In speech perception, an acoustic signal is mapped onto phoneme categories. It is well known that infants have perceptual sensitivities not only for phonemes in their parents’ language but also for non-native phonemes (Streeter, 1976). Perhaps the auditory capacity and the perceptual mapping are fundamentally language-universal. However, it appears that perception is constrained by language-specific experience (Kuhl, 1992; Best, 1995; Flege, 1995; Johnson, 2004). Thus, the perception pattern of non-native sounds provides much insight into language-specific aspects of the native phonology of a language. In exploring the influence of language-specificity on perception, it is important to directly compare languages having distinct sound systems.

This study is concerned with the English [æ]-[ɛ] and [æ]-[a] distinctions for speakers of Canadian English, Korean and Japanese. Japanese has 5 short vowels /a, e, i, o, u/ whereas Standard Korean includes 7 contrastive monophthongal vowels /a, e, i, o, u, ɯ, ʌ/. Thus, /æ/ is not contrastive either in Japanese or in Korean. Further, it seems that the non-back and low regions in vowel space of the two languages are quite similar in terms of density and average formant frequencies of phonemes. By utilizing vowel stimuli with varying frequencies of the first (F1) and second (F2) formants, I explore how the language-universal perceptual sensitivities are modified, and which acoustic cue the speakers were more attentive to for vowel categorization.

In investigating the role of language-specific experience on vowel categorization, an identification test was conducted. English nonsense words of the form CV-s were created. All the items contained one of the three vowels /æ, e, a/ preceded by either /h/ or /f/. By testing nonsense words, the influence of existing loans or frequency effects can be obviated. Most of the possible responses are nonce words in Korean and Japanese as well. The onset fricatives, particularly /h/, were chosen in order to minimize possible effects of a consonant on the acoustic characteristics of a following vowel. One male speaker of American English read a list containing the six target items together with twelve fillers. Two sets of synthesized sounds for each onset consonant were created, by manipulating F1 and F2 of the English vowel [æ] obtained in the recording session: F1 continuum and F2 continuum. Specifically, F1 of the vowel was lowered by 30 Hz steps (ranging from 800 Hz to 620 Hz), and F2 by 60 Hz (ranging from 1540 Hz to 1180 Hz), yielding 7 stimuli for each continuum. As a reference for the range of formant continua, the F1 and F2 values of [a] and [ɛ] obtained in the recording were used. It is worth noting that the duration of stimuli was carefully controlled, since it is widely acknowledged that duration can be a critical cue for the vowel contrast. The duration of all the manipulated stimuli in this study was 85 ms, which was the average duration of [a] and [ɛ] for the speaker.

The target items and fillers were shuffled and presented in a random order. Native speakers of English were asked to choose the word that contains the closest vowel that they heard. Korean and Japanese participants were instructed to choose the closest sequences that they heard among given choices. For Korean and Japanese, choices were provided in Korean (Hangul) or Japanese orthography (Katakana). Twenty-eight English, thirty-four standard Korean, and twenty-eight standard Japanese monolingual speakers participated in a forced-choice vowel identification task. Participants ranged in age from 18 to 39 years at the time of the test. All were born and grew up in the respective linguistic target areas and had no history of speech or hearing impairment. All Korean and Japanese speakers had received English language education in school, but had not lived abroad.

Figure 1 shows the percentage of vowel responses depending on F1 (left) and F2 (right). For both the F1 continuum and F2 continuum, the [æ] responses increase gradually as each formant frequency become higher. It appears that the change of the perceived vowel from [ɛ] to [æ] requires relatively higher F1 or lower F2, compared to American English, presumably due to the Canadian vowel shift (Labov, 1991; Clarke et al., 1995). Nonetheless, the results reveal that both the F1 and F2 cues of the ranges tested are perceptually salient to the speakers of Canadian English.

A crosslinguistic Study of Vowel Categorization: Data from Canadian English, Korean and Japanese

Hyun Kyung Hwang
National Institute for Japanese Language and Linguistics
hwang@ninjal.ac.jp
Figure 1. Canadian English speakers’ vowel perception by varying F1 (left) and F2 (right)

Figure 2 presents the percentages of [e] responses depending on F1 and F2 values by the speakers of Korean and Japanese. The two language groups exhibit fairly distinct perception patterns.

Figure 2. Percentages of [e] perception by varying F1 (left) and F2 (right)

For the F1 continuum, no appreciable change of vowel perception is observed in Korean, indicating that the stimuli were invariably perceived as [e]. Unlike Korean, sounds on the F1 continuum are negatively correlated with the percentages of [e] responses in Japanese, suggesting that Japanese speakers are sensitive to vowel height, which is signaled by F1. Interestingly, completely opposite patterns are observed for the F2 continuum; Korean speakers exhibit a gradual decrease of [e] responses, suggesting sensitivity to the backness of a vowel, whereas Japanese speakers do not show a noticeable change in perceived vowel. These results suggest that language-specificity can modify the sensitivity to different acoustic cues in perceptual processing. The current study considers two possible language-specific factors—the L1 phonology and phonetic variations—in directing native speakers’ attention to different acoustic cues.

References