10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

# Word goodness affects the L1-dependent ability to store $F_0$ contrasts

*Y. Asano*[1], *H. Mitterer*[2]

[1]Research institute of Media and Communication, Hokkaido University
[2]Neuroscience, University of Malta
yuki.asano@imc.hokudai.ac.jp

## Abstract

The ability to store prosodic information is known to be modulated by the use of prosodic contrasts in one's first language (L1). We tested to what extent this L1-dependent ability can vary along the levels of word goodness of the stimuli (a word, a pseudoword respecting L1 phonotactics; henceforth native pseudoword, or a pseudoword violating L1 phonotactics; henceforth nonnative pseudoword). Three L1 groups; Mandarin, Japanese and German listeners participated into an online adaptive version of the Sequence Recall Task presenting sequences of between 2–9 stimuli (with high-low or low-high $F_0$ contrasts) expressed on the word goodness. Mandarin listeners showed the highest ability to store $F_0$ contrasts of words, followed by pseudowords with native and nonnative phonotactic structures, the least ability in the control segmental condition (, whose stimuli differed not in tone but in their segmental make-up). The outcome by German listeners was the opposite. Mandarin listeners seem to store $F_0$ information together with words, while Germans process $F_0$ and segments separately, and their ability to store $F_0$ contrasts was interfered with the word goodness. Japanese did not show any differences among the conditions, supporting their insensitivity to $F_0$.

**Index Terms**: Sequence Recall Task, $F_0$ contrasts, word goodness, L1

## 1. Introduction

Listeners' L1 is known to influence the ability to process and store prosodic information depending on the use of prosodic contrasts in the respective language [1, 2, 3]. For example, French speakers, as opposed to Spanish speakers, showed difficulties in storing pseudowords differing only in the location of stress [2]. Here we examine whether this L1-dependent ability to store prosodic contrasts be equally applied to stimuli differing in their word goodness. The term *word goodness* is used here to describe the level how a pseudoword sounds like a real word. To this aim, we created stimuli contrasting in $F_0$ that were varied in three word goodness conditions: an existing word (with the highest word goodness), a pseudoword respecting L1 phonotactics, and a pseudoword violating L1 phonotactics (with the lowest word goodness). We conducted an online adaptive version of a Sequence Recall Task (SRT) that had been used to investigate the effect of language on the ability to store nonnative sounds [1, 4], testing Mandarin Chinese (henceforth Mandarin), Tokyo-Japanese (henceforth Japanese) and German listeners. These languages in focus differ from each other in the linguistic levels at which $F_0$ contrasts are used, and in the manner of its use as follows: First, $F_0$ can be used at different linguistic levels. In tone languages, such as in Mandarin, each syllable is assigned to one of four tones and substituting the tone of a syllable changes the meaning of the word or morpheme. A pitch-accent language, such as Japanese, exhibits the limited use of $F_0$ to distinguish words in a manner similar to stress [5]. Finally, in an intonational language, such as in German, $F_0$ is primarily used at the intonational level without primarily bearing lexical $F_0$. Second, the richness of the variety of $F_0$ contrast also changes across languages. German exhibits a richer variety [6] than Japanese, because the limited $F_0$ pattern of a $F_0$ accent restricts its intonational patterns. These different linguistic levels at which $F_0$ contrasts are used and its different manner of use influence the sensitivity to $F_0$ contrasts and learning of words contrasting in $F_0$ showing an advantage for L1 listeners of a tone language over others [2, 7]. At the same time, L1 listeners of a tone language are less sensitive to $F_0$ at the intonational level as lexical and intonational $F_0$ compete with each other [8]. No study up to date investigated storing of $F_0$ contrasts differing in the word goodness.

The word goodness and $F_0$ processing is worth investigating to study whether $F_0$ processing in native, nonnative speech and non-speech (such as in music without vocal information) is on a continuum and to what extent the L1-dependent ability to store prosodic contrasts can be transferred on this continuum toward non-speech. This is because $F_0$ is used to describe sounds both in music and language [9], and is indispensable in all languages, although its contrast is used at different linguistic levels and in different manners across languages as described before.

Memory capacity of $F_0$ contrasts in linguistic and non-linguistic $F_0$ (the latter produced using the speech synthesis program MBOLA by defining phonemes with the respective language packages, durations and $F_0$ places) has been studied in [7] adopting the same experimental paradigm testing the same language groups. The study showed that the L1-dependent ability to store linguistic $F_0$ contrasts could fully be applied to non-linguistic $F_0$ contrasts. We examine whether this is also true for native or nonnative pseudowords. If the word goodness is advantageous in storing $F_0$ contrasts, the largest memory capacity in the word condition will be found followed by that in the native pseudoword condition, finally by that in the nonnative pseudoword condition. If the word goodness of the stimuli compete with storing process of $F_0$ contrasts, the largest memory capacity in the nonnative pseudoword condition will be shown followed by the native pseudoword condition, finally by the existing word condition. The facility or impedance in storing $F_0$ can occur through the activation of semantic meanings of the stimuli or of intonational meanings. The comparison between a tone language (Mandarin) and an intonational language (German) enables us to examine whether $F_0$ is primarily perceived at the lexical or at the intonational sentence level in our experimental paradigm and how this influence the memory capacity measured. The comparison between Mandarin and Japanese enables us to examine differences between a tone and an accentual language.


10.21437/SpeechProsody.2020-96

# 2. Experiment

## 2.1. Methods

### 2.1.1. Participants

Twenty-four native listeners of Mandarin listeners (f = 12, aged between 20–32, mean = 26.4), of Standard German (f = 14, aged between 18–26, mean = 22.6) and, of Japanese (f = 14, aged between 20–36, mean = 27.1) took part for a small fee. The participants were identical with those tested in [7]. All participants were recruited in Tübingen. None of them was L1 bilingual nor had learnt any tone or pitch-accent language as a second language (henceforth L2).

### 2.1.2. Materials

For the word condition, minimally-paired existing words that contrasted in $F_0$ were selected in Chinese ( *ai*-High/Low, henceforth HL, meaning love and *ai*-Low/High, henceforth LH, meaning cancer) and in Japanese ( *ame*-HL meaning rain and *ame*-LH meaning candy). Since a $F_0$ contrast alone in German does not distinguish words, a disyllabic word ( *Imbiss* meaning snack) was selected and varied in its $F_0$ (HL vs. LH). For the native pseudoword condition, a pseudoword *numi* was selected that follow the phonotactic structures in all three languages in focus. For the nonnative pseudoword condition, a pseudoword *bsbli* was selected that violates the phonotactic structures of all three languages in focus. For the segmental condition, a minimal pair of non-words involving a segmental contrast (*muku* vs. *munu*) was constructed. The stimuli were produced using the speech synthesis program MBOLA by defining phonemes using the respective language packages, durations and $F_0$ places.

Pitch values in the word, native and nonnative pseudoword conditions were first defined as H=280 Hz, L=150 Hz for HL stimuli and L=170 Hz, H=280 Hz for LH stimuli, as words with these values sounded most natural in the linguistic $F_0$ condition. Then, these values were varied by multiplying them by .95, .97, .99, 1.00, 1.01, 1.03 and 1.05 resulting in seven tokens with slightly different $F_0$ values. $F_0$ values in the segment condition were consistently 200 Hz, resulting in a flat $F_0$ contour. The duration of each stimulus was 500 ms in all conditions.

### 2.1.3. Procedures

An online adaptive version of SRT presented sequences of 2–9 stimuli (e.g. a sequence of HL and LH). The task was to reproduce each sequence by typing the associated keys "1" and "2" in the correct order. The experiment was presented with MatLab on a desktop computer. The language of the experiment was English for all language groups. Each condition started with a training session, in which they were trained to associate one of the members with key "1" and the other member with key "2". Then, they were invited to listen to a single, randomly presented token from each set by pressing either key as often as they wished. After they indicated having learnt this two-way classification, participants moved on to an identification task. At this stage, they heard two stimuli in succession and were asked to respond by pressing "1" or "2", after which either "correct!" or "incorrect!" was displayed on their screen. This procedure was repeated until three correct sequential responses had been given. After passing this training session, participants entered the experimental session. In each language group, half of the participants were first tested with the segmental contrast. The order of the other three $F_0$ conditions was counterbalanced. The number of the stimuli varied between 2 and 9 according to the participant's previous response; if the answer was correct, the next trial contained N+1 of stimuli, if incorrect, N−1. The maximal number of stimuli from the same member in a sequence was three.

## 2.2. Results

The mean achieved length of the last three stimuli was extracted for each condition and participant to analyse participants' language and condition-dependent memory capacity of the respective contrast. In the following analysis, statistical results from LMER models are first reported and then descriptive mean values and 95% CI error bars in plots. The full LMER-model with mean achieved length as a dependent measure, language group and condition as fixed factors and participant as a random factor showed significant interactions (overall $p < .001$). The between-condition analyses showed that, in the segment condition, all language groups did not differ from each other ($\beta$ = -.24, *SE*= .47, *t*= -.5, *p*= .9 between Mandarin and Japanese, $\beta$= .44, *SE*= .46, *t*= 1.0, *p*= .6 between Mandarin and Germans, $\beta$= -.67, *SE*= .47, *t*= -1.5, *p*= .3 between Japanese and Germans). In the word condition, Mandarin listeners outperformed Japanese, followed by Germans ($\beta$= 2.73, *SE*= .46, *t*= 6.0, $p < .001$ between Mandarin and Germans, $\beta$= 1.66, *SE*= .46, *t*= 3.6, $p < .01$ between Mandarin and Japanese, $\beta$= 1.10, *SE*= .46, *t*= 2.4, *p*= .05 between Japanese and Germans). Also in the native pseudoword condition, Mandarin and Japanese did not differ from each other ($\beta$ = .90, *SE* = .46, *t* = 1.9, *p* = .1). Mandarin listeners showed larger memory capacity than Germans ($\beta$ = 2.0, *SE* = .46, *t* = 4.3, $p < .001$). Japanese performance was slightly better than that of German listeners ($\beta$ = 1.01, *SE* = .46, *t* = 2.3, *p* = .06, approaching significance). In the nonnative pseudoword condition, only Mandarin and German listeners differed from each other ($\beta$ = 1.38, *SE* = .46, *t* = 3.0, $p < .01$, but $\beta$ = .69, *SE* = .47, *t*= 1.5, *p*= .3 between Mandarin and Japanese, $\beta$ = -.69, *SE* = .47, *t* = -1.5, *p* = .3 between Japanese and Germans).

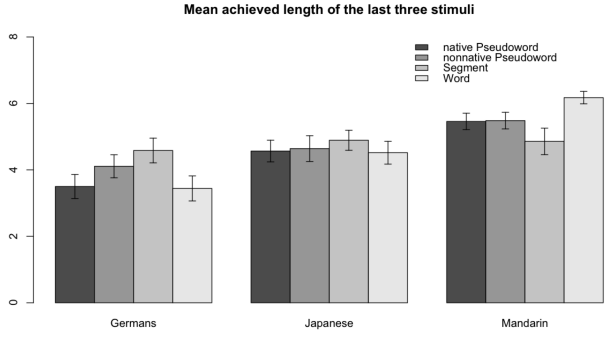


Figure 1: *Mean achieved lengths of the last three stimuli and 95% CI bars for each condition and L1..*

In the between-language analyses, Mandarin listeners showed larger memory capacity in the word condition than in the segmental condition ( $\beta$ = 1.15, *SE* = .33, *t* = 3.5, $p < .01$), while other conditions did not differ from each other. German listeners showed the opposite tendency showing larger memory capacity in the segment condition than in the word condition ( $\beta$ = 1.14, *SE* = .32, *t* = 3.5, $p < .01$), and than

in the nonnative pseudoword condition ( $\beta$ = 1.08, *SE* = .32, $t$ = 3.3, $p$ <.01). Other conditions did not differ from each other. Japanese listeners did not show any difference between the conditions.

## 3. General Discussion

Our study investigated whether the L1-dependent memory capacity to store $F_0$ contrasts can be equally transferred to $F_0$ contrasts of the stimuli differing in the word goodness.

Mandarin listeners generally showed larger memory capacity in storing $F_0$ contrasts regardless of the word goodness than Japanese. The least memory capacity was shown by German listeners. The use of lexical $F_0$ contrasts alone, regardless of a tone or a pitch-accent language, was more advantageous than exhibiting $F_0$ contrasts at the intonational level to store $F_0$ contrasts of disyllabic stimuli. At the same time, the use of $F_0$ contrasts in a tone language was more advantageous than in an accentual language, so that larger memory capacity for Mandarin listeners was found than for Japanese. The richness of $F_0$ patterns at the intonational level was not advantageous in this study, so that Germans did not outperform Mandarin nor Japanese. Besides the three $F_0$ conditions, the segmental condition perfectly served as a control while showing no differences between the language groups. Mandarin listeners seem to be sensitive to $F_0$ in general thanks to their L1 experience [7, 2, 10, 11, 12].

The patterns of memory capacities shown in each group along the word goodness were different: Mandarin listeners showed the largest memory capacity for existing words, followed by that for the both pseudoword conditions, which did not differ from each other. This suggests that $F_0$ is an essential lexical entity that is processed together with segmental information. The lexical advantage shown in this group suggests that, in the word condition, they accessed these minimally paired stimulus words and activated both of them. In this sense, the memory tested in this condition can be also regarded as semantic memory (love vs. cancer). Further, the violence of the native phonotactic structures did not impede their $F_0$ processing, meaning that it just mattered whether the stimuli were words or pseudowords. They were still able to concentrate on $F_0$ contrasts without being violated by the nonnative phonotactic structures. Their general sensitivity to $F_0$ may be the reason for this result.

On the contrary, our German listeners showed that the lower the word goodness was, the larger became memory capacity, suggesting that segmental processing outweighed $F_0$ processing, because the former impeded the latter. This claim further supports the finding that their memory capacity in the segmental condition was higher than in other conditions. Why was their $F_0$ memory impeded by the segmental processing? We propose that the German listeners may have activated more intonational meanings while activating the stimuli of higher word goodness and this made the $F_0$ storing process more difficult.

How about the Japanese listeners? Interestingly, they showed any significant differences between neither conditions. The results indicate that, differently from our Mandarin and German listeners, they seem not to have accessed semantic meanings either intonational meanings. One of the reasons why they might have not accessed semantic meanings surely lies in the experimental paradigm. They had to memorise $F_0$ contrasts under time pressure without being forced to distinguish semantic meanings. Japanese listeners had been shown to be slower and less accurate in activating minimally paired words contrasting in $F_0$ than Mandarin listeners [13].

The Limiting-Domain Hypothesis [14] claims that the syllable is a unit of processing in lexical access. According to this hypothesis, $F_0$ of a tone language can effectively be used for lexical access while $F_0$ of an accent language cannot be, although both are perturbations of $F_0$ and acoustically identical. In the latter case, $F_0$ information is completed too late for spoken word recognition.

Further, our Japanese listeners did not even show any impedance by the violence of native phonotactic structures. The comparison between the German and Japanese listeners' performance again supports the claim that German listeners may have activated intonational meaning in processing $F_0$ of the stimuli with higher word goodness, while Japanese did not. In summary, we claim that the data indicate the following: Mandarin may have activated lexical meanings, while Germans intonational ones, but Japanese remained inactivated in neither meanings.

Despite the fact that both Japanese and Mandarin language exhibit lexical $F_0$, these two groups seem to encode $F_0$ contrasts differently because the underlying mental representations of a Japanese pitch accent and Mandarin tones may be very different [13]. It was proposed that $F_0$ information itself is not stored with a word in the Japanese mental lexicon, being instead stored solely as an accent position without $F_0$ information, whereas in the Chinese mental lexicon, $F_0$ movement itself is represented together with words. In other words, Japanese listeners only (have to) pay attention to the position of a fall, but not to the whole contours. This propose also explains why a great range of dialectal varieties exists in Japanese and why Japanese dialects exhibit opposite pitch contrast patterns to each other [15, 16]. The fact that Japanese listeners only have to pay attention to a position of an pitch accent, but not $F_0$ itself leads to the insensitivity to $F_0$ frequently reported in the study comparing Japanese and Mandarin listeners [2, 13]. Also, the previous experiments investigating the question whether $F_0$ is used for spoken-word recognition in Mandarin and Japanese show that Mandarin listeners clearly rely on $F_0$ information [10, 11, 12], while Japanese listeners showed controversial results [17, 18, 14, 19, 20]. Therefore, the existence of lexical $F_0$ contrasts in one's lexicon alone does not seem to necessarily create listeners' higher sensitivity to that contrast, but the crucial aspect is how this contrast is used in spoken word recognition.

Whereas the German listeners were better than Japanese at learning non-words with Chinese $F_0$ contrasts in [2], our German participants did not show larger memory capacity than our Japanese did. This may lie in our experimental conditions. First, the $F_0$ contours used in this experiment resembled existing Japanese $F_0$ patterns (HL vs. LH), whereas in [2], Japanese participants had to learn nonnative tonal patterns (e.g., fall-rise-fall; tone 3). Second, the mental mechanism of learning new words and contours associated with pictures in [2] seems to be different to that of memorizing orders of limited patterns of sequences. To examine this assumption, future research should examine the memory capacity of $F_0$ patterns that do not exist in Japanese lexicon (Tone 3 in Mandarin). The authors in [2] claimed that the rich intonational structure of German was more advantageous in learning non-native $F_0$ contrasts than the restricted lexical $F_0$ contrasts of Japanese and the fewer utterance-level $F_0$ contrasts available in Japanese to explain why higher ability to learn words differing in $F_0$ was found for Germans than for Japanese. In this study however, we

argue that the activation of intonational meaning impeded the storage of the stimuli.

[8] showed that $F_0$ is primarily perceived at the lexical level for listeners of tone languages but at the intonational sentence level for listeners of non-tone languages. They also suggested that lexical tone is the primary factor in reducing the sensitivity to $F_0$ cues at the sentence level for listeners of a tone language. In our experiment, the activation of intonational meaning was disadvantageous, at least while memorising stimuli presented without contexts (because it is difficult to speedy process and store whether the stimulus was e.g., "question" or "statement").

We aimed at investigating the influence of word goodness on storing $F_0$ contrasts in order to examine whether native, nonnative and non-speech are on a continuum in its processing or storage. While [7] did not show any difference in the memory capacity between the linguistic $F_0$ condition (corresponding to our word condition) and non-linguistic $F_0$ one, our study shows difference between the three linguistic $F_0$ conditions in the Mandarin and German listener groups. Comparing the outcome in [7] and our study, Mandarin listeners showed less difficulties in storing $F_0$ contrasts without linguistic information than $F_0$ contrasts of pseudowords. This means that the aforementioned impedance through the word goodness was not applied to the non-linguistic $F_0$ processing. For the German outcome, the same was true that the activation of intonational meaning impeded the processing of non-linguistic $F_0$ contrasts, but less was so for the pseudowords.

With respect to storing $F_0$ contrasts of disyllabic stimuli, our results together with [7] show that pitch processing in native, nonnative speech and processing of non-linguistic pitch may not be on a continuum. The word goodness mattered differently according to how $F_0$ is used in listeners' L1. For Mandarin listeners, the word goodness mattered in a sense that higher word goodness was more advantageous in storing $F_0$ contrasts, because it aided lexical access. Although the results in linguistic and non-linguistic pitch condition were at the same level, the underlying mechanism may be different. In the non-linguistic pitch condition, they sorely relied on their general sensitivity to $F_0$ that had been developed through their L1 experience with $F_0$. Prior experience with a tone system is known to be advantageous in differentiating $F_0$ not only in language, but also in music [21, 22, 23]. The application of their general sensitivity to $F_0$ was also the case in both pseudoword conditions. However, they somehow had to process segmental information at the same time, so their memory capacity decreased compared to that in the word and non-linguistic pitch conditions. For German listeners, the word goodness also mattered in sense that higher word goodness accelerated the activation of intonational meanings, which did not aid them to store $F_0$ contrasts effectively, but rather impeded it. For this group, the non-linguistic pitch processing in a sense of storing disyllabic stimuli contrasting in $F_0$, does not appear definable on a continuum from native, nonnative speech to non-speech. Assumably, the underlying mechanism in the linguistic and non-linguistic pitch storage in this group may be different, but this proposal cannot be further examined in this study, for which future research is needed. For Japanese listeners, the word goodness did not influence their performance, suggesting that they may have separated segmental and $F_0$ processing from each other, and processing $F_0$ at a more abstract level, regardless of $F_0$ of speech or non-speech.

# 4. Acknowledgements

We thank the English Department at the University of Tübingen for the budget and sources provided for testing participants.

# 5. References


[1] E. Dupoux, S. Peperkamp, and N. Sebastián-Gallés, "A robust method to study stress "deafness". A robust method to study stress "deafness"," *Journal of Acoustical Society of America*, vol. 110, no. 3, pp. 1606–1618, 2001.

[2] B. Braun, T. Galts, and B. Kabak, "Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation," *Second Language Research*, vol. 30, no. 3, pp. 323–350, 2014.

[3] Y. Asano, "Localising foreign accents in speech perception, storage and production," Ph.D. dissertation, University of Konstanz, Konstanz, Germany, 2016.

[4] S. Peperkamp and E. Dupoux, "A typological study of stress 'deafness'," *Laboratory Phonology*, vol. 7, pp. 203–240, 2002.

[5] L. M. Hyman, "How (not) to do phonological typology: The case of pitch-accent," *Berkeley Phonology Lab Annual Report*, pp. 654–685, 2007.

[6] S. Baumann, M. Grice, and R. Benzmüller, "GToBI – a phonological system for the transcription of German intonation," in *Prosody 2000: Speech recognition and synthesis*, ser. 21–28, S. Puppel and G. Demenko, Eds. Adam Mickiewicz University, 2001.

[7] Y. Asano and H. Mitterer, "Storing linguistic and non-linguistic pitch contrasts," in *Proceedings of the 14th conference of Phonetik und Phonologie des deutschsprachigen Raums*, Wien, Austria, 2019.

[8] J. Liang and V. J. van Heuven, "Chinese tone and intonation perceived by L1 and L2 listeners," in *Tone and Tunes: Experimental studies in word and sentence prosody*, C. Gussenhoven and T. Riad, Eds. Berlin & New York: Mouton De Gruyter, 2007, vol. 2, pp. 27–61.

[9] L. Jäncke, "The relationship between music and language," in *Frontiers in Psychology*, vol. 3, no. 123, doi: 10.3389/fpsyg.2012.00123, 2012.

[10] C.-Y. Lee, "Does horse activate mother? processing lexical tone in form priming," *Language and Speech*, vol. 50, no. 1, pp. 101–123, 2007.

[11] X. Li, Y. Yang, and P. Hagoort, "Pitch accent and lexical tone processing in chinese discourse comprehension: an erp study." *Brain Research*, vol. 1222, pp. 192–200, 2008.

[12] N. Poss, T.-H. Hung, and U. Will, "The effects of tonal information on lexical activation in Mandarin," in *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*, K. C. Marjorie and H. Kang, Eds., vol. 1. Columbus, Ohio: The Ohio State University, 2008, pp. 205–211.

[13] Y. Asano and S. Yuen, "Asymmetric representations of lexical pitch," in *Abstraction, Diversity and Speech Dynamics*, 2017.

[14] D. L. Walsh, "Limiting-domains in lexical access: Processing of lexical prosody," in *University of Massachusetts Occasional Papers in Linguistics 19: Linguistics in the Laboratory*, M. Dickey and S. Tunstall, Eds. Amherst: GLSA, 1993, pp. 133–155.

[15] B. Frellesvig, *A history of the Japanese language*. Cambridge University Press, 2010.

[16] T. J. Vance, *An Introduction to Japanese Phonology*. State University of New York Press, 1987.

[17] T. Otake, G. Hatano, A. Cutler, and J. Mehler, "Mora or syllable? Speech segmentation in Japanese," *Journal of Memory and Language*, vol. 32, no. 2, pp. 258–278, 4 1993.

[18] K. Tamaoka, N. Saito, S. Kiyama, K. Timmer, and R. G. Verdonschot, "Is pitch accent necessary for comprehension by native Japanese speakers? –an ERP investigation," *Journal of Neurolinguistics*, vol. 27, no. 1, pp. 31–40, 2014.

[19] A. Cutler and T. Otake, "Pitch accent in spoken-word recognition in Japanese," *Acoustical Society of America*, vol. 105, no. 3, pp. 1877–1888, 1999.

[20] N. Minematsu and K. Hirose, "Role of prosodic features in the role of prosodic features in the human process of perceiving spoken words and sentences in Japanese," *J. Acoust. Soc. Jpn.*, vol. 16, pp. 311–320, 1995.

[21]

[22] A. Krishnan, J. T. Gandour, and G. Bidelman, "The effects of tone language experience on pitch processing in the brainstem."

[23] R. J. Giuliano, P. Q. Pfordresher, E. M. Stanley, S. Narayana, and N. Y. Y. Wicha, "Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change," *Frontiers in Psychology*, vol. 2, no. 146, 2011.