Individual differences in cue weights are stable across time: The case of Japanese stop lengths[a)]

Kaori Idemaru[b)]
Department of East Asian Languages and Literatures, University of Oregon, Eugene, Oregon 97403

Lori L. Holt
Department of Psychology and Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15123

Howard Seltman
Department of Statistics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

(Received 8 November 2011; revised 9 October 2012; accepted 16 October 2012)

Speech categories are defined by multiple acoustic dimensions, and listeners give differential weighting to dimensions in phonetic categorization. The informativeness (predictive strength) of dimensions for categorization is considered an important factor in determining perceptual weighting. However, it is unknown how the perceptual system weighs acoustic dimensions with similar informativeness. This study investigates perceptual weighting of two acoustic dimensions with similar informativeness, exploiting the absolute and relative durations that are nearly equivalent in signaling Japanese singleton and geminate stop categories. In the perception experiments, listeners showed strong individual differences in their perceptual weighting of absolute and relative durations. Furthermore, these individual patterns were stable over repeated testing across as long as 2 months and were resistant to perturbation through short-term manipulation of speech input. Listeners own speech productions were not predictive of how they weighted relative and absolute duration. Despite the theoretical advantage of relative (as opposed to absolute) duration cues across contexts, relative cues are not utilized by all listeners. Moreover, examination of individual differences in cue weighting is a useful tool in exposing the complex relationship between perceptual cue weighting and language regularities. © *2012 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4765076]

PACS number(s): 43.71.An, 43.71.Es, 43.70.Mn [MSV] Pages: 3950–3964

I. INTRODUCTION

Speech perception is complex. One of the reasons is that multiple acoustic dimensions define speech categories, requiring integration of information across dimensions. For example, the voicing distinction in English stops (e.g., [b] versus [p]) may give an appearance of a simple phonetic contrast. However, examined acoustically, as many as 16 dimensions covary with this distinction (Lisker, 1986). This is emblematic of speech categories; multiple acoustic dimensions typically covary with phonetic category distinctions (Coleman, 2003; Dorman *et al.*, 1977, for stop place of articulation; Jongman *et al.*, 2000, for fricative place of articulation; Hillenbrand *et al.*, 2000, for tense and lax vowels; Kluender and Walsh, 1992, for fricative/affricate distinction; Lisker, 1986, for stops voicing; Polka and Strange, 1985, for liquids).

Whereas any of these dimensions may inform phonetic categorization, they are not necessarily perceptually equivalent. Some acoustic dimensions play a more important role in determining category membership of a sound than do others. To distinguish [b], [d], and [g], for example, English listeners make greater use of differences in formant transitions than frequency information in the noise burst that precedes the transitions although each dimension reliably covaries with these consonant categories (Francis *et al.*, 2000). In the voicing distinction between stop consonants at the syllable initial position such as [ba] versus [pa], English listeners rely primarily on the duration of voice onset time (VOT) and use fundamental frequency (F0) of the following vowel as a secondary source of information (Abramson and Lisker, 1985; Francis *et al.*, 2008; Idemaru and Holt, 2011). In a vowel distinction, tense [i] and lax [ɪ] are differentiated acoustically by both spectral and temporal acoustic dimensions. However, English listeners rely much more on the spectral dimension than the temporal dimension in categorizing these vowels (e.g., Hillenbrand *et al.*, 2000). Thus while exploiting multiple acoustic dimensions to inform phonetic categorization, listeners give greater perceptual weight to some dimensions. This has been referred to as perceptual cue weighting (e.g., Holt and Lotto, 2006; Francis *et al.*, 2008). Understanding what determines perceptual cue weighting is fundamental to understanding speech perception.

Cue weighting has been proposed to arise, at least in part, from distributional characteristics of the input (Holt and Lotto, 2006; Francis *et al.*, 2008; Toscano and McMurray, 2010).

[a)]A portion of this work was presented in "Relational timing or absolute duration? Cue weighting in the perception of Japanese singleton vs. geminate stops," Proceedings of the 16th International Congress of Phonetic Sciences, Saarbruecken, Germany, August 2007.

[b)]Author to whom correspondence should be addressed. Electronic mail: Idemaru@uoregon.edu

 0001-4966/2012/132(6)/3950/15/$30.00

Holt and Lotto (2006) argue that adaptive listeners will tune perceptual weighting of acoustic dimensions to the distributional regularities of the input to maximize categorization accuracy but that constraints from sensory processing, cognition, and previous experience may interact to influence the extent to which listeners achieve idealized perceptual weighting. For example, all other things being equal, dimensions very well correlated with category identity ought to be more strongly perceptually weighted than those less predictive of category identity. Such differences in *informativeness* might occur as a consequence of categories' distributional regularities. If, for example, category distributions do not overlap much along a particular acoustic dimension, then the dimension is highly informative about category identity and is likely to be strongly perceptually weighted. Holt and Lotto (2006) investigated this question by studying the extent to which distributional informativeness affected perceptual cue weights as listeners learned novel, arbitrary nonspeech auditory categories. They observed a role for informativeness but found that it interacted with other factors such as the inherent perceptual salience of a dimension and the task in which listeners were engaged.

In the domain of speech categorization, studies of perceptual cue weighting typically have examined phonetic categories for which there is a robust difference in the informativeness of the acoustic dimensions under investigation. Take the distinction of English [l] and [ɹ], for example. Acoustically, the onset frequency of the third formant (F3) is the single best predictor of English talker's intended [l] and [ɹ] productions (Yamada and Tohkura, 1992; Iverson *et al.*, 2003; Ingvalson *et al.*, 2011; Lotto *et al.*, 2004). The onset frequency of the second formant (F2) is also a predictor, but a substantially weaker one (Lotto *et al.*, 2004). This is mirrored in perception in that F3 is given most perceptual weight by native English listeners (Yamada and Tohkura, 1992). When listeners categorize sounds that span from [l] to [ɹ] varying along the dimensions of F3 and F2, responses are best correlated with the stimulus value along the F3 dimension and are weakly correlated along the F2 dimension (Ingvalson *et al.*, 2011). Similarly, in production of syllable-initial English stop voicing (e.g., [ba] versus [pa]), VOT is the single best predictor and F0 of the following vowel is a secondary weaker one (Lehiste and Peterson, 1961; Raphael, 2005; Holt and Wade, 2004). In perception, VOT is more heavily weighted as the primary cue (Abramson and Lisker, 1985; Francis *et al.*, 2008; Idemaru and Holt, 2011).

These cases demonstrate that when there is a robust differentiation among acoustic dimensions as a function of informativeness of category membership, the more informative dimension is most heavily weighted in speech categorization. It is equally important to investigate phonetic categories for which acoustic dimensions' informativeness is relatively equivalent. Without a robust bias in acoustic informativeness, perceptual cue weight may be balanced across dimensions because of the parity in informativeness, it may be variable across listeners as either cue will lead to accurate categorization, or it may be biased for reasons other than the informativeness of acoustic dimensions. Thus by studying such cases, it may be possible to unmask other factors contributing to perceptual cue weighting for speech categories such as how subtle distributional regularities in spoken language relate to perceptual cue weighting, how the computational demands introduced by different cues may influence perceptual weighting, and how resilient or flexible cue weight may be to short-term acoustic variability in the signal. This study investigates perceptual cue weighting when there is parity among acoustic dimensions in terms of their informativeness, a situation that has received a little attention (Holt and Lotto, 2006; Francis *et al.*, 2008; Toscano and McMurray, 2010). We further aim to investigate the extent to which there are significant individual differences in perceptual cue weights among native listeners.

To this end, the singleton and geminate distinction in Japanese stop consonants presents an excellent example. Singleton and geminate stops (e.g., [t] and [tt]) in Japanese, and in many other languages, are distinguished primarily by the duration of stop closure. However, segmental duration is heavily influenced by speaking rate.[1] As a result, absolute stop closure duration provides imperfect information because across different speaking rates, the stop closure durations corresponding to the singleton and geminate categories overlap considerably. Said another way, the informativeness of absolute duration for singleton/geminate categorization is reduced due to variability from speaking rate. This has been observed for durational contrasts in consonants and vowels in languages including English, Italian, Icelandic (Miller and Baer, 1983; Miller and Liberman, 1979; Pickett *et al.*, 1999; Pind, 1999; Port and Dalby, 1982; Boucher, 2002) and Japanese (Fujisaki, 1979; Hirata and Whiton, 2005; Idemaru and Guion-Anderson, 2010).

Given that speaking rate variability undermines the informativeness of absolute duration in signaling durationally differentiated phonetic categories across speaking rates, relative duration has been proposed as a higher-order dimension that is more stable across variable speaking rate (Kohler, 1979; Pickett *et al.*, 1999; Pind, 1999; Port and Dalby, 1982). Relative duration is typically expressed in the form of durational ratios between a target speech segment such as the absolute stop duration and the duration of a neighboring segment(s), reflecting a kind of inherent context-dependent normalization for rate changes. Studies have demonstrated that relative duration does better predict rate-dependent phonetic category membership for speech productions than absolute duration (Kohler, 1979; Pickett *et al.*, 1999; Pind, 1986, 1999; Port and Dalby, 1982).

However, there is also evidence that the extent to which absolute duration differentiates (or fails to differentiate) phonetic categories across speaking rate varies across languages. For example, the absolute duration of geminate stops is reported to be three times as long as the duration of singleton stops in Japanese (Han, 1994; Idemaru and Guion, 2008), whereas geminate stops in Italian are only about two times as long as singleton stops (Ham, 2001). The robust absolute duration difference in Japanese may mean that both relative duration and absolute duration are adequate at categorizing singleton and geminate stop productions in Japanese across speaking rates. In support of this, Idemaru and Guion-Anderson (2010) showed that absolute duration (stop duration) was sufficient to categorize 87% of native Japanese singleton and geminate stops, whereas relative duration

(durational ratio of stop to the previous syllable) categorized 93% of singleton and geminate stops produced by six speakers across three distinct speech rates. Thus although the informativeness of the cues, measured as their classification accuracy, is uniformly high, relative duration is slightly more informative; however, it is unclear whether this small difference in informativeness is perceptually significant.

In categorizing Japanese singleton and geminate stops, an ideal observer using the full extent of the information available in the input would rely somewhat more on relative duration due to its slight advantage in informativeness. However, it is unclear whether listeners behave as ideal observers. It is possible, for example, that listeners exploit absolute duration despite its lower informativeness. As a unidimensional acoustic cue not requiring integration of information across the utterance, it may confer a computational processing advantage. Or, instead, listeners may be promiscuous in their cue use, committing to neither dimension and exhibiting high variability in perceptual cue weighting given the parity in informativeness.

The issue of individual differences in cue weighting was noted in earlier studies (e.g., Haggard *et al.*, 1970) and has gained attention in more recent research (Kong and Edwards, 2011; Allen *et al.*, 2003; Shultz *et al.*, 2012; Raizada *et al.*, 2010). In particular, Shultz *et al.* (2012) and Kong and Edwards (2011) showed that whereas listeners consistently weighted VOT more than F0 in categorizing stop voicing, there was considerable individual variation in the extent with which listeners used F0. This seems to reflect the F0's secondary status in informativeness to categorization relative to VOT (e.g., Abramson and Lisker, 1985). It can be considered that relatively small difference that F0 makes for category informativeness when VOT is available leads to individual variation in weighting of F0 as a perceptual cue. Along the same lines, we predict that in categorizing Japanese singleton and geminate stops, listeners exhibit individual variation across the use of relative and absolute durations that do not vary greatly in their informativeness. Examining individual differences in perceptual cue weighting in a situation where two acoustic dimensions provide similar informativeness provides an opportunity to better understand listeners' sensitivity to distributional statistics of fine-grained acoustic dimensions defining speech categories.

In the experiments that follow, we have adopted methods and approaches recently applied to studies of perceptual cue weighting (Holt and Lotto, 2006) and Japanese geminates (Idemaru and Guion-Anderson, 2010) to investigate these issues. We explore the strength of perceptual cue weights (Experiment 1), investigating whether individual listeners' weights are relatively stable across time (Experiment 2) and whether they are resistant to perturbation (Experiment 3). Finally, we investigate whether the individual patterns of perceptual cue weights are related to the talker's own speech production patterns (Experiment 4).

II. EXPERIMENT 1—PERCEPTION

In Experiment 1, listeners categorized synthesized Japanese words spanning from *seta* (with a singleton) and *setta* (with a geminate) in the dimensions of absolute and relative durations. Durational parameters were manipulated so that the absolute and relative durations varied from singleton to geminate values, allowing us to assess listeners' relative use of each dimension in categorizing the stops.

A. Methods

1. Participants

Thirty-five (19 females; ages, 21–35 yr, mean = 30 yr) native Japanese listeners participated for a small payment. Participants were born in various regions of Japan (with the largest group, $N = 11$, from Tokyo or its surrounding areas). All listeners resided in the U.S. at the time of testing. Length of residency in the U.S. ranged from 1 month to 9 yr (mean = 2 yr, 2 months). All listeners reported normal hearing. The data from two female participants were excluded from subsequent analyses due to substantial early exposure to a foreign language.[2] In addition, a technical problem occurred while testing one of the participants, and he could not complete the experiment; his data were excluded from the analysis.

2. Stimuli

The experiment used Japanese words *seta* and *setta* (Idemaru and Guion-Anderson, 2010; Idemaru and Guion, 2008) and methods from auditory category-learning experiments (Holt and Lotto, 2006). The stimulus space was defined by absolute duration (the duration of stop closure) and relative duration of stop closure (the durational ratio of stop closure to the previous CV syllable, [se]) to investigate the effect of these dimensions in categorization [Fig. 1(a)].

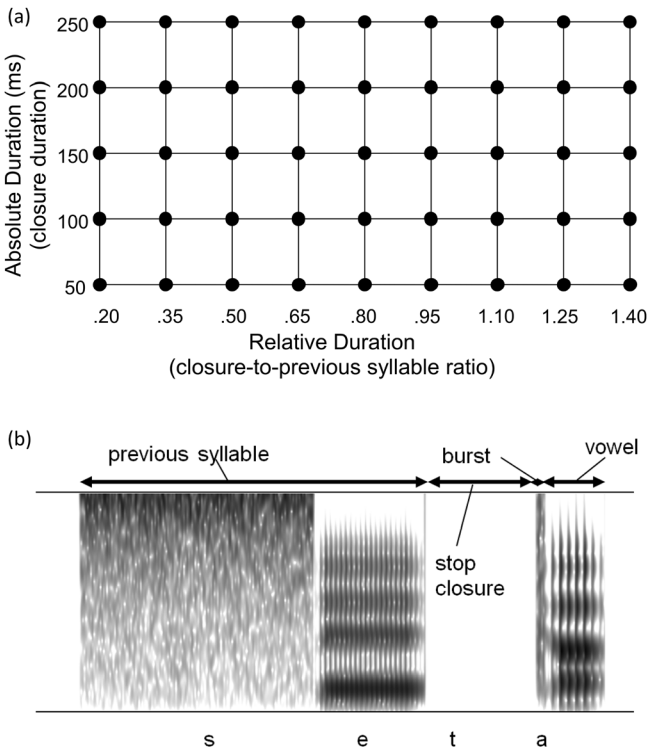

FIG. 1. (a) The acoustic space defining the experimental stimuli. Each dot represents a single stimulus sound. Absolute duration was defined as the duration of stop closure in milliseconds. Relative duration was defined as the durational ratio of stop closure to the previous syllable ([se]). (b) Sample spectrogram of a stimulus.

Mora (instead of syllable) is a term more consistent with the Japanese phonology (Vance, 1987). However, the term syllable will be used here for convenience and simplicity.

One of the endpoints, *setta,* is a lexical item in Japanese meaning "hurried," whereas the other, *seta,* is a non-lexical item. Although endpoint tokens mismatched on their lexical status are known to introduce a lexical bias (Ganong, 1980) on speech categorization, Idemaru and Guion-Anderson (2010) used the same stimulus pair with native Japanese listeners and report a very small bias (about 10%) in the direction of the non-lexical item. The acoustic structure of these tokens provides for clean acoustic analyses and segmentation and has the benefit of being directly relatable to the previous research of Idemaru and Guion-Anderson (2010). Therefore this pair was selected for the perceptual targets.[3]

In a strict sense, absolute duration and relative duration defined here are not independent. Absolute duration is simply duration of stop closure, whereas relative duration is the duration of the stop closure relative to a context duration (here defined as the previous syllable duration). Thus relative duration likewise depends upon stop closure duration. The claim that has been made previously is that *relative* perception of this duration with respect to the rate of adjacent speech provides an acoustic correlate more robust to variability in speaking rate (Kohler, 1979; Port and Dalby, 1982; Pickett *et al*., 1999; Pind, 1999). Reliance on relative versus absolute acoustic cues, although not independent, has been proposed as two distinct perceptual strategies.

The landmarks for measuring component durations are illustrated in Fig. 1(b). Absolute duration (stop closure duration) varied from 50 to 250 ms in five 50 ms steps. The endpoints, 50 and 250 ms, spanned exaggerated values of stop closure duration outside those of the typical Japanese voiceless singleton and geminate stops (mean singleton = 78 ms, mean geminate = 225 ms, in Idemaru and Guion-Anderson, 2010). Relative duration (the durational ratio of stop closure to previous syllable) varied from 0.20 to 1.4 in nine 0.15 steps. These ratio values also exaggerated typical Japanese singleton and geminate stop values (mean singleton = 0.42, mean geminate = 1.08, in Idemaru and Guion-Anderson, 2010). The dots in Fig. 1(a) illustrate the stimuli defined across absolute duration and relative duration.

The stimuli were synthesized using KLATTWORKS (McMurray, 2000). The two-dimensional (2-d) acoustic space [Fig. 1(a)] determined the durations for [se] (previous syllable) and [t] (stop). The [a] duration was determined by the stop-to-vowel durational ratio (2.00) reported by Idemaru and Guion-Anderson (2010) as a value unbiased either for singleton or geminate. The duration of [s] within [se] was determined to be 68% of the [se] duration, and the duration of [e] to be 32% of the [se] duration based on the production data reported by Idemaru and Guion-Anderson (2010).

The frication noise for [s] was synthesized using parameter values proposed by Klatt (1979). The F1 through F6 frequencies were 320, 1390, 2530, 3250, 3700, and 4900 Hz with the parallel tract amplitude (A1–A6) set as zero for the first five formants and 52 dB for F6. Amplitude of frication noise (AF) was set as 70 dB for the duration of the [s].

To synthesize the vowels [e] and [a], the steady state F1, F2, and F3 frequencies were taken from the acoustic study of Japanese vowels by Keating and Huffman (1984). In each stimulus, the F1 and F2 frequencies varied across the first 20 ms, rising from 276 to 476 Hz and 1515 to 1715 Hz for [e], respectively. For [a], F1 increased from 432 to 632 Hz and F2 decreased from 1663 to 1374 Hz, characteristic of vowels following [t]. This formant transition was determined using the locus equation of Sussman, McCaffrey, and Matthew (1991). The F3 frequencies, 2500 for [e] and 2383 for [a], were steady-state across the vowel. Amplitude was 40 dB at the onset of [e], then increased linearly to 60 dB across the first 20 ms of [e] and decreased to 40 dB in the last 20 ms of the [e]. Amplitude then transitioned to 0 dB where it remained for the duration of the stop, after which it increased linearly to 60 dB across the first 20 ms of [a] and decreased to 40 dB in the last 20 ms of the [a]. It was not possible to maintain these transitions for vowels with durations less than 40 ms. For these vowels, duration of the transitions was shortened (e.g., 10 ms) and the duration of the steady state was also shortened. Fundamental frequency (F0) was 160 Hz for [e] and 100 Hz for [a] within the typical range of male values (Idemaru and Guion, 2008). Amplitude and F0 correlate with Japanese stop length production (Idemaru and Guion, 2008); however, ambiguous values were chosen so that there was no acoustic bias. A 10-ms stop burst was excised from a natural production of *seta* by a male native Japanese speaker and was inserted before [a].

3. Procedure

Seated in individual sound-attenuated booths and wearing headphones (Beyer DT-150), listeners categorized 20 repetitions of each of the 45 stimuli (900 trials) by pressing response buttons labeled "seta" and "setta" in Japanese orthography. Stimulus presentation and response collection were under the control of E-PRIME (Psychology Software Tools, Inc.).

4. Statistical analysis

Although logistic regression analysis has been proposed for analyzing speech perception response data (Nearey, 1990; Benkí, 2001; Morrison, 2007) and provides a statistically rigorous and promising method, the approach is ruled out here by the fact that our data exhibit within-subject correlation and response asymptotes other than zero and 100. Therefore local polynomial nonparametric regression (LPNR; Loader, 1999), a standard statistical tool applied to data that does not conform to a known parametric shape, was used.

Nonparametric regression uses techniques to fit a smoothed curve to the data scatter plot. Unlike logistic regression, nonparametric regression does not assume the shape of the regression line. Rather, it derives the shape from the data. In the case of LPNR, instead of attempting to fit the curve to the data points all at once, a small window of analysis is applied across the independent variable(s) obtaining a local regression fit for the corresponding local dependent values. LPNR further uses kernel density

estimation, a smoothing technique, so that local averaging is done with weighting such that observations closer to the center of the analysis window are weighted more. An important advantage of this smoothing technique is that a large number of observations (this could be the entire set of observations) is used to make a prediction of the dependent variable. However, this is done so that the observations closer to the center of the analysis window contribute more to the prediction.

To understand the application of this statistical technique to perceptual cue weighting in speech categorization, it is useful to consider how the categorization data fall within an acoustic space (Fig. 1). For each of the points marking a stimulus in the 2-d acoustic space, there is a percentage of geminate responses for each listener that is thought of as coming out of the plane of the plot toward the reader in the "z axis," thus forming a 3-d data scatter plot. Here, a Gaussian kernel was applied around each x-y coordinate (where x and y were absolute duration and relative duration, respectively) in the data scatter plot. The outcome values (percentage geminate responses) associated with all the x-y observations within the analysis window were averaged with kernel smoothing, producing a fitted value of the outcome. The analysis window was moved across the x and y dimensions obtaining the locally fitted values across the entire acoustic stimulus space defined by the ranges of x and y dimensions. The entire set of observations was used in this case to make a prediction regarding the dependent variable, percent geminate responses. Furthermore, this was done so that the observations closer to the center contributed more to the prediction. Technically, this process is repeated for a range of variances (bandwidths) of the kernel, and the one with the best cross-validation score is used.[4]

Thus the resulting outcome was a predicted percent geminate response across the entire stimulus space for each listener, which allowed us to estimate the perceptual geminate-singleton category space in relation to the acoustic space defined by relative and absolute duration for each listener. We defined the geminate area as a region in the perceptual space where geminate responses were greater than 80% and the singleton area as an area in which geminate responses were fewer than 20%. The centers of geminate and singleton categories were then defined as the centroid of these areas, i.e., for both absolute and relative duration, the mean duration for all test points in the region was calculated. The line connecting the geminate category center to the singleton category center for each listener describes the positional relationship between the category centers within the stimulus space. The angle of this line (where a horizontal line pointing to the right is zero and angles clockwise from this origin are negative and counterclockwise angles are positive) provides a single value (angle) reflecting the relative weights that each listener placed on absolute duration and relative duration in perception of the geminate and singleton sounds.

There have been a few other methods for computing the influence of multiple acoustic dimensions on speech perception (Escudero and Boersma, 2004; Holt and Lotto, 2006). Our preliminary analysis found strong, statistically significant correlations among the cue weight indices provided by LPNR and methods proposed by Escudero and Boersma (2004) and Holt and Lotto (2006) (Pearsons r's > 0.9), indicating that LPNR, as well as the other two methods, captures some features of perceptual cue weighting in speech categorization.

B. Results

As discussed earlier, LPNR provides predicted percent geminate responses for the entire stimulus space for each listener. Figure 2 shows obtained predicted percent geminate responses for two listeners [Figs. 2(a) and 2(b)] as well as predicted pattern summarized for all listeners [Fig. 2(c)]. As the legend [Fig. 2(d)] indicates, darker areas show more singleton responses and lighter areas indicate more geminate responses.

As exemplified by Fig. 2, visual inspection of the predicted percent geminate response patterns revealed evidence of both consistent patterns and strong individual differences in perceptual weighting. For example, the two listeners in Fig. 2 perceptually divided the acoustic space for geminate and singleton categories in very different ways. Listener 1 [Fig. 2(a)] categorized geminate and singleton stops primarily on the dimension of relative duration, whereas Listener 2 [Fig. 2(b)] categorized the two sounds primarily on the dimension of absolute duration. The lines (angle) in the figure connect the center of their geminate category and the center of singleton category, providing an intuitive means of understanding which acoustic dimension most affected speech categorization. An examination of angle lines of individual listeners in Fig. 2(c) shows that although the location of the singleton center varied substantially, the location of the geminate center was relatively consistent across listeners. This may be due to the lexical status of the word including the geminate.

Figure 3(a) shows the distribution of the angle values (ranging from -78 to -201), with Fig. 3(b) illustrating the meaning of angle values. There is a small peak in the frequency distribution around -170 and a larger peak around -140. The rest were scattered between -70 and -120. This suggests that some listeners primarily used relative duration (those whose angle values were around -170), some primarily used absolute duration (those whose angle values were between -70 and -110), and yet others used both dimensions fairly equally (those whose angle values were around -140).

C. Discussion

In categorizing Japanese singleton and geminate stops, native Japanese listeners showed considerable variability in their use of absolute versus relative duration. Some listeners primarily rely on relative duration, others use mostly absolute duration, and yet others use the two dimensions fairly equally. The results here demonstrated that a slight advantage in informativeness for relative duration did not translate into this dimension weighted more heavily across the board. Cue weighting, thus, is not dictated solely by informativeness. The

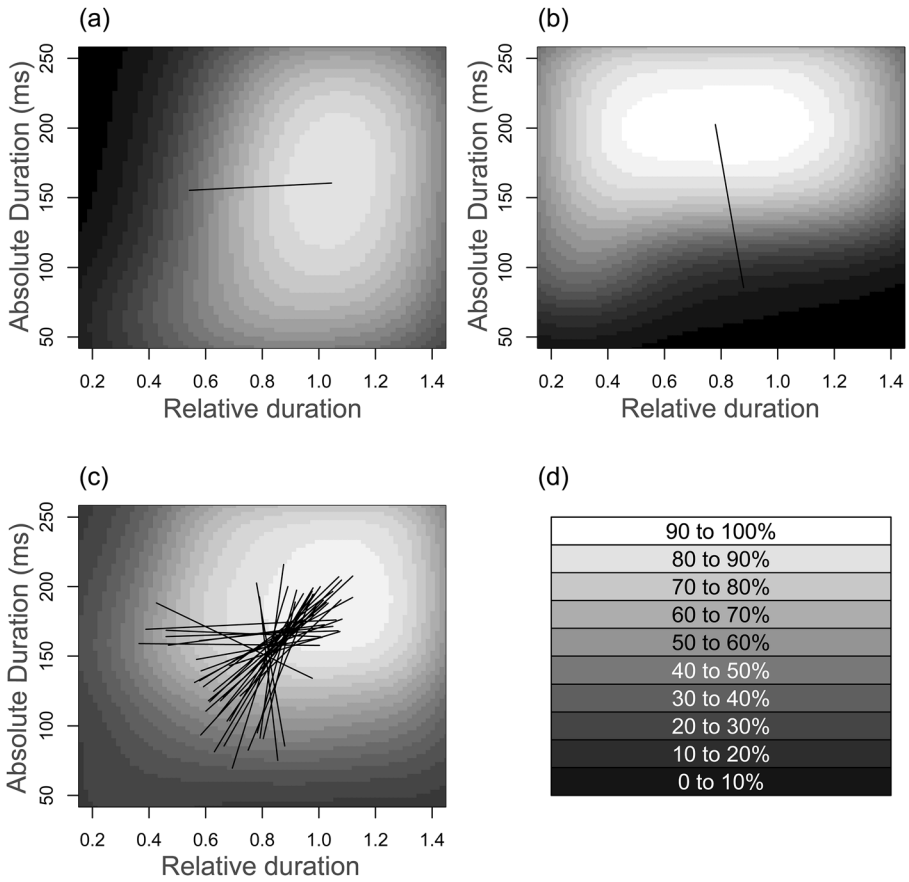

FIG. 2. Categorization responses for two individual listeners [(a) = Listener 1; (b) = Listener 2], and categorization responses summarized for all listeners (c). The white area indicates the predicted geminate category area, and the black area is the singleton category area. The line connects the center of geminate category and the center of singleton category. The last panel (d) shows the mapping between the gray-scale and percent geminate response.

present results can be interpreted in the context of another study of Japanese geminate and singleton stops. In Idemaru and Guion-Anderson (2010), participants categorized stimuli for which context duration varied while stop duration was constant and perceptually ambiguous. In other words, relative duration varied across values typical of singletons and geminates, whereas absolute duration remained perceptually ambiguous. In this study, all listeners used relative duration: Perception of stop length changed between singleton and geminate as a function of the context duration preceding the stop. Thus when relative duration is the only reliable acoustic information, Japanese listeners can use it for categorization. The results of this study thus mirror previous research demonstrating listeners' perceptual reliance on relative durations (Kohler, 1979; Port and Dalby, 1982; Pind, 1986, 1999; Pickett *et al.*, 1999). However, the present results demonstrate that when both relative duration and absolute duration information is available, listeners show large individual differences in perception. It is perhaps unsurprising that both relative duration and absolute duration are used by Japanese listeners with similar frequency, given the parity in informativeness of the two dimensions (87% accurate prediction by absolute duration, and 93% by relative duration, Idemaru and Guion-Anderson, 2010). These results suggest that the small difference in informativeness of relative duration and absolute duration is not highly significant for perception. The slightly better informativeness of relative duration (Hirata and Whiton, 2005; Idemaru and Guion-Anderson, 2010) did not translate into the across-the-board primacy of relative duration in perception.

It has been widely assumed that due to the acoustic variability introduced by different speaking rates, relative duration is better than absolute duration for sound categorization (e.g., Pind, 1986). However, this expectation must be conditioned by the extent to which absolute duration is undermined by increased variability in the critical segmental duration for a particular language. We have demonstrated that in the case of the Japanese stop length contrast, in which absolute duration approximates the informativeness of relative duration, the perceptual role of absolute duration does not diminish among many listeners.

Furthermore, the finer-grain LPNR analyses employed here demonstrate that there were listeners all across the spectrum with some listeners primarily relying on relative duration, others using mostly absolute duration, and yet others using the two dimensions fairly equally. It is important to note that if only the group data were considered, it would be concluded that relative and absolute durations are weighted almost equivalently (mean angle = −143.5, indicating nearly the mid-point between strong reliance on relative duration and strong reliance on absolute duration), failing to expose the extensive individual differences in listeners' relative perceptual weighting of absolute and relative duration. The current findings stress the importance of examining perceptual patterns at the individual level.

The informativeness of both absolute and relative duration across rate variability in Japanese (Idemaru and Guion-Anderson, 2010), and the extensive individual differences we observe in listeners' reliance on the two sources of information for singleton and geminate stop categorization

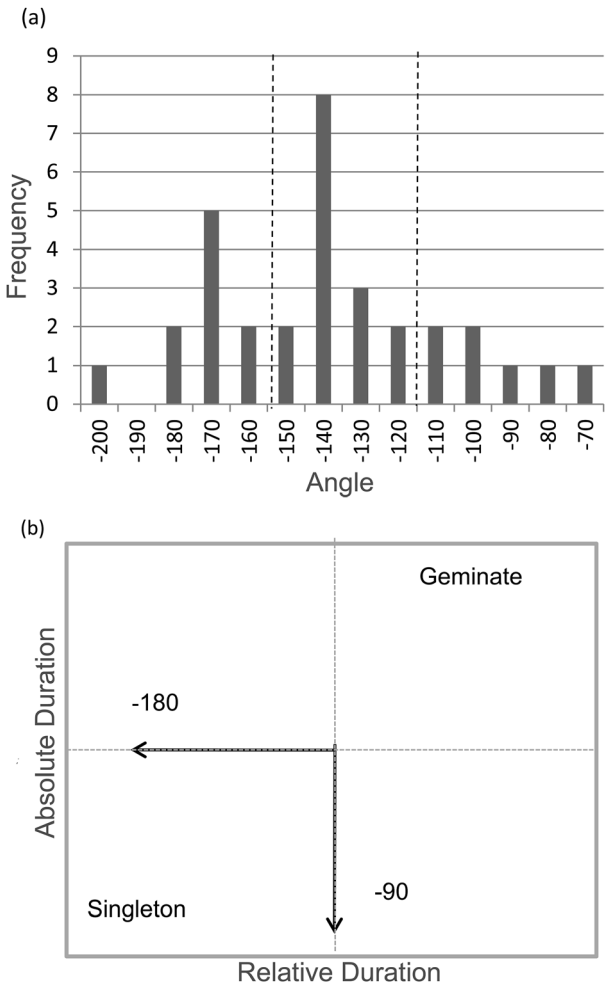

FIG. 3. (a) Distribution of angle values. The dotted lines separate values that indicate primary use of relative duration (from −190 to −160), mixed use of the two durations (from −150 to −120), and primary use of absolute duration (from −110 to −70). (b) Assignment of angle values. The angle value of −180 indicates that singleton and geminate stops were categorized along the dimension of relative duration, whereas the angle −90 indicates categorization along the dimension of absolute duration.

provide an opportunity to investigate the stability of the perceptual weight listeners give to acoustic cues. The fact that many listeners used both relative and absolute duration with differential weight in the current study may simply reflect promiscuous, unsystematic use of the two dimensions. The parity in informativeness may allow listeners to switch readily between the two in perception without strong stable individual patterns across time. Or, perhaps because of long-term regularities in their own speech production or listening experience, listeners may exhibit relatively more stable individual differences across time. Experiment 2 re-tested some of the Experiment 1 participants to investigate this issue.

III. EXPERIMENT 2—PERCEPTUAL STABILITY

Of the 35 listeners of Experiment 1, 23 returned for the second test. At least a 3-wk interval (mean = 59 days, range = 27–140 days) separated the two testing sessions. The experimental stimuli and procedure were identical to Experiment 1.

A. Results and discussion

LPNR was applied to the geminate responses. The angle values characterizing singleton versus geminate perceptual cue weights were calculated. To examine whether the listeners' perceptual cue weight was consistent between the initial test and the retest, the angle values from Experiments 1 and 2 were examined for their relationship. If perceptual cue weight is relatively consistent across time, we would expect a positive correlation between the angle values of Experiments 1 and 2, whereas if differential weighting of absolute and relative duration in Experiment 1 reflects unsystematic use of the two sources of information, there should be no relationship. In fact, there was a strong and statistically significant correlation between the two sets of angle values, $r = 0.69$, $P < 0.001$ (Fig. 4). There was one listener (indicated by a * symbol in Fig. 4), who showed substantially different angle values across two tests (−178 in Experiment 1 and − 69 in Experiment 2), showing strong reliance on relative duration in Experiment 1 and strong reliance on absolute duration in Experiment 2. When this listener was excluded from analyses, the correlation improved, $r = 0.88$, $P < 0.001$. These results indicate that most listeners make consistent use of absolute and relative information in categorizing Japanese singleton and geminate stops.

To ensure that the response pattern did not emerge as a result of learning through the perceptual task, geminate response to the first presentation of the 49 stimuli were correlated with relative duration and absolute duration in the stimuli (Holt and Lotto, 2006). Relative correlation coefficients of the two dimensions based on the very first presentation of the stimuli showed highly consistent response pattern with the overall response pattern (21 of 23 listeners showing a consistent preference for the relative or absolute dimension).

In Japanese, relative and absolute durations are similarly informative of singleton and geminate category membership across rate variability (Idemaru and Guion-Anderson, 2010), presenting a situation in which listeners potentially could categorize with high accuracy using either dimension or using the dimensions unsystematically. Experiment 1 evidenced considerable individual differences in listeners'

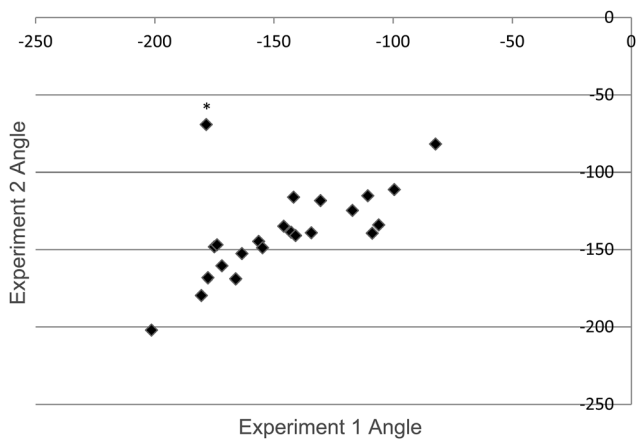

FIG. 4. Scatter plot showing the angle values of each listener obtained in Experiments 1 and 2. The starred data point indicates a listener who switched cue weighting across two experiments.

relative reliance on the two dimensions. Experiment 2 demonstrated that these individual differences observed in the categorization task were stable across time. Experiment 3 investigated this further in a new task by examining the extent to which these perceptual patterns were resistant to short-term perturbation.

IV. EXPERIMENT 3—RESISTANCE TO PERTURBATION

Holt and Lotto (2006) conducted a series of experiments in which listeners learned to categorize non-speech sounds varying in a 2-d acoustic space. One of the experiments demonstrated that passive exposure to variability along an acoustic dimension led listeners to make greater use of this dimension in a later categorization task than another group of listeners without such pre-exposure. Thus exposure to variability across an acoustic dimension was sufficient to shift listeners' perceptual cue weights toward the dimension.

We exploited this method to examine whether Japanese listeners' pattern of cue weighting was resistant to perturbation. If the individual patterns of perceptual cue weighting are robust, they may remain stable after exposure to variability across the less-preferred acoustic dimension. If the patterns are flexible, exposure to acoustic variability across the less-preferred dimension will increase the perceptual weight of the less-preferred acoustic dimension in subsequent categorization responses.

In this experiment, Japanese listeners who weighted absolute duration more in Experiment 1 and 2 were exposed to stimuli varying from *seta* to *setta* only in the relative duration (their less-preferred dimension) prior to a categorization test; those who weighted relative duration more in Experiment 1 and 2 were exposed to stimuli varying only in the absolute duration (their less-preferred dimension) prior to the test. The value of the other dimension, the dimension that the listeners relied more in Experiments 1 and 2, was held constant in the stimuli at a value acoustically ambiguous for category membership.

A. Method

1. Participants

Twenty participants, a subset of those who participated in both Experiment 1 and 2, returned for a third test. The individual mean angle values from Experiments 1 and 2 were used to group the participants. Those participants with mean angle value was less than −135 (the middle point between exclusive reliance on relative and absolute duration, see Fig. 3) were grouped as relative duration listeners ($N = 12$), and those with mean angle values greater than −135 were grouped as absolute duration listeners ($N = 8$). At least a 3-wk interval (mean = 41 days, range = 21–85 days) separated the testing sessions between Experiments 2 and 3.

2. Stimuli

A set of relative duration exposure stimuli was created with a consistent absolute duration (stop duration) of 125 ms and the relative duration varying in nine steps of 0.15 from 0.20 to 1.40. A set of absolute duration exposure stimuli was created with the consistent relative duration of 0.65 and the absolute duration varying in nine steps of 25 ms from 50 to 250 ms. These constant values approximate the acoustically ambiguous boundary values found in acoustic studies (e.g., Idemaru and Guion-Anderson, 2010). These exposure stimuli are indicated in Fig. 5 with filled circles. These stimuli ranged from *seta* to *setta* on each dimension of relative and absolute duration.

In addition, nine test stimuli were defined (illustrated as gray circles in Fig. 5) as varying in relative duration across three steps of 0.15 straddling the boundary value (0.65), and in absolute duration in three steps of 25 ms straddling the boundary value (125 ms). These small numbers of ambiguous tokens around the boundary regions were selected as test stimuli so that the effect of their variability, if any, would not override the effect of exposure-stimuli variability. Six stimuli, gray circles with black dots in Fig. 5, were used as both exposure and test stimuli.

3. Procedure

The procedure comprised eight cycles of an exposure block followed by a test block. During exposure blocks, listeners simply listened to 10 randomized presentations of the nine exposure stimuli (black dots in Fig. 5) varying in the acoustic dimension opposite their most heavily weighted perceptual dimension in the previous experiments. During test blocks, listeners categorized five randomized repetitions of each of the nine test stimuli (gray circles in Fig. 5) by pressing response buttons labeled "seta" and "setta" in Japanese orthography. The absolute duration listeners were exposed to acoustic variability across the relative duration dimension in the exposure block, whereas relative duration listeners were exposed to acoustic variability across the absolute duration dimension in the exposure block. Other than the inclusion of the exposure blocks, the apparatus and the procedure of this experiment were identical to Experiments 1 and 2.

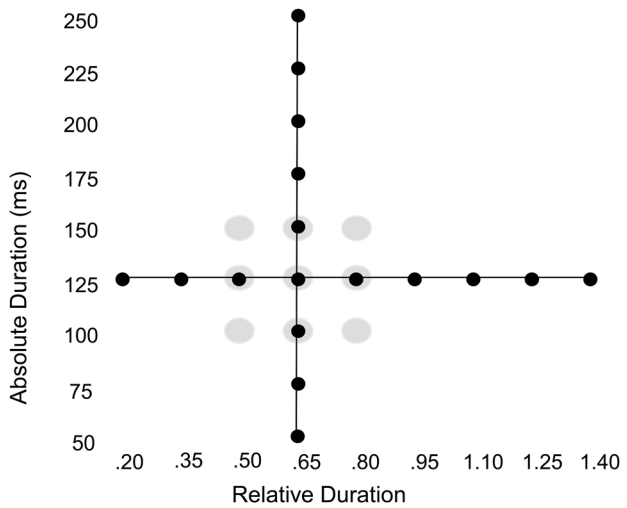

FIG. 5. Exposure stimuli (black dots) and test stimuli (gray circles) used for Experiment 4.

B. Results

To examine the effect of exposure to the less-preferred acoustic dimension, a geminate response difference score was obtained for each listener by subtracting the number of geminate responses to the lowest level of the dimension (e.g., 100 ms of absolute duration) from the number of geminate responses to the highest level of the dimension (e.g., 150 ms of absolute duration). This difference score (cue influence, maximum = 15) was used as an index of the influence of each dimension on categorization (e.g., Escudero and Boersma, 2004). Figure 6 shows the mean cue influence for the preferred dimension and exposed dimension across eight experimental blocks for two groups of listeners, absolute (a) and relative duration (b) listeners. It is noted that there was more variability in the cue influence scores among the absolute duration listeners. However, both groups maintained their bias toward their preferred dimension through the course of experiment. Furthermore, it appears that the use of the less-preferred dimension (i.e., exposed dimension) did not increase from the beginning to the end of the experiment.

To statistically examine the effect of exposure to the less preferred dimension, cue influence scores at the beginning of the experiment (Block 1) and at the end of the experiment (Block 8) were compared. A $2 \times 2 \times 2$ repeated-measure ANOVA with group (relative duration listeners and absolute duration listeners) as a between-subject factor, dimension (preferred and exposed) and block (first and last) as within-subject factors indicated a significant main effect of dimension, $F(1,18) = 32.150$, $P < 0.001$, and a significant interaction between dimension and group, $F(1,18) = 6.591$, $P < 0.05$. While the primary interest of the current experiment was an effect of block (first versus last block), none of the analyses involving this factor was statistically significant [main effect of block, $F(1,18) = 1.406$, $P = 0.251$; block*group, $F(1,18) = 3.471$, $P = 0.79$; block*dimension, $F(1,18) = 0.501$, $P = 0.488$), indicating that the response pattern did not change in a reliable way as a result of exposure.

Post hoc tests examining the dimension × group interaction, collapsing for block, showed that whereas the difference between the preferred dimension and the exposed dimension was statistically significant for relative duration listeners, $t(11) = 7.294$, $P < 0.025$ (alpha adjusted for multiple comparisons), the difference did not reach a statistical significance for absolute duration listeners in either block, $t(7) = 1.744$, $P = 0.125$. The results for absolute duration listeners may be due to a small sample size ($N = 8$) and the possibility that this group happened to include listeners with a more balanced reliance on the two acoustic dimensions. Whereas seven of 12 relative duration listeners' angle values were between −150 and −170, six of eight absolute duration listeners' angle values were between −120 and −130, somewhat closer to the middle angle value of −135. Rather large differences in variability across the two groups, indicated by the differences in the heights of error bars in Fig. 6, might be an indication of this. If the absolute duration group included more mixed cue listeners, one would expect higher variability in cue influence scores.

However, more importantly, the response patterns for both groups did not show an influence of the acoustic dimension to which they were exposed. This suggests the fairly stable manner in which listeners maintain the use of the absolute and relative dimensions, particularly for the relative duration listeners, to resist the perturbation of acoustic context encountered in this experiment.

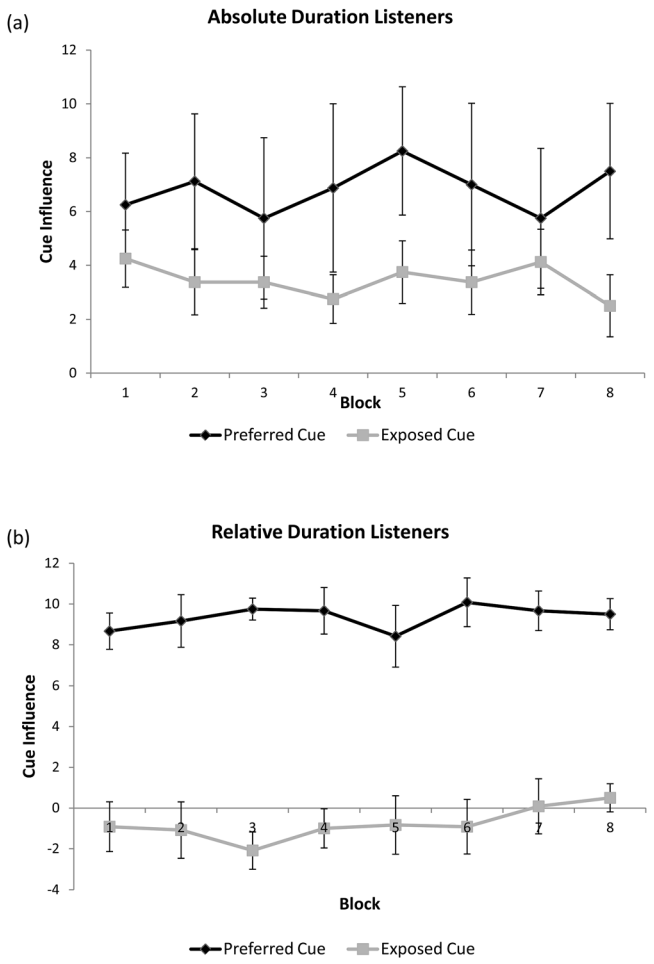

FIG. 6. Mean cue influence scores for relative duration and absolute duration cues across the eight experimental blocks for (a) absolute duration listeners and (b) relative duration users. Error bar shows one standard error of the mean. Higher values of cue influence indicate more reliance on the cue.

C. Discussion

In this experiment, there were alternating blocks of passive listening and singleton and geminate stop categorization. In the passive listening blocks, participants listened to *seta* and *setta* varying only in the less-preferred acoustic dimension (as identified in Experiments 1 and 2). This method of exposing listeners to a range of acoustic variability along a less-weighted acoustic dimension was found to shift perceptual cue weights in previous research (Holt and Lotto, 2006). However, here the manipulation had no effect even after eight iterations of the exposure-categorization blocks. The listeners maintained the perceptual cue weight that characterized their performance in Experiments 1 and 2 over the course of the experiment.

These results further indicate the robustness of individual differences in Japanese listeners' perceptual cue weighting for

absolute and relative duration. What distinguishes this experiment from that of Holt and Lotto (2006) is that Japanese listeners had long-term experience with the test sounds. In Holt and Lotto (2006), the stimuli were novel, artificial sounds that participants were trained to categorize in the same experimental session. Therefore it is possible that long-term experience with Japanese speech categories affects perceptual cue weighting in such a way that individuals' pattern of cue weights is not easily perturbed.

V. EXPERIMENT 4—RELATING PERCEPTION TO PRODUCTION

The preceding experiments demonstrate that Japanese listeners show extensive individual differences in their use of absolute and relative duration in singleton and geminate stop categorization with remarkably consistent patterns across time. The high informativeness of both relative and absolute duration across rate variability in Japanese singleton and geminate stop categories likely contributes to observed variability in perceptual cue weight across listeners. However, the similarity in high informativeness between the two dimensions is based on the group analysis. Applying the proposal of Holt and Lotto (2006) at the individual speaker/listener level that adaptive listeners will tune their perception to the distributional regularities of the input to maximize categorization accuracy, it could be hypothesized that individual listener's perceptual cue weight reflects detailed distributional statistics of the specific language environment that the listener is in. By this view, one possibility is that listener's use of relative versus absolute duration in perception reflects the distributional patterns across the dimensions in their own speech productions. If individual speaker's production may reflect small biases in the informativeness of relative and absolute dimensions, this regularity may influence perceptual cue weight. To evaluate this hypothesis, Experiment 4 measured the acoustics of singleton and geminate stops produced by the speakers who participated in Experiment 2 to examine the relationship between perception and production.

A. Method

1. Participants and stimuli

The same 23 individuals who participated in Experiment 2 also participated in the speech production study. Test words were disyllabic (e.g., [sepa], [seppa]). The first syllable was always [se], the second consonant was the target stop ([p], [t], [k], [pp], [tt], [kk]), and three non-high vowels ([e], [a], [o]) were used following the consonant. High vowels were not included because in Japanese they are likely to be devoiced after a voiceless stop, and they can cause affrication of the preceding [t]s. The resulting speech materials consisted of 9 minimal pairs (18 words), which were mostly nonce words.[5] These words could be produced either as high-low (HL) or low-high (LH) pitch pattern for words with a singleton and HLL or LHH pitch pattern for words with a geminate. The HL (singleton) and HLL (geminate) were selected to control for the pitch pattern. All test words were embedded in the carrier phrase, "*sokowa____ to yomimasu*" ("*That is read____*").

2. Procedure

Participants read aloud the sentences from the five randomized lists of 18 test words in each of three speaking rate conditions: Normal, slow, and fast, in that order. For the normal rate, participants were instructed to speak at a relaxed and comfortable tempo. For the slow rate, they were instructed to speak carefully and clearly. For the fast rate, instruction was to speak as fast as possible without making errors. These speaking rates were demonstrated by the first author prior to each recording session, and before recording began participants practiced speaking at all rates. This procedure was modeled after previous research (Magen and Blumstein, 1993; Kessinger and Blumstein, 1998; Hirata and Whiton, 2005). All the utterances were recorded at a 22.05 kHz sampling rate with 16-bit quantization in a sound-attenuated booth, using a flash digital recorder (Maranz PMD670) and a microphone (Electro Voice RE).

3. Measurements

The durations of [s], [e], stop closure, VOT, the second vowel of the test words, as well as the duration of the entire sentence were measured. The measures were made using waveform and spectrographic displays generated by PRAAT acoustic analysis software (version 4.4.30; Boersma and Weenink, 2006). The segmentation procedures described below followed those described by prior studies (Peterson and Lehiste, 1960; Klatt, 1976; Lahiri and Hankamer, 1988; Hankamer *et al*., 1989; Ham, 2001) and applied to the measurement of Japanese stop sounds by Idemaru and Guion (2008) and Idemaru and Guion-Anderson (2010).

The vowel duration was measured from the first complete cycle of periodic oscillation to the last complete cycle in the waveform. Onset of periodicity in the waveform and voicing energy of a time-locked spectrogram were referred to in order to determine the onset of the vowel. Closure duration was measured from the end of the last periodic cycle of the preceding vowel to the onset of stop burst. VOT was measured from the onset of the visible burst in the waveform to the onset of the first complete periodic cycle in the following vowel. Last, duration of the sentence was measured from the left edge of frication noise of the first segment in *sokowa* to the right edge of frication noise of the last segment, in *yomimasu*. All subjects devoiced the sentence-final [u].

The first author conducted measurements for approximately one-third of the speech data, and a native-Japanese research assistant who was trained by the first author measured the remaining data. A set of 300 randomly selected test words was measured by both researchers to examine consistency of the durational measurements. An analysis of the two sets of 1500 observations (300 words $\times$ 5 segments, [s], [e], stop closure, VOT and the final vowel) showed a strong and statistically significant correlation ($r = 0.98$, $P < 0.001$).

It should be noted that some errors and non-typical productions were observed. Speakers occasionally skipped a word. Due to a technical problem, several words in the last

repetition cycle of two speakers were not retrieved successfully. As a result of such errors, there were a total of 52 missing word tokens (26 singletons; 26 geminates). Some speakers occasionally produced a test word that was different from the word shown on the list (i.e., producing a wrong vowel or stop) or a test word with a wrong pitch pattern. The first author and research assistant, both native speakers of Japanese, listened for such production errors. When a potential error involved the length of a stop (e.g., the word sounding like *setta* when the actual test word was *seta*), tokens which both the first author and the research assistant rated as an error were excluded from the analysis. Eleven tokens were excluded as production errors, of which seven were errors of stop length. Eighty-seven words (47 singletons; 40 geminates) were produced with wrong LH or LHH pitch pattern or a flat pitch pattern. Although it has been reported that the pitch in Japanese does not affect segmental duration (Pierrehumbert and Beckman, 1988), these productions were excluded from the acoustic analysis.

In addition, speaking at a fast rate affected some of the speakers such that some segments were not produced likely due to a stronger degree of co-articulation. There were 10 instances of singleton stops with no observable stop closure or VOT produced by three speakers. In other words, the segment was produced as a glide. Of the 10 instances of glided stops, seven were produced by one speaker. Glided stops were also observed in Idemaru and Guion-Anderson (2010). In these cases, zero was recorded for the duration of stop closure and VOT. The value of zero for the glided stop duration made it impossible to compute duration ratio (relative duration), which takes the consonant duration as the numerator. Therefore, the measurements of 10 words including glided stops were excluded from the analysis of segmental duration and ratio.

Of a possible 6750 productions (18 words $\times$ 5 repetitions $\times$ 3 rates $\times$ 25 speakers), there were 6698 utterances collected (52 skipped words). The total duration of each of these 6698 utterances was measured to provide a manipulation check that participants really varied their speaking rate in response to instructions. The collected utterances included 108 production errors as described earlier (87 wrong pitch, 11 wrong word, and 10 glided). These production errors were excluded and the measurement data from the remaining 6590 words (97.6% of the data collected) were submitted to the segmental analysis.

B. Results

Mean sentence durations across five repetitions of the sentence containing each of the 18 test words were analyzed to examine whether the speakers produced the speech data in distinct speech rates. The mean sentence durations were analyzed by using a linear mixed effects model with speaking rate as a fixed effect and speaker as a random effect with repeated measure on rate. The analysis indicated that the effect of speaking rate was significant, $F(2,22) = 79.901$, $P < 0.001$, and sentence durations spoken in normal rate was longer than those spoken in fast rate, $t(22) = -8.396$, $P < 0.001$, and shorter than those spoken in slow rate, $t(22) = 10.679$, $P < 0.001$. These results confirmed that speakers did indeed produce the utterances in three distinct speaking rates.

Using the segmental measurements, absolute duration values (sum of stop closure duration and VOT) and relative duration values (sum of closure duration and VOT divided by sum of the first consonant and vowel durations) were computed for each token produced by each speaker. These data were submitted to discriminant analysis (DA) to examine the ability of each acoustic dimension to classify the stimuli accurately as singleton or geminate (according to the category intended by the speaker). Separate DAs run on the group data with absolute duration and relative duration indicated that each dimension independently classified the singleton and geminate categories fairly accurately (80% by absolute duration, 90% by relative duration, and both discriminant functions were statistically significant at $P < 0.001$). These results were similar to 87% and 9% found in Idemaru and Guion-Anderson (2010). The somewhat higher classification scores found in the previous research were likely due to smaller acoustic variability (and thus higher classification accuracy) arising from a smaller number of speakers ($N = 6$) compared to the number of speakers in the current study ($N = 23$).

To further examine the classification accuracy of absolute and relative duration at the individual level, the absolute and relative duration values were submitted to a multiple discriminant analysis for each speaker. Multiple discriminant analysis builds a discriminant function using multiple predictor variables (i.e., absolute and relative duration) and provides structure coefficients to indicate the degree of uncontrolled association of each predictor variable with the singleton versus geminate categorization. Structure coefficients in discriminant analysis are, thus, analogous to Pearson's coefficients in correlation analysis. Holt and Lotto (2006) proposed using correlation coefficients between acoustic cue dimensions and categorization responses as a way to compute relative perceptual cue weight. Applying this method to speech production data, relative production cue weights could be computed based on acoustic cue dimensions and the strength of association between each dimension and intended speech categories (i.e., structure coefficients). The values of structure coefficients for absolute duration and relative duration, therefore, were obtained and normalized to sum to one for each speaker. These values provide a quantitative estimate of the relative weight of the absolute versus relative duration for categorizing singleton versus geminate speech productions. Whereas absolute and relative duration each classifies the stop categories highly accurately (80% and 90%, respectively), normalized structure coefficients indicate *relative* classification strength of each acoustic dimension.

Across listeners, the mean production cue weights for absolute duration (0.39) and relative duration (0.61) indicated better classification of rate-variable speech productions by relative duration. The weight for relative duration was greater than that of absolute duration for all speakers (ranging from 0.50 to 0.70), except for one speaker who showed an equal weight for both dimensions. The difference in the

means was statistically significant, $t(22) = -9.579$, $P < 0.001$. This is expected, however, because relative duration is a rate-normalized duration, by definition, and thus should better accommodate speaking rate variability than absolute duration (raw values). As such, one would expect more efficient categorization of productions on the basis of a less variable criterion (i.e., relative duration). The results here simply confirm the argument that relative durations are less variable than absolute durations in speech productions varying in rate and thus present a potentially more reliable perceptual cue (e.g., Pind, 1999). In independent DA analyses, we found compatible classification accuracy for absolute and relative durations. Here, however, comparison of their classification accuracy against each other indicates the advantage of relative duration over absolute duration.

Due to the normalization inherent in relative dimensions, relative duration might be a better classifier than absolute duration for the singleton and geminate productions for all participants. However, if perception is related to production at the individual level, there should be a quantitative difference in the relative duration bias in speech production as a function of participants' perceptual bias for relative duration. In other words, relative duration listeners may show a greater degree of relative duration bias in speech production than absolute duration listeners.

A correlation analysis was conducted on the production and perception cue weights across all participants. The normalized structure coefficient values for relative duration were used as production cue weights and the angle values from Experiment 2 were used as perception cue weights. Experiment 2 angle values (instead of Experiment 1 values) were selected here because the speech data for the current study and perception data for Experiment 2 were collected on the same day.

The results of the correlation analysis indicated, however, that there was no association between the production and perception cue weights ($r = 0.25$, $P = 0.249$) (Fig. 7). In addition, the production cue weight for relative duration was compared between relative duration listeners ($N = 14$) and absolute duration listeners ($N = 9$), using the angle value

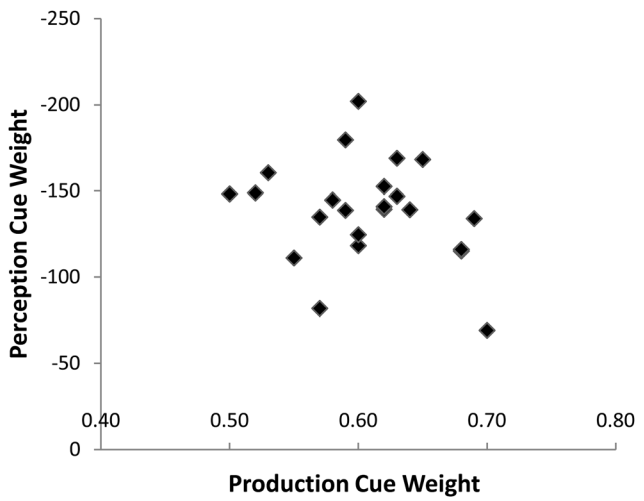

FIG. 7. Scatter plot showing the production cue weight and perception cue weight for each speaker/listener.

(−135) obtained from category response data (Experiments 1 and 2) to divide the participants into relative duration listeners (their mean angle smaller than −135) and absolute duration listeners (their mean angle greater than −135). The comparison using a *t*-test showed some trend of difference, $t(21) = 1.922$, $P = 0.068$. However, the mean of the production cue weight for relative duration was greater for absolute duration listeners than relative duration listeners (absolute duration listeners: M = 0.63, SE = 0.06; relative duration listeners: M = 0.59, SE = 0.05). These results demonstrate that the relative informativeness of absolute and relative duration in an individual talker's speech productions is not related to that talker's use of the dimensions in speech categorization.

C. Discussion

Experiments 1 and 2 showed robust individual differences in the cue weight for the perception of Japanese singleton and geminate stops: Some listeners tend to rely upon absolute duration, others primarily use relative duration, and yet others make use of both. Moreover, these patterns are quite consistent over time. Given that both dimensions provide fairly reliable independent acoustic information signaling category membership of Japanese geminate stops, one may hypothesize that individual differences may be reflected in individual differences in speech production. However, the results of Experiment 4 demonstrate that this is not the case.

Perceptual pattern (examined by a categorization task) was not related to production pattern (examined by a sentence elicitation task) in terms of the weight participants placed on relative and absolute durations in singleton and geminate stop categorization in Japanese.

The results of this production study suggest that individual differences in the weight of absolute and relative durations in the perception of Japanese stop length were not due to corresponding variability in the acoustic realization of these two types of durations at the individual level. It remains, therefore, unclear how the observed individual differences in cue weights come about. As a *post hoc* analysis, an association was examined between the listeners' cue weight and age, geographic region where they spent most of their lives, accent pattern identified for the region (Hirayama, 1957), time spent in the U.S., and gender. None of these factors showed a statistically significant correlation with listeners' perceptual cue weights.

VI. CONCLUSIONS

The current study examined perceptual weighting of two acoustic dimensions with similar informativeness, exploiting the absolute and relative durations that are nearly equivalent in signaling Japanese singleton and geminate stop categories. In categorizing Japanese singleton and geminate stops, an ideal observer tuned to the distributional regularities of acoustic dimensions would rely more on relative duration for its slight advantage in informativeness. Such an observer would achieve 90%–93% accuracy as opposed to 80%–87% accuracy (Experiment 4 and Idemaru and Guion-Anderson, 2010). Here we have demonstrated that in this case for which two acoustic dimensions are closely matched

in informativeness, there are large individual differences in perceptual cue use across listeners. We observed listeners all across the spectrum with some listeners primarily relying on relative duration, others using mostly absolute duration, and yet others using the two dimensions fairly equally.

Although relative duration has been proposed as a better classifier of rate-dependent speech categories (e.g., stop voicing, singleton and geminate stops, and short and long vowels) with the implication that it is also a better perceptual cue (Kohler, 1979; Port and Dalby, 1982; Pickett *et al.*, 1999; Pind, 1999), there has been little examination of how perceptual weighting of acoustic dimensions may be influenced by distributional regularities of a particular language (relative informativeness of the dimensions). The current results demonstrate that the role of relative duration for the perception of rate-dependent speech categories is conditioned by the distributional regularities of the specific sound contrast in the specific language. More specifically, whether relative duration is a better dimension to rely on for the perception of rate-dependent speech categories depends on the degree to which rate differences undermine informativeness of absolute duration relative to that of relative duration. These results suggest the importance of considering the distributions of available acoustic information within the specific language environment and highlight the value of attention to individual differences across listeners in understanding perceptual cue weighting.

Native Japanese listeners exhibited extensive individual differences in their use of relative duration versus absolute duration. Simply examining the group averages would have concealed these rich individual patterns and would have lead us to believe that relative and absolute duration are equivalently weighted in Japanese speech perception, consistent with the similar ability of the two acoustic cues to accurately inform classification of Japanese speech productions. On the contrary, listeners' category response data showed an intriguing and rich pattern of individual differences.

It would be natural to speculate that the individual cue weighting pattern observed here is due to the similar informativeness of the two acoustic dimensions and that the relative parity of the information allows listeners to freely use either source of information, perhaps varying in which information they use across time. Whether listeners use either of the dimensions separately or integrates the two in any weighting function, they would achieve at least 80% successful categorization. Nevertheless, the current study demonstrates that perceptual cue weighting patterns are remarkably consistent across time and exhibit some resistance to short-term perturbations of cue informativeness in the local acoustic speech context. Although it is reasonable to speculate that the consistency and robustness arise from long-term representations of the speech categories developed through the specific and detailed acoustic distributions experienced by listeners, it is not yet clear how such strikingly stable individual differences emerge. An obvious suspect is listeners' own speech productions. However, the extensive acoustic speech production analyses of Experiment 4 demonstrate that there is no direct relationship with a talker's own elicited speech and individual differences in perceptual cue weighting in speech categorization. Moreover, factors such as listeners' geographic region, time spent in the U.S., and gender were not related to individual differences in perception.

Other acoustic cues may provide a supporting role in categorizing the singleton versus geminate stop contrast in Japanese, thus reducing the need for relative duration. In a previous study, both F0 and intensity of the surrounding vowels were shown to correlate with singleton and geminate categorization (Idemaru and Guion, 2008). Those factors were neutralized in this study by using unbiased values in sound synthesis. However, one hypothesis is that listeners more sensitive to non-durational features such as F0 and intensity of the vowels surrounding the stop use them as secondary cues to support absolute duration instead of relying on relative duration. There is evidence suggesting that a presence or lack of a secondary acoustic correlate influences the way singleton and geminate stops are produced (Engstrand and Krull, 1994). A prediction for future research for Japanese singleton versus geminate stop contrast is that listeners who tend to rely on absolute duration may also make greater use of F0 and intensity compared to relative duration listeners.

Research is still scarce with regard to individual differences in perceptual cue weighting. A few available studies, which have examined stop voicing contrasts, found a wider difference in the reliance on F0 in the vowel following the stop relative to the reliance on VOT (Haggard *et al.*, 1970; Massaro and Cohen, 1976; Kong and Edwards, 2011; Shultz *et al.*, 2012). In particular, Kong and Edwards (2011) found that, in perceiving [da] versus [ta], a quarter of their listeners were categorical listeners, another quarter gradient listeners, and the rest were somewhere in between. Only the gradient listeners used F0 information in the stop voicing categorization. F0 in stop voicing distinction is less informative and a secondary perceptual cue, whereas VOT is the more informative and primary (e.g., Abramson and Lisker, 1985). It is possible that this secondary status of F0 allows greater individual differences in perceptual weighting of this dimension: Perhaps the stake is low whether to optimally use the F0 or not for the result of categorization. The similarly high informativeness (i.e., classification accuracy) of relative and absolute duration perhaps affords Japanese listeners whether or not to integrate the contextual duration to derive a relative duration for a slightly better outcome.

The current findings have cross-linguistic implications. Recall that the informativeness of relative and absolute duration varies across languages. The singleton-to-geminate stop closure ratio is 1 to 2 in Italian, whereas it is 1 to 3 in Japanese (Han, 1994; Idemaru and Guion, 2008, for Japanese; Esposito and Di Benedetto, 1999; Ham, 2001, for Italian). In languages, such as Italian, where the singleton and geminate durational contrast is not robustly differentiated by absolute stop closure, greater category overlap is predicted along the dimension of absolute duration across speaking rates. Such category overlap, in turn, undermines the informativeness of absolute duration. One may hypothesize that this would increase the advantage of relative duration in categorizing singletons and geminates in Italian. Furthermore, in Italian, the vowel duration is shorter before a

geminate than before a singleton stop (Smith, 1993; Esposito and Di Benedetto, 1999), whereas in Japanese it is longer before a geminate than before a singleton stop (Kawahara, 2006; Idemaru and Guion, 2008). This inverse covariation of the vowel and stop durations in Italian can work to enhance the singleton and geminate stop contrast, further increasing the informativeness of relative duration. The positive covariation in Japanese, however, undermines the informativeness of relative duration. Therefore another prediction for future research is that in languages for which informativeness of absolute and relative durations differs significantly (e.g., Italian), listeners may rely more heavily on relative duration than we observed here with Japanese listeners.

The current study has employed laboratory experimental tasks with highly controlled stimuli and elicitation procedure. Whereas the current methodology allowed necessary fine control of acoustic dimensions of interest, it will be important to replicate the study with more ecologically valid setting as well. Thus further research is warranted to verify whether the lack of production-perception relationship is observed in different tasks and with more extended speech materials. This study, nonetheless, demonstrates the importance of examining perceptual cue weighting in the context of the distributional regularities of the acoustic dimensions in question and highlights the importance of examining cue weighting at the individual level. The pattern of perceptual cue weighting can vary greatly across individuals and is remarkably consistent within individuals across time. Whereas the source of the individual differences remains to be discovered in future research, it seems likely to be related to how acoustic information is distributed in the specific language environment.

ACKNOWLEDGMENT

This work was in part supported by grants from the National Institute of Deafness and other Communication Disorders (NIH 5R01DC004674) and the National Science Foundation (BCS 0746067) to L.L.H. We thank two anonymous reviewers for their valuable comments and Shigeto Kawahara for his helpful comments on an earlier version of the manuscript. We also thank Christi Gomez and Sung-Joo Lim for running experiments and Shuhei Okumura for conducting acoustic analysis.

[1]The term "speaking rate" is used in this study to refer to global speaking tempo as well as local speaking rate that may vary due to pragmatic, discourse and sociolinguistic factors.

[2]One participant lived in Germany for a year when she was 9 years old; the other lived in the U.S. from age 9 to 14. The data from these participants were also excluded from Experiments 2–4.

[3]As Japanese allows short and long vowel contrasts when not adjacent to a geminate consonant, *seeta* and *setaa* can also be potential labels for the stimulus words. However, the construction of the stimulus is driven by the variation of stop closure duration from the acoustic data (Idemaru and Guion-Anderson, 2010) and corresponding variation in the preceding syllable duration. As such, this study focuses on listeners' perception of consonant duration as a function of absolute and relative stop duration.

[4]The computer code using statistical package R developed for this analysis can be found at: http://www.stat.cmu.edu/hseltman/langSmooth/langSmooth.R (Last viewed 7/18/2012).

[5]Six of 18 words are lexical words in Japanese when produced in the HL or HLL pitch pattern: *Seppa*, "driven into a corner"; *setta*, "hurried"; *sette*, "being in hurry"; *seto*, type of porcelain; *setto*, loan word meaning "set"; and *seko*, person or place name. Previous research found trends of durational covariations among stops and adjacent segments either using mixed list of real and nonce words (Homma, 1981; Kawahara, 2006) or lexical words alone (Campbell, 1999; Han, 1994). Those results together suggest that the effects of stop length on its own duration and on adjacent segments would be present regardless of the lexical status of the word.

Abramson, A. S., and Lisker, L. (**1985**). "Relative power of cues: F0 shift versus voice timing," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic, New York), pp. 25–33.

Allen, J. S., Miller, J. L., and DeSteno, D. (**2003**). "Individual talker differences in voice-onset-time," J. Acoust. Soc. Am. **113**, 544–552.

Benkí, J. R. (**2001**). "Place of articulation and first formant transition pattern both affect perception of voicing in English," J. Phon. **29**(1), 1–22.

Boersma, P., and Weenink, D. (2006). "PRAAT: Doing phonetics by computer," version 4.4.30.

Boucher, V. J. (**2002**). "Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates," Percept. Psychophys. **64**, 121–130.

Campbell, N. (**1999**). "A study of Japanese speech timing from the syllable perspective," J. Phonet. Soc. Jpn. **3**, 29–39.

Coleman, J. (**2003**). "Discovering the acoustic correlates of phonological contrasts," J. Phon. **31**(3-4), 351–372.

Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. J. (**1977**). "Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues," Percept. Psychophys. **22**(2), 109–122.

Engstrand, O., and Krull, D. (**1994**). "Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion," Phonetica **51**, 80–91.

Escudero, P., and Boersma, P. (**2004**). "Bridging the gap between L2 speech perception research and phonological theory," Stud. Second Lang. Acquis. **26**(04), 551–585.

Esposito, A., and Di Benedetto, M. G. (**1999**). "Acoustical and perceptual study of gemination in Italian stop," J. Acoust. Soc. Am. **106**, 2051–2062.

Francis, A. L., Baldwin, K., and Nusbaum, H. C. (**2000**). "Effects of training on attention to acoustic cues," Percept. Psychophys. **62**(8), 1668–1680.

Francis, A. L., Kaganovich, N., and Driscoll-Huber, C. (**2008**). "Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English," J. Acoust. Soc. Am. **124**(2), 1234–1251.

Fujisaki, H. (**1979**). "On the modes and mechanisms of speech perception: Analysis and interpretation of categorical effects in discrimination," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Ohman (Academic, New York), pp. 177–189.

Ganong, W. F. (**1980**). "Phonetic categorization in auditory word perception," J. Exp. Psychol. **6**(1), 110–125.

Haggard, M., Ambler, S., and Callow, M. (**1970**). "Pitch as a voicing cue," J. Acoust. Soc. Am. **47**, 613–617.

Ham, W. H. (**2001**). *Phonetic and Phonological Aspects of Geminate Timing* (Routledge, New York), pp. 205–244.

Han, M. S. (**1994**). "Acoustic manifestations of mora timing in Japanese," J. Acoust. Soc. Am. **96**, 73–82.

Hankamer, J., Lahiri, A., and Koreman, J. (**1989**). "Perception of consonant length: Voiceless stops in Turkish and Bengali," J. Phon. **17**, 283–298 (**1989**).

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (**2000**). "Some effects of duration on vowel recognition," J. Acoust. Soc. Am. **108**(6), 3013–3022.

Hirata, Y., and Whiton, J. (**2005**). "Effects of speaking rate on the single/geminate stop distinction in Japanese," J. Acoust. Soc. Am. **118**, 1647–1660.

Hirayama, T. (**1957**). *Nihongo Oncho no Kenkyu (A Study of Japanese Accent)* (Meiji Shoin, Tokyo), pp. 132–170.

Holt, L., and Wade, T. (**2004**). "Non-linguistic sentence-length precursors affect speech perception: Implications for speaker and rate normalization," in *Proceedings of From Sound to Sense: Fifty+ Years of Discoveries in Speech Communication*, June, MIT, pp. C49–C54.

Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**(5), 3059–3071.

Homma, Y. (**1981**). "Durational relationship between Japanese stops and vowels," J. Phon. **9**(3), 273–281.
Idemaru, K., and Guion, S. (**2008**). "Acoustic covariants of length contrast in Japanese stops," J. Int. Phonet. Assoc. **38**(2), 167–186.
Idemaru, K., and Guion-Anderson, S. (**2010**). "Relational timing in the production and perception of Japanese singleton and geminate stops," Phonetica **67**(1–2), 25–46.
Idemaru, K., and Holt, L. L. (**2011**). "Word recognition reflects dimension-based statistical learning," J. Exp. Psychol. Hum. Percept. Perform. **37**(6), 1939–1956.
Ingvalson, E. M., McClelland, J. M., and Holt, L. L. (**2011**). "Predicting native English-like performance by native Japanese speakers," J. Phon. **39**, 571–584.
Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**(1), 47–57.
Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**(3), 1252–1263.
Kawahara, S. (**2006**). "A faithfulness ranking projected from a perceptibility scale: The case of [+ voice] in Japanese," Language **82**(3), 536–574.
Keating, P., and Huffman, M. (**1984**). "Vowel variation in Japanese," Phonetica **41**(4), 191–207.
Kessinger, R. H., and Blumstein, S. E. (**1998**). "Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies," J. Phon. **26**(2), 117–128.
Klatt, D. H. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.
Klatt, D. H. (**1979**). "Synthesis by rule of segmental durations in English sentences," in *Frontiers of Speech Communications Research*, edited by B. Lindblom and S. Ohman (Academic, New York), pp. 287–299.
Kluender, K. R., and Walsh, M. A. (**1992**). "Amplitude rise time and the perception of the voiceless affricate/fricative distinction. *Attention*," Percept. Psychophys. **51**(4), 328–333.
Kohler, K. J. (**1979**). "Dimensions in the perception of fortis and lenis plosives," Phonetica **36**(4–5), 332–343.
Kong, E. J., and Edwards, J. (**2011**). "Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking," in *Proceedings of International Congress of Phonetic Sciences*, August, Hong Kong, China, pp. 1126–1129.
Lahiri, A., and Hankamer, J. (**1988**). "The timing of geminate consonants," J. Phon. **16**, 327–338.
Lehiste, I., and Peterson, G. E. (**1961**). "Some basic considerations in the analysis of intonation," J. Acoust. Soc. Am. **33**(4), 419–425.
Lisker, L. (**1986**). "'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang. Speech **29**(1), 3–11.
Loader, C. (**1999**). *Local Regression and Likelihood* (Springer, New York), pp. 1–233.
Lotto, A. J., Sato, M., and Diehl, R. L. (**2004**). "Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/," in *Proceedings of From Sound to Sense: 50+ Years of Discoveries in Speech Communication*, June, MIT, pp. C381–C386.
Magen, H. S., and Blumstein, S. E. (**1993**). "Effects of speaking rate on the vowel length distinction in Korean," J. Phon. **21**, 387–409.
Massaro, D. W., and Cohen, M. M. (**1976**). "The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction," J. Acoust. Soc. Am. **60**, 704–717.
McMurray, B. (**2000**). "KLATTWORKS: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research," (computer program).
Miller, J. L., and Baer, T. (**1983**). "Some effects of speaking rate on the production of /b/ and /w/," J. Acoust. Soc. Am. **73**(5), 1751–1755.
Miller, J. L., and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**(6), 457–465.
Morrison, G. S. (**2007**). "Logistic regression modeling for first and second language perception data," in *Segmental and Prosodic Issues in Romance Phonology*, edited by M. J. S. Solé, P. Prieto, and J. Mascaró (John Benjamins, Amsterdam), pp. 219–236.
Nearey, T. M. (**1990**). "The segment as a unit of speech perception," J. Phon. **18**(3), 347–373.
Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693.
Pickett, E. R., Blumstein, S. E., and Burton, M. W. (**1999**). "Effects of speaking rate on the singleton/geminate consonant contrast in Italian," Phonetica **56**(3–4), 135–157.
Pierrehumbert, J., and Beckman, M. (**1988**). "Japanese tone structure," Ling. Inq. Monog. **15**, 1–282.
Pind, J. (**1986**). "The perception of quantity in Icelandic," Phonetica **43**(1–3), 116–139.
Pind, J. (**1999**). "Speech segment durations and quantity in Icelandic," J. Acoust. Soc. Am. **106**, 1045–1053.
Polka, L., and Strange, W. (**1985**). "Perceptual equivalence of acoustic cues that differentiate /r/ and /l/," J. Acoust. Soc. Am. **78**, 1187–1197.
Port, R. F., and Dalby, J. (**1982**). "Consonant/vowel ratio as a cue for voicing in English," Percept. Psychophys. **32**(2), 141–152.
Raizada, R. D. S., Tsao, F. M., Liu, H. M., and Kuhl, P. K. (**2010**). "Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences," Cereb. Cortex **20**(1), 1–12.
Raphael, L. J. (**2005**). "Acoustic cues to the perception of segmental phonemes," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Malden, MA), pp. 182–206.
Shultz, A. A., Francis, A. L., and Llanos, F. (**2012**). "Differential cue weighting in perception and production of consonant voicing," J. Acoust. Soc. Am. **132**(2), EL95–EL101.
Smith, C. (**1993**). "Prosodic patterns in the coordination of vowel and consonant gestures," Haskins Lab. Rep. Speech Res. **115–116**, 45–55.
Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). "An investigation of locus equations as a source of relational invariance for stop place categorization," J. Acoust. Soc. Am. **90**(3), 1309–1325.
Toscano, J. C., and McMurray, B. (**2010**). "Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics," Cogn. Sci. **34**(3), 434–464.
Vance, T. J. (**1987**). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY), pp. 56–76.
Yamada, R. A., and Tohkura, Y. (**1992**). "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners," Percept. Psychophys **52**, 376–392.