Title

Frequency Effects in Cross-linguistic Stop Place Perception: A Case of /t/-/k/ in Japanese and English

Permalink

https://escholarship.org/uc/item/5qr980vk

Journal

UC Berkeley PhonLab Annual Report, 3(3)

ISSN

2768-5047

Authors

Kataoka, Reiko
Johnson, Keith

Publication Date

2007

DOI

10.5070/P75qr980vk

UC Berkeley Phonology Lab Annual Report (2007)

Frequency Effects in Cross-Linguistic Stop Place Perception: A Case of /t/ - /k/ in Japanese and English

Reiko Kataoka and Keith Johnson

1. Introduction

The study described in this paper is an attempt to answer the question ‘Which aspect of speech perception is altered by linguistic experience and how this alteration is done?’ As Strange and Jenkins stated, “[t]he knowledge of a language possessed by a normal adult is a product of many years of exposure to a specific language environment” (1978: 125). Therefore, if linguistic knowledge influences speech perception, then adult listeners would perceive speech sounds in language-specific ways. Evidence of experience-based speech perception has been accumulated from both cross-linguistic and within language studies. One source of evidence is the phenomena called categorical perception (Liberman, et al. 1957, 1961a, 1961b), which is characterized by difficulty in discriminating acoustically similar patterns in the single phonemic category and near perfect discrimination for acoustic patterns that straddle different categories. Categorical perception has been replicated with listeners of various language groups, and it seems to be an undeniable fact that linguistic experience influences listeners’ ability to discriminate speech stimuli. However, exactly which aspect of linguistic experience is responsible for language-specific speech perception is not well understood. This is the primary question addressed in this paper. Also, it is often reported that listeners react to speech sounds differently depending on whether they use a continuous mode or a categorical mode of memory (see, for example, Pisoni 1973). Thus, it is of interest to test whether experience-based perception is elicited when the listening task calls for auditory acoustic perception.

The goal of this study is to obtain answers of these two questions. The first question was treated by testing whether a particular aspect of linguistic structure—phoneme frequency—influences speech perception; and the second question was treated by testing whether the linguistic knowledge has any effect in an auditory acoustic perception task. The rest of the paper is organized as follows. Section 2 provides reviews of previous studies on the experience-based auditory perception, and gives rationales for the particular ways current experiment is designed. Section 3 describes the experimental study, provides the results, and discusses the results. Finally, Section 4 discusses the implications of some of the findings from current study for theories of speech perception.

2. Background

2.1. Experience-based speech perception

One of the prime examples of experience-based speech perception is categorical perception, a phenomena originally defined by Liberman, Harris, Hoffman, and Griffith (1957). Categorical perception can be demonstrated in an experiment that uses a series of synthetic consonant-vowel stimuli, ranging across two or more initial consonant categories (e.g. ten-equal-step /da/-/ga/ continuum), and involves identification and discrimination tasks. One of the defining characteristics of categorical perception is predictability of discrimination accuracy from the identification result. In its strongest form, categorical perception predicts that listeners can discriminate the two stimuli so long as they identify the two stimuli belonging to the separate categories. This form of categorical perception has never been demonstrated empirically. In its weaker form, in which the defining characteristic is coincidence of the point of maximum ambiguity in the identification function with the peak in discrimination, categorical perception has been replicated in many studies involving listeners of various linguistic backgrounds (Repp, 1984). Although these studies have also revealed that discriminability of within-category stimuli varies considerably depending on the task and stimuli, at least, it has been strongly suggested that a listener's ability to discriminate the speech sounds is closely related to the phonological system that the listener has acquired.

The relationship between discriminability of speech sounds and listener's phonological system has been often investigated in cross-linguistic studies. Such studies have shown that monolingual adult listeners have difficulties discriminating speech sounds when they are acoustically similar and both represent the same phoneme in their native languages. For example, monolingual Japanese adult listeners perform poorly in discriminating two English liquids [ɹ] (voiced alveolar approximant) and [l] (voiced alveolar lateral approximant), which are not contrastive phonemes in Japanese (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura, 1975; MacKain, Best, and Strange, 1981; Yamada and Tohkura, 1992). English speaking adults have their own difficulties in discriminating Hindi [t̪] (voiceless unaspirated dental stop) and [ʈ] (voiceless unaspirated retroflex stop) (Werker, Gilbert, Humphrey, and Tees, 1981). In the study of Werker, et al., the authors noted that all infants of English speaking parents except for one could distinguish [t̪] from [ʈ], and concluded to attribute poor performance of adult listeners to "developmental decrease in speech discriminatory abilities" (354) for this particular contrast.

Evidence of cross-linguistic difference in speech perception has been found not only in a quantitative difference in perception (e.g. discrimination accuracy) but also in a qualitative difference in perception such as the use of acoustic cues. Lisker and Abramson (1970, Abramson and Lisker 1970) compared the voice onset time (VOT) (Lisker and Abramson, 1964) perception by speakers of English, Spanish, and Thai using VOT continua as experimental stimuli. They found that the locations of phoneme boundaries and the associated peaks in discriminability varied among the groups, and these locations matched the typical VOT boundaries used in the groups' own languages. Keating, Mikoś, and Ganong (1981) also reported that English speaking listeners drew phonemic boundary at higher VOT than Polish listeners, as expected from the way two languages realize voicing distinction.

Not only may the same listeners react differently to the different types of stimuli but the same stimuli may elicit different response from the different subject groups. Beddor and Strange (1982) compared identification and discrimination of Hindi oral-nasal vowel contrast by American and Hindi listeners. Interestingly, American listeners did not exhibit 'poorer' discrimination than Hindi listeners. They could discriminate not only two sounds across category but also two sounds in an oral category. That is, American listeners exhibited continuous perception—a qualitatively different response than that of Hindi listeners, whose response was to exhibit categorical perception.

2.2. Perceptual magnet effect

A mechanism underlying the formation of experience-specific perception has been proposed by Kuhl (1991). In a series of experiments, Kuhl demonstrated that: 1) when rating within-category variants of a vowel /i/, adult listeners judge a certain instance of /i/ vowel as a 'good' (prototypic) instance and another as a 'poor' (nonprototypic) instance; and 2) both adult and infant listeners exhibited lower discrimination accuracy when the task involved prototypic vowel as the referent vowel than when the task involved nonprototypic vowel as the referent. Kuhl explained the observed results in terms of what she called the 'perceptual magnet effect'. The perceptual magnet effect causes a warping of perceptual space in a way that acoustic patterns near phonemic category prototypes are perceived as more similar than acoustic patterns that has the same distance on the acoustic scale but reside further away from the prototypes. The perceptual magnet effect thus hypothesizes that discriminability of speech sounds is a function of perceived category 'goodness': The 'better' category members the sounds are, the harder to discriminate them. Being different from the original formulation of categorical perception, the perceptual magnet effect captures gradient nature of category goodness judgment and thus nicely accounts for the gradient discrimination accuracy, which has been a problem of categorical perception.

The perceptual magnet effect was observed from American and Swedish infants as young as six months of age (Kuhl, Williams, Lacerda, Stevens, and Lindblom, 1992), and under different experimental conditions (i.e. with different Inter Stimulus Intervals) (Iverson and Kuhl, 1995). However, the authors noted that the location of the best stimuli and highest perceptual clustering varied across experiments and therefore acknowledged the possibility for the perceptual magnet effect to be influenced by experimental context.

Recently, Guenther and Gjaja (1996) proposed a neural network model that accounts for the perceptual magnet effect. The model predicted that formation of reduced or heightened sensitivity to the heavily exposed within-category stimuli is determined by the distribution of the training stimuli, not by the training method itself. However, a later study by Guenther, Husain, Cohen, and Shinn-Cunningham (1999) showed that changing training regimen alone could alter an effect on the subjects' sensitivity to the training stimuli. This result lead the authors to re-formulate the model and state that 'categorical' training leads to less number of neurons coding the most frequently encountered stimuli; and 'discrimination' training, to more number of neurons.

Studies on the perceptual magnet effect thus have offered mixed results. On one hand, they have clearly shown that listeners are able to tell 'good' members from 'poor' members of a given phonemic category. Also, within a language group, adult listeners draw phonemic boundary consistently, at the similar location across listeners (Kuhl, 1991). However, how directly the category goodness judgment is related to discrimination accuracy is not clear. Also, whether the perceptual magnet effect is task-specific and/or stimuli-specific phenomenon or not seems to be an open question.

2.3. Effect of task and stimuli

As a variety of experimental methods have been used in the investigation, it has become increasingly clearer that obtained results vary depending on the details of the experimental set-up and the type of stimuli used in the experiment. In the very early stage of speech perception research, Liberman, et al. (1961a, b) demonstrated that listeners exhibited categorical perception, or better discrimination across phoneme category than within category, when synthesized 'speech' stimuli were presented but did not exhibit the same response patterns when synthesized stimuli with inverted formant structure, which were assumed to be heard as 'non-speech', were presented. As other examples of stimuli influencing the perception, Pisoni and Lazarus (1974) cited Fujisaki and Kawashima (1970) and Pisoni (1971, 1973), which demonstrated that listeners tend to perceive vowels categorically if they are short in duration or in the context of a steady-state reference vowel, or noncategorically if they are long in duration or presented in isolation.

On the other hand, the same stimuli may elicit different mode of perception from the same subject if the experimental procedure is altered. Pisoni and Lazarus (1974) demonstrated that the same speech stimuli were perceived in more nearly categorical mode when presented in ABX test but were perceived in more nearly continuous mode when presented in 4IAX test. The authors offered a possible mechanism underlying observed task effect in terms of memory load. They reasoned that the ABX task places greater demand on short-term memory, which is much less in 4IAX task; therefore, the listeners rely on phonetic rather than auditory coding in order to respond in the ABX discrimination test.

All these results imply that listeners are more likely to employ either categorical or continuous mode of perception depending on whether the task calls for linguistic knowledge or not. ABX test and the stimuli resembling the linguistically meaningful acoustic patterns as in the case of speech stimuli or short stimuli embedded in contexts are likely to call for linguistic knowledge, and thus elicit categorical perception. On the other hand, 4IAX test and linguistically meaningless stimuli seem to skip access to linguistic knowledge, and thus elicit continuous perception. The question to ask is which aspect of linguistic knowledge is responsible in modulating the mode of perception and how this modulation is done.

2.4. Role of lexicon

The answer to the above question might come from 'lexical distance' model of speech perception (Johnson, 2003). The lexical distance model is a mathematical representation of human auditory system that computes the perceived distance between two acoustic stimuli. The model assumes: 1) language-universal perceptual map for speech, and 2) language-specific perceptual modulation via contact through lexical memory, allowing for both language-independent and language-specific auditory perceptual responses to emerge depending on the perceptual task. The model also assumes a lexicon that stores rich and detailed phonetic exemplars. Johnson tested these hypotheses by comparing responses from English-speaking and Dutch-speaking listeners on the two tests: One was an AX discrimination task with a short Inter Stimulus Interval (100 ms), which has a low memory load and thus eliminates lexical access; and the other was a similarity rating task, which involves lexical knowledge. Since the stimuli were VCV syllables wherein medial C was one of six fricatives [f θ s ʃ x h] ([θ] and [x] is a non-native phoneme for Dutch subjects and English subjects, respectively), a subject group main effect in the result was evidence of language-specific perception. The author found no subject group difference in the response pattern in the speeded AX task but did found the group difference in the similarity rating task, thus concluded that 'lexical distance' model was correct.

2.5. Phoneme frequency

The involvement of the lexicon in speech perception may be evidenced by showing the effect of various aspects of lexical information such as lexical status (word or non-word) (Ganong, 1980; Fox, 1984; Whalen, Best, and Irwin, 1997;), neighborhood density (how many other words of similar form a given word has) (Luce and Pisoni, 1998; Sussman, Dee, and Curcio, 2003) and word/phoneme frequency (how often a given word/phoneme occurs in a language). The current study compares speech perception of two different groups of listeners—native speakers of English and native speakers of Japanese—and explores the effect of phoneme frequency of the two languages on speech perception.

Yoneyama (2000) investigated structural aspects of the Japanese lexicon using an electronic dictionary (Sanseido Shinmeikai Dictionary, 1983). One of the findings was that, among voiceless stops, "[k] is the most frequent[1] stop among the three voiceless stops, and [t] is the second most frequent segment[2] followed by [p]". This makes an interesting contrast between Japanese and English because in English [t] is the most frequent segment (Yoneyama, 2000). Later, Yoneyama, Beckman and Edwards (2003) reported that Japanese children acquire phoneme /k/ earlier than /t/, while the order of acquisition is reversed for English acquiring children. Therefore, it is of interest to know if phoneme frequency has any effect on ways adult Japanese- and English-speaking listeners perceive /t/ and /k/.

3. Experiment

3.1 Hypotheses and tasks

Several hypotheses were tested in the current study. The first is whether different modes of perception are elicited depending on the task. Two tasks were employed for this end: The first task was 4IAX discrimination test, which has low memory load; and another task was determining phoneme boundary along the stimulus continuum, which involves consulting lexicon. These tasks are described in detail in the Section 3.2. I also attempted to answer the question addressed in Section 2.2—whether the perceptual magnet effect is task specific or not. Since the perceptual magnet effect hypothesizes that discriminability is inversely correlated to perceived category 'goodness' (Kuhl 1991; Kuhl, et al. 1992; Iverson and Kuhl 1994), I tested this hypothesis by comparing the discrimination accuracy with 'goodness' rating. Finally, in order to test for a phoneme frequency effect, experimental stimuli were constructed using Japanese and English /k/-/t/ continua. The perceptual magnet effect predicts that, due to phoneme frequency, English speakers will have the /t/ region more severely warped than /k/

---

[1] Also, [k] is the most frequent consonant. Among all consonants, [k]'s share is 15%; [t]'s share is 6%; and [p]'s share is 1%.

[2] Japanese /t/ is realized as [tʃ] before /i/ and as [ts] before /u/, contributing to the reduction of phone frequency. However, the relative frequency difference between [t] and [k] remains the same even phoneme frequency of /t/ ([t], [tʃ] before /i/, and [ts] before /u/) and /k/ are compared (Yoneyama, 2000).

region in their perceptual spaces, and Japanese speakers' perceptual spaces are altered in an opposite way. Thus, phoneme frequency effect might manifest itself in two different ways. One is on phoneme boundaries: English-speaking listeners might draw phoneme boundary closer to /k/ end than Japanese-speaking listeners do. The other is on discrimination accuracy: English-/Japanese-speaking listeners might exhibit poorer discrimination for stimuli involving [t]-like/[k]-like segment than for stimuli involving [k]-like/[t]-like segment. To test these hypotheses, an experiment that consists of three tasks—1) 4IAX discrimination task, 2) phonemic boundary decision task, and 3) category goodness rating task—was conducted.

3.2. Method

3.2.1. Participants

30 native speakers of American English ('English listeners') and 35 native speakers of Japanese ('Japanese listeners') participated in the experiment to serve as listeners. All American participants were undergraduate students of UC Berkeley. Japanese participants were undergraduate students, graduate students, and visiting scholars of UC Berkeley, and family members of them. All of them reported to have normal hearing. Participants received $10 for participation[3].

3.2.2. Block 1: 4IAX discrimination task

3.2.2.1. Construction of basic stimuli

Ten tokens of the four VCV sequences /aka/, /ata/, /oko/, and /oto/ were produced by a female native speaker of American English and a female native speaker of Japanese and digitally recorded in a sound-treated room using a head-mounted microphone and a Marantz PMD670 Solid State recorder at a 44.1-kHz sampling rate and 16-bit quantization. The rest of the procedures in stimuli construction were done by using Praat (Boersma and Weenik 2002) through LPC analysis and PSOLA resynthesis. First, one token of /aka/, /ata/, /oko/, and /oto/ from both speakers' tokens was selected in a way that all eight tokens have similar total duration and intensity[4]. Next, English tokens /aka/ and /ata/ were used to create a ten-step /aka/-/ata/ continuum. To do so, each CVC sequence was first divided, at zero-crossing, into four parts: initial vowel (V1), consonantal closure (Closure), burst plus aspiration (Burst), and final vowel (V2). Then, four pairs of corresponding parts (e.g. V1s of /a̲ka/ and /a̲ta/, Closures of /ak̲a/ and /at̲a/, etc.) were adjusted so that the pair of V1 are identical in duration, intensity, and F0 contour; so are the pair of V2; the pair of C are identical in duration; and pair of Burst are identical in duration and intensity. Next, a ten-step continuum was created four each of the four parts by adding two corresponding parts, each of which was one of the ten graded proportions in terms of intensity, as shown in (1):

---

[3] A few Japanese participants refrained from receiving compensation.

[4] All tokens were then adjusted to have the identical values of total duration and intensity. Still the similar tokens were selected in the first step to minimize the amount of modification and thus ensure the naturalness of the stimuli.

(1) Ratio of the two original parts for a ten-step continuum

9/9 of V1 of /aka/ + 0/9 of V1 of /ata/ → step 0 of V1 (same as V1 of /aka/)
8/9 of V1 of /aka/ + 1/9 of V1 of /ata/ → step 1 of V1
↓ ↓ ↓
1/9 of V1 of /aka/ + 8/9 of V1 of /ata/ → step 8 of V1
0/9 of V1 of /aka/ + 9/9 of V1 of /ata/ → step 9 of V1 (same as V1 of /ata/)

Three more ten-step continua were created using the rest of the parts—Closure, Burst, and V2. Finally, the V1, Closure, Burst, and V2 for each step number were concatenated to reconstruct a VCV sequence, yielding ten-step /aka/ -/ata/ continuum. In the rest of the paper, each of the ten stimuli will be referred to as 'stimulus 0', 'stimulus 1', and etc., stimulus 0 being same as /VkV/ token and stimulus 9 being same as /VtV/ token.

Three more ten-step /VkV/-/VtV/ continua were created—one of English /o/ (i.e. /oko/-/oto/ continuum) and one each of Japanese /a/ and /o/, yielding four /VkV/ - /VtV/ continua: English /aka/ - /ata/, English /oko/ - /oto/, Japanese /aka/ - /ata/, and Japanese /oko/ - /oto/. The total duration of a single stimulus of each set is as flows: English /a/ stimulus - 519 ms, English /o/ stimulus - 525 ms, Japanese /a/ stimulus - 514 ms, and Japanese /o/ stimulus - 519 ms.

3.2.2.2. Presentation of stimuli

For the discrimination task, a four-interval two-alternative forced-choice (4IAX) procedure was used. On each trial a subject heard two pairs of stimuli (thus, 'four intervals') with a longer Inter Stimulus Interval (ISI) (500 ms) between the pairs and shorter ISI within a pair (200 ms). One pair is always of identical stimuli ('same pair'): These are two repetitions of a single stimulus from the ten-step continuum. In the other pair ('different pair'), the same stimulus is paired with another stimulus which is two steps away in the continuum. For a given stimulus pair, there were eight possible combinations of the order of stimuli in a different pair and the order of two pairs (thus, eight different kinds of 4IAX trials, as illustrated in Table 1 below.

Table 1: A list of four intervals created from a given stimulus pair. In the left column, a letter A and B denote two different stimuli. In the right column, examples of 4IAX trials using a Stimulus 0 in the place of A are provided. In both columns, a Different Pair is bold faced.

Combination	Examples (Stimulus 0 as A)
<**AB** – AA>	<**0,2** – 0,0>
<**AB** – BB>	<**0,2** – 2,2>
<**BA** – AA>	<**2,0** – 0,0>
<**BA** – BB>	<**2,0** – 2,2> (first pair is the Different Pair)
<AA – **AB**>	<0,0 – **0,2**>
<AA – **BA**>	<0,0 – **2,0**>
<BB – **AB**>	<2,2 – **0,2**>
<BB – **BA**>	<2,2 – **2,0**> (second pair is the Different Pair)

Since two stimuli in the different pair were two steps apart, there were eight kinds of different pairs (i.e. <0,2> … <7,9>) obtained from a ten-step stimulus continuum. Thus, in total of 64 kinds of 4IAX (8 pairs x 8 combinations) were obtained from a given continuum to be presented as experimental stimuli.

3.2.2.3. Procedure

Participants were randomly assigned to one of two conditions. For one group, stimuli only from /aka/-/ata/ continuum were presented (/a/ condition); and in the other, from /oko/-/oto/ continuum (/o/ condition). This condition remains the same for all three blocks. block 1 consists of two sub-blocks. In one sub-block only English stimuli were presented; and in the other, only Japanese stimuli. The order of sub-block was randomized within groups. Within a sub-block, each of the 64 stimuli was repeated twice, yielding 128 trials. The trials were given in a different random order for each subject.

The task for the listeners was to listen to the 128 stimuli (per sub-block), which was played binaurally through headphones, and, after each stimulus was played, determine which pair was the different pair and make a response by pressing a button on a five-button box, which was placed in front of the listener. They were asked to press the leftmost button if the first pair is the different pair; and the rightmost button if the second pair is the different pair. They were told to guess when they were not sure which pair was the different pair. If no response was made within 500 ms after the offset of the fourth Interval, then ‘no response’ was recorded for that trial and next stimulus was played automatically. If the response was made within 500 ms, then the next stimulus was played immediately after the response is made. Listeners were encouraged to take a break between trials as wished. Presentation of the stimuli and response collection were controlled by E-Prime.

3.2.3. Block 2: phoneme boundary decision task

3.2.3.1. Presentation of stimuli

Every token of the ten-step continua was presented in a boundary determination task, following the ‘up/down’ procedure described by Ganong and Zatorre (1980). In this method, the ten stimuli from a given continuum are presented as an ascending or descending series, one step at a time, with an ISI of 1000 ms. Out of eight series (2 talker languages x 2 vowels x 2 orders, four involving the vowel /a/ were presented to /a/ group of listeners, and the other four involving the vowel /o/ were presented to the /o/ group of listeners. The four series were presented in a different random order for each listener in each group.

3.2.3.2. Procedure

The task of the listeners in block 2 was to listen to the four series of stimuli, and within each series, determine when the medial consonant shifted from one category to the other and respond by pressing a button on the five-button box. They were told that the series consists of the stimuli that they already heard in block 1 and that the stimuli were ordered so that the medial consonant shifts from /k/ to /t/ (or from /t/ to /k/) somewhere within the series. They were instructed to press the middle button in the button box the moment the middle consonants shifts. If no response was made before the end of the tenth stimulus, then 'no response' was recorded for that trial and the next series was played automatically. When the response was made, the presentation of the series stopped, and play back of the next series began. Presentation of the series and response collection was controlled by E-Prime.

3.2.4. Block 3: category goodness rating task

3.2.4.1. 'Goodness' rating

As a way of measuring subjects' opinion on the category 'goodness' of the medial consonant, a nine-point goodness rating scale was employed (see Figure 1). A nine-point scale was chosen because intrarater reliability (i.e. the consistency with which one listener rates a number of different stimuli) and interrater reliability (i.e. consistency with which different listeners rate a number of different stimuli) are near maximum when 7-11 scale categories are used (Grether and Stroh, 1973). To help listeners maintain consistency in rating across stimuli, descriptions—'Very easy', 'Pretty easy', 'Neutral', 'Pretty hard', and 'Cannot tell' – were provided alongside the point 9, 7, 5, 3, and 1, respectively. The choice of descriptions (instead of 'Excellent', 'Good', and etc.) was due to the expected uniformity of criterion used for rating. It was felt that there would be considerable variation in the criteria used in evaluating 'goodness' of a speech sound but telling 'how easy it is to identify the sound' would be based upon similar criteria across listeners. Thus, listeners were instructed to evaluate each consonant in terms of how easy it is to tell the consonant /t/ as /t/ (or /k/ as /k/).

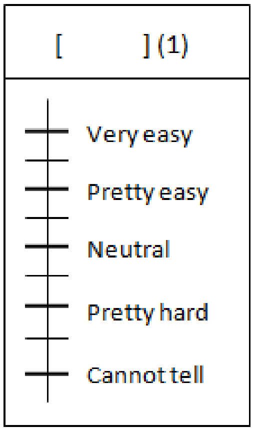

Figure 1. Consonant goodness rating scale

3.2.4.2. Presentation of Stimuli

Each token in the ten-step continua served as experimental stimuli. Stimuli were presented with an ISI of 3000 ms, during which response was to be made. Out of forty stimuli (2 talker languages x 2 vowels x 10 steps), twenty using the vowel /a/ were presented to the /a/ group of listeners, and the other twenty using the vowel /o/ were presented to the /o/ group of listeners. Each of the twenty stimuli was repeated five times so as to reduce spurious errors that may affect ratings (Grether and Stroh, 1973), yielding 100 trials. A beep (a 1000 Hz sine wave plus white noise) of 300 ms was played after every ten trials to help listeners keep their place on answer sheets. The trials were given in a different random order for each listener. Presentation of the stimuli was controlled by E-Prime.

3.2.4.3. Procedure

The task of the listeners in this block was to listen to one hundred stimuli and, for each stimulus, write down the medial consonant (either *k* or *t*) and give the goodness rating for the medial consonant by circling over one of the nine horizontal lines on the scale. They were told that all stimuli were the ones they had already heard in the previous parts of the experiment, and instructed to write their response as soon as possible after each stimulus is played. They were also told that a short beep would be played after every tenth stimulus. After the experimental session, the responses were hand-copied onto the database.

3.3. Data Analysis

3.3.1. D-prime (d')

The raw scores of 4IAX task were transformed into d-prime (d') values. D' is a measure of sensitivity developed within the signal detection theory (SDT), which indicates the observer's ability to discriminate a signal-plus-noise stimulus from a noise-alone stimulus and computed on a basis of rates of Hit and False Alarm (Macmillan and Creelman, 2005). From data collected in the 4IAX task, in which signal represents 'acoustic difference in the 1st pair', observer's response was coded either as Hit, Miss, False Alarm, or Correct Rejection, as shown in Table 2. A Hit rate was the ratio of the number of Hits to the number of Hits plus Misses, and a False Alarm rate was the ratio of the number of False Alarms to the number of False Alarms plus Correct Rejections (see Table 3). Finally, the value for d' was the Z value of the Hit-rate minus that of the False-Alarm-rate (see Table 3).

Table 2: Stimulus-response matrix for the 4IAX discrimination paradigm

Presentation	Response "1st pair"	Response "2nd pair"
<AB – AA> <AB – BB> <BA – AA> <BA – BB>	Hit	Miss
<AA – AB> <AA – BA> <BB – AB> <BB – BA>	False Alarm (FA)	Correct Rejection (CR)

* signal: difference is in the '1st' pair

Table 3: calculation for Hit rate, False Alarm rate, and d' value

Hit rate = #Hit / (#Hit + #Miss)
FA rate = #FA/ (#FA + #CR)
D' = Z (Hit rate) – Z (FA rate)

3.3.2. Apex of a parabolic curve

As a way of defining the point of highest discriminability and lowest goodness rate on a stimulus continuum, an apex of a parabolic curve fitting the paired data (i.e. stimulus numbers and corresponding mean d'/rating) was obtained by using the response in the discrimination task and rating task by English /a/ group, English /o/ group, Japanese /a/ group, and Japanese /o/ group for each of English and Japanese stimuli. To do so from the 4IAX discrimination data, each stimulus pair was re-coded using the number of 'skipped stimuli'. Thus, for example, the stimulus pair 0-2 was re-coded as 1, the stimulus pair 1-3 was re-coded as 2, and so on. A parabolic curve was defined by using the three data points near the apex (i.e. highest d' and two adjacent d' values/lowest rate and two adjacent rates) in such a way as to minimize the sum of the squared residuals. An example of three data-points plot and a parabolic least-squares fit is shown in Figure 2.

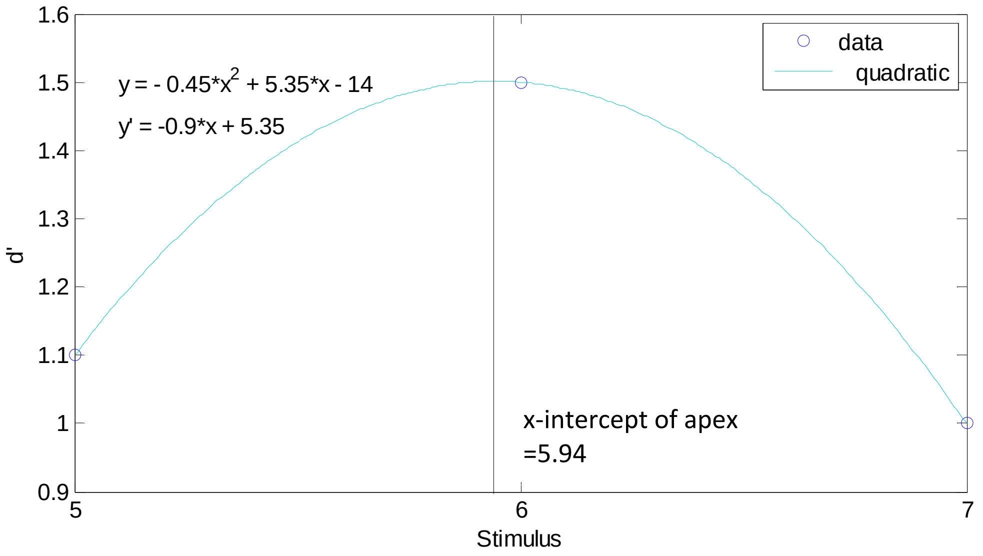

Figure 2. Example of parabolic least-squares fit by using three data points, showing a location of apex on a stimulus continuum. Three data points ((5, 1.1), (6, 1.5), (7, 1.0)) are English /a/ group's mean d' values for Japanese stimuli 4-6, 5-7, and 6-8 pairs.

From the obtained quadratic function of the parabolic fit, x-intercept of apex (i.e. a root of the first derivative of the quadratic function) was computed. The x-intercept of apex indicates the center of most discriminable pair or the point of lowest goodness rating on the stimulus continuum, which would be then compared by listener groups[5].

3.3.3. Statistical Tests

Initial observation of the data revealed that responses of the /a/ group of listeners were qualitatively and quantitatively different from those of /o/ group of listeners. Thus, the data analysis (plots and statistical tests) were performed separately for the data obtained from /a/ and /o/ groups.

For the 4IAX data obtained in block 1, Repeated Measure ANOVA was conducted for the following factors: between group factor – Listener's Language (2 levels: English and Japanese); and within group factors – Talker's Language (2 levels: English and Japanese) and Stimulus (8 levels: Stimulus 0-2 pair to Stimulus 7-9 pair). Mauchly's tests revealed that the current data does not satisfy the sphericity condition along some of the factors. In such cases, the Greenhouse-Geisser corrections were applied to produce valid F-ratios (Keppel and Wickens, 2004: 378).

For the boundary data obtained from block 2, two-tailed t-tests were performed to test the Listener's Language effect. Four separate t-tests—for English /a/, English /o/, Japanese /a/, and Japanese /o/ continua—were performed. Because the question asked in the t-test—whether English listeners and Japanese listeners draw phonemic boundary at different location on the [k-t] continuum or not—was a primary question, no family-wise error rate correction was employed (Keppel and Wickens, 2004: 113). Levene's tests revealed that the some of the current data do not satisfy the equal variance condition. Corrections were applied both on the calculation of the t-statistics and degree of freedom to produce valid t-statistics and p-values.

3.4. Results and discussion

3.4.1. Block 1 - 4IAX discrimination

Figure 3 provides the line plots of mean d' for each stimulus pair (0-2 pair to 7-9 pair) in each talker's language (English and Japanese) by each listener group (English /a/, English /o/, Japanese /a/, and Japanese /o/ group of listeners) together with vertical lines representing the location of d' apex by the two listener groups (data are presented in Appendix). Overall, English listeners and Japanese listeners responded to the stimuli in a similar manner in that the highest d' for both groups was for the same stimulus pair in a given continuum (i.e. 2-4, 3-5, 5-7, and 2-4 pair, in English /a/, English /o/, Japanese /a/, and Japanese /o/ continuum, respectively), and low scores were for the stimulus pairs from the both ends of continua. The similar response patterns by the two language groups are most clearly observed in Japanese /a/ continuum. Figure 3B shows the two plots exhibiting almost identical profiles.

---

[5] It might have been desirable to characterize the peak location in each listener's d' function, however the individual d' data are quantized fairly roughly (due to the rather small number of observations) and as a consequence a peak finding algorithm cannot be reliably used on individual data.

Interestingly, the d' peaks fall at drastically different locations between English and Japanese /a/ continua (on 2-4 pair and 5-7 pair, respectively). The difference in the location of d' peak is present in /o/ continua, as well (3-5 pair in English and 4-6 pair in Japanese continuum). However, the magnitude of the gap is much smaller in these continua. Location of d' peak seems to be a function of the particular stimulus continuum, indicating perhaps that place cues were more or less strong in the particular token chosen as end points. This factor is of less interest than the comparison of different subject groups for the same stimuli.

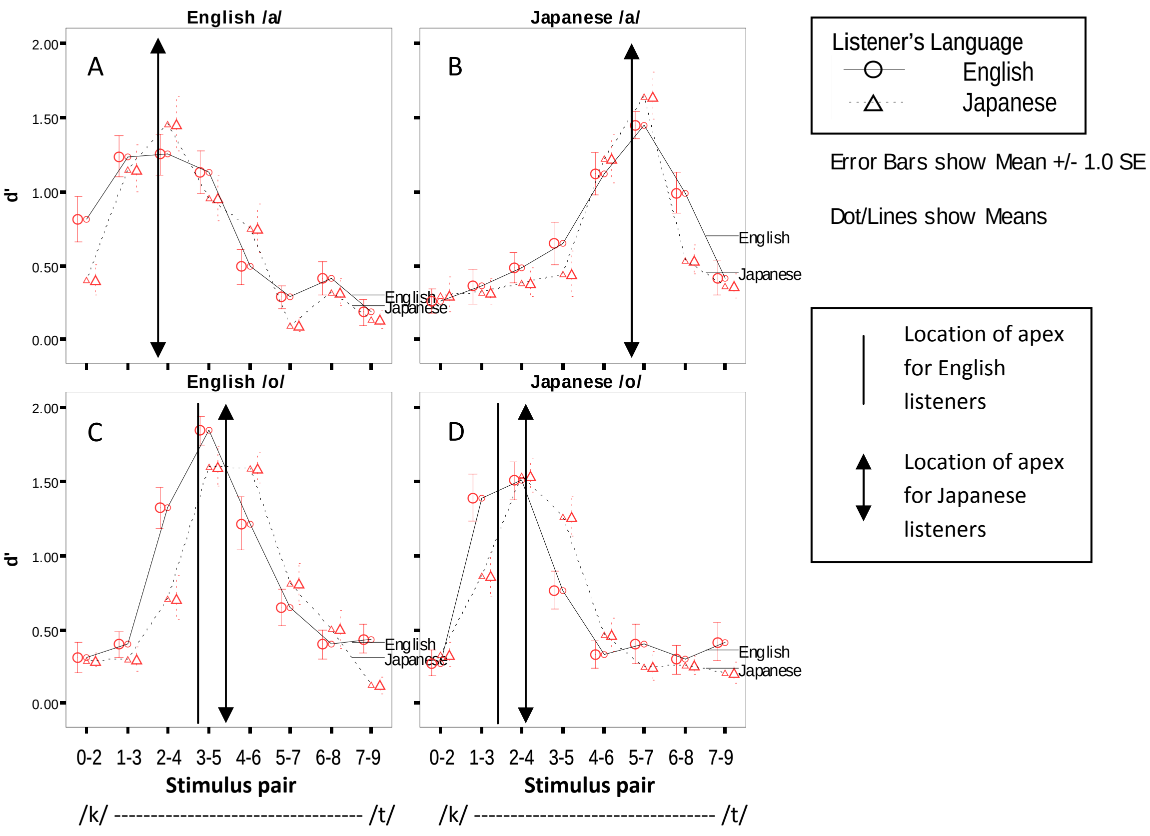

Figure 3. Line plots of mean d-prime for stimulus pair 0-2 to 7-9.
(A): English /a/ pairs (B): Japanese /a/ pairs
(C): English /o/ pairs (D): Japanese /o/ pairs

Two vertical lines indicate the stimulus location of apex of d' function for each group of listeners.
(In A and B, the distance of the two lines are too short to be shown. The location of apex for English listeners is slightly leftward in A and slightly rightward in B.)

Another point of interest is that English and Japanese listeners exhibited more similar response patterns in /a/ condition than in /o/ condition. The similarity is reflected in the similar profiles for the two groups in the Figure 3A and 3B. In /o/ condition, on the other hand, English group scored better for stimuli of near the /k/ end of the continuum than Japanese group did, as reflected in the diverging profiles in the Figure 3C and 3D. The same patterns were observed in the stimulus locations of apex, as well. The two groups' x-intercepts of apex were much closer in /a/ condition than in /o/ condition.

The results of the statistical tests were as follows. As for the main effect, the Listener's language did not have significant main effect in either /a/ or /o/ condition [F (1, 26) = 0.91, p = 0.34 in /a/ condition, F (1, 35) = 0.97, p = 0.33 in /o/ condition]; that is, the mean d' for each listener group, when data were pooled, ignoring all other factors, were not significantly different from each other. Talker's language had significant main effect only in /o/ condition: In this condition, the mean d' is lower for Japanese stimuli than for English stimuli [F (1, 26) = 0.0, p = 0.83 in /a/ condition, F (1, 35) = 10.91, $p < 0.01$ in /o/ condition] (see Figure 3). As expected, there was a significant effect of stimulus in both /a/ and /o/ conditions [F (5.3, 136.5) = 14.7, $p < 0.01$ for /a/ condition, F (4.2, 145.8) = 54.84, $p < 0.01$ for /o/ condition].

As for the interactions, there was a significant Stimulus x Listener's Language interaction only in /o/ condition: In this condition, English listeners scored higher for the stimulus pair of relatively lower number, while Japanese listeners did so for the stimulus pair of relatively higher number [F (5.3, 136.5) = 1.49, p = 0.19 in /a/ condition, F (4.2, 145.8) = 3.34, $p < 0.05$ in /o/ condition]. This result suggests that the two listener groups differ in the location of highest sensitivity on the /oko/-/oto/ continua (see Figure 4).

In both /a/ and /o/ conditions, there were significant Stimulus x Talker's Language interactions [F (4.9, 128.4) = 43.42, $p < 0.01$ in /a/ condition, F (4.7, 164.7) = 26.74, $p < 0.01$ in /o/ condition]. In /a/ condition, Japanese stimuli elicited higher d' for the stimulus pair of relatively higher number comparing with English stimuli; In /o/ condition, on the other hand, it was English stimuli that elicited higher d' for the stimulus pair of relatively higher number (see Figure 5).

Finally, only in /o/ condition, there was a significant three-way Talker's Language x Stimulus x Listener's Language interaction [F (4.9, 128.4) = 1.55, p = 0.18 in /a/ condition, F (4.7, 164.7) = 3.56, $p < 0.01$ in /o/ condition]; that is, discriminability differed across stimulus pairs and relative discriminability of each stimulus pair differs between two listener groups with different patterns in English and Japanese stimuli. For example, as Figures 3C and 3D display, English listeners scored higher than Japanese listeners for some of the stimulus pairs, but the two groups' relative performances were reverse for other stimulus pairs. Also, there was a clear listener group difference in d' for the most discriminable English stimulus (3-5 pair), while for the most discriminable Japanese stimulus (2-4 pair) d' were comparable between the two groups.

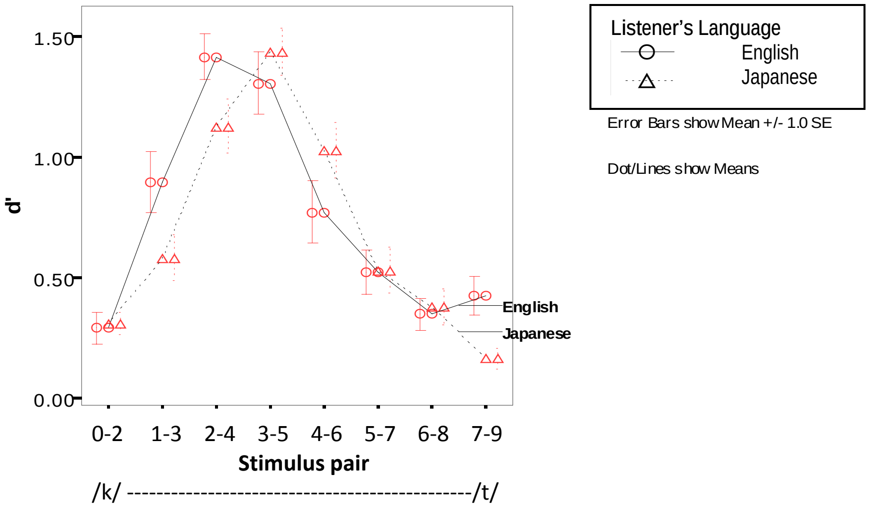

Figure 4. Line plots of the Mean d' showing Stimulus x Listener's Language interaction in the /o/ condition.
Data are pooled for other variables (i.e. Talker's Language).

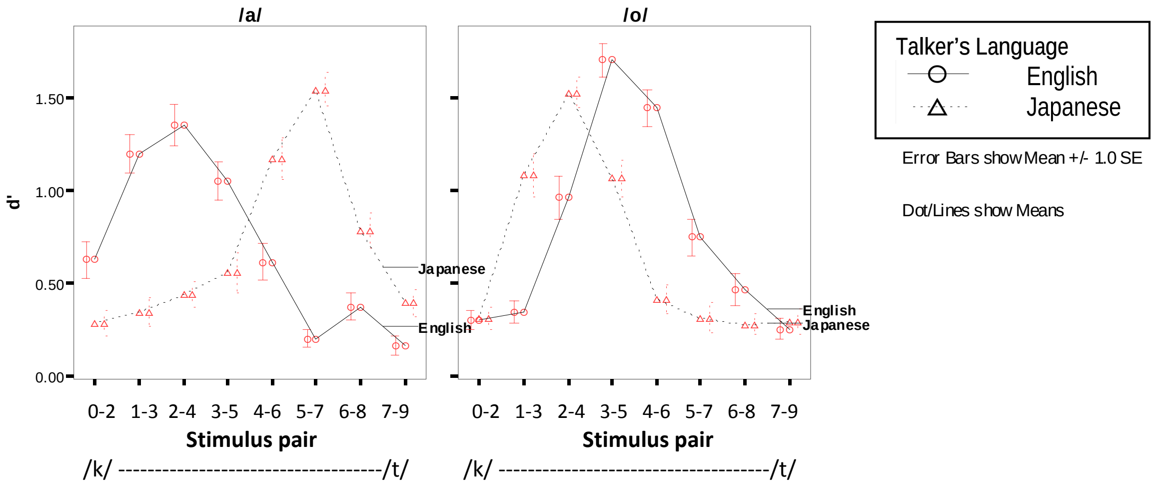

Figure 5. Line plots of the Mean d' showing Stimulus x Talker's Language interaction in the /a/ condition (left panel) and in the /o/ condition (right panel).
Data are pooled for other variables (i.e. Listener's Language).

In summary, there were significant effects of the factors regarding the stimuli: by splitting the stimuli along vowel dimension, we observed that the /a/ stimuli and /o/ stimuli elicited different response patterns. Similarly, by using stimuli spoken by English and Japanese native speakers, we observed that the English stimuli and Japanese stimuli elicited the different responses. On the other hand, there was no significant main effect of the listener's language factor. The effect of listener's language was not seen in the overall d' score, but in the location of the highest d' on the continua; that is, d' peak for the two groups were different, with the peak for the English group being slightly toward /k/ end comparing with the Japanese group's peak location. These results suggest that assumed difference in the listener's linguistic experience (due to the different phoneme frequency of /t/ and /k/ in the two languages) did not influence discriminability of the stimuli as a whole, but did influence the location of highest sensitivity revealed in the 4IAX discrimination task.

3.4.2. Block 2 – phoneme boundary decision

Table 4 provides the phonemic boundary locations drawn by the English and Japanese group of listeners for the English /aka/-/ata/ English /oko/-/oto/, Japanese /aka/-/ata/, and Japanese /oko/-/oto/ continua. In the same table, the stimulus locations of the apex of d' function from block 1 are presented for the purpose of comparison of the two results. As expected from categorical perception and the perceptual magnet effect, the phonemic boundary and the location of the d' peak (i.e. the center of the most accurately discriminable pair) coincide or nearly coincide. Figure 6 provides the plots of the same data. In three out of the four conditions, Japanese listeners drew the boundary closer to the /t/ end than the English listeners did. The t-tests revealed that the boundary location was significantly different between the two groups in all but Japanese /a/ continuum [$t(21.6) = -2.3$, $p < 0.05$ in English /a/ condition; $t(33.2) = -2.1$, $p < 0.05$ in English /o/ condition; $t(25.5) = 0.60$, $p = 0.55$ in Japanese /a/ condition; and $t(31.4) = -3.4$, $p < 0.01$ in Japanese /o/ condition]. These results are consistent with the results of the 4IAX discrimination task. There is a systematic difference in the two groups of listeners such that Japanese listeners identify more of the tokens as /VkV/ than do English listeners.

Table 4: Phonemic boundary for the English and Japanese continua (columns) by English and Japanese listeners (nested columns) in /a/ and /o/ conditions (rows).
Within each thick-lined cell are phonemic boundary and location of the apex of d' (in square bracket).

	English continua Boundary in stimulus # [Stimulus location of d' apex]		Japanese contiunua Boundary in stimulus # [Stimulus location of d' apex]	
Vowel	E Group (n=15)	J Group (n=13)	E Group (n=15)	J Group (n=22)
/a/	3.0 [2.83]	3.5 [2.87]	5.9 [5.9]	5.7 [5.8]
/o/	4.1 [3.96]	4.7 [4.5]	3.2 [2.65]	4.1 [3.25]

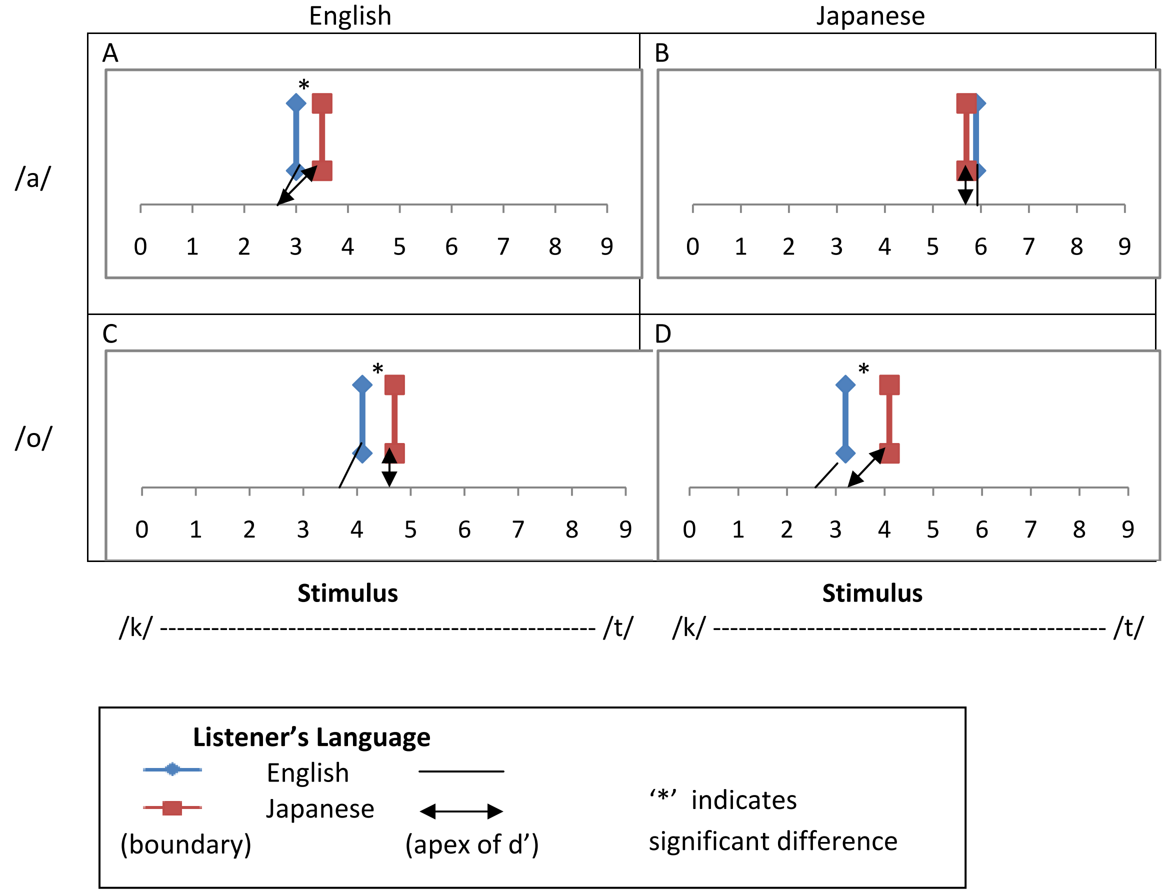

Figure 6. Phoneme boundary (and location of apex of d') along /VkV/ - /VtV/ continuum drawn by English and Japanese Group.
(A): English /a/ continuum (B): Japanese /a/ continuum
(C): English /o/ continuum (D): Japanese /o/ continuum

3.4.3. Block 3 – category goodness rating

Figure 7 provides the line plots of the mean goodness rating for each stimulus (0 to 9) in each talker's language (English and Japanese) by each listener group (English /a/, English /o/, Japanese /a/ and Japanese /o/ group of listeners) together with the location of d' apex by the two listener groups (summary data are presented in Appendix). Overall, the two groups of listeners rated the stimuli in a similar manner in that both groups gave relatively higher ratings for the stimuli at the both ends of the continua and lower ratings for the stimuli in the middle of the continua. Also, both groups gave generally lower rating to the stimuli of the foreign language. This pattern suggests that both groups heard some kind of 'foreignness' or 'badness' from the Stimuli of the non-native language. Interestingly, the English listeners gave much lower rating for Japanese stimuli than Japanese listeners did for English stimuli. This is illustrated by that the gaps between the two plots in Figure 7B and 7D are much wider than those in Figure 7A and 7C.

The gaps were particularly large at stimulus 9 of Japanese /a/ and /o/, suggesting that Japanese /ata/ and /oto/ sounded particularly 'bad' to the ears of the English speaking listeners. As was the case for the discrimination data, rating data also exhibited listener group difference in the location of apex of rating function. Japanese listener's locations of rating apex were closer to the /t/ end than that of English listeners in all but Japanese /aka/-/ata/ continuum. For this continuum, location of d' apex found in the block 1 and phoneme boundary found in the block 2 were also closer to the /k/ end for Japanese listeners than for English listeners.

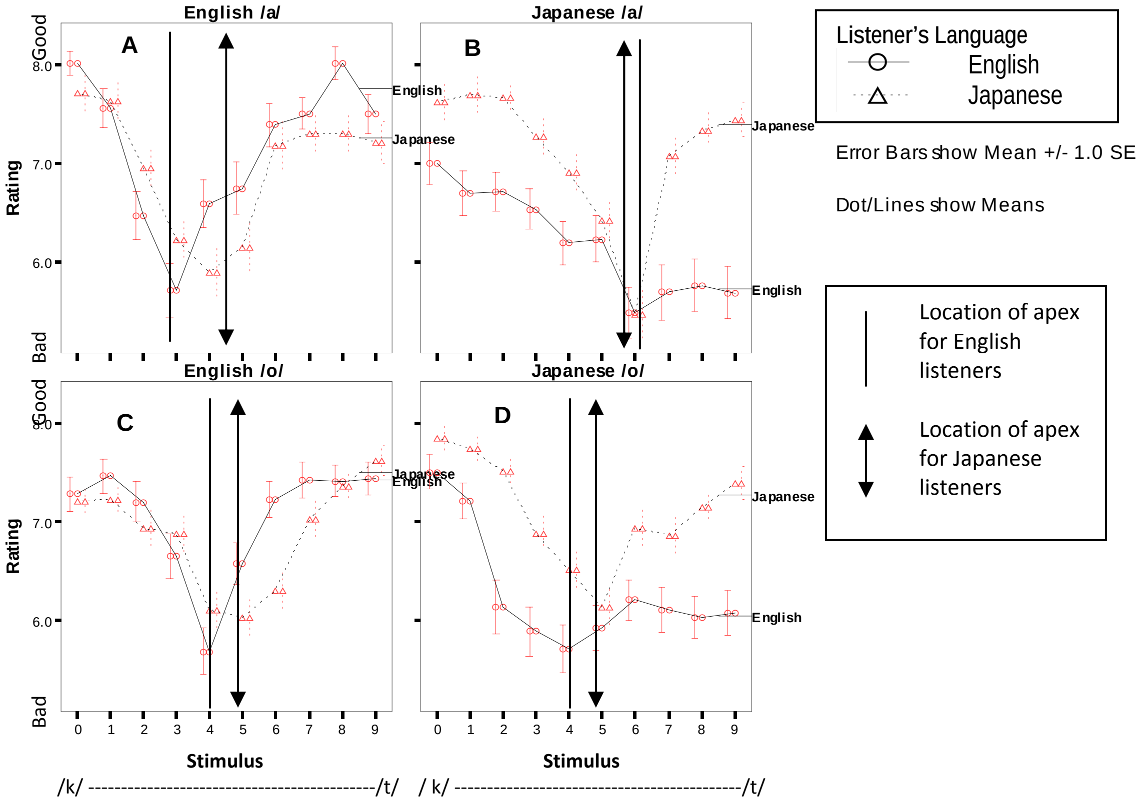

Figure 7. Line plots of mean Rate for stimulus 0 to 9.
(A): English /a/ series (B): Japanese /a/ series
(C): English /o/ series (D): Japanese /o/ series

The vertical lines overlaid on the line plots indicate the location of apex for each listener group.

The results of the statistical tests were as follows. As for the main effect, the Listener's Language effect was not significant in either /a/ or /o/ condition [*F* (1, 26) = 0.71, *p* = 0.41 in /a/ condition, F (1, 35) = 0.57, p = 0.46 in /o/ condition]; that is, the mean rating for the groups were not significantly different from each other. Talker's language also did not have significant main effect [F (1, 26) = 3.20, p = 0.09 in /a/ condition, F (1, 35) = 3.74, p = 0.06 in /o/ condition]. As expected, there was a significant effect of stimulus for both /a/ and /o/ group of listeners [*F* (5.2, 136.3) = 11.35, $p < 0.01$ for /a/ condition, F (5.2, 182.5) = 21.68, $p < 0.01$ for /o/ condition].

As for the interactions, there was a significant Stimulus x Listener's Language interaction only in /o/ condition [F (5.2, 136.3) = 1.80, p = 0.11 in /a/ condition, F (5.3, 182.5) = 2.39, $p < 0.05$ in /o/ condition]: In this condition, the rating profiles of the two groups diverge in a way that the Japanese rating function shifts upward and rightward from that of English group (see Figure 8). In both /a/ and /o/ conditions, there were significant Stimulus x Talker's Language interactions [F (4.2, 109.8) = 10.78, $p < 0.01$ in /a/ condition, F (5.4, 190.6) = 4.88, $p < 0.01$ in /o/ condition]. However, the pattern was different in the two conditions: In /a/ condition, the rating function for Japanese stimuli was somewhat shifted to lower and rightward position from the rating function for English stimuli; In /o/ condition, on the other hand, the two plots had more similar profiles except that Japanese stimuli toward /k/ end had higher ratings than English counterpart and English stimuli toward /t/ end had much higher rating than Japanese counterpart (see figure 9). Finally, there was no significant three-way Talker's Language x Stimulus x Listener's Language interaction [F (4.2, 109.8) = 2.05, p = 0.09 in /a/ condition, F (5.4, 190.6) = 1.28, p = 0.27 in /o/ condition].

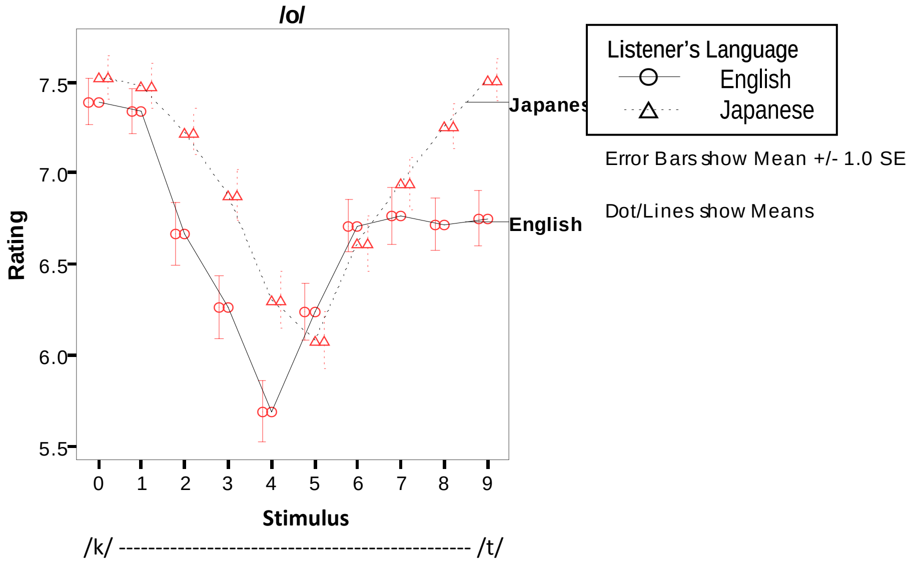

Figure 8. Line plots of the mean rating showing Stimulus x Listener's Language interaction in the /o/ condition.
Data are pooled for other variables (i.e. Talker's Language).

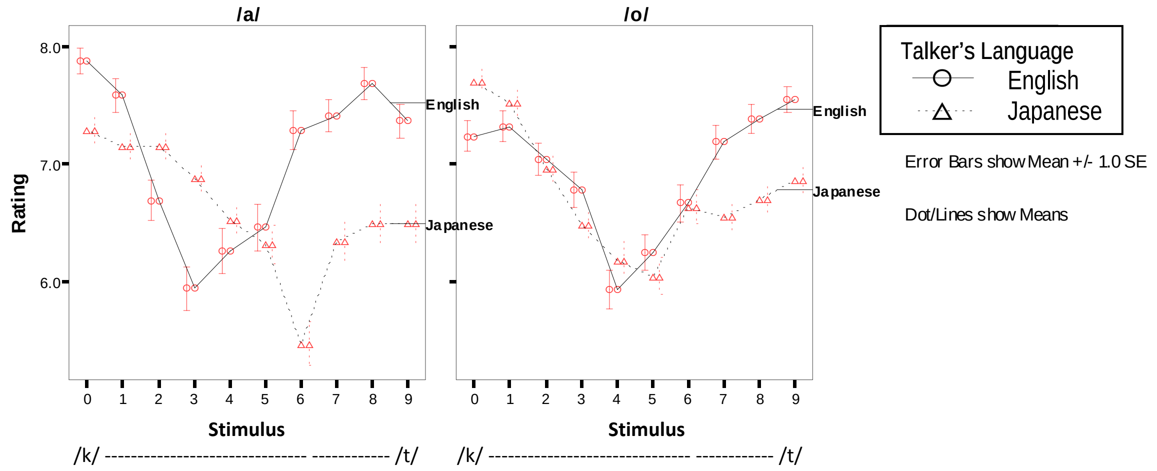

Figure 9. Line plots of the mean rating showing Stimulus x Talker's Language interaction in the /a/ condition (left panel) and in the /o/ condition (right panel). Data are pooled for other variables (i.e. Listener's Language).

In summary, stimuli toward the end of the continua tended to receive higher ratings than the stimuli in the middle. English listeners gave lower ratings than Japanese listeners to /o/ stimuli: English listener's low ratings for Japanese /o/ stimuli seem to contribute to this result. Japanese stimuli near the /t/ end of the continuum received lower ratings than English counterparts: Again, particularly lower ratings given by English listeners for these Stimuli seem to contribute this result. Although not a direct support, a significant Listener's Language x Stimulus interaction suggests that the two groups are different in the location of valley of the rating profiles at least in /o/ condition; that is, the worst stimulus on the continua differs for the two groups. This difference is also indicated by the location of rating apex. In this respect, it seems to be safe to conclude that English listeners and Japanese listeners rated stimuli differently. These results seem to show two different rating behaviors of the listeners. On one hand, English and Japanese listeners gave similar relative rating for each stimulus; that is, the two groups had similar judgments on which stimulus is better/worse than which. On the other hand, the two groups were different in absolute score they gave to each stimulus, and the gap was particularly large for Japanese stimuli on the /t/ end of the continuum.

As a summary of the entire experiment, the results from the three blocks are presented in Figure 10. Also, Table 5 provides the Stimulus that elicited the highest d' in the 4IAX task, phoneme boundary obtained in the up-down task, and the stimulus that received the lowest rating in the goodness rating task for each vowel condition (row), for each talker's language (column) by each listener group (nested column). These are three different indices for the location of the /k/-/t/ boundary on the continuum, and the theory of categorical perception predicts that these three measures coincide. As shown in Figure 10, these three measures roughly coincide in all conditions; that is, the peak of d' function, the valley of rating function, and the phoneme boundary all fall in the same location on the continua. The only apparent exception is Japanese listeners' behavior in Japanese /a/ condition (shown in the lower right panel in Figure 10A; in this condition, the Phonemic boundary seems to fall more leftward position than expected from the locations of the peak mean d' and valley of mean rating.

The perceptual magnet effect hypothesizes that d' is predictable from the goodness rating. This seems to be supported by small but nonetheless significant correlation[6] between rating and d' in both /a/ and /o/ conditions [$r = -0.21$, $t(26) = -4.58$, $p < 0.01$ in the /a/ condition, and $r = -0.13$, $t(33) = -3.13$, $p < 0.01$ in the /o/ condition]. However, a closer look at the results raises some doubt on this interpretation of the correlation. For the stimuli at or near the phonemic category boundary, dicriminability scores were high and goodness ratings were low, in accordance with the paerceptual magnet effect hypothesis. However, within category, the hypothesized predictability of discrimination performance from the goodness rating seemed to be rather weak. For example, there was a large listener group difference in the ratings for the Japanese stimuli near the /t/ end but the d' scores were not particularly different between the groups for these stimuli. The small correlation between the d' and ratings might be entirely due to the match of high d' and low ratings for the stimuli at or near the phonemic boundaries, which supports categorical perception rather than the perceptual magnet effect.

As for the effect of listener's language, statistical tests for the d' and rating showed a Listener's Language x Stimulus interaction. This indicates that Japanese and English listeners differed in the location of peak d' and valley of goodness rating. In fact, the t-tests testing the Listener's Language effect in phoneme boundary position revealed a significant effect in all conditions except for Japanese /a/ condition. These results suggest that the linguistic knowledge strongly influences not only categorization of auditory stimuli but also sensitivity to the acoustic difference of stimuli and subjective opinion on the acoustic pattern as well.

Table 5: Comparisons of the three indices for the /k/-/t/ boundary. Within each thick-lined cell are: Stimulus # of the highest d' in block 1 (top), phonemic boundary in block 2 (middle), and stimulus # of the lowest rating on block 3 (bottom).

Vowel	English Stimuli [Stimulus # of the highest d' in Block1] Phoneme boundary in Block2 (Step # of the lowest Rate in Block3)		Japanese Stimuli [Stimulus # of the highest d' in Block1] Phoneme boundary in Block2 (Step # of the lowest Rate in Block3)	
	English Group	Japanese Group	English Group	Japanese Group
/a/	[3] 3.0 (3)	[3] 3.5 (4)	[6] 5.9 (6)	[6] 5.7 (6)
/o/	[4] 4.1 (4)	[4/5] 4.7 (5)	[3] 3.2 (4)	[3] 4.1 (5)

---

[6] Correlation coefficients were calculated by using d' data from all pairs (i.e. 0-2 pairs to 7-9 pairs) and rating data only from stimulus 1 to 8. Thus, the obtained r values do not reflect the rating for stimulus 0 and stimulus 9.

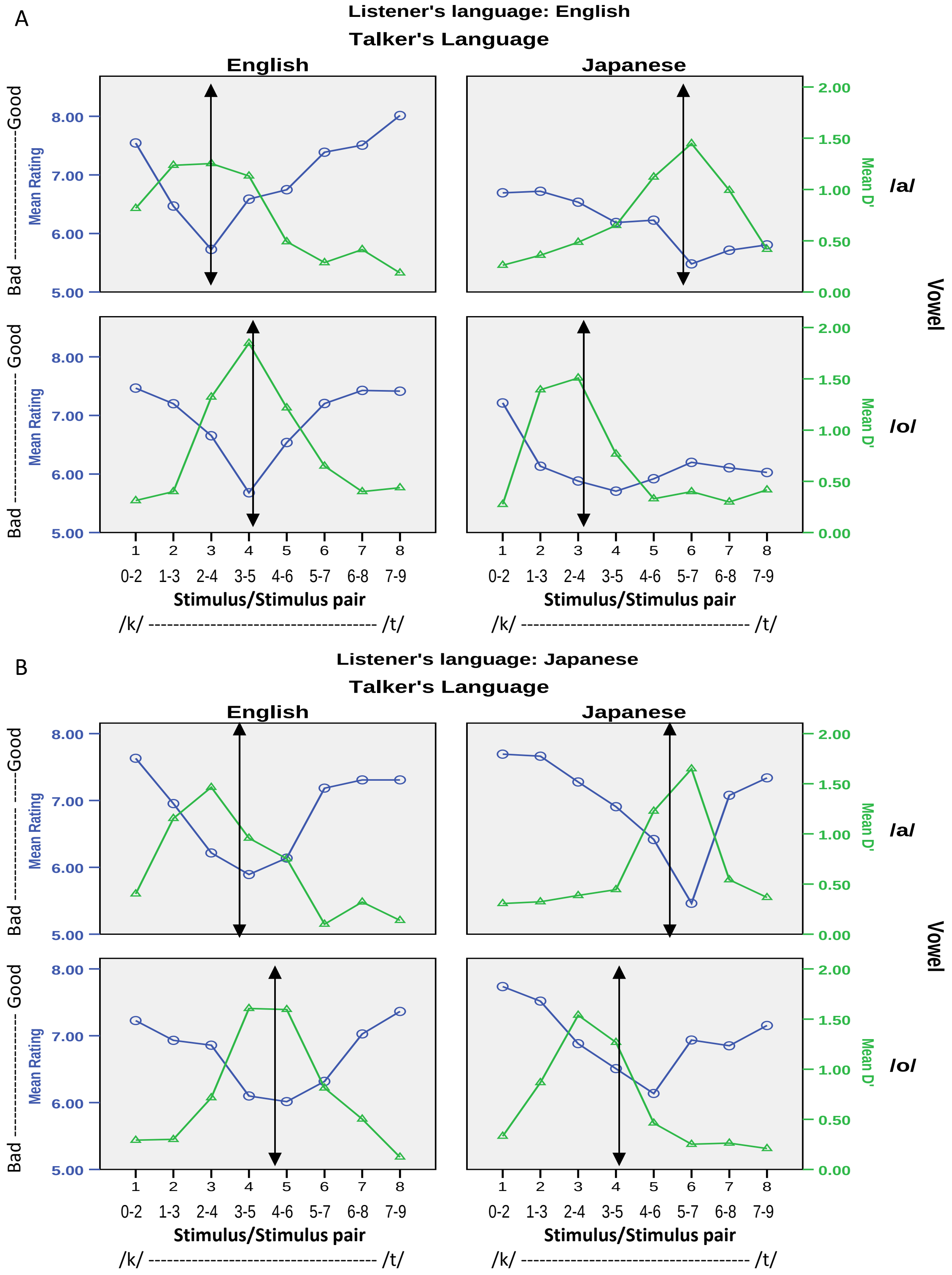

Figure 10. Line plots of mean ratings, mean d' values, and phoneme boundaries (vertical line)

3.5. Interim summary

The experimental study reported in this paper was aimed at investigating: 1) how phoneme frequency influences speech perception by English and Japanese speaking listeners; and 2) whether acoustic-auditory representations of speech sounds are shaped by the perceptual magnet effect. The results were somewhat mixed.

As for the effect of phoneme frequency, Guenther and Gjaja's (1996) neural network model for the perceptual magnet effect predicts that more frequent phoneme leads to much greater magnet effect than less frequent phoneme. Since /k/ is more frequently used than /t/ in Japanese, and the frequency distribution pattern is opposite in English, it was expected that Japanese listeners possess more heavily warped perceptual space near /k/ than English listeners do and therefore draw boundary in a way to have larger /k/ region on the acoustic scale than English listeners do. This expectation was tested in the block 2 and met in three out of four conditions, with Japanese /a/ condition as an exception. It is thus not unreasonable to conclude that the perceptual space used in speech perception is influenced by phoneme frequency.

This result implies that one aspect of the phoneme frequency effect on speech perception is a bias on phoneme categorization. The low goodness rating scores for the tokens near the phoneme boundary indicates that the listeners perceive these stimuli as 'poor' or ambiguous instances of a category. The different boundary locations for Japanese and English listeners then suggest that when listeners encounter ambiguous stimuli, they are biased to perceive them as member of the more frequent phoneme.

Whether the frequency effect was brought about by warping of perceptual space or not is not answered by the result of block 2 alone because the observed frequency effect could occur with or without warping. The perceptual space for frequent phonemes could be larger than for less frequent phonemes, or it could have more greatly warped region for frequent phonemes than for less frequent phonemes. The effect of warping needs to be tested against discrimination sensitivity and goodness ratings.

The d' and goodness rating data clearly indicate that two sounds that straddle a phoneme boundary are particularly easy to discriminate and these peripheral members are perceived as particularly poor members of a category. However, as mentioned in the previous section, within-category d' and goodness rating do not seem to correlate well. Overall, the data obtained from the current study seem to be explained sufficiently in terms of phoneme frequency effect and categorical perception, but not necessarily of the perceptual magnet effect.

4. General discussion

There is growing body of study showing the effect of the lexicon on speech perception. Much of effort has been devoted in the investigation of the effect of lexical status on identification of acoustically ambiguous speech stimuli (Ganong, 1980; Connine, 2004; Whalen, et al., 1997) and the effect of neighborhood density on speed and accuracy of spoken word recognition (Luce and Pisoni, 1998; Sussman, et al., 2003). The current study adds another piece of evidence that the linguistic knowledge stored in the mental lexicon is reflected in the way listeners perceive speech, by showing that the difference between the ways English and Japanese speaking listeners react to the speech stimuli is in accord with differences of phoneme frequency in the two languages.

The fact that language-specific perception was observed in the 4IAX task, which has low-memory load and therefore expected to by-pass lexical access, has significant implications for the theory of speech perception because many approaches of speech perception assume that the acoustic-auditory capacity underlying speech perception is not influenced by linguistic experience (Johnson, 2003; see also papers in Hume and Johnson, 2001) but available evidence is mixed on this point. For example, recent neuroscience studies offer evidence of speech non-specific, thus universal, auditory processing (see review by Scott and Johnsrude, 2003). Other studies, however, offer evidence that linguistic experience alters auditory processing even at the brain stem level (Krishnan, et al., 2005). The results of the current study do not support a general physiological view of speech perception. The fact that language-specific phoneme frequency affects d' in the 4IAX task suggests that low-level auditory processing is influenced by linguistic experience.

The results also cast some doubt on the current interpretation of the 'perceptual magnet' concept. In its original formulation (Kuhl, 1991; Iverson and Kuhl, 1995) the perceptual magnet effect was introduced as a prototype-based effect. Later, Guenther and Gjaja (1996) proposed a neural model that replaces 'prototype' with the most frequently exposed stimuli in a given language; and 'category goodness' , with the number of cells in the brain's auditory map which are preferentially activated by input. However, the current study did not show a phoneme frequency effect on discriminability in the way Guenther and Gjaja's neural map model predicts. English listeners did not show higher sensitivity for within-category /VtV/ stimuli vs. within-category /VkV/ stimuli, nor did the Japanese listeners showed higher sensitivity for /VtV/ Stimuli than for /VkV/ Stimuli. Therefore, the link between frequency of exposure to the certain acoustic pattern and warping in the perceptual space is questioned. Guenther, et al. (1999) also discussed the possibility for dissociation between frequency distribution of input and changes in the auditory map. We need more empirical data to better understand the underlying mechanism of the observed listeners' response.

Finally, it has to be pointed out that difference in phonemic boundary may or may not reflect listeners' knowledge of phoneme frequency in their native language. This could be caused solely by the listeners' knowledge of the detailed acoustic realization of the native phonemes. For example, Best and Strange (1992) compared Japanese and English listeners' perceptions of English approximants using a synthesized /w/-/j/ continuum, and observed that Japanese listeners drew a phonemic category boundary more to the /j/ end compared with American listeners. This difference in the boundary location was explained in terms of the phonetic differences between the English and Japanese /w/. The Japanese /w/ involves less lip-rounding than English counterpart; thus, Japanese listeners perceived a stimuli with intermediate level of lip-rounding, which English listeners perceive as /j/, as /w/. However, in the case of /t/ and /k/, there is no basis to assume that Japanese [t] is similar to English [k] in any of the phonetic properties. Therefore, although there is a possibility of attributing the observed boundary location difference to the different acoustic properties of English and Japanese [t]s and [k]s, at this point it seems more sensible to attribute the differences to the phoneme frequency difference in the two languages.

References

Abramson, A.S. & Lisker, L. (1970) Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences.* pp. 569-573.

Beddor, P. S. & Strange, W. (1982) Cross-linguistic study of perception of the oral-nasal distinction. *Journal of Acoustical Society of America*, 71: 1551-1561.

Best, C.T. and Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20: 305-330.

Connine, C.M. (2004) It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review,* 11: 1084-1089.

Fox, R.A. (1984) Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10: 526-540.

Fujisaki, H., & Kawashima, T. (1970). Some Experiments on Speech Perception and a Model for the Perceptual Mechanism. *Annual Report of the Engineering Research Institute Raculty of Engineering*. University of Tokyo, 29: 207.

Ganong, W.F., III. (1980) Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance,* 6: 110-125.

Ganong, W.F., III and Zatorre, R.J. (1980) Measuring phoneme boundaries four ways. *Journal of the Acoustical Society of America*, 68: 431-439.

Grether, C.B. and Stroh, R.W. (1973) Subjective Evaluation of Differential Pulse-Code Modulation Using the Speech "Goodness" Rating Scale. *IEEE transactions on audio and electroacoustics*, AU-21: 179-184.

Guenther, F.H. and Gjaja, M.N. (1996) The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, 100: 1111-1121.

Guenther, G.H., Husain, F.T., Cohen, M.A., and Shinn-Cunningham, B.G. (1999) Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, 106: 2900-2912.

Hume, E. & Johnson, K. eds. (2001) *The Role of Speech Perception in Phonology*. New York: Academic Press.

Iverson, P. & Kuhl, P.K. (1994) Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97: 553-562.

Johnson, K. (2003). Cross-linguistic Perceptual Differences Emerge from the Lexicon. *Proceedings of the 2003 Texas Linguistics Society Conference*. pp. 26-41.

Keating, P. A., Mikoś, M. J. & Ganong III, W.F. (1981) A cross-language study of range of Voice Onset Time in the perception of initial stop voicing. *Journal of the Acoustical Society of America*, 70: 1261-1271.

Keppel, G. and Wickens, T. (2004) *Design and Analysis: A Researcher's Handbook* (4th ed.) Upper Saddle River, NJ.: Prentice Hall.

Kuhl, P.K. (1991) Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics,* 50: 93-107.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992) Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, 255: 606-608.

Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25: 161-168.

Liberman, A.M., Harris, K.S., Hoffman, H.S. & Griffith, B.C. (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54: 358-368.

Liberman, A.M., Harris, K.S., Eimas, P.D., Lisker, L. & Bastian, J. (1961) An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, 4: 175-195.

Liberman, A.M., Harris, K.S., Kinney, J.A. & Lane, H.L. (1961) The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 61: 378-388.

Lisker, L., & Abramson, A.S. (1970) The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*. pp. 563-567.

Lisker, L., & Abramson, A.S. (1964) A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20: 384-422.

Luce, P.A. and Pisoni D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19: 1-36.

MacKain, K. S., Best, C. T., and Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2: 369-90.

Macmillan, N.A. and Creelman, C.D. (2005) *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ.: Lawrence Erlbaum Associates.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., & Fujimura, O. (1975) An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophysics,* 18: 331-340.

Pisoni, D.B. (1971) On the Nature of Categorical Perception of Speech Sounds. Status *Report on Speech Research Haskins Labs*, SR-27: 101.

Pisoni, D.B. (1973). Auditory and Phonetic Memory Codes in the Discrimination of Consonants and Vowels. *Perception and Psychophysics*, 13: 253.

Pisoni, D.B. & Lazarus, J.H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 55: 328-333.

Repp, B.H. (1984) Categorical Perception: Issues, methods, and findings. In N. Lass (ed.) *Speech and language*: *Advances in Basic Research and Practice*. Vol. 10 (pp. 244-335). Orlando, FL: Academic Press.

Scott, S.K. and Johnsrude, I.S. (2003) The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences,* 26: 100-7.

Strange, W. & Jenkins, J. (1978). Role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick (Eds.), *Perception and experience* (pp. 125-169). New York: Plenum Press.

Sussman, J.E., Dee, D., and Curcio, D. (2003). Neighborhood effects on speech-perception abilities of adults and children aged 6-9 years. Journal of the *Acoustical Society of America*, 113: 2256-2256.

Werker, J.F., Gilbert, J.H.V., Humphrey, K., and Tees, R.C. (1981) Developmental aspects of cross-language speech perception. *Child Development*, 52: 349-53.

Whalen, D.H., Best, C.T., & Irwin, J.R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of phonetics*, 25: 501-528.

Yamada, R.A. & Tohkura, Y. (1992) The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. Perception and Psychophysics, 52: 376-392.

Yoneyama, K. (2000) *Phonological neighborhoods and phonetic similarity in Japanese word recognition*. Unpublished Ph. D. thesis, Ohio State University.

Yoneyama, K., Beckman, M. & Edwards, J. (2003) Phoneme frequency and acquisition of lingual stops in Japanese. Submitted to *Language and Speech*.

Acknowledgements

This research was supported in part by Grant R01DC004421-05 from the National Institute of Health for the second author.

We would like to thank Grant McGuire and Ronald Sprouse for their technical assistance, audience members of Phonology Phorum for valuable discussions, Yoko Hasegawa for helpful comments for revising this paper. Errors are ours.

Appendix

I. Links to the audio files

To hear the /k/ to /t/ continua that were used in this experiment, please click the links below:

Japanese, vowel /a/, aka-ata
Japanese, vowel /o/, oko-oto
English, vowel /a/, aka-ata
English, vowel /o/, oko-oto

II. Summary data from block 1

Table 1: Results of the discrimination task for each one of the English and Japanese Stimuli (columns) by English and Japanese groups (nested rows) in /a/ and /o/ conditions (rows). Within a cell the upper values are mean d-primes and lower values are standard deviations. The highest d-prime for each group for a given continuum is bold-faced.

			English Stimuli								Japanese Stimuli							
	Stimulus #		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
a	E	M	0.8	1.2	**1.3**	1.1	0.5	0.3	0.4	0.2	0.3	0.4	0.5	0.7	1.1	**1.5**	1.0	0.4
	(n=15)	SD	0.6	0.5	0.5	0.6	0.5	0.3	0.4	0.3	0.3	0.5	0.4	0.6	0.6	0.4	0.5	0.5
	J	M	0.4	1.2	**1.5**	1.0	0.8	0.1	0.3	0.1	0.3	0.3	0.4	0.4	1.2	**1.7**	0.5	0.4
	(n=13)	SD	0.4	0.6	0.7	0.5	0.6	0.2	0.3	0.2	0.4	0.3	0.3	0.5	0.6	0.6	0.3	0.3
o	E	M	0.3	0.4	1.3	**1.9**	1.2	0.7	0.4	0.4	0.3	1.4	**1.5**	0.8	0.3	0.4	0.3	0.4
	(n=15)	SD	0.4	0.3	0.6	0.4	0.7	0.5	0.4	0.4	0.3	0.6	0.5	0.5	0.4	0.5	0.4	0.5
	J	M	0.3	0.3	0.7	**1.6**	**1.6**	0.8	0.5	0.1	0.3	0.9	**1.5**	1.3	0.5	0.3	0.3	0.2
	(n=22)	SD	0.2	0.4	0.7	0.6	0.5	0.7	0.6	0.3	0.4	0.7	0.5	0.6	0.5	0.5	0.3	0.3

Table 2: Stimulus location of apex of d' function, computed from the mean d' values in Table 1, for the English and Japanese stimuli (columns) by English and Japanese groups (sub-columns) in /a/ and /o/ conditions (rows). The values in parenthesis are d' at apex.

Vowel	English Stimuli		Japanese Stimuli	
	English Group	Japanese Group	English Group	Japanese Group
/a/	2.8 (1.3)	2.9 (1.5)	5.9 (1.5)	5.8 (1.7)
/o/	4.0 (1.9)	4.5 (1.7)	2.7 (1.6)	3.3 (1.5)

III. Summary data from block 3

Table 3: Mean Goodness Rating task for each English and Japanese Basic Stimulus (columns) by English and Japanese Groups (nested rows) in /a/ and /o/ conditions (rows). Within a cell the upper values are mean and lower values are Standard Deviation.
The lowest Rate for each continuum is bold-faced.

Stimuli Language	Vowel	Subject Group	Stat	Stimulus #									
				0	1	2	3	4	5	6	7	8	9
English	a	English	M	8.0	7.6	6.5	**5.7**	6.6	6.7	7.4	7.5	8.0	7.5
		(n=15)	SD	1.0	1.7	2.1	2.4	2.1	2.3	1.9	1.4	1.4	1.7
		Japanese	M	7.7	7.6	7.0	6.2	**5.9**	6.1	7.2	7.3	7.3	7.2
		(n=13)	SD	1.6	1.7	1.9	1.9	2.4	2.3	2.2	1.8	1.7	1.7
	o	English	M	7.3	7.5	7.2	6.7	**5.7**	6.6	7.2	7.4	7.4	7.4
		(n=15)	SD	1.6	1.5	1.8	2.0	2.1	1.8	1.5	1.6	1.4	1.5
		Japanese	M	7.2	7.2	6.9	6.9	6.1	**6.0**	6.3	7.0	7.4	7.6
		(n=22)	SD	1.9	1.9	2.0	2.1	2.3	2.2	2.4	2.1	1.9	1.6
Japanese	a	English	M	7.0	6.7	6.7	6.5	6.2	6.2	**5.5**	5.7	5.8	5.7
		(n=15)	SD	1.8	1.9	1.7	1.8	1.9	2.0	2.2	2.4	2.3	2.3
		Japanese	M	7.6	7.7	7.7	7.3	6.9	6.4	**5.5**	7.1	7.3	7.4
		(n=13)	SD	1.8	1.7	1.4	1.8	1.9	1.8	2.1	1.7	1.5	1.4
	o	English	M	7.5	7.2	6.1	5.9	**5.7**	5.9	6.2	6.1	6.0	6.1
		(n=15)	SD	1.5	1.6	2.3	2.2	2.1	2.0	1.8	2.0	1.9	2.0
		Japanese	M	7.8	7.7	7.5	6.9	6.5	**6.1**	6.9	6.9	7.2	7.4
		(n=22)	SD	1.5	1.7	1.8	2.1	2.4	2.4	2.1	2.2	1.8	1.8

Table 4: Stimulus location of apex of rating function, computed from the mean rating values in Table 3, for the English and Japanese stimuli (columns) by English and Japanese groups (sub-columns) in /a/ and /o/ conditions (rows). The values in parenthesis are rating at apex.

Vowel	English Stimuli		Japanese Stimuli	
	E Group (n=15)	J Group (n=13)	E Group (n=15)	J Group (n=22)
/a/	2.9 (5.7)	4.1 (5.9)	6.3 (5.5)	5.9 (5.5)
/o/	4.0 (5.7)	4.7 (6.0)	4.0 (5.7)	4.8 (6.1)