ダスビー	/dasɯbiː/	HLLL
パスビー	/pasɯbiː/	HLLL
バスビー	/basɯbiː/	HLLL
カスビー	/kasɯbiː/	HLLL
ガスビー	/gasɯbiː/	HLLL

[1]H = High pitch; L = Low pitch.

Acoustic Analyses

The duration of the post-voiced-stop vowel and post-voiceless-stop vowel differed by less than 20ms (an arbitrary number, see Table 2).

Table 2
Durations of post-stop vowels in each item by talker (ms)

	tasɯbi:	dasɯbi:	pasɯbi:	basɯbi:	kasɯbi:	gasɯbi:
Female	90	86	82	87	74	81
Male	71	78	73	81	74	78

The f0 contours of the post-stop vowels were analyzed with Praat. Prosody Pro (Xu, 2013) divided the post-stop vowels into 50 equal time intervals and extracted the mean f0 values of each time interval. The averaged f0 contours on the post-stop vowels from each talker are shown in Figure 2.

The male talker produced the f0 contours which have been reported in previous production studies (Gao & Arai, 2019); the onset f0 of the post-voiceless-stop vowels is higher than that of the post-voiced-stop vowels. The female talker, however, produced deviant patterns from what previous studies have reported. The onset f0 of the post-voiced-stop vowels is higher than that of the post-voiceless-stop vowels. These abnormal patters observed in the female talker might have been caused by the difficulty in producing the pitch-accent pattern of the non-word items. To avoid this potential confound, we removed the data points obtained with the stimuli from this talker from the analysis.

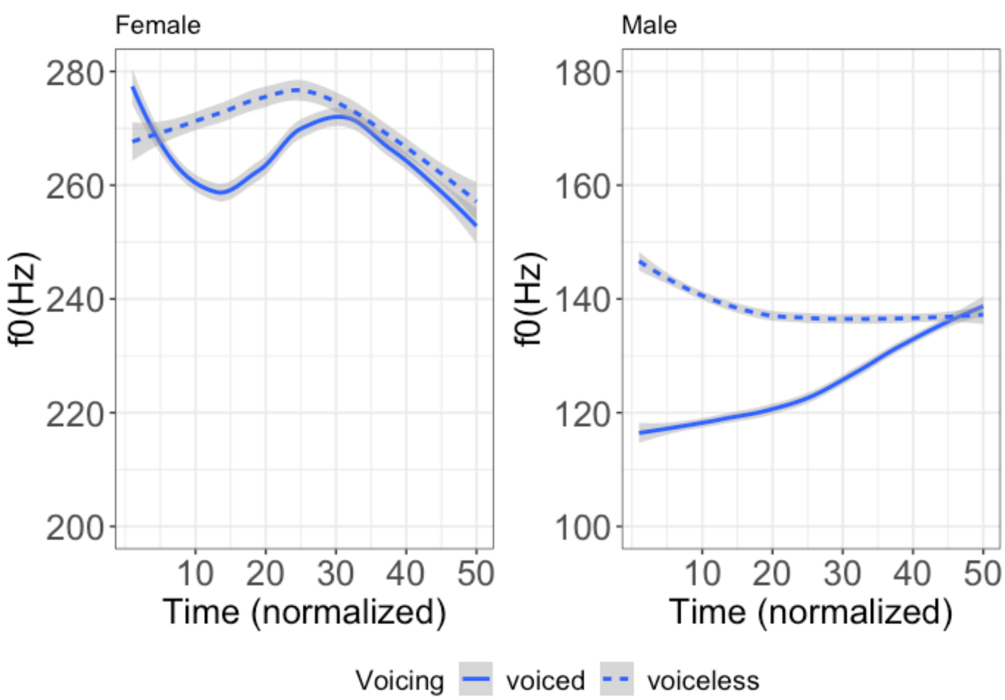

Figure 2. Time normalized pitch (f0) contours of the post-stop vowels (voiced: solid lines, voiceless: dashed lines) produced by the female (left) and male (right) talkers, with 95% confidence intervals indicated by shading.

Gates

Three gates of each item were created. The first gate is the items without the first CV, i.e., /sɯbiː/. The second gate is the items without the first stop consonant, i.e., /asɯbiː/. This second fragment is *phonemically* the same across the non-word items, but *phonetically* different in the pitch patterns (i.e., post-voiceless-stop vowels have a higher pitch than post-voiced-stop ones). Thus, the second fragments beginning with voiced stops (/b, d, g/) are *phonetically* similar and vice versa. The third gate is the full non-word items without any manipulations, e.g., /tasɯbiː/, /dasɯbiː/.

To create the three gates, segmental boundaries (1) between the first stop and the following vowel, and (2) between the end of the first CV, and the onset of /s/ were annotated on Praat. For (1), the onset of voicing was identified as "an onset of periodic waveform and low-frequency voicing energy on the spectrogram" (Dmitrieva et al., 2015, p. 83). For (2), the onset of /s/ was identified as an onset of high-frequency noise on the spectrogram. These annotations were based on visual inspections and auditory judgment.

After the creation of the three gates for each item, a Praat script normalized the amplitude (Winn, 2020). The amplitude was set at 75db. To avoid unwanted effects of glottal stops that are *artificially* created due to clipping (for the second gate), a pure tone masker was added at the beginning of every stimulus, i.e., every stimulus begins with the pure tone. The pure tone was created on Praat (Duration: 300ms; Sampling frequency: 44100.0 Hz; Tone frequency: 100.0 Hz; Amplitude: 0.05 Pa).

Task and Procedure

The task was a two-alternative forced choice gating task (Grosjean, 1996). The task had six blocks. At the beginning of each block, two options written in Japanese orthography (*katakana*) appeared on the computer's screen which remained on the screen throughout the block. Participants heard a stimulus and selected one of the two options by pressing a key on a keyboard. There were three option pairs: bilabial, alveolar, and velar, by place of articulation (POA). The block of each POA pair was repeated twice (3 POA x 2 repetitions). Each block had 24 trials (3 gates x 2 voicing x 2 sex x 2 repetitions). Each voicing pair was repeated two times within the same block. The talker's sex was alternated every trial to avoid juxtapositions of the same stimuli. The presentation of each gate was randomized.

Additionally, the side of the two options on the screen was swapped between each block of the same POA pairs. For example, if /tasɯbi/ was on the right side and /dasɯbi/ on the left in one block, the other block had /tasɯbi/ on the left side and /dasɯbi/ on the right side. This is to avoid such a case where participants constantly press the same key when they cannot identify the word (especially for /sɯbi/ and /asɯbi/).

In total, there were 144 trials (6 blocks x 24 trials). There were 400ms between each trial. The participants saw a fixation mark before each trial. The task took approximately 7-10 minutes. At the beginning of the task, participants were familiarized with the task with 6 trials similar to the test trials, except that the two options (/tekoseː/ vs. /dakoseː/) differed in the first syllable vowel. The task was created in PsychoPy (version 2022.2.5). Participants saw the following instructions:

> "You will see two options in *katakana* on the screen. You will hear a part of either option. Your task is to guess which words you just heard, and then select one of the options."

Participants were tested individually in a quiet room after providing consent. Participants first completed the gating task and then a background questionnaire (Qualtrics). The whole session lasted approximately 15 minutes. The procedures were carried out in accordance with applicable laws and institutional guidelines and were approved by the relevant institutional committee(s).

RESULTS

Table 3 shows the mean accuracy scores and SD for each gate for each L1-group. Gate 1 and Gate 3 are not the test conditions but the results for these gates manifest the validity of this gating task. In Gate 1, both typical cues and pitch cues were unavailable, so the accuracy scores should be around 50% (a chance level). The mean accuracy scores (SD) for the English, Japanese, and Mandarin groups are 43.8% (13.8), 51.4% (8.3), and 44.3% (19.8). In Gate 3, both all cues were available. The mean accuracy scores (SD) for the English, Japanese, and Mandarin groups are 97.5% (3.5), 100.0% (0.0), and 97.9% (3.2). Every group was able to identify voicing categories when all cues were available.

Table 3
Mean accuracy scores (% correct identification) and SD by gates for each first language group

Gate	L1	% Correct	SD
1	English	43.8	13.8
(/sɯbiː/)	Japanese	51.4	8.3
	Mandarin	44.3	19.8
2	English	75.8	10.4
(/asɯbiː/)	Japanese	81.0	10.0
	Mandarin	84.9	12.0
3	English	97.5	3.5
(e.g., /**t**asɯbiː/ or /**d**asɯbiː/)	Japanese	100	0.0
	Mandarin	97.9	3.2

Gate 2 is the test gate. At Gate 2 (/asɯbiː/), pitch cues were available. If the % correct identification at this gate is above 75%, that indicates that listeners were able to identify voicing categories without voicing cues reliably. The mean accuracy scores (SD) for the English, Japanese, and Mandarin groups are 75.8% (10.4), 81.0% (10.0), and 84.9% (12.0).

Regarding the RQ1, the results indicate that all L1 groups were able to identify voicing categories better than chance relying only on pitch cues in the post-stop vowels. Regarding the RQ2, as the *Extension Hypothesis* predicted, the hierarchical order was Mandarin, Japanese, and English.

By Voicing Category

Performance as a function of voicing category was analyzed (for Gate 2 only). The mean accuracy scores (SD) for English, Japanese, and Mandarin groups respectively were 94.2% (5.6), 99.1% (2.8), and 94.8% (9.9), for 'voiced' trials. For voiceless trials, it was 57.5% (17.3), 63.0% (19.1), and 75.0% (18.4). Figure 3 visually reveals the hierarchy predicted by the *Extension Hypothesis*. Also, the results suggest that only Mandarin participants were able to identify both categories reliably based on pitch cues.

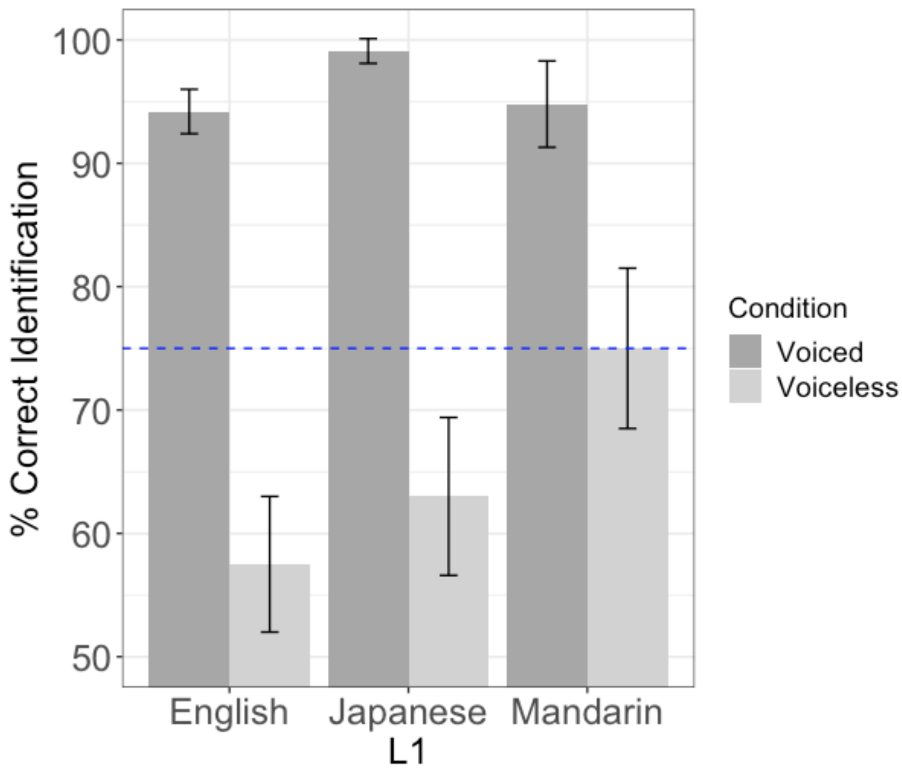

Figure 3. Mean accuracy scores and SE for each first language group at Gate 2 (/asɯbiː/) by voicing categories. A blue dashed line shows 75%. Error bars show +/- 1 SE.

DISCUSSION

To examine potentially beneficial interaction between the L1 lexical function of pitch and pitch cues for voicing contrasts, three hypotheses (Figure 1) were tested with a gating task.

The results supported the *Extension Hypothesis*; the more pitch differentiates lexical items in the L1, the more sensitive to pitch people become even in segmental contrast. The results rejected the *Pitch Saliency Hypothesis*, implying that perceptual sensitivity characterized by L1 phonological system matters to use pitch cues phonologically.

Both interpretations indicate that the lexical function of pitch in L1 matters, implying that future L2 perception models may benefit from incorporating the possibility that learners can repurpose sensitivity to cues in L1 to help them with L2 phonological acquisition. Learners may be able to harness patterns of regularities in the L2 input to perform distinctions reliably, even though this is not the same features that L1 listeners use. This creative recycling of *seemingly* redundant cues by L2 learners may be worth exploration.

The results also rejected the *Attenuation Hypothesis*, implying that tone language speakers attenuate pitch perturbation effect only in production not in perception, in harmony with Francis et al. (2006).

There is another scenario: cue-*re*weighting (Holt & Lotto, 2006). The previous interpretations focused only on how L1 predicts the degree of pitch cue usage, not on the relationship (cue-weighting) between pitch and other cues. Results imply that Mandarin learners weigh pitch cues more strongly than other cues, unlike L1 Japanese and English speakers.

The relationship between the nature of voicing contrast ('voicing' vs. 'aspiration' languages) and the L1 lexical function of pitch may interactively influence their cue-weighting strategies. L1 Japanese/English speakers can probably rely on non-pitch cues to identify voicing categories, resulting in down-weighting of pitch cues because it is redundant. Hence, they do not have to *re*weight cues. However, Mandarin speakers cannot rely on non-pitch cues (because of the merger on voicing cues) and need to *re*weight cues: down-weight non-pitch cues and up-weight pitch cues. Learned sensitivity attuned to the L1 tonal system may help such up-weighting of pitch. The results of such cue-*re*weighting might reflect their superior performance in the gating task which examined the phonological use of pitch cues, not of other cues.

As a limitation, post-stop vowel duration may have created a bias towards 'voiced', leading to higher accuracy in 'voiced' condition. Duration of post-voiced-stop vowels are longer than that of post-voiceless-stop vowels although care was taken to minimize the difference. Further research should neutralize this potential confound. Additionally, this study only investigated word-initial position. Other prosodic positions (e.g., intervocalic) should be explored to fully understand to what degree pitch cues are used in each position.

Pedagogical Implications

High accuracy score in Gate 3 imply that L1 Mandarin learners can identify voicing categories in word-initial positions where pitch differences between categories is more reliably categorical than word-medial positions (Gao & Arai, 2019), implying that instruction should focus on word-medial positions.

Additionally, the results imply that 'non-native-like' perception is fine, insomuch as learners can identify voicing categories even though the cue-weighting of 'primary' and 'redundant' cues deviates from that of NS. In other words, pathways (acoustic processing) for an identification goal do not have to be the same between NS and learners. In the spirit of the Intelligibility Principle (Levis, 2020), therefore, the goal of perception training might not necessarily have to be 'native-like' (against the Nativeness Principle). For example, in any perceptual training paradigm, the acoustic target for voicing contrasts does not always have to be on pre-voicing/aspiration but can be on pitch. Perceptual sensitivity learned in L1 should be leveraged if helpful.

ACKNOWLEDGEMENTS

We appreciate colleagues in L2 Psycholinguistics lab at Indiana University, and the audience at the 14th PSLLT conference for their valuable comments.

ABOUT THE AUTHORS

Keiji Iwamoto (kiwamoto@iu.edu) is a PhD student in the Department of Second Language Studies at Indiana University, working on L2 phonological acquisition, sound change (cue-reweighting), and pronunciation instruction.

Isabelle Darcy (idarcy@indiana.edu**)** is a professor in the Department of Second Language Studies at Indiana University (USA) and publishes on pronunciation pedagogy, second language phonology, and the bilingual mental lexicon.

REFERENCES

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer (Version 6.2.04)*. In http://www.praat.org/

Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics, 72*, 52-65, https://doi.org/10.1016/j.wocn.2018.11.002

Dmitrieva, O., Llanos, F., Shultz, A., A., & Francis, A., L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics, 49*, 77-95, https://doi.org/10.1016/j.wocn.2014.12.005.

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

Francis, A. L., Ciocca, V., Wong, V. K., & Chan, J. K. (2006). Is fundamental frequency a cue to aspiration in initial stops?. *The Journal of the Acoustical Society of America, 120*(5 Pt 1), 2884–2895. https://doi.org/10.1121/1.2346131

Gao, J., & Arai, T. (2019). Plosive (de-)voicing *f*0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics, 77*, 1-33. https://doi.org/10.1016/j.wocn.2019.100932

Grosjean, F. (1996). Gating. *Language and cognitive processes, 11*(6)*,* 597–604.

Holt, L., L. & Lotto, A., J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America 119*, 3059-3071. https://doi.org/10.1121/1.2188377

Kang, Y., & Hirayama, M. (2023). Age- and gender-based variation in the perception of voicing contrast in Tokyo Japanese. *Proceedings of the 20th International Congress of Phonetic Sciences*, Prague, Czeck Republic.

Kirby, J., P. (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics, 71.* 326-354. https://doi.org/10.1016/j.wocn.2018.09.009

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation, 6*(3), 310-328. https://doi.org/10.1075/jslp.20050.lev

Schaefer, V., & Darcy, I. (2014). Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology, 5*(4), 489–522.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America, 93*(4), 2152-2159. https://doi.org/10.1121/1.406678

Winn, M. B. (2020). Scale intensity of all sounds in a directory. *GitHub*. https://github.com/ListenLab/Praat/blob/master/Scale_intensity_all_sounds_in_folder_v1.txt

Xu, Y. (2013). *ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis*. In Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France. 7-10.