Attention to Critical Acoustic Features for L2 Phonemic Identification and its Implication on L2 Perceptual Training

Yoon Hyun Kim[1] and Jung-Oh Kim[2]

[1]Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, Korea
[2]Department of Psychology, Seoul National University, Seoul, Korea
toithaka@gmail.com, jungokim@plaza.snu.ac.kr

Abstract

This study examined whether native speakers of Japanese could attend to critical acoustic features while identifying lenis and aspirated among Korean alveolar stops. Most Japanese participants had studied Korean for more than one year at a university language center in Korea. Native speakers of Korean were also tested with the same task for comparison. Korean participants discriminated the phonemic contrast according to both VOT and F0 or just F0. In contrast, Japanese participants identified lenis and aspirated mostly based on VOT information. They correctly identified stimuli of the phonemes which a speaker produced distinctively in terms of VOT. When stimuli weren't noticeably different in VOT, they confused the two phonemes. Unlike Korean participants, they hardly considered F0 information. This result suggests that some training materials, although they were produced by native speakers, can't lead L2 learners to catch critical acoustic information of L2 phonemes. If learners can identify L2 sounds accurately without attention to critical features, they may stick to wrong information in the sounds.

Index Terms: L2 phonemic identification, phonemic training, acoustic feature, attention, Korean

1. Introduction

In many languages including Japanese, VOT (Voice Onset Time) is a critical acoustic feature by which two-way contrast of stop consonants can be differentiated. Stops are perceived categorically as voiced or voiceless according to the information [1, 2]. To discriminate voiced and voiceless stops listeners should attend to VOT while ignoring variance of other acoustic features. However, not all languages have two-way contrast. Stops of Korean are subdivided into three voiceless phonemes: lenis, aspirated, and fortis. They are distinguished by at least two acoustic features, VOT and F0 (fundamental frequency) of the initial point of vowel. Native speakers of Korean make use of both the information to discriminate the three-way contrast [3, 4].

Native speakers of Japanese can't discriminate them easily [5]. They confuse particularly between lenis and aspirated more than lenis-fortis pair or aspirated-fortis pair. The reason is related to perceptual assimilation of the three phonemes by Japanese people. Lenis and aspirated are assimilated mainly into Japanese voiceless stop. But, in case of relatively well-identified fortis, Japanese people assimilate the phoneme into Japanese voiceless or voiced with similar ratio [6]. The perceptual assimilation may depend on which acoustic cues are attended to. Iverson et al. [7] showed that insensitivity on F3 was related to /r/-/l/ confusion of Japanese people. Japanese participants assimilated stimuli into Japanese /r/ though the stimuli were varied according to F3 and identified as /r/ or /l/ by native speakers of English. Native speakers of Japanese didn't attend to F3 though the cue was critical for differentiating /r/ from /l/. In the same way, it seems that the different critical features between Japanese and Korean lead the Japanese to confusion of the contrast of Korean phonemes. To identify three-way contrast of Korean stops listeners need to attend to both VOT and F0. However, Japanese people may adhere to attend to only VOT information though it is sufficient only for the identification of Japanese stop contrast .

This study investigated whether native speakers of Japanese who learned Korean attend to critical acoustic information when they need to identify lenis and aspirated of Korean alveolar stop. For the goal, we synthesized stimuli containing conflicting acoustic cues. For example, one stimuli consisted of lenis' VOT and aspirated's F0. If listeners identify it as lenis more than aspirated, it suggests that VOT information is weighted more than F0 when they identify the stimuli. Native speakers of Korean were also tested for comparison.

2. Method

2.1. Participants

Participants were 24 native speakers of Korean (14 females, 10 males) and 26 native speakers of Japanese (25 females, 1 male). All Korean participants were in their twenties. They were university students attending an introductory lecture of psychology and participated in the experiment for course credits. Japanese participants were attending language courses for Korean at a university language center in Korea. They had been staying in Korea for more than 1 year except for four participants (at least more than 3 months). Three participants were in Korea for five years or more. 15 Japanese participants were in their twenties and 6 participants were in thirties. And 4 participants were aged between 16 and 20 years. 1 participant was in forties.

2.2. Materials and Instruments

Lenis (/ta/) and aspirated (/t^ha/) of Korean alveolar stop were recorded by two male speakers. The speakers (symbolized by "I" or "K") were in their twenties and spoke Standard Korean (Seoul dialect). 8 stimuli (2 speakers * 2 phonemes * 2 tokens) were recorded. In addition, another 8 stimuli were synthesized from the recordings. The number of stimuli was 16 in total. By manipulating F0 of /ta/ new stimuli were synthesized to have the same F0 as /t^ha/. Thus, these synthesized stimuli have conflicting information; VOT of them is the same as /ta/ and F0 is the same as /t^ha/. In the same manner, 4 stimuli of /t^ha/ were manipulated to have the same value as F0 of /ta/. These

stimuli have also conflicting information in that VOT is the same as /t^{h}a/ and F0 is the same as /ta/. Table 1 shows VOT and F0 of all stimuli (See also Figure 1).

Table 1: *VOT and F0 of stimuli*

Stimulus	Token 1		Token 2	
	VOT (ms)	F0 (Hz)	VOT (ms)	F0 (Hz)
I_T	34	125	47	114
I_Th	103	150	97	156
I_T_Th	34	150	47	156
I_Th_T	103	125	97	114
K_T	45	162	32	161
K_Th	57	188	63	188
K_T_Th	45	188	32	188
K_Th_T	57	162	63	161

* I_T, I_Th, K_T, and K_Th are original stimuli. Both I_T and K_T are labels of lenis, /ta/. I_Th and K_Th are labels of aspirated, /t^{h}a/. I_T_Th and K_T_Th represent that VOT of the stimulus is the same as /ta/ and F0 is the same as /t^{h}a/ while I_Th_T and K_Th_T represent that VOT is the same as /t^{h}a/ and F0 is the same as /ta/. I and K represent the speakers who recorded the stimuli.

Stimuli were recorded with Adobe Audition (v. 1.0) using a PC equipped with microphone (Shure SM48) connected by an audio interface (TASCAM US122) in a sound-attenuated recording room. They were digitized at a sampling rate of 22050 Hz with 16-bit quantization. The recorded stimuli were manipulated using Praat (version 4.0.53) to synthesize the stimuli. Stimuli were presented through a headphone (Sennheiser HD212Pro) in a sound-attenuated recording room. Desktop computers (Pentium 4) equipped with a portable sound card (Sound Blaster Audigy 2 Nx) were used for data collection.

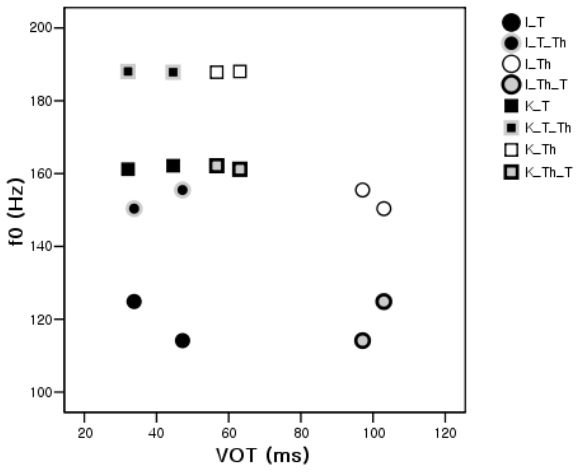

Figure 1: *VOT and F0 of stimuli*

2.3. Procedure

A two-alternative forced choice task was used. Participants were instructed to judge whether it was perceived as /ta/ or /t^{h}a/. They were asked to press one of two keys labeled as '다' or '타' according to their judgment. '다' and '타' are Korean characters representing /ta/ and /t^{h}a/ respectively. After a '+' sign was displayed on the monitor for 1 second, a stimulus was presented through a headphone. After participant responded, a next trial began with appearing '+' symbol again on the monitor. Stimuli were presented randomly. There was no time pressure. The experiment in which Korean participated was composed of 5 blocks of trials (each block consisted of 16 trials). Japanese participants accomplished 3 blocks of 16 trials.

3. Results

3.1. Native speaker of Korean

Rates of /ta/ response for 8 categories of stimulus were reported in Figure 2. Because participants were asked to choose either /ta/ or /t^{h}a/, the response rate of /ta/ is opposite of the rate of /t^{h}a/ response. On I_T and K_T participants responded as /ta/ with a rate of 98.8 % (SD 3.38) and 98.3 % (SD 3.81) respectively. The stimuli were hardly identified as /t^{h}a/. Because both I_T and K_T were /ta/, recorded by speaker I or speaker K, it is reasonable that Korean participants responded almost correctly for those stimuli. Similarly, they responded as /ta/ with a rate of almost 0% (Mean 1.3%, SD 3.38) for I_Th and K_Th which were /t^{h}a/ stimuli unmodified. These response rates mean that participants identified stimuli of I_Th and K_Th as /t^{h}a/ mostly.

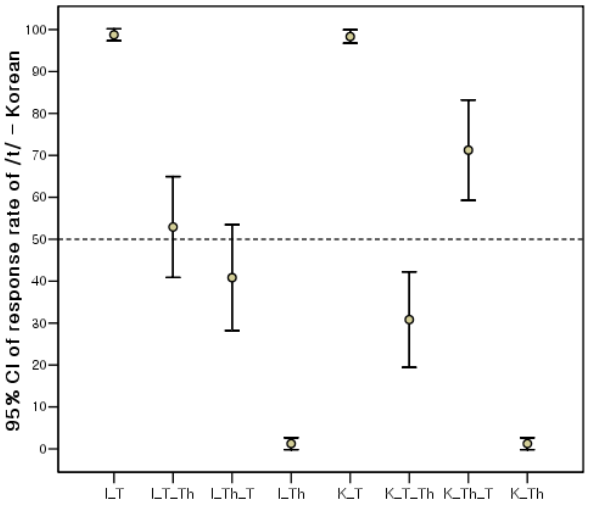

Figure 2: */ta/ response of Korean participants*

In case of I_T_Th, I_Th_T, K_T_Th and K_Th_T, participants responded depending on speaker and acoustic features of stimulus. For I_T_Th and I_Th_T, Korean participants responded as /ta/ with about 50% (MEAN 52.9%, SD 28.5 for I_T_Th and MEAN 40.8%, SD 29.9 for I_Th_T). This result suggests that the two acoustic features, VOT and F0, were considered while they identified the stimuli. If participants attended to VOT and ignored F0 of the stimuli, they would identify I_T_Th as /ta/ because I_T_Th comprised VOT of /ta/. On the contrary, I_Th_T will be perceived as /t^{h}a/ because VOT of the stimulus is the same as / t^{h}a/. However, the mean rate of /ta/ response was around the chance level. In other words, participants didn't ignore F0 when discriminating lenis-aspirated contrast.

Identification rates of K_T_Th and K_Th_T were different from those of I_T_Th and I_Th_T , of which the speaker was different. The result showed that participants identified K_T_Th and K_Th_T based on F0 more than VOT. K_T_Th

was identified as /ta/ with a rate of 30.8% (SD 27.0). That is, participants identified the stimulus as /t^ha/ with a rate of about 70%. On K_Th_T, response rate of /ta/ was 71.3% (SD 28.3). This means that their identification was influenced by F0 more than by VOT in that their responses were more accorded with F0 of the stimuli than VOT.

However, the response pattern was different according to the speaker of stimulus. In case of the stimuli recorded by speaker I, participants considered both VOT and F0 with similar weight for identification. But, on the stimuli of speaker K, they seemed to consider F0 more seriously than VOT. The reason can be related with different values of the acoustic features. VOT difference between lenis and aspirated of speaker I was much larger (about 60 ms) than that of speaker K (about 20 ms).

In sum, Korean participants identified lenis-aspirated contrast considering both VOT and F0. But, when VOT wasn't evident, they identified the stimuli according to F0.

3.2. Native speaker of Japanese

Japanese participants identified I_T stimuli as /ta/ with a mean rate of 94.4% (SD 14.7) and I_Th as /ta/ with a rate of 4.5% (SD 9.4). Considering only this result, Japanese participants seemed to identify Korean lenis-aspirated almost like native speakers of Korean. But, result of K_T and K_Th identification exposed that such conclusion might be hasty. Though the mean response rate of /ta/ for K_T was 89% (SD 19.8), response rate of /ta/ for K_Th was around chance level (Mean 52.8%, SD 31.9). This result showed that participants couldn't discriminate lenis-aspirated contrast recorded by speaker K (Figure 3).

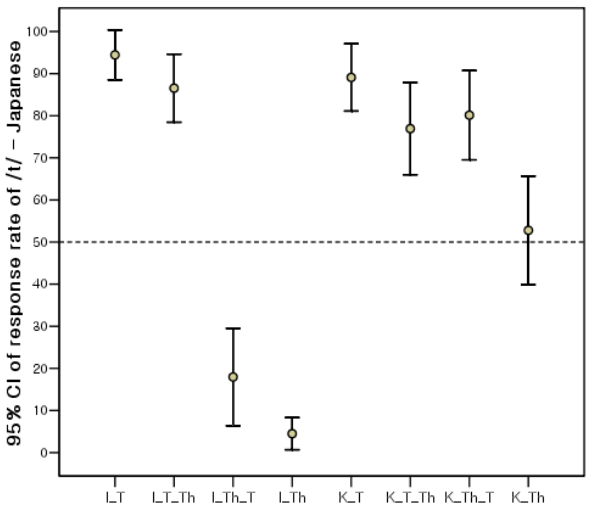

Figure 3: */ta/ response of Japanese participants*

Why did Japanese participants respond accurately to the stimuli recorded by speaker I and not to the stimuli recorded by speaker K? One clue for that could be found in responses of I_T_Th and I_Th_T. In case of I_T_Th, the mean rate of /ta/ response was 86.5% (SD 20.0). And, the mean rate of /ta/ response was 18% (SD 28.6) for I_Th_T. VOT of I_T_Th which was mainly identified as /ta/ was the same as that of /ta/ stimuli. Likewise, I_Th_T which was identified as /t^ha/ over 80% consisted of VOT of /t^ha/. Thus, the result suggests that Japanese participants identified the stimuli mostly based on VOT. Likewise high rate of /ta/ response for K_T_Th (76.9% SD 27.1) and for K_Th_T (80.1% SD 26.3) seemed to be caused also by attending more to VOT information. VOT of /t^ha/ produced by speaker K was similar to VOT of /ta/ recorded by speaker K and speaker I (Figure 1). However, Korean participants considered F0 of the stimuli more in that case.

4. Discussion

This study examined whether native speakers of Japanese used critical acoustic information while they identified Korean alveolar lenis or aspirated stop. Native speakers of Korean were also tested in order to compare the results. Korean participants identified lenis-aspirated contrast attending to both cues, VOT and F0 (for example, stimuli recorded by speaker I in Figure 2). If VOT cue wasn't salient, they identified stimuli relying on F0 information (for example, stimuli recorded by speaker K in Figure 2). Japanese participants solved the problem differently from Korean participants. They depended mostly on VOT information (see Figure 3). That is, even though Japanese participants could identify some stimuli correctly (for example, I_T and I_Th in Figure 3), they identified the phonemes in different ways from the native speakers of Korean. Thus, Japanese participants got into trouble when they had to identify the stimuli of which VOT didn't give enough information (for example, stimuli recorded by speaker K in Figure 3).

The result suggests that L2 phonemic training program needs to select training materials cautiously. Training materials can affect whether learners of unfamiliar L2 phonemes may shift their attention to critical acoustic cues for the identification during training. For example, stimuli recorded by speaker I can be identified easily without attention to F0. Thus, if they are used in the training, Japanese learners of Korean might adhere to their wrong perceptual strategy. In this respect, high-variability phonetic training approach that uses various materials produced by many speakers can be a reasonable choice [8]. However, this approach may not be enough by itself [e.g., 9]. Though Japanese participants learned /r/ and /l/ they didn't shift attention to the critical acoustic cue, F3, when identifying them. Even in the program it may need to exclude any materials that can be correctly identified with inappropriate acoustic cues.

5. Acknowledgements

This work was a part of the unpublished doctoral dissertation of Yoon Hyun Kim

References

[1] Lisker, L., and Abramson, A. "A cross-language study of voicing in initial stops: acoustical measurements", Word, 20, 384-422, 1964.

[2] Cho, T., and Ladefoged, P. "Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics", 27, 207-229, 1999.

[3] Kim, M., Beddor, P., and Horrocks, J. "The contribution of consonantal and vocalic information to the perception of Korean initial stops", Journal of Phonetics, 30, 77-100, 2002.

[4] Kang, K., and Guion, S. "Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals", J. Acoust. Soc. Am., 119, 1672-1683, 2006.

[5] Kim, Yoon Hyun and Kim, Jung-Oh "The effect of the Japanese' perceptual learning of Korean alveolar stop consonants on discrimination of Korean bilabial stop

consonants", The Korean Journal of Experimental Psychology, 17, 499-507, 2005.

[6] Jeehyun Kim "Japanese adults' perceptual categorization of Korean three-way distinction", Unpublished thesis for master degree, Seoul National University, Seoul, Korea, 2005.

[7] Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. "A perceptual interference account of acquisition difficulties for non-native phonemes", Cognition, 87, 47-57, 2003.

[8] Logan, J., Lively, S., and Pisoni, D. "Training Japanese listeners to identify English/r/and/l: A first report", J. Acoust. Soc. Am., 89, 874-886, 1991.

[9] Iverson, P., Hazan, V., & Bannister, K. "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults", J. Acoust. Soc. Am., 118, 3267-3278, 2005.