

# Stability in perceiving non-native segmental length contrasts

Yuki Asano

Department of Linguistics, University of Konstanz, Konstanz, Germany

yuki.asano@uni-konstanz.de

## Abstract

Previous studies have demonstrated that listeners show high sensitivity in discriminating non-native segmental length contrasts thanks to auditory memory [1, 2]. We tested the limits of discriminating Japanese consonantal length contrasts with three groups of listeners (German learners of Japanese, German non-learners and Japanese natives) under increasing task demands. We increased memory load through a longer inter-stimulus interval (= ISI) (2500ms vs. 300ms) and added psycho-phonetic complexity (trials with task-irrelevant pitch falls that occurred simultaneously with the consonant vs. with monotonous pitch).

Results showed very good discrimination in all groups when the task demands were the lowest. With increasing task demands, only the non-natives' discrimination abilities decreased: non-learners were strongly affected by both ISI and pitch, while learners only by pitch. The psycho-phonetic complexity of the stimuli had a stronger impact on performance than the increased memory load.

Our findings suggest that L2 learners can establish novel phonological representations, but the ability to use them can be applied only under favourable listening conditions with no distracting acoustic information. The non-native listeners' reduced sensitivity under increasing task demands appears to be the reason why even advanced learners still face difficulties in natural learning situations.

**Index Terms:** task demands, discrimination, non-native length contrast, Japanese, German, geminate

## 1. Introduction

The discrimination of non-native length contrasts is expected to be difficult for L2 listeners when such contrasts do not exist in an L2 lexicon (theoretical supports for segmental contrasts see e.g. [3, 4]). Indeed, better performances have been oft found in L1 listeners rather than in L2 listeners [5, 1, 6]. At the same time however, despite this prediction, even non-native listeners have occasionally been reported to discriminate non-native prosodic contrasts fairly better than expected, given the lacking L2 phonological category in a listeners' L1. Hayes and Masuda [1] and Wilson *et al.* [2] report that even non-learners who have no consonantal length contrasts in their L1 could discriminate them without any exposure to the L2, by simply relying on auditory memory [7] and absolute durations [8, 2]. We postulate that these findings do not contradict each other, but do relate to task demands.

Under challenging situations, speech perception becomes more demanding. This is because the listener's cognitive load, operationalized as attention control (efficient attention shift among foregrounding and backgrounding of task-relevant and -irrelevant information [9]) and memory load becomes higher in such situations (e.g. [10, 11]). That appears to be all the

more true when it comes to L2 perception. Previous studies show that the difficulty of listening to speech in noise or with greater talker variability is more exacerbated in L2 perception when compared to L1 perception [12, 13, 14, 15]. Given that the cognitive resources in L2 speech processing are fewer due to reduced L2 proficiency in comparison to native speakers [12], L2 perception is expected to be more "vulnerable" under such demanding listening conditions.

Werker and Logan [16] for instance show a performance decrease in the L2 listeners' discrimination of non-native segmental contrasts once an ISI becomes longer. ISI is known to increase task demands affecting the memory load (the longer the higher) and involving different levels of speech processing (the longer the more higher phonological perception) [17, 18, 19, 7, 20, 16]. Their findings suggest that L2 listeners had difficulties in discriminating non-native segmental contrasts once memory load increased and the processing tapped into the phonological one. We will investigate non-native listeners' limitations in discriminating non-native segmental *length* contrasts (and not segmental contrasts) under increasing task demands by increasing ISI [16] and by adding another dimension of task demands, psycho-phonetic complexity relating to attention control.

A discrimination task is a suitable method to investigate said issue, because it appears to require low task demands under a certain condition (see the aforementioned fairly good non-native performance) and it is therefore easy to enhance task demands in other conditions. In the current study, we conducted a speeded same-different-task to discriminate consonantal length contrasts testing Japanese natives, German learners and non-learners. Consonantal length is lexically contrastive in Japanese, but not in German. We expect therefore that the discrimination of consonantal length contrasts will be exacerbated to a greater extent only for German learners and even more for German non-learners under increasing task demands. Increasing task demands in our study are defined along two dimensions: attention control through psycho-phonetic complexity of the stimuli (e.g. [9, 21]) and memory load (e.g. [22]), the capacity to hold memory for a limited period of time.

psycho-phonetic complexity was increased by adding a task-irrelevant pitch movement. We built two stimuli conditions, one with pitch falls that occurred simultaneously with the consonant and the other with flat (namely monotonous) pitch. Attention control in the former condition is expected to be higher, because ignoring the task-irrelevant pitch movement is required in this condition. In case listeners cannot ignore it, they need to process both pitch and length simultaneously in the consonantal length contrast. In the flat pitch condition, listeners have one prosodic cue less to process (only duration). In case of an unsuccessful attention control, the trials with a falling pitch are expected to become psycho-phonetically more complex and more difficult for the discrimination of consonantal length contrasts than those with a flat pitch.

Previous studies consistently report that the complexity of speech perception in a bad speech quality or in unfavourable circumstances (e.g. in noise) impairs successful speech perception [10, 11, 23] and that is more the case for L2 perception [12, 13, 14]. It may be assumed that L2 listeners' are less successful in ignoring or shutting down the task-irrelevant information, e.g. noise in background. In our study we therefore hypothesize that the psycho-phonetic complexity of the stimuli constitutes an impediment to a greater extent for L2 listeners than for L1 listeners, because the former are less likely to ignore the task-irrelevant pitch movement. Moreover, this additional falling pitch movement mirrors the phonetic form of a Japanese lexical falling pitch accent. The study will therefore contribute to understanding the L2 perception of a language with lexical tone or pitch accent.

The other dimension, memory load, was manipulated by increasing the duration of ISI (in one condition 300ms and in another condition 2500ms). In order to eliminate the risk of backward masking, ISI should be longer than 250ms [17, 24, 25]. Around 250ms after the offset of a sound, information is recognized at the sensory level, but is not yet identified or categorized [26]. The discrimination ability increases rapidly between 100 and 500ms and falls gradually as the ISI increases further [17, 7, 20]. The decrease after 500ms may be interpreted as the effect of gradually decaying auditory information in short-term memory. In order to test the acoustic comparison of successive stimuli without risking a backwards masking effect, we decided to use 300ms in one condition (slightly longer than 250ms) to ensure the acoustic trace was available. Then, it is claimed that this uncategorized acoustic information is maintained for a while (approximately 2000ms) in the precategorical acoustic storage [26] or in the short-term phonological storage as an auditory image [27, 28]. To keep our experiment to a duration that would not cause the participants to diminish their concentration or motivation, but at the same time to make sure that the processing taps into the categorical level after 2000ms [27, 28, 26], we decided to use 2500ms as the long ISI.

At the segmental level, numerous studies have investigated the effect of the durations of ISIs on the discrimination of non-native *segmental* contrasts. With a short ISI with which researchers hypothesize that the lexical knowledge or phonological category is not consulted, no difference between L1 and L2 listeners was found. When the ISI became long, however, the L1 listeners' performance surpassed the L2 listeners' one. No study has been carried out so far that explores the discrimination of non-native *segmental length* contrasts by varying ISIs. We assume that segmental length contrasts are perceived relatively in a larger unit in relation to the adjacent words than segmental contrasts. Therefore, it might be more difficult to discriminate two stimuli with a segmental length contrast by directly relying only on auditory memory.

## 2. Experiment

### 2.1. Methods

#### 2.1.1. Participants

Twenty-four native Japanese participants (henceforth= JNs, 10 male, 20-31 years), 24 native German who were not learning Japanese (= non-learners, henceforth= GNs, 8 male, 19-30 years) and 48 German learners of Japanese (henceforth= GLs, 30 male, 20-34 years) took part for a small fee. They were unaware of the purpose of the experiment. None of the learners had prior training in Japanese phonology.

#### 2.1.2. Materials

We used disyllabic triplets that differed segmentally only in the length of the first vowel or in the length of the second consonant (e.g., [pʊnʊ], [pʊ:nʊ], [pʊn:ʊ]). Twenty-one triplets were evaluated in a pretest with Japanese and German native listeners (different from those of the main experiment) to select only stimuli that did not activate a word via phonological analogy. Participants were presented with one stimulus at a time and were required to write down the first word coming to their mind. We analysed the responses of 24 Japanese natives and 24 German natives separately. From there, six non-word triplets with the lowest association strength in both groups were selected (word association rate between 29.8 % and 45.3%, mean = 34.5%, while the one of all 21 triplets ranged between 29.8% and 100.0%, mean = 52.3%). The selected stimuli differed in manner of articulation and voicing of the medial consonant (= phon), *punu*, *gunu*, *gupu*, *gubu*, *zusu*, *sifu*. The materials were recorded by a female speaker of Japanese in two pitch conditions; high flat pitch and falling pitch (with a pitch fall during the medial consonant pitch tracks, see Figure 1). Each stimulus was recorded six times in order to have different tokens of the same type (N = 216). To ensure the same pitch between the two stimuli presented together (= A and X), the pitch was manipulated by using a method which is based on the representation of  $F_0$  contours with B-splines [29] and on a smooth warping of the time axis allowing us to move selected time boundaries to desired positions (see [30]). More specifically 6 tokens of each triplets of a given phon (thus N = 18 for each phon) realized in the same pitch pattern were aligned on the average pitch across tokens (in the flat pitch: average = 1.3 semitones, range = 1.0 – 1.6 semitones; in the falling pitch: average = 13.0 semitones, range = 10.5 – 16.4 semitones). Finally, a female native speaker of Japanese and a male native speaker of German selected the most naturally sounding tokens as experimental items (for each phon, N = 3 for the stimuli with a long vowel or a long consonant, N = 4 for the stimuli with short vowels and consonants). There was no disagreement on the decisions.

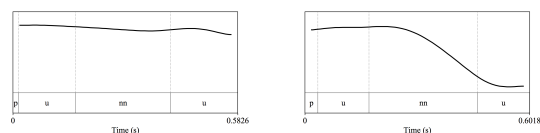


Figure 1: Pitch track of geminate stimuli in the flat and falling pitch condition.  $F_0$  range is shown between 100 and 350 Hz.

To verify the durational differences of the selected stimuli, mean durations from the three different tokens of the respective long and short consonants and vowels were analyzed. A linear mixed effects regression model with the critical vowel or consonant *duration* as dependent measure and *pitch* (flat vs. falling), *segmental condition* (short vs. long vowel or consonant) as fixed factors and *phon* as a random factor including random slopes for the fixed factors [31, 32] showed a significant interaction for vowel contrasts ( $p < 0.01$ ), but not for consonantal contrasts ( $p > 0.1$ ). Then we analyzed the durations in the flat and falling pitch conditions separately for vocalic contrasts. Results of paired t-tests showed that, on average, long vowels in the flat pitch condition were 3.3 times longer than short vowels ( $t(5) = 20.0$ ,  $p < 0.001$ ) and those in the falling pitch condition were 3.0 times longer ( $t(5) = 28.1$ ,  $p < 0.001$ ). For consonant contrasts, we analyzed the durations in the flat

and falling pitch conditions together. Results of paired t-tests showed that geminates were on average 3.2 times longer than singleton consonants in the flat pitch condition ( $t(10) = 25$ ,  $p < 0.001$ ). These duration measurements ensured that the acoustic criteria for the length distinction in vowels and consonants were met (ratio for Japanese vowels was approximately 3.2:1, [33] and for consonants 3.2:1 [34, 35]).

To compare the spectral quality for long and short vowels /u:/ and /u/, the first and second formants at the midpoint of the vowel were automatically extracted. A linear mixed effects regression model with *formant* as dependent measure and *pitch* (flat vs. falling), *formant condition* (F1 vs. F2) as fixed factors and *phon* as a random factor including random slopes for the fixed factors [31, 32] showed that an interaction approached significance ( $p = 0.07$ ). Results of paired t-tests separately in the two pitch conditions showed no difference in F1 nor in F2 ( $p$ -values were overall  $> 0.15$ ). On average, F1 was 501.3 Hz for long vowels and 484.8 Hz for short vowels in the flat pitch condition and 467.9 Hz for long vowels and 429.9 Hz for short vowels in the falling pitch condition. F2 was 1499.0 Hz for long vowels and 1530.8 Hz for short vowels in the flat pitch condition and 1510.0 Hz for long vowels and 1495.9 Hz for short vowels in the falling pitch condition.

### 2.1.3. Procedures

A speeded AX-task was used to test the subjects' sensitivity for consonantal length contrasts. One base list was assembled by presenting all possible pairings of the stimuli ( $N = 84$ ). The short ISI was 300ms ISI (session 1) and the long one 2500ms ISI (session 2). In both sessions, the intertrial-interval was 1000ms. Each trial began with a beep of 44100 Hz (500ms). No feedback was provided during the experiment. Each session began with 10 training trials using the phones that were not used as the experimental ones (*guna* and *puna*) followed by a pause before the experimental session started. The test lasted approximately 20 minutes. The order of presentation was automatically randomized using *Presentation* (Neurobehavioral Systems).

## 2.2. Results

### 2.2.1. Sensitivity to contrast: $d'$ scores analyses

Participants' sensitivity to the contrasts was calculated using  $d'$  [36]. We calculated  $d'$  scores for each participant for each of the consonantal and vocalic length contrasts, flat and falling pitch as well as short and long ISI. We normalized them by subtracting those for vocalic length contrasts (as baseline) from those for consonantal length contrasts (note that there were no differences in the  $d'$  scores for vocalic length contrasts across *language groups*, overall  $p > 0.2$ ). The value of 0 in the plot means that the  $d'$  for consonantal and vocalic length contrasts were almost the same. A linear mixed effects regression model with  $d'$  scores as dependent measure and *language groups*, *pitch* (flat vs. falling), *ISI* (short vs. long) as fixed factors and *participants* as a random factor including random slopes for the fixed factors [31, 32] showed a significant three-way interaction ( $p < 0.003$ ). To investigate the nature of this interaction, the data were split by *pitch*. In the flat pitch condition, an interaction was found between *language group* and *ISI* (GNs'  $d'$  scores decreased in the long ISI condition, but the  $d'$  scores of the other two groups did not,  $p < 0.03$ ). JNs' and GLs'  $d'$  scores did not differ from each other. In the falling pitch condition, a main effect of *language groups* was found (the JNs'  $d'$  scores higher than the GLs' ones,  $p < 0.01$ , the GLs' ones higher than the

GNs' ones,  $p < 0.01$ ); see Figure 2. In order to compare the  $d'$  scores between the two plots, we analyzed the  $d'$  scores for both pitch conditions in each *language groups*. JNs'  $d'$  scores were not affected by *pitch* or *ISI* (overall  $p > 0.7$ ). In the GLs' data, the effect of *pitch* approached significance (flat  $>$  falling,  $p = 0.09$ ). The GNs' data showed a significant interaction between *pitch* and *ISI* ( $d'$  scores only in the flat pitch condition decreased in the long ISI condition,  $p < 0.03$ ).

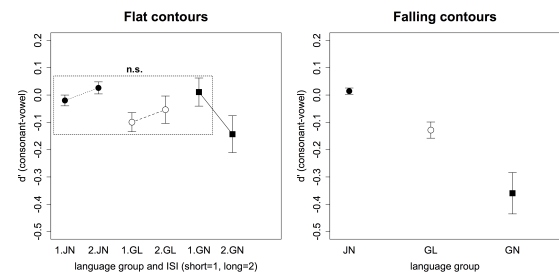


Figure 2: Results of the normalized  $d'$  scores

### 2.2.2. Processing difficulty: reaction time analyses

The reaction time (RT) analyses were performed to investigate task difficulty for the discrimination of *length conditions* (vowel vs. consonant) [37, 38]. We only analysed RTs in trials with different pairs. To account for participant-specific RT-differences, we normalized the raw RT data in the following way: We discarded RTs longer than 2000ms and aggregated the data for each *participant*, *pitch*, *length condition* and *ISI*. Then, the averaged RTs for vocalic length contrasts were subtracted from those for consonantal length contrasts.

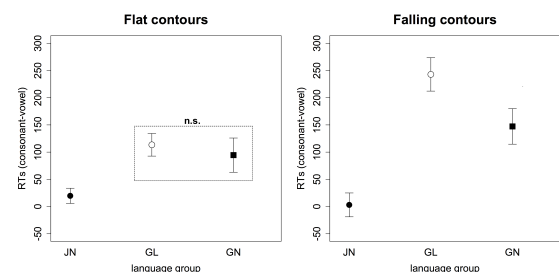


Figure 3: Results of the normalized RTs

In the following, we will describe the results of a linear mixed effects regression model with normalized RTs as dependent variable, *language groups*, *pitch* and *ISI* as fixed factors and *participants* as a random factor including random slopes for the fixed factors. Results showed a main effect of *language groups* (JNs faster than other two groups,  $p < 0.001$ ) and an interaction between them (only GLs slower in the falling condition than JNs and GNs were,  $p < 0.03$ ); see Figure 3. In order to compare the data between the two plots in Figure 3, we further analysed the RTs in both two pitch conditions in each *language groups* separately (*pitch* as a fixed factor and *participants* as a random factor including random slopes for the fixed factor). The analysis revealed that the JNs' and GNs' RTs in the two pitch conditions did not differ ( $p > 0.9$ ,  $p > 0.2$  respectively), while those of GLs did ( $p < 0.001$ ).

### 3. Discussion

To investigate the (in)stability of the discrimination of non-native prosodic contrasts, we conducted discrimination tasks with various conditions that may impair the processing.

We found a consistent effect of pitch on the discrimination of consonantal length contrasts in the  $d'$  score analyses: the discrimination of consonantal length contrasts presented in falling pitch was more difficult for learners than in flat pitch and this especially for non-learners. The findings consistently confirm that the psycho-phonetic complexity affected only the non-natives' discrimination ability, indicating the non-natives' decreasing attention control along the increasing psycho-phonetic complexity. Additionally, the difference between the learners' and the non-learners'  $d'$  scores suggests a positive L2 learning effect in establishing higher attention control and in obtaining a more stable L2 processing of non-native prosodic contrasts.

The RT analyses also showed longer RTs in the learners' group when the pitch was falling. Contrary to the results of the  $d'$  scores, the non-learners' RTs did not change in both pitch conditions, suggesting that the psycho-phonetic complexity did not impact on a *perceivable* task difficulty for non-learners just as for native listeners. In this way, however, the learners' performance regarding the RTs became worse than the non-learners' one in the falling pitch condition. Such a similar tendency was also found in Altmann *et al.* [5] that investigated the non-native (= German) discrimination ability of Italian consonantal length contrasts. They found higher  $d'$  scores for learners than non-learners, but no significant difference between the learners' and non-learners' RTs. We postulate that learners and non-learners applied different strategies for the task. When the stimuli became psycho-phonetically more complex, the direct auditory comparison of the stimuli became more difficult. Therefore listeners started to rely on phonological knowledge. Such learners' indecisiveness demonstrates that the learners already had a degree of access to a newly forming category and thus needed to decide between two possible mapping representations. Non-learners on the contrary, required a shorter time for the decision without the "selection" processing. Another possible interpretation would be the following: this difference between the learners and non-learners was found to be independent from the two ISI conditions, suggesting a direct co-activation or mapping of phonological representations from acoustic speech signals [39]. Here again, since learners established L2 phonological representations, they needed more time to select one between the L1 and the L2 phonological representations, but non-learners did not. We assume that this difference was found only in the increasing psycho-phonetic complexity, because this temporal effect of the selection was relatively small. In future experiments further attention needs to be paid to the underlying mechanisms for non-learners and learners in future experiments.

As for the other dimension of increasing task demands, memory load, native-like good performance by learners and even non-learners was observed only in the condition with the lowest task demands (short ISI and flat pitch). In the long ISI condition, however, the non-learners'  $d'$  scores decreased. Such decrease is the evidence that they had difficulties in discriminating non-native consonantal length contrasts, when the memory load increased and once the processing tapped into the phonological one. Learners were not affected by the increasing memory load just as Japanese natives were not. This effect of memory load was found in a so to say "greenhouse condition" for the consonantal length contrasts, when it was not disturbed with another prosodic cue, pitch, but the processing only dealt with

length contrast without any effects of pitch. Once pitch came into play, the effect vanished. Recall that the  $d'$  scores in the falling pitch condition were generally lower than those in the flat pitch condition in both the learners' and the non-learners' group, while those of native listeners did not change between the two pitch conditions. Therefore, the reason for not finding the temporal decrease in the falling pitch condition may be explained with a floor effect.

Taken together, the effect of psycho-phonetic complexity was stronger than the one of memory load. Both learners and non-learners were distracted by the stimuli with greater psycho-phonetic complexity, because their attention control became lower. As learners become more familiar with the situation, attention demands were eased and automatic processes developed [21]. Therefore, our finding suggests that learners, who even established the phonological representations of non-native consonantal length contrasts still were not successful in automatising the L2 speech processing fully.

The task demands controlled in our study may be translated into various distracting factors in our natural speech perception. Therefore, the performance decreases found in our study suggest the instability of the L2 perception under various task demands in our daily life. Moreover, the performance decrease due to the psycho-phonetic complexity suggests that the Japanese lexical pitch movement (e.g. a falling lexical pitch accent or an initial low [40]) makes the perception of non-native consonantal length contrasts more difficult.

### 4. Conclusions

We examined the discrimination of non-native consonantal length contrasts under increasing task demands. Even non-learners without any exposure to the L2 could discriminate them in the lowest task demands, simply by comparing two stimuli at the auditory level. However, such a reliance on the auditory comparison could not last long. Once the ISI became longer and memory load became higher, so that the auditory information began to be processed phonologically, their discrimination ability decreased and differed from the one by learners or native listeners. The learners' performance did not differ from the native listeners' one. This result suggests that the exposure to the L2 helped them to establish the phonological representations of non-native consonantal length contrasts. However, even learners who were not distracted by the increasing memory load could not overcome the performance decrease due to the increasing psycho-phonetic complexity of the stimuli. It was difficult for both learners and non-learners to ignore the task-irrelevant pitch and to focus on their attention only to the task-relevant information. The finding indicates the difficulty to automatise the L2 processing even after establishing or being exposed to the L2 categories. To summarize, the non-natives' performance was native-like good only under favourable listening conditions with no distracting acoustic information and with lower memory load.

In our natural listening situations, there are numerous distracting factors that might impair L2 perception. The overall decreases indicate why L2 perception remains difficult in the daily-life situations, despite the still successful exercise in L2 class room situations.

### 5. Acknowledgements

The study was funded thanks to a grant from the Young Scholar Funds at the University of Konstanz.

## 6. References

- [1] R. Hayes-Harb and K. Masuda, "Development of the ability to lexically encode novel 12 phonemic contrasts," *Second Language Research*, vol. 24, no. 1, pp. 5–33, 2008.
- [2] A. Wilson, H. Kato, and K. Tajima, "Native and non-native perception of phonemic length contrasts in japanese: Effects of speaking rate and presentation context," *The Journal of the Acoustical Society of America*, vol. 117, p. 2425, 2005.
- [3] C. T. Best, G. W. McRoberts, and E. Goodell, "Discrimination of nonnative consonant contrasts varying in perceptual assimilation to the listeners native phonological system," *Journal of Acoustical Society of America*, vol. 109, pp. 775–794, 2001.
- [4] J. E. Flege, M. J. Munro, and I. Mackay, "Effects of age of second-language learning on the production of english consonants," *Speech Communication*, vol. 16, pp. 1–26, 1995.
- [5] H. Altmann, I. Berger, and B. Braun, "Asymmetries in the perception of non-native consonantal and vocalic length contrasts," *Second Language Research*, vol. 28, no. 4, pp. 387–413, 2012.
- [6] B. Kabak, T. Reckziegel, and B. Braun, "Timing of second language singletons and geminates," in *ICPhS in Hong Kong, 17-21 August 2011*, 2011, pp. 994–997.
- [7] D. B. Pisoni, "Auditory and phonetic memory codes in the discrimination of consonants and vowels," *Perception & Psychophysics*, vol. 13, pp. 253–260, 1973.
- [8] K. Tajima, K. Hiroaki, R. Amanda, A.-Y. Reiko, and M. Kevin G., "Training english listeners to perceive phonemic length contrasts in japanese," *The Journal of the Acoustical Society of America*, vol. 123, pp. 397–413, 2008.
- [9] T. Isaacs and P. Trofimovich, "Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech," *Applied Psycholinguistics*, vol. 32, pp. 113–140, 2011.
- [10] K. M. Dallett, "Intelligibility and short-term memory in the repetition of digit strings," *Journal of Speech and Hearing Research*, vol. 7, pp. 362–368, 1964.
- [11] P. A. Luce, T. C. Feustel, and D. B. Pisoni, "Capacity demands in short-term memory for synthetic and natural speech," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 25, no. 1, pp. 17–32, 1983.
- [12] M. Antoniou, P. C. M. Wong, E. Ingvalson, and S. Wang, "Cognitive factors contribute to speech perception: Implications for sound change actuation," in *Poster presented at Workshop on Sound Change Actuation*. Chicago, USA: University of Chicago, 2013.
- [13] A. Cutler, M. Cooke, M. L. G. Lecumberri, and P. D., "L2 consonant identification in noise: Cross-language comparisons," in *Proceedings of Interspeech*, Antwerp, Belgium, 2007, pp. 1585–1588.
- [14] M. L. G. Lecumberri and M. Cooke, "Effect of masker type on native and non-native consonant perception in noise," *Journal of Acoustical Society of America*, vol. 119, pp. 2445–2454, 2006.
- [15] M. Sonu, H. Kato, K. Tajima, R. Akahane-Yamada, and Y. Sagisaka, "Non-native perception and learning of the phonemic length contrast in spoken japanese: training korean listeners using words with geminate and singleton phonemes," *Journal of East Asian Linguistics*, vol. 22, no. 4, pp. 373–398, 2013.
- [16] J. F. Werker and J. S. Logan, "Cross-language evidence for three factors in speech perception," *Perception & Psychophysics*, vol. 37, no. 1, pp. 35–44, 1985.
- [17] N. Cowan and P. Morse, "The use of auditory and phonetic memory in vowel discrimination," *Journal of Acoustical Society of America*, vol. 79, pp. 500–507, 1986.
- [18] E. Gerrits, "The categorisation of speech sounds by adults and children: a study of the categorical perception hypothesis and the development weighting of acoustic speech cues," Ph.D. dissertation, Universiteit Utrecht, 2001.
- [19] K. Johnson, "Cross-linguistic perceptual differences emerge from the lexicon," in *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception*, G. Agwuele, W. Warren, and S.-H. park, Eds. Somerville, MA: Cascadia Press, 2004, pp. 26–41.
- [20] M. Schouten and A. J. Van Hoesen, "Modelling phoneme perception: I. categorical perception," *Journal of Acoustical Society of America*, vol. 92, pp. 1841–1855, 1992.
- [21] B. McLaughlin, T. Rossman, and B. McLeod, "Second language learning: An information-processing perspective1," *Language Learning*, vol. 33, no. 2, pp. 135–158, 1983.
- [22] A. Baddeley and B. A. Wilson, "Prose recall and amnesia: implications for the structure of working memory," *Neuropsychologia*, vol. 40, no. 10, pp. 1737–1743, 2002.
- [23] A. L. Francis and H. C. Nusbaum, "Effects of intelligibility on working memory demand for speech perception," *Attention, Perception, & Psychophysics*, vol. 71, no. 6, pp. 1360–1374, 2009.
- [24] T. Imada, R. Hari, N. Loveless, L. McEvoy, and M. Sams, "Determinants of the auditory mismatch response," *Electroencephalogr. Clin. Neurophysiol.*, vol. 87, no. 3, pp. 144–53, 1993.
- [25] E. Sussman, "Integration and segregation in auditory scene analysis," *Acoustical Society of America*, vol. 117, no. 3, pp. 1285–1298, 2005.
- [26] R. G. Crowder and J. Morton, "Precategorical acoustic storage (pas)," *Perception & Psychophysics*, vol. 5, pp. 365–373, 1969.
- [27] A. D. Baddeley, *Working memory*. Oxford: Oxford University Press, 1986.
- [28] A. Baddeley, "The episodic buffer: a new component of working memory?" *Trends in Cognitive Sciences*, vol. 4, no. 11, pp. 417–423, 2000.
- [29] C. de Boor, *A practical guide to splines*. New York: Springer, 2001.
- [30] M. Gubian, Y. Asano, S. Asaridou, and F. Cangemi, "Rapid and smooth pitch contour manipulation," in *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 2013, pp. 31–35.
- [31] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 4 2013.
- [32] I. Cunnings, "An overview of mixed-effects statistical models for second language researchers," *Second Language Research*, vol. 28, no. 3, pp. 369–382, 2012.
- [33] S. Akaba, "An acoustic study of the japanese short and long vowel distinction," 2008.
- [34] M. S. Han, "Acoustic manifestations of mora timing in japanese," *Acoustical Society of America*, vol. 96, no. 1, pp. 73–82, 1994.
- [35] Y. Homma, "Durational relationships between japanese stops and vowels," *Journal of Phonetics*, vol. 9, no. 3, pp. 273–281, 1981.
- [36] N. Macmillan and C. Creelman, *Detection theory: a user's guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2005.
- [37] J. Borràs-Comes, V. M. M., and P. Pioto, "The role of pitch range in establishing intonational contrasts in catalan," in *Proceedings of the 5th International Conference on Speech Prosody*, Chicago, 2010, pp. 1–4.
- [38] F. Tomaschek, H. Truckenbrodt, and I. Hertrich, "Processing german vowel quantity: Categorical perception or perceptual magnet effect?" in *Proceedings of the 17th International Congress of the Phonetic Sciences*, 2011, pp. 2002–05.
- [39] I. Darcy, L. Dekydtspotter, R. A. Sprouse, J. Glover, C. Kaden, M. McGuire, and J. H. Scott, "Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in 11 english– 12 french acquisition," *Second Language Research*, vol. 28, no. 1, pp. 5–40, 2012.
- [40] S. Haraguchi, *The tone pattern of Japanese: An autosegmental theory of tonology*. Tokyo: Kaitakusha, 1977.