Journal of Phonetics 50 (2015) 81–98

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

Cross-language similarity and difference in quantity categorization of Finnish and Japanese

Kenji Yoshida[a,*], Kenneth J. de Jong[b,c], John K. Kruschke[c,d], Pia-Maria Päiviö[e]

[a] School of Culture, Media and Society, Waseda University, 1-24-1 Toyama, Shinjuku-ku, Tokyo 162-8644, Japan
[b] Department of Linguistics, Indiana University, Memorial Hall 322, 1021 East 3rd Street, Bloomington, IN 47405-7005, USA
[c] Cognitive Science Program, Indiana University, 819 Eigenmann Hall 819, 1900 East 10th Street, Bloomington, IN 47406-7512, USA
[d] Department of Psychological and Brain Sciences, Indiana University Bloomington, 1101 East 10th Street, Bloomington, IN 47405-7007, USA
[e] Department of Scandinavian Studies and Linguistics, University of Copenhagen, Njalsgade 120, 2300 København, Denmark

ARTICLE INFO

Article history:
Received 14 June 2011
Received in revised form
16 December 2014
Accepted 26 December 2014
Available online 12 March 2015

Keywords:
Quantity categorization
Cue co-variation
Language-specific cues
Word-level prosody
Finnish
Japanese

ABSTRACT

The present study investigates how listeners of Finnish and Japanese, languages with very similar contrasts in plosive quantity (short vs. long), use language-specific phonetic knowledge of acoustic attributes which covary with closure duration. A fully-crossed perceptual experiment on consonant quantity categorization was conducted with 22 Finnish and 20 Japanese listeners, using natural speech stimuli with systematically manipulated closure durations. Stimuli were created from Finnish and Japanese productions of both long and short plosives. In the naturally produced stimuli, the duration of the vowel preceding the target consonant was affected by language-specific word prosody patterns. The results showed an across-the-board effect of the original quantity of the produced stimulus, with the stimuli created from words with a geminate consonant tending to be perceived as geminate, irrespective of the language of the listeners or talkers, suggesting a strong influence of cues besides closure duration that are shared by the two languages. However, Japanese listeners were more heavily affected by the acoustic cues concomitant to the quantity contrast in their native language, likely due to robust language-specific vowel duration effects in the Japanese productions. Word prosody, besides creating subtle shifts in category boundary for both language groups, created confusions in the listener responses, especially when the language-specific word-level prosodic effect is localized in the vowel preceding the target word, and when the listener is confronted with stimuli from the other language. Thus, some aspects of the quantity contrast are remarkably similar between the two languages, and listeners from one language group are attuned to attributes found in the other language. However, other attributes appear to be language specific, and such durational differences may interfere with categorization in an unfamiliar language.

1. Introduction

1.1. Introduction: cross-linguistic similarity and differences in plosive quantity

Research in phonetics over the last 25 years has pervasively found language-specific differences in the production of and hence perception of phonological elements, from vowel differences measured in the vowel plane (e.g., Escudero, Simon, & Mitterer, 2012; Ferrari Disner, 1977, 1983; Pols, van der Kamp, & Plomp, 1969; Strange, Bohn, Trent, & Nishi, 2004; Terbeek, 1977; Terbeek & Harshman, 1971), to consonant place differences (e.g., Akpanglo-Narty, 1982; Maddieson, 1980), to differences in laryngeal specifications (e.g., Caramazza, Yeni Komshian, Zurif, & Cabone, 1973; Keating, Mikos, & Ganong, 1981). Indeed, the thrust of this large body of research is no longer whether languages can differ idiosyncratically, but rather how languages are constrained from varying idiosyncratically. Besides being important for phonetic theory in general, this observation has pragmatic import for speakers dealing in multiple languages, raising the question of whether and to what extent the phonetic capabilities developed for one language can be carried over into a second language without diverging from the native population.

* Correspondence to: 7-16-7 Asahigaoka, Tajimi-city, Gifu 507-0071, Japan. Tel./fax: +81 572 29 3600.
E-mail addresses: kenjiyo.work@gmail.com (K. Yoshida), kdejong@indiana.edu (K.J. de Jong), kruschke@indiana.edu (J.K. Kruschke), hng786@hum.ku.dk (P.-M. Päiviö).

0095-4470/$ - see front matter © 2015 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.wocn.2014.12.006

The current study compares native perceptual capabilities in speakers of two languages with respect to a distinction that has been shown to be very similar across the two languages, stop quantity. According to the survey by Ham (2001: 213), Finnish and Japanese have productions with a fairly close measured singleton-to-geminate duration ratio (2.14 and 2.25). Aoyama (2001) examined listeners of Finnish and Japanese, finding similar perceptual treatment of variation in consonant closure duration by listeners of both languages. She conducted a word identification experiment on both Finnish and Japanese speakers, with an identical procedure for the two listener groups: a series of 10 speech stimuli in a continuum from /hana/ to /hanna/ (created by incrementally deleting the nasal consonant from the original production of *hanna*) were presented to each listener for word identification (*hana* or *hanna*). In this study, stimuli were made from speech by speakers of the listeners' language, a Finnish speaker for Finnish listeners and a Japanese speaker for Japanese listeners. The results reveal a clear sigmoidal identification function of quantity identification as a function of nasal duration. Aoyama (2001: 59) concludes that “the categorical boundary between single and geminate nasals seems to be very similar in Finnish and Japanese.” The authors of the current paper submitted these identification data (Tables 14, 15, 18, 19 and Appendix IV from Aoyama, 2001) to a logistic regression with acoustic duration of /n-nn/ in the speech stimuli as the predictor for each of the two listener groups, respectively. While the slopes of the logistic regression are somewhat steeper with the Finnish listeners (Finnish = 0.129, Japanese = 0.116), the identification boundaries for the two groups are extraordinarily similar (Finnish = 106 ms, Japanese = 107 ms).

This similarity suggests that there may be some stereotypical durational property rooted either in production or perception that can be used by these different language communities. Either coordinative effects in production tend to independently create distributions which are similar in the two languages, or the perceptual systems have some propensity for detection of certain durational properties that makes the two groups similar. What either of these scenarios suggests is that, despite cross-linguistic differences, it is possible that the active perceptual skills developed via experience with one language might be neatly transferred to a second language.

The current study probes the perceptual system of speakers of Japanese and Finnish more in-depth, examining how they deal with variation in the signal which covaries with the quantity-marking closure duration. While closure duration is always the primary determinant of quantity, previous studies have demonstrated the importance of other acoustic properties than closure duration, such as formant values of adjacent vowels, fundamental frequency (*f*0), and intensity (Idemaru & Guion, 2008, for Japanese production; Doty, Idemaru, & Guion, 2007; Isei-Jaakkola, 2010; Järvikivi, Aalto, Aulanko, & Vainio, 2007; Järvikivi, Vainio, & Aalto, 2010; O'Dell, 2003; Richardson, Leppänen, Leiwo, & Lyytinen, 2003, for Finnish productions; and Kingston, Kawahara, Chambless, Mash, & Brenner-Alsop, 2009; Ofuka, Mori, & Kiritani, 2005, for perceptual perturbation of the previous vowel duration, and Nagano-Madsen, 1992 for perceptual perturbation of the *f*0 contour such that a falling *f*0 contour has a very strong tendency to induce “long” perception for Japanese). O'Dell (2003) presented modified productions of the Finnish contrast, finding a moderate effect of the original produced word (*katoa* vs. *kattoa*) for Finnish listeners, even when the temporal organization of the entire test word was modified to be neutral. This suggests a range of additional covariates besides segment duration, such as the formant dynamics of the transition into the closure, the strength of the burst for plosives, and possibly even subtle differences in the spectra of the bursts corresponding to place of articulation differences.

Such variation would seem to come in two types. The first are aspects of the acoustics surrounding the closure which arise from the local dynamics and aerodynamics of producing a contrast in duration. For example, longer closure durations tend to create larger pressure differentials between the oral cavity and outside air, which then lead to stronger bursts for geminate than for singleton consonants (e.g., as in Doty et al., 2007; Local & Simpson, 1999). Since such variation seems to inhere in the facts of production, one would expect that experience with quantity in one language would form a solid basis for exploiting such covariation in the second language.

The second type of covariation is that introduced by the larger phonetic system in a particular language, so that secondary cues begin to explicitly covary with the primary cue. For example, studies of Japanese have shown that the vowel preceding a geminate consonant is longer before geminates than before singletons. This is the opposite of the commonly found compensatory tendency for a vowel to be shorter before a longer consonant. Maddieson (1985: 94) argued that the compensatory tendency is virtually universal and acknowledged Japanese as an “apparent counter example”. Following Port, Dalby, and O'Dell (1987: 1582), we will refer to the pattern found in Japanese as “anti-compensation”. This pattern of variation in vowel duration has been replicated repeatedly in production studies and seems very robust (Aoyama, 2001; Campbell, 1999; Ofuka et al., 2005; Port et al., 1987). This anti-compensating variation in vowel duration has also been demonstrated to be a perceptual cue for consonant quantity for Japanese (Kingston et al., 2009; Ofuka et al., 2005). Kingston et al. (2009) investigated categorization and discrimination of Japanese quantity by Japanese, Norwegian, Italian and English listeners. Japanese listeners shifted the singleton/geminate identification threshold to a shorter closure duration value when the preceding vowel is longer. This shift was opposite from that found in the behavior of Norwegian and Italian listeners (Kingston et al., 2009: 305).

The results in previous acoustic studies on Finnish are, by contrast, mixed in this regard. While Maddieson (1985: 90) drew Finnish as an example where vowels are shorter before geminate consonants, there are also studies suggesting that vowel duration before singleton/geminate in Finnish is mildly “anti-compensating” (Lehtonen, 1970: mean vowel duration of 64 ms before singletons vs. 77 ms before geminates; Ylitalo, 2009: 60 ms before singletons vs. 70 ms before geminates). Also in the production study in Aoyama (2001), Finnish speakers exhibit only a 13% increase (90–102 ms) in the duration of the vowel preceding the word-medial consonant from /hana/ to /hanna/, whereas Japanese speakers show a 45% increase (53–77 ms) for the same minimal pair. Myers and Hansen (2005: 329) also found an anti-compensating pattern for Finnish vowels but report that “this effect varied from speaker to speaker”. In sum, the pattern of anti-compensation in consonant and the preceding vowel seems more robust and systematic in Japanese than in Finnish, and thus is a language-specific pattern.

One complication with respect to this anti-compensating pattern is that fact that both the vowel and the consonant are also affected by general effects of prosodic position in word (e.g., Amano & Hirata, 2010; Port et al., 1987), and overall tempo (e.g., Amano & Hirata, 2010; Idemaru & Guion, 2010; Pickett, Blumstein, & Burton, 1999). Dealing with such variation has led various researchers to investigate these larger patterns of covariation with an eye toward finding some "relational invariance" (Pickett et al., 1999). These studies tend to look at the duration of segments in regions surrounding the consonant, e.g., the ratio of duration of the singleton/geminate consonant and the entire word (CW ratio, Amano & Hirata, 2010; Hirata & Whiton, 2005) or ratio with respect to the preceding mora (Idemaru & Guion, 2010), which is claimed to be constant under variation due to speech rate. Based on such findings, Amano and Hirata (2010: 2057) anticipated that acoustic information is rather localized within "two syllables between which the quantity distinction occurs". Important for the present discussion, to the extent that tempo variation is not language specific, such relational normalization in one language would be expected to be evident in listeners from both language groups, and would constitute a perceptual mechanism which would naturally generalize from one language to the other language.

There are, however, factors that create differences in the covariate in one language, which are clearly not general to both languages. For example, vowel duration in Finnish has been shown to be sensitive to prosodic position in the word in a way that apparently is idiosyncratic to Finnish. Vowels in the second syllable of a word (V_2) vary in duration depending on the syllable structure of the word-initial syllable. The phonemically short vowel in the second syllable of the word is considerably longer when the first syllable is CV, as compared to when the first syllable is CVC or CVV (e.g., Lehtonen, 1970; Suomi, 2005; Suomi, Toivonen, & Ylitalo, 2003; Ylitalo, 2009). The extent of this lengthening is so large that the term "half-long" has been traditionally applied to it.[1] For example, the second, phonologically short vowel in a form like *kana* is "half-long", being about 1.5 times longer than that of the first vowel, whereas V_2 in a form like *kanta* is considerably shorter (Suomi, 2005: 297). Such effects are not reported in the numerous studies of Japanese vowel durations.

1.2. The current study

The current study investigates how aspects of the context of closure duration influence quantity categorization for two groups with different native languages. Our primary interest is in those differences that covary with phonological contrast of quantity (short vs. long) in consonants. We further assume that the covariation enhances the perceptual distinctiveness of contrasts in phonological quantity and thus forms part of speaker's phonetic knowledge, as discussed for the [voice] contrast in Kingston and Diehl (1994). As reviewed in the previous section, part of this covariation is likely to be language-specific. Specifically, the pattern of anti-compensation between closure duration and the duration of preceding vowel is more robust and consistent in Japanese than in Finnish. To the extent that such covariation is language-specific, one would expect the perceptual sensitivity to the differences in the acoustic effects in geminate categorization to be language-specific as well. Hence, categorization behavior might be very similar across the languages in one condition, but diverge in situations where the covariate may have some different value.

In addition, we investigate the perceptual relevance of acoustic differences related to word prosody. Specifically, as also reviewed in the preceding section, the pattern of variation in acoustic duration in V_2 (half-long vs. short) depending on the word-initial syllable structures is found only for Finnish. Experience specific to the language would lead listeners to perceptually compensate for such variation. On the contrary, when these effects are unfamiliar to the listener, i.e., when found in an unfamiliar language, we would not expect compensation for these factors, and expect that such factors will present something of a challenge for speakers acquiring that second language.

To operationalize these conditions, the current perceptual study collected Finnish and Japanese speech material consisting of segments generally considered to be similar across the two languages and modified the closure duration of the target plosive. Contextual variation is obtained by creating stimuli from both geminate and singleton productions. In such stimuli, contextual variation inherent in the production of the geminate and singleton consonants is retained across the range of closure durations, and so is orthogonal to the duration variable. Such variation would include vowel duration anti-compensation in the Japanese productions, as well as minor differences in burst-strength and formant dynamics, etc., which are expected to be located around the closure. Contextual variation which is language-specific is examined by presenting stimuli with different bases to listeners of the other language. To the extent that these production covariates differ across the two languages and such differences are not prefigured in the experience of the listeners, we do not expect the listeners to exhibit sensitivity in their categorization behavior.

In addition, variation due to word prosody was examined by obtaining base stimuli with the target consonant produced in forms with different word prosody. In all stimuli, the target consonant was in the word medial position where the word-level prosody is likely to affect the duration of the vowel preceding it. For Finnish, this was done in two ways. The structure of the word-initial syllable either did or did not have a coda consonant (CV or CVC). Thus, in the CV-initial forms, the vowel preceding the target consonant is in the "half-long position", and so is considerably longer than that in the CVC-initial forms (see Table 2, Section 2.1.3). In addition, a third condition added another syllable between the target and the beginning of the word, in order to move the initial stress in the Finnish forms away from the target. The point of both of these manipulations is to probe the effect of inducing Finnish-specific natural production differences into the prosodic dynamics of the stimuli. We expect Finnish listeners to naturally compensate for such

[1] Word-level temporal organization has been reported to differ across Finnish dialects. According to the most recent and systematic documentation by Ylitalo (2009), while the half-long vowel is characteristically observed in northern and eastern Finnish dialects, the second vowel is not considered half-long in Tampere dialects (Central Häme dialect; Research Institute for the Language of Finland, 2006). However, Ylitalo's study nevertheless indicates that all three dialects of Standard Spoken Finnish have the same direction of difference in V_2 duration depending on the structure of word-initial syllable, though the magnitude varies across the three varieties (comparing CV.CV.(X) and CVC.CV.(X) words in Tables 5 and 11; Ylitalo, 2009).

Table 1
Six nonsense experimental words (3 minimal pairs contrasting with /p/ vs. /pp/). Target consonants /p/ or /pp/ are in bold. Syllable breaks are indicated by periods.

Syllable structure	CV (*mata*)-initial	CVC (*manta*)-initial	CV (*manata*)-initial
Target *p~pp* in	3rd σ/3rd μ	3rd σ/4th μ	4th σ/4th μ
Singleton	*ma.ta.**p**a.na*	*man.ta.**p**a.na*	*ma.na.ta.**p**a.na*
Geminate	*ma.ta**p**.**p**a.na*	*man.ta**p**.**p**a.na*	*ma.na.ta**p**.**p**a.na*

differences in the prosodic pattern surrounding the target consonants. However, since Japanese does not have the "half-long effect", nor does it have extensive stress variation (Beckman, 1986), we expect such variation to exhibit an (inappropriate) influence on the categorization of the Japanese listeners, or at least to create confusion for them.

2. Methods

2.1. Design of perceptual experiment

A fully-crossed non-word categorization experiment was conducted with Finnish and Japanese speakers as listeners. One female Finnish speaker and one female Japanese speaker provided the original speech materials for the acoustic stimuli with different word-prosodic structures. The closure epoch in the target consonants (/p/ or /pp/) in three word-prosody contexts were replaced by silent intervals with 7 different durations in equal steps. These acoustic stimuli are mixed together to yield a stimulus set consisting of 84 stimuli (2 Talker's Languages $\times$ 2 Bases $\times$ 3 Word Prosody $\times$ 7 steps). This set of acoustic stimuli was presented to both Finnish and Japanese listeners for non-word categorization. The response choices were the nonsense words containing either singleton /p/ or geminate /pp/ (see Table 1). Each listener listened to each stimulus 9 times, thus providing 756 responses in total.

2.1.1. Stimulus bases

Table 1 lists the 6 nonce words used for the non-word identification experiment. All of them are phonotactically legal but not real words in either language. Nonce words were used in order to minimize the influence of lexical status and frequency of existing words in quantity categorization (Moreton & Amano, 1999).[2] For all 6 words, the target consonant to be categorized either as /p/ or /pp/ is preceded by a syllable /ta/. The target words have either singleton or geminate consonants. The primary purpose was to elicit production differences inhering in the acoustics around the target closure duration for the geminate and singleton consonants. These production differences would also include covarying patterns in the vowel duration, such as the anti-compensation pattern in Japanese. As the durations in Table 2 show, the Japanese tokens did exhibit an anti-compensating pattern.

As the columns of Table 1 indicate, there are three types of word prosody. The first two differ in the structure of the word-initial syllable, being either CV or CVC (*mata*- or *manta*-initial condition). The vowel /a/ preceding the target consonant /t/ is in the "half-long position" for the *mata*-initial condition, and so it is expected that Finnish talkers produce a longer vowel as compared with the *manta*-initial condition, which was indeed the case (see Table 2). The third type of word structure was included by adding one more open syllable /na/ before /ta/ (*manata*-initial condition, 4th column of Table 1). In this case, the vowel preceding the target consonant is in the "half-long position", unlike the CVC-initial (*manta*-initial) condition. The vowel preceding the target consonant /p/ or /pp/, i.e., the vowel in the 3rd syllable /ta/, was produced with a comparable duration to the one in the *mata*-initial condition (see Table 2). The purpose of including this pair is to examine the effect of moving the word-initial stress in Finnish away from the target consonant.

A word-final syllable /na/ was added to all the test words to avoid the influence of larger variability of production and lower perceptual sensitivity of segmental duration in word final position (Kato, Muto, Tsuzaki, & Sagisaka, 2003). This would also discourage Finnish listeners from identifying the vowel of the target syllable /a/ as a word final case marker (indicating *partitive*). The consonant /p/ was chosen for the target consonant because bilabial consonants are known to have shorter VOT and less salient burst releases than the consonants with other places of articulation, reducing differences in noise transients between the geminate and singleton originals. An anonymous reviewer, however, pointed out that there is a potential problem caused by a special status of /p/ in Japanese phonology. Due to historical change, /p/ has been argued to have disappeared from the phonological inventory of Japanese and has been reintroduced fairly recently through loanword adaptation. A singleton /p/ is still disallowed for native words, which might skew the categorization of the experimental words toward geminate. However, because of the word forms involving /p/, the experimental words should sound like Western loan words for the Japanese participants, where singleton /p/ is completely legal. These possible sources of bias will have to be borne in mind in interpreting the current data. The vowels are held constant to /a/ to reduce the effect of intrinsic durational differences in vowels. The consonants were made to vary within a word to avoid articulatory difficulties.

2.1.2. Recording of the stimulus bases

One female Finnish speaker (born 1981, Imatra, Eastern Finnish dialect region) and one female Japanese speaker (born 1977, Kawasaki, a city adjacent to Tokyo) provided the original material for the speech stimuli. Neither of them has any professional training

[2] Idemaru and Guion (2010: 30) reviewed previous studies of Japanese geminates (some of which used nonse words) and concluded that generally the same sort of temporal organization is observed "regardless of the lexical status of the word".

Table 2
Pattern of acoustic duration for the vowel preceding the target consonant, the range of closure duration, and the step interval in milliseconds.

(a) Finnish stimuli				
Structure		*mata*+target	*manta*+*target*	*manata*+*target*
Preceding V (ms)	before /p/	69	46	65
	before /pp/	64	47	66
Closure duration (ms)		75–148	76–159	77–151
Step interval (ms)		12.1	13.8	12.4
(b) Japanese stimuli				
Structure		*mata*+target	*manta*+*target*	*manata*+*target*
Preceding V (ms)	before /p/	65	68	68
	before /pp/	95	103	98
Closure duration (ms)		69–141	71–144	74–135
Step Interval (ms)		12.0	12.2	10.1

in linguistics or phonetics. For both Finnish and Japanese, the target words were embedded in a frame sentence, Finnish: *Kouvolan uusi* (target) *on mahtava* 'Kouvola's new (target) is great', Japanese: *Tugiwa* (target) *to iimasu*. 'Next I say (target)', respectively. The two talkers produced the 6 utterances with 4 other unrelated utterances 6 times in randomized orders. The Finnish speaker produced all the tokens with word-initial primary stress and with no obvious secondary stress. The Japanese speaker was instructed to produce all the tokens as unaccented forms (with no lexical pitch accent) and did so as judged by the first author, who is a native speaker of Japanese. The speech materials were recorded digitally with a sampling rate of 22,050 Hz and 16 bit quantization.

2.1.3. Creation of speech stimuli

Acoustic measurements of segmental duration were carried out for all the samples provided by the Finnish and Japanese talkers. The mean segmental duration was computed for all the segments of the 6 experimental words (e.g., /m, a, t, a, p, a, n, a/ for /matapana/). Then Pearson product moment correlation coefficients were computed between the vector of segmental durations and that of the mean durations for each of the 6 samples. The sample with the highest correlation coefficient, i.e., the sample whose pattern of segmental duration is the closest to the means, was chosen as the “base token” for each word and for each speaker from which speech stimuli were created. These recordings are included with the on-line version of this article.

The closure epoch of the target consonant /p/ was eliminated and replaced with silent intervals of 7 different lengths. The range from the shortest to the longest intervals was determined by the mean values of /p/ and /pp/ for each minimal pair and the duration of the silence was varied from the shortest (/p/) to the longest (/pp/) in equal steps. Therefore the smallest and largest values and the intervals between steps are slightly different across the three minimal pairs, as summarized in Table 2 (4th and 5th rows). The portion of the speech signal between the closure release and the vowel onset (VOT) was not eliminated from the original token: i.e., the closure duration was measured between the offset of preceding vowel to the closure release. All acoustic measurements, segmentation of original speech stimuli and insertion of silence were carried out in Praat Ver. 5.1.08. (Boersma & Weenink, 2009).

Table 2 reveals that the durations of the vowels preceding the target plosives vary as expected from previous studies. For the Finnish speaker, the preceding vowel is longer for *mata*- and *manata*- and shorter for *manta*-initial conditions. The difference between the two vowels was around 20 ms in our recordings, which is substantially smaller than what has been reported as “half-long” in the literature on Finnish phonetics (around 60 ms or larger). An anonymous reviewer has suggested that this may be because the speaker in the present experiment does not represent the variant of Finnish where typical case of “half-long” vowels is found. While this may be the case, most of the previous acoustic observations of the half-long vowel have been made in shorter words such as *kata*, *kanta*, *mipatu*, *mippatu*. The smaller range is thus may be due to the compression of the segmental duration in longer words. For the Japanese speaker, the preceding vowels are longer before geminates, again as expected. Idemaru and Guion (2008: 183) questioned the perceptual relevance of the preceding vowel for the quantity contrast in Japanese, which they found to be fairly small. In our stimuli, the magnitude was about 30 ms, which is comparable to the values in Ofuka et al. (2005). The Finnish speaker, as expected, did not exhibit anti-compensation. The duration of the vowels is not exactly equal across conditions. However, since the focus of the current study was to examine perceptual effects of natural variation, we decided not to change any part of the original token except for the closure duration. The differences in duration across the stimuli were quite small relative to previous work. For example, in Kingston et al. (2009: 301), the duration of the preceding vowel was varied by 40 ms (from 49 to 89 ms), or about 82%, to achieve a shift of threshold of about 15 ms. In our stimuli, the largest difference of the vowel preceding the target consonant between Finnish /mata-/ and /manata-/ is 4 ms (69 vs. 65 ms or about 6%). All stimuli were equalized in peak intensity at −1.0 dB using the “normalize” function of Audacity 1.3.

2.2. Participants

Finnish speakers were recruited in Helsinki, Finland from the subject pool of the Department of Speech Sciences and the Department of Linguistics at the University of Helsinki. There were 23 participants, but the data from 1 participant were discarded due to equipment failure, leaving the data from 22 participants (female = 17, male = 5; age: 20–52, median age = 30.5). Among them, 14 participants were from the greater Helsinki region, 2 from the Southern area outside of Helsinki, 2 from the Western and 4 from the Eastern area of Finland (as indicated in Research Institute of the Languages in Finland, 2007). The difference of vowel duration depending on the structure of the word-initial syllable has also been replicated for Finnish speakers recruited at the University of Helsinki in Vainio, Järvikivi, Aalto, and Suni (2010), Table VI. Thus, it would be safe to assume that this durational variation is a part of the phonetic experience of all of the Finnish participants in the present study.

Japanese speakers were recruited in Japan from the subject pool of Showa Music Academy in Kawasaki City and Hiroshima University in Hiroshima City. There were 22 participants, but the results from 2 were discarded due to equipment failure, leaving the data from 20 participants (female = 16, male = 4; age: 19–30, median age = 20.5). All Japanese participants are from regions where mora-based dialects are spoken (as indicated in Jo'o, 1977: 135).

No speakers reported abnormal speech or hearing abilities and no Finnish speakers had any experience of learning Japanese, or vice versa. All the participants were paid for their participation.

2.3. Experimental procedure

For Finnish listeners, experiments were conducted in a quiet lab facility in the Department of Speech Science at Helsinki University and, for Japanese listeners, in a quiet meeting room in Showa Music Academy and Hiroshima University. The experimental session began with a short speech production task, where participants were asked to read the sentences used for eliciting the recordings from which the stimuli were made. This was intended to familiarize participants with the nonce words they were going to hear. Following this, a 2AFC (2-alternative forced choice) perception test was carried out beginning with a short practice (with 4 unrelated words by the same Finnish and Japanese listeners) to familiarize the participants with the task. Speech stimuli were presented one at a time monaurally through headphones, and participants were asked to choose the ear with which they listened to the stimuli.[3] Upon presentation of each of the stimuli, two visual alternatives were shown horizontally on a computer screen from which participants had to choose. The alternatives were the pair of nonsense words (as in Table 2) written in Roman script for Finnish and *hiragana* for Japanese listeners. The participants were instructed to press the left or right *alt* key to report which alternative the stimulus sounds more similar to. Participants had to respond before a timeout of 3 s after the onset of the speech stimulus for their responses to be recorded. The inter-stimulus interval was 3000 ms and then the next speech stimulus was presented immediately. There were 9 blocks in total. In each trial within each block, each of the 84 stimuli was presented in isolation (without the frame sentence) in randomized orders (thus, there were 9 responses per each stimulus). The timing of the presentation of acoustic and visual stimuli, stimulus randomization and recordings of the categorization responses and reaction times were controlled by DMDX (Forster & Forster, 2003). The randomization of the order of stimulus presentation was automatically done by DMDX for each block for each participant. The 84 stimuli were presented continuously (without a break) within each block, with untimed breaks between blocks. The entire session lasted from 50 min to 1 h.

2.4. Statistical analysis: Bayesian hierarchical logistic regression

The task performed by our listeners is a two-alternative forced choice between singleton /p/ and geminate /pp/ categories. The primary characterization of the categorization behavior is in terms of (a) the *threshold* of closure duration that is required for the listener to change from singleton to geminate responses, and (b) the rapidity of that change as a function of closure duration, which we call the *gain*. The probability of geminate categorization is modeled as a logistic function of closure duration, with the form

$$p(y = /\mathrm{pp}/) = \mathrm{logistic}[\beta(CD-\theta)] \equiv \frac{1}{1+\exp(-\beta(CD-\theta))} \tag{1}$$

where y is the categorical response, CD is closure duration, β is the gain, and θ is the threshold. The parameters have intuitive meanings. The threshold θ is the closure duration at which the probability of geminate response switches from below 50% to above 50%, and the gain β represents the rapidity of that switch as a function of closure duration.

We used a hierarchical model in which every participant's data were described by an individual-level logistic function, and the individual logistic parameters were simultaneously modeled by higher group-level distributions. The hierarchical approach accommodates individual differences while mutually informing the individual estimates via the higher-level group distribution.

The details of the hierarchical model are as follows. The singleton/geminate categorization data from the ith individual were modeled as a logistic function of closure duration, as in Eq. (1), with the gain and threshold parameters decomposed into influences of Base, Word Prosody, and Talker Language:

$$p(/\mathrm{pp}/) = \mathrm{logistic}\left[\beta^{[i]}_{B,WP,TL}\left(CD-\theta^{[i]}_{B,WP,TL}\right)\right] \tag{2}$$

where the gain parameter $\beta^{[i]}_{B,WP,TL}$ for the ith individual can take on different values depending on the Base B (geminate or singleton), the Word Prosody WP (*mata*-, *manta*-, or *manata*-), and the Talker Language TL (Japanese or Finnish), and where the threshold parameter $\theta^{[i]}_{B,WP,TL}$ can also take on different values in the same way.

[3] The sound presentation system available at the time of the experiment allowed only a single channel playback. There appears to be no significant tendencies related to the choice of the ear to which the stimuli were presented.

In principle, we could estimate separate $\beta^{[i]}_{B,WP,TL}$ and $\theta^{[i]}_{B,WP,TL}$ values for every possible combination of the factors B, WP, and TL. However, our interest in this study is limited to two interactions involving B × TL and WP × TL crossed with listener language, as discussed below in Section 2.5. Therefore, to include the theoretically interesting interactions in the model, the gain parameter from Eq. (2) is expanded as

$$\beta^{[i]}_{B,WP,TL} = \beta^{[i]}_{0} + \beta^{[i]}_{B} + \beta^{[i]}_{WP} + \beta^{[i]}_{TL} + \beta^{[i]}_{B\times TL} + \beta^{[i]}_{WP\times TL} \tag{3}$$

where $\beta^{[i]}_{0}$ is a baseline and the other terms indicate deflections due to main effects and interactions of the manipulated factors. For example, $\beta^{[i]}_{TL[FIN]}$ represents the change from baseline gain when TL is Finnish, whereas $\beta^{[i]}_{TL[JPN]}$ represents the opposite change from baseline gain when TL is Japanese. By the same logic, the threshold parameter in Eq. (2) is expanded analogously as

$$\theta^{[i]}_{B,WP,TL} = \theta^{[i]}_{0} + \theta^{[i]}_{B} + \theta^{[i]}_{WP} + \theta^{[i]}_{TL} + \theta^{[i]}_{B\times TL} + \theta^{[i]}_{WP\times TL}. \tag{4}$$

As is standard in ANOVA-style decompositions such as these, the deflections from baseline in Eqs. (3) and (4) were constrained to sum to zero, so that the baseline and deflections could be identified (otherwise any constant could be added to the baseline and subtracted from deflections to yield an equivalent model). Thus, $\beta^{[i]}_{TL=[FIN]} + \beta^{[i]}_{TL[JPN]} = 0$, hence $\beta^{[i]}_{TL[JPN]} = -\beta^{[i]}_{TL[FIN]}$. Analogously, $\beta^{[i]}_{WP[mata-]} + \beta^{[i]}_{WP[manta-]} + \beta^{[i]}_{WP[manata-]} = 0$, hence $\beta^{[i]}_{WP[manata-]} = -\beta^{[i]}_{WP[mata-]} + -\beta^{[i]}_{WP[manta-]}$. The interaction deflections were constrained to sum to zero within factors. For example, $\theta^{[i]}_{B[p],TL[FIN]} + \theta^{[i]}_{B[pp],TL[FIN]} = 0$ and $\theta^{[i]}_{B[p],TL[FIN]} + \theta^{[i]}_{B[p],TL[JPN]} = 0$. Altogether, each individual's data were described by 8 independent threshold parameters plus 8 independent gain parameters.

The individual parameters were modeled as coming from higher-level distributions for each listener language (*LL*). For example, the baseline thresholds for all individuals within $LL=1$ (Japanese) or $LL=2$ (Finnish) were modeled as being normally distributed around a central tendency for that group:

$$\theta^{[i]}_{0} \sim \mathrm{normal}(\theta^{LL(i)}_{0}, \sigma^{LL(i)}_{0}) \tag{5}$$

where the tilde symbol "~" means "is distributed as." The parameter $\theta^{LL(i)}_{0}$ is the mean of the group-level normal distribution for the relevant *LL*, and $\sigma^{LL(i)}_{0}$ is the standard deviation of the group-level normal distribution. The standard deviation, $\sigma^{LL(i)}_{0}$, describes the spread of individual-level baselines within the listener language. Analogous group-level distributions were used on all the other parameters. For example,

$$\theta^{[i]}_{B=1} \sim \mathrm{normal}(\theta^{LL(i)}_{B=1}, \sigma^{LL(i)}_{B=1}) \tag{6}$$

Altogether, for each *LL*, there were 8 group-level threshold-mean parameters plus 8 group-level gain-mean parameters, plus the corresponding 16 standard deviations.

We are primarily interested in the group-level parameters. A key benefit of the hierarchical approach is that individual differences are explicitly accommodated, while simultaneously mutually informing each other and the group-level description. We do not collapse the data across individuals and ignore individual differences, nor do we model each individual independently and ignore the informativeness of other individuals' data. The hierarchical model simultaneously describes individuals and the relation between them.

There was one additional parameter included in the model, to accommodate "oops" errors for trials when participants may have been mentally distracted or otherwise off task. The model assumes that the final response probability is a mixture of the logistic probability given by Eq. (2) and a 50-50 random guessing response, with the parameter α being the mixture weight on the random response. This aspect of the model accommodates the rare occurrence of responses that are contrary to the asymptotic predictions of the logistic function at very short or long closure durations, which could otherwise only be fit by spuriously reducing the gain parameter.

The 16 parameters per individual plus 32 parameters per listener language plus 1 guessing parameter were estimated using Bayesian methods, which yield a complete joint distribution of credible parameter–value combinations, given the data. From this distribution we glean not only the most credible parameter values, but also any trade-offs among parameters, and details of the range of credible parameter values. Among various qualities of Bayesian methods (e.g., Kruschke, 2010a, 2010b, 2013, 2015), a major feature is their flexibility for specifying complex hierarchical models and seamlessly providing parameter estimates. The posterior distributions provide complete information for making decisions about parameter values, without need for sampling distributions from null hypotheses and assumptions about sampling intentions needed to compute *p* values (e.g., Kruschke, 2011, 2015).

Bayesian inference reallocates credibility across parameter values, starting from a prior distribution. In order to present here as neutral an analysis as possible, we did not use previous results (Aoyama, 2001) to inform the prior. Instead we chose vague priors that allowed a very broad range of values relative to the scale of the data, whereby the prior has minimal influence on the posterior distribution. Top-level constants in the prior were set to make the prior properly scaled to the data and simultaneously broad. For example, the individual baseline thresholds within a listener language come from a group-level normal distribution for which the group-level mean, $\theta^{LL(i)}_{0}$, was given a normal prior with mean of 106 and standard deviation of 20, which is wide compared to the variation in the data, and the group-level standard deviation of the individual thresholds, $\sigma^{LL(i)}_{0}$, was given a gamma prior with mean 10 and standard deviation 10, which again is wide compared to the variation in the data. Some top-level parameters were set to reasonable constants to facilitate MCMC convergence after extensive runs that freely estimated the parameters revealed credible modal values. Results did not qualitatively differ for different nearby values of the top-level constants. For example, the guessing parameter, α, was set to 0.01. The program that specifies the model details is available upon request to the first author.

Bayesian inference yields a unique posterior distribution, given the data and the prior, but its exact mathematical form is analytically intractable in this complex of a model. The posterior distribution is accurately approximated by a large representative sample of parameter values randomly drawn from the posterior distribution. The random representative parameter values are generated by a class of algorithms known as Markov chain Monte Carlo (MCMC) methods, as implemented in the software JAGS (Plummer, 2003) accessed from the computing language R (R Development Core Team, 2009). The MCMC methods take a random walk through high-dimensional parameter space, at each step landing on a combination of parameter values that are jointly credible. The probability of sampling any particular combination of parameter values is exactly proportional to its posterior probability. The trajectory of a random walk through parameter space is called a "chain." For our analysis, we ran 3 independent chains, with a 10,000 step burn-in period. The total number of recorded steps was 12,000, thinned every 50th step to reduce the saved file size. The chains mixed well with effective sample sizes (ESS: compensating for autocorrelation) of the chains for most parameters being at least several thousand (see Kruschke, 2015 for further details concerning these methods).[4]

2.5. Predictions concerning the directions of the effects

Since the predicted patterns of results are so involved, this section presents the direction of the effects we might expect to find. In general, we expect a reduction in the slope of the logistic function (β) to indicate that some manipulation has either reduced gain to the durational manipulation, or has generally introduced confusion into the identification process. By contrast, shifts in threshold (θ) would indicate that the listeners are taking the contextual information into account, and integrating them into their evaluation of the closure duration.

First, concerning TL (talker language) and LL (listener language), we expect overall, based on previous work, to find a similar threshold for the two languages, and for this to generalize to stimuli from the other language. Expressed in terms of the model parameters, this means we predict that $\theta_{LL[FIN]}{\approx}\theta_{LL[JPN]}$. The threshold, though, may be influenced by interaction effects with B (Base) and WP (Word Prosody), which will appear as $\beta_{B\times TL}$ and $\beta_{WP\times TL}$ that are non-zero. Depending on the degree to which the overall sound of productions differ across the languages, one might expect to find an interaction effect of TL and LL on slope, if such differences have the effect of generally sowing confusion in the listeners who are unused to the novel production effects. In terms of the model parameters, this means we predict that $\beta_{LL[FIN],\ TL[FIN]}>\beta_{LL[FIN],TL[JPN]}$ and $\beta_{LL[JPN],\ TL[JPN]}>\beta_{LL[JPN],TL[FIN]}$.

If the two languages differ specifically in how the quantity contrast is articulated, we expect this effect to be localized in the TL × B interaction. The Base effect indicates in general the effect of production differences between geminate and singleton consonants. If consistent across the languages, one would expect B to shift thresholds for both language groups in the direction of the produced base form consonant for each of the listener groups, i.e., $\theta_{B[p]}-\theta_{B[pp]}$ would be positive (earlier threshold when Base form is geminate) for all the four combinations of LL and TL. However, if such production effects are language specific, such as the anti-compensation effect in Japanese, one would expect to find such shifts only when LL and TL match in language. So, for example, Japanese listeners are expected to be influenced by their knowledge of the anti-compensating pattern in their native language (Section 2.5). Previous results in this regard are mixed. Amano and Hirata (2010: 2053) found no significant effect of the original materials for Japanese listeners. By contrast, O'Dell (2003) found a moderate effect of the original word (*katoa* vs. *kattoa*) for Finnish listeners, even when the temporal organization of the entire test word was modified to be neutral.

Finally, we expect prosodic modulation effects to create effects in the TL × WP interaction for each of the listener groups. Specifically, Finnish listeners, because they have experience with the prosodic modulation of the "half-long vowel," are not expected to be affected by this variation in the Finnish stimuli. By contrast, when Japanese listen to Finnish stimuli, a WP effect is expected if they interpret these effects as anti-compensating for the geminate-singleton contrast. Specifically, the boundaries should be earlier for the *mata*- and *manata*-conditions as compared with the *manta*-condition, because of the longer preceding vowel for the former two conditions. It is also possible that a mismatch in talker and listener language will cause confusion, and hence a WP-specific reduction in slope values when TL and LL do not match.

3. Results

3.1. Overall summary

Among the 31,752 trials in total (756 responses × 42 participants), there were 121 timeout errors from 21 participants (Finnish = 13, Japanese = 8), which account for approximately 0.4% of the total trials. All of the remaining responses (31,631 in total) were submitted to the statistical analysis. Bayesian analysis behaves seamlessly when there are missing data or unbalanced designs.

Figs. 1 and 2 show the mean proportion of /pp/ identification for each of the LL (Listener Language) × TL (Talker Language) × B (Base) and LL × TL × WP (Word Prosody) combinations. The figures also display logistic curves that are fit to the means. The curves are merely illustrative; the Bayesian analysis produced an entire distribution of credible logistic curves from fitting all the data. The data points are not aligned at the same time points on the *x*-axis because the range of closure duration and time steps are slightly

[4] O'Dell (2003: Section 4.1.2) also applied Bayesian statistical analysis to Finnish quantity perception data, with MCMC sampling using WinBUGS to evaluate a model employing a normal cumulative distribution function.

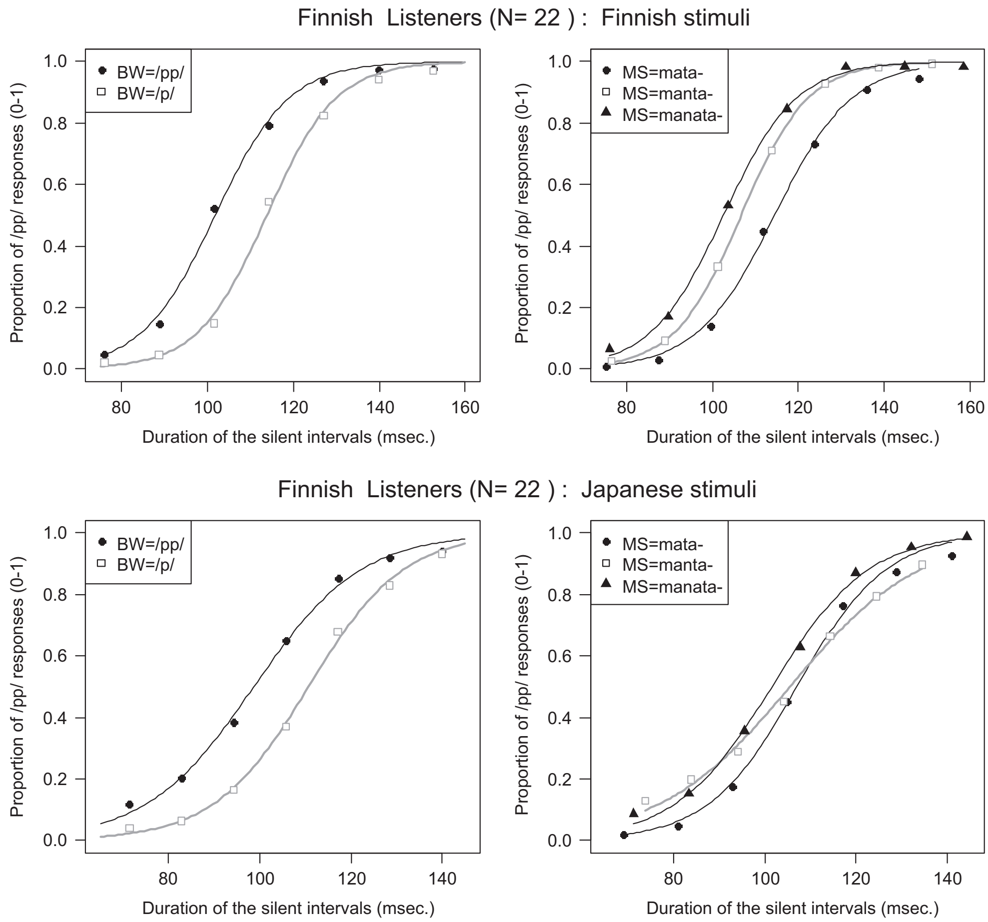

Fig. 1. Proportion of geminate identification by Finnish listeners (means for all trials from all listeners). Top panels: Finnish talker; Bottom: Japanese talker. Left panels: effects of original production having a geminate vs. a singleton consonant. Right panels: effects of the three different word-prosody conditions. Coding symbols indicate observed means, and curves are for the sigmoid function fitted to the means.

different between Finnish and Japanese talkers and across MS conditions (see Table 2). For the statistical analysis, the raw acoustic durations (in milliseconds) of the stop closure were used so that comparisons across TL, B and WP conditions are possible.

In general, we find good logistic curves as a function of closure duration for all of the listener groups and conditions, as is evident in all of the figure panels. We also find a persistent Base production (B) effect for both listener groups listening to both languages, so it appears that there are robust acoustic effects besides consonant closure duration in both languages, and these differences are readily interpretable by listeners from the other language. This Base effect is particularly obvious in the Japanese listeners hearing the Japanese productions, suggesting that some Japanese-particular covariant, such as anti-compensation, is particularly salient within this language. Word Prosody effects (WP) are less obvious across the board, but are particularly in evidence in the Finnish listeners, hearing Finnish productions. Here, there is an apparent threshold shift in the half-long *mata*-condition toward more singleton responses. This is not expected, since such production variations are not due to the quantity distinction in the first place, and also since vowel durations were very similar for the *mata*- and *manata*-conditions. Japanese listeners exhibited smaller threshold shifts across WP conditions, and were affected in the opposite direction from the Finnish listeners by the Finnish stimuli. Here, this half-long *mata*-condition tended to elicit more geminate responses, as would be expected from their interpretation of these effects as indicating anti-compensation for the following plosive.

To explore the results more rigorously, the following sections report the Bayesian estimates for the slope β and threshold θ for each of the LL × TL × B and LL × TL × WP combinations and examine whether there are differences as a function of the three factors and their interactions. We summarize the marginal posterior distribution on each parameter by reporting its mean value, the lower and upper bounds of its 95% HDI (Highest Density Interval), and the proportion of its posterior that is greater or smaller than 0. The 95% HDI indicates an interval on the parameter being estimated that spans 95% of the distribution, such that every point inside the interval has higher probability than any point outside the interval. We examine the 95% HDI to see whether it excludes zero. (See Kruschke, 2011, 2015, for discussion of these methods, including the issue of using a region of practical equivalence around zero.) For example, for the overall slope β_0, the mean of the posterior for Finnish listeners was 0.167, with the 95% HDI spanning the interval [0.146, 0.188], and all (100%) 12,000 representative parameter values (posterior) were greater than 0. Therefore, it is highly credible that the overall slope is positive, i.e., overall, the listeners do have positive gain to closure duration and the probability of a geminate response does increase as closure duration gets longer. Complete numerical details of all parameters are summarized in Appendices A and B.

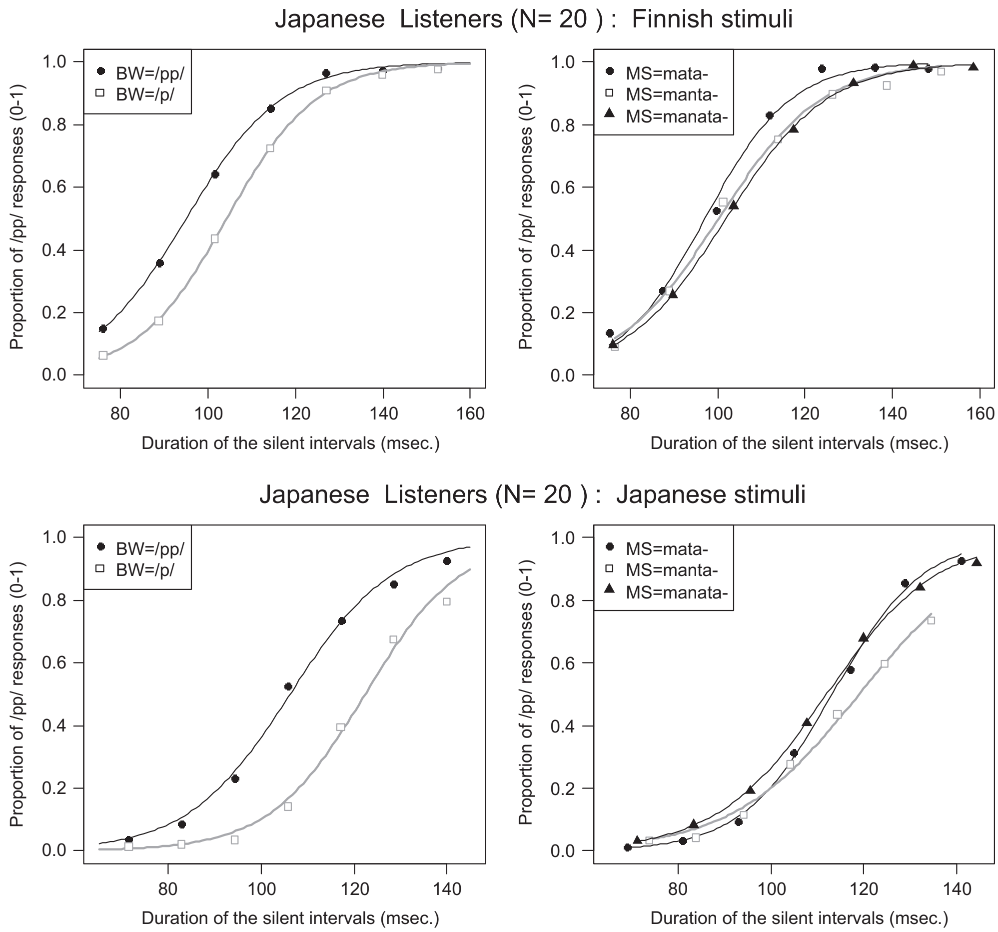

Fig. 2. Proportion of geminate identification by Japanese listeners (means for all trials from all listeners). Top panels: Finnish talker; Bottom: Japanese talker. Left panels: effects of original production having a geminate vs. a singleton consonant. Right panels: effects of the three different word-prosody conditions. Coding symbols indicate observed means, and curves are fitted values as indicated in the text.

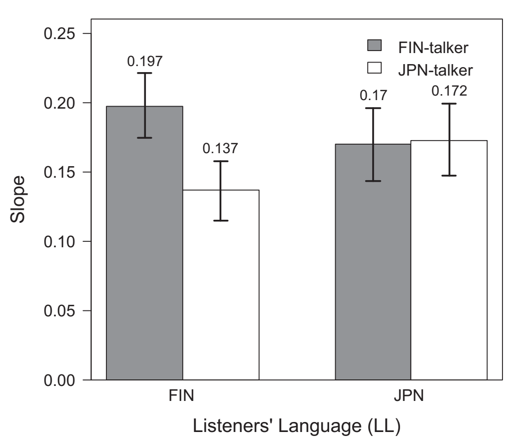

Fig. 3. Gain parameters β (slopes on the logistic function) as a function of Listener (Finnish/Japanese) and Talker (Finnish/Japanese) estimated with Bayesian logistic regression. The thick bars indicate the means, and the thin vertical bars indicate the 95% HDI of the posterior distribution. The leftmost bar is $\beta_{LL[FIN],TL[FIN]}$ as defined in Eq. (5).

3.2. *Listener Language×Talker Language*

3.2.1. *Slope*

Fig. 3 summarizes the estimated slopes (gain parameters) of singleton/geminate categorization (β in Eqs. (1)–(3)) as a function of listeners' language (LL) and talker's language (TL). These are obtained as the linear combinations of the posterior samples following Eq. (3). For example, the estimated slope for Finnish listeners hearing Finnish talkers is computed as the baseline slope due to being a Finnish listener, plus the increment due to listening to a Finnish talker (or minus the increment for listening to a Japanese talker), which is formally denoted as

$$\beta_{LL[FIN],TL[FIN]} = \beta_{0LL[FIN]} + \beta_{TL[FIN]} \tag{7}$$

As can be seen in Fig. 3, for Finnish, the slopes are steeper when participants listen to stimuli created from a talker of their own native language. This is suggested by generally steeper slopes of the logistic curves in the top two panels, as compared with the bottom two panels of Fig. 1. On the other hand, for Japanese listeners (Fig. 2), no clear difference can be found in steepness of the slopes depending on the talker's language.

The Bayesian posterior distribution provides details regarding the magnitude of this disadvantage. For the Finnish listeners, we computed the differences in slope between TL conditions, i.e., the slope for Finnish talkers minus the slope for Japanese talkers, $\beta_{LL[FIN],TL[FIN]}-\beta_{LL[FIN],TL[JPN]}$, at every one of the 12,000 jointly credible parameter combinations in the posterior. The distribution of the differences computed in this way has a mean of 0.061, with 95% HDI spanning [0.045, 0.075], and 100% of the differences greater than 0. On the other hand for the Japanese listeners, the differences in slope between TL conditions had a mean of −0.002, with 95% HDI spanning [−0.187, 0.128], and 36.3% of the differences greater than 0. In this way, we can confirm that there is disruption in geminate perception caused by Japanese productions for Finnish listeners that is credibly different (larger) than zero, while the same disruption by the Japanese productions is not observed (not credibly different from zero) for Japanese listeners.

We can also examine whether our results in Fig. 3 replicate those of Aoyama (2001, Section 4.7), where it is concluded that "the bandwidth of the categorical boundary" is narrower for Finnish listeners than Japanese, suggesting sharper quantity categorization of the Finnish group. In Aoyama (2001) Finnish and Japanese listeners were presented only with the stimuli created from their own languages and then the results were compared. In the present study, this corresponds to comparing Finnish listeners hearing a Finnish talker: $\beta_{LL[FIN],TL[FIN]}$ (upper left in Fig. 3), with Japanese listeners hearing a Japanese talker: $\beta_{LL[JPN],TL[JPN]}$ (upper right in Fig. 3). The Bayesian posterior distribution indicates a larger slope for Finnish but the uncertainty of the estimated difference is larger relative to its magnitude: The difference, $\beta_{LL[FIN],TL[FIN]}-\beta_{LL[JPN],TL[JPN]}$, has a mean of 0.025, with 95% HDI spanning [−0.012, 0.058] and 92.3% greater than 0. This trend agrees with the results in Aoyama (2001).

Next we examine whether one of the two listener groups is more finely attuned to the production of their own native language. To examine this, we have to compare the "effect sizes" of the advantage of being a native speaker in quantity categorization, i.e., the differences in sharpness of categorization function when two listener groups listen to the same materials. This amounts to comparing the two gray bars, and also two white bars in Fig. 3, and then comparing the two differences to see whether the difference is larger for one of the listeners groups.[5] For Finnish stimuli, while the mean slope is steeper for Finnish listeners than Japanese listeners, the uncertainty of the estimated difference is large relative its magnitude and the 95% HDI includes zero: The difference, $\beta_{LL[FIN],TL[FIN]}-\beta_{LL[JPN],TL[FIN]}$, has a mean of 0.028, with 95% HDI=[−0.008, 0.062], with 94% greater than 0. For Japanese stimuli, there is small but credibly non-zero difference, suggesting an advantage for Japanese listeners perceiving Japanese stimuli: The difference, $\beta_{LL[JPN],TL[JPN]}-\beta_{LL[FIN],TL[JPN]}$, has a mean of 0.036, with 95% HDI spanning [0.003, 0.070], and 97.9% greater than 0. Finally, the difference of the two differences just computed above is approximately zero. The difference of the differences just computed above, $(\beta_{LL[JPN],TL[JPN]}-\beta_{LL[FIN],TL[JPN]})-(\beta_{LL[FIN],TL[FIN]}-\beta_{LL[JPN],TL[FIN]})$, has a mean of 0.008, with 95% HDI spanning [−0.060, 0.072], suggesting there is no notable difference in the extent of L1 advantage between the two listener groups.

Note that we do not make any correction for conducting such comparisons more than once, such as the various methods of *p*-value adjustment that are usually deemed mandatory for multiple comparisons in ordinary (null-hypothesis significant testing) statistical tests. This is because the Bayesian posterior distribution is unaffected by the number of comparisons the researcher wants to make (see Kruschke, 2011, Section 11.4, and Kruschke, 2015).

3.2.2. Thresholds

Turning to the location of the category boundary, Fig. 4 summarizes the 50% thresholds estimated with Bayesian logistic regression (θ in Eqs. (1), (2) and (4)) as a function of LL and TL. For Finnish listeners, we find a clear effect of TL. The threshold duration is lower for Japanese stimuli: the difference, $\theta_{LL[FIN],TL[FIN]}-\theta_{LL[FIN],TL[JPN]}$, has a mean of 3.39 ms, with 95% HDI spanning [1.57, 5.27], and 100% larger than 0, suggesting that Japanese stimuli sound more like geminates to Finnish listeners as compared with Finnish stimuli. By contrast, we find a TL effect in the opposite direction for Japanese listeners. The threshold is lower for Finnish stimuli: $\theta_{LL[JPN],TL[FIN]}-\theta_{LL[JPN],TL[JPN]}$ has a mean of −14.8 ms, with 95% HDI spanning [−16.8, −12.8], and 100% smaller than 0, suggesting that Finnish stimuli sound more like geminates to Japanese listeners. The TL effect is much larger for Japanese listeners. Furthermore, the threshold is lower for Finnish listeners than Japanese listeners when hearing the stimuli of their own language: $\theta_{LL[FIN],TL[FIN]}-\theta_{LL[JPN],TL[JPN]}$ has a mean of −6.28 ms., with 95% HDI spanning [−11.5, −0.50], and 98.7% smaller than 0.

3.3. Listener Language×Talker Language×Base

3.3.1. Slopes

For both listener groups, no credible changes are found in slopes as a function of TL, Base Word (B), nor their interactions, suggesting quantity categorization is largely unaffected by these factors. See Appendix A for details.

[5] We thank an anonymous reviewer for pointing this out.

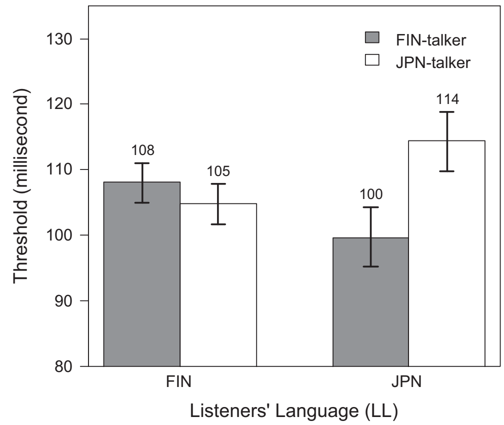

Fig. 4. 50% threshold parameters (singleton/geminate boundary) for singleton/geminate categorization as a function of Listener (Finnish/Japanese) and Talker (Finnish/Japanese) estimated with Bayesian logistic regression. The thick bars indicate the means, and the thin vertical bars indicate the 95% HDI of the posterior distribution.

3.3.2. Thresholds

Turning to whether language specific production differences are evident in the categorization boundary locations, we now turn to the thresholds. Fig. 5 summarizes the estimated threshold as a function of LL, TL and B. For both listener groups, and for both Finnish and Japanese stimuli, the singleton/geminate threshold is always lower when the base production is /pp/. That is, listeners tend to categorize the target consonant as a geminate with shorter closure duration when the stimuli were created from geminate originals: For Finnish listeners, the difference between B-conditions, $\theta_{LL[FIN],B[p]}-\theta_{LL[FIN],B[pp]}$, has a mean of 11.8, with 95% HDI spanning [9.9, 13.6], and 100% greater than 0. For Japanese listeners, the difference between B-conditions, $\theta_{LL[JPN],B[p]}-\theta_{LL[JPN],B[pp]}$, has a mean of 12.5, with 95% HDI spanning [10.6, 14.5], and 100% greater than 0. However, the sizes of these B effects are not very different: The difference of the differences just computed above, $(\theta_{LL[FIN],B[p]}-\theta_{LL[FIN],B[pp]})-(\theta_{LL[JPN],B[p]}-\theta_{LL[JPN],B[pp]})$, has a mean of −0.8, with 95% HDI spanning [−3.4,2.0]. Furthermore, we can compare the sizes of the Base effects of the four combinations of LL and TL. The B effect is larger when Japanese listeners listen to the Japanese talker. The B effect is larger when both LL and TL are Japanese than for all the other three cases. For example, the difference of B effects, $(\theta_{LL[JPN],TL[JPN],B[p]}-\theta_{LL[JPN],TL[JPN],B[pp]})-(\theta_{LL[FIN],TL[FIN],B[p]}-\theta_{LL[FIN],TL[JPN],B[pp]})$, has a mean of 4.0, with 95% HDI spanning [0.4, 7.8]. All the other pairwise comparisons of the B effects among the other three LL × TL combinations are not credibly non-zero. By contrast, the B effect is smaller when LL is Japanese and TL is Finnish than the mean of the B effects of the other three LL × TL combinations. The difference of the B effect for LL[JPN],TL[FIN] and the mean of the other B effects, (B effect $\theta_{LL[FIN],TL[FIN]}$ + B effect $\theta_{LL[FIN],TL[JPN]}$ + B effect $\theta_{LL[JPN],TL[JPN]}$)/3−(B effect $\theta_{LL[JPN],TL[FIN]}$), has a mean of 3.9, with 95% HDI spanning [0.7, 7.0].

For Japanese listeners hearing the Japanese talker's stimuli, the B effect could partly be attributed to the anti-compensation pattern in the duration of preceding vowel. However, the across-the-board B effect cannot be accounted for this way, since the anti-compensation pattern is only in the Japanese talker's stimuli. Also, even though anti-compensation patterns are not obvious in Finnish productions, the Japanese geminate productions are categorized more as geminates by the Finnish listeners. The B effect is in the same direction across all combinations of LL and TL, suggesting that both Finnish and Japanese stimuli pervasively contain properties besides the closure duration which indicate the identity of the base word, and this information is accessible to both listener groups.

3.4. Listener Language×Talker Language×Word Prosody

3.4.1. Slopes

Fig. 6 summarizes the estimated slopes as a function of LL, TL, and WP. Out of the 12 pairwise contrasts between the WP conditions within each combination of LL and TL, only 2 were credibly different from zero. The slope is shallower when Finnish listeners listened to a Japanese talker and WP is *manta*- rather than the other two WP conditions. The difference, $\beta_{LL[FIN],TL[JPN],WP[mata\text{-}]}-\beta_{LL[FIN],TL[JPN],WP[manta\text{-}]}$, has a mean of 0.05, with 95% HDI spanning [0.03, 0.07], and 100% greater than 0 and the difference, $\beta_{LL[FIN],TL[JPN],WP[manta\text{-}]}-\beta_{LL[FIN],TL[JPN],WP[manata\text{-}]}$, has a mean of −0.05, with 95% HDI spanning [−0.07, −0.03], and 100% smaller than 0. As we saw in Section 3.2.1, Finnish listeners' sensitivity to change in closure duration is significantly lower for Japanese stimuli than for Finnish stimuli, suggesting that, overall, the quantity distinction in the Japanese stimuli is more difficult to perceive than that of their L1. The current result further reveals that the Japanese productions are particularly disruptive when WP is *manta*-. One of the possible explanations is that language-specific word-prosody production differences between the two languages especially created problems for Finnish listeners (Table 2). The *manta*-condition is the “non-half-long” condition in the Finnish stimuli, and was characterized by a vowel preceding the target consonant being shorter than those found in the Japanese stimuli, and those found in the Finnish words from the other Word Prosody conditions. The lack of such a short vowel in the Japanese stimuli appears to be disruptive for the Finnish listeners.

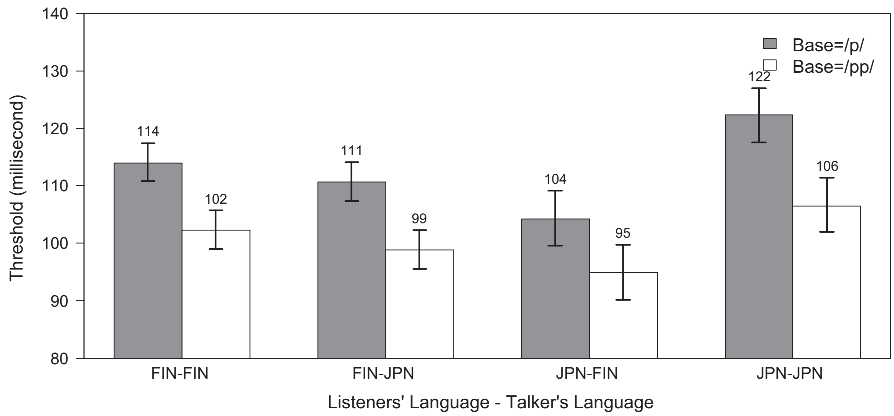

Fig. 5. 50% threshold parameters (singleton/geminate boundary) for singleton/geminate categorization as a function of Listeners, Talkers and Base (whether the stimuli were created with singleton or geminate base words). The thick bars indicate the means, and the thin vertical bars indicate the 95% HDI of the posterior distribution.

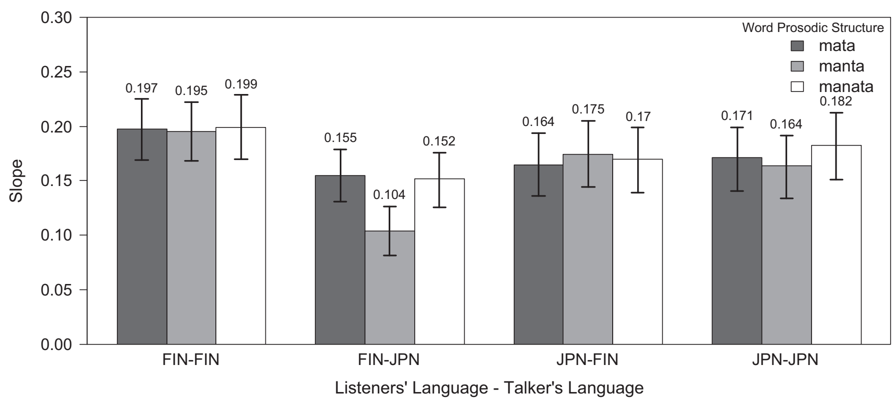

Fig. 6. Gain parameters (steepness of singleton/geminate categorization function) as a function of Listener Language, Talker Language and Word Prosody. The thick bars indicate the means, and the thin vertical bars indicate the 95% HDI of the posterior distribution.

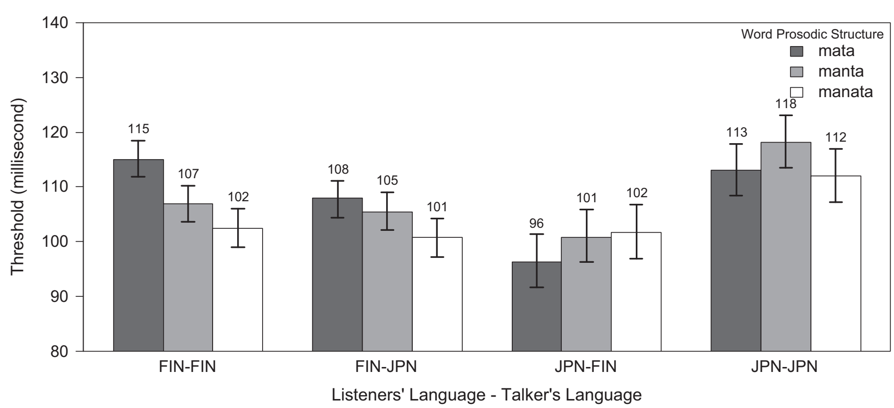

Fig. 7. 50% threshold parameters (singleton/geminate boundary) for singleton/geminate categorization as a function of Listener (Finnish/Japanese) and Talker (Finnish/Japanese) and Word Prosody. The thick bars indicate the means, and the thin vertical bars indicate the 95% HDI of the posterior distribution.

3.4.2. Thresholds

Fig. 7 summarizes the estimated threshold as a function of LL, TL and WP. Out of all 12 pairwise contrasts between WP conditions within each combination of LL and WP, there are 8 differences that are credibly non-zero as summarized in Table 3. The results reveal that the contextual effects created by Word Prosody are not totally canceled out, yielding various shifts in threshold, even with listeners listening to stimuli from their own language.

For Finnish listeners hearing the Finnish talker, the threshold is the earliest for *manata*- and the latest for *mata*- and intermediate for the *manta*-initial condition. The difference between the *mata*- and *manta*-conditions might be explained as a local contextual effect of the preceding vowel: because of a longer preceding vowel for *mata*-, the consonant needs to be longer to be perceived as geminate, suggesting that Finnish listeners expect a compensating, rather than an anti-compensating pattern of the preceding vowel. However, the higher threshold for *manata*- cannot be explained this way, since the preceding vowel is as long as that of *mata*- in this

Table 3
Pairwise contrasts of 50% threshold between the WP conditions within each of the four combinations of Listeners and Talker that are credibly non-zero.

LL	TL	Comparisons	Mean diff.	95% HDI		Posterior >0	Posterior <0
				Lower	Upper		
Finnish	Finnish	*mata-* vs. *manta-*	8.2	6.1	10.1	100.0	0.0
		mata- vs. *manata-*	12.6	9.8	15.5	100.0	0.0
		manta- vs. *manata-*	4.5	1.5	7.3	99.9	0.1
Finnish	Japanese	*mata-* vs. *manta-*	2.5	0.2	4.8	98.4	1.6
		mata- vs. *manata-*	7.2	4.1	10.1	100.0	0.0
		manta- vs. *manata-*	4.7	1.5	7.7	99.8	0.2
Japanese	Finnish	*mata-* vs. *manta-*	−4.5	−6.9	−2.3	0.0	100.0
		manta- vs. *manata-*	−5.4	−8.5	−2.2	0.0	100.0
Japanese	Japanese	*mata-* vs. *manta-*	−5.1	−7.4	−2.9	0.0	100.0
		mata- vs. *manata-*	6.2	3	9.3	100.0	0.0

condition. For Finnish listeners hearing the Japanese talker, the same pattern of boundary shift between the three WP conditions is observed, with smaller overall size of the shift (12.6 ms vs. 7.2 ms).

For Japanese listeners hearing the Japanese talker, credible differences are between the *manta-* and the other two conditions, the former being later in boundary than the others, whereas no credible difference is found between the *mata-* and *manata*-conditions. This pattern is unexpected since there is no significant variation in duration of the preceding vowel as a function of the WP condition for the Japanese stimuli. Unlike Finnish listeners, a different pattern of boundary shift is found when Japanese listeners hearing the Finnish talker. The threshold is earlier for *mata-* than the other two conditions. This is as expected if the longer preceding vowel in *mata*-condition enhances geminate perception. However, the *manata*-condition does not pattern together with the *mata-* condition, even though the preceding vowel is longer for this condition, too.

4. Discussion

4.1. General aspects in quantity categorization

The current study investigated the role of different sorts of covariate effects on quantity categorization of a continuum of plosive durations. The main point of this study was to examine the degree to which such covariates affect this categorization in a language-specific manner. Of interest is the degree to which they are available to listeners of another language, and hence induce shifts in their categorization functions. Alternatively, since such covariates are also affected by language-specific effects which are unrelated to the quantity contrast, such language-specific effects might also cause confusion for native listeners of a different language.

Japanese and Finnish were chosen based on previous observations of a great similarity in the quantity contrast in the two languages. Overall, the current results reinforce this impression. Categorization functions for listeners of both languages listening to stimuli made from both languages all exhibit very clear and sharp categorization functions. Further, though the locations of the category boundaries (in terms of milliseconds of closure duration) are not exactly identical, they are, in general, fairly closely matched in most conditions. Mean estimated thresholds for the various conditions tend to be within a 15 ms difference, and most striking of all, the deflection of the estimated listener group parameter is 0.4 ms (compare the mean θ_0 values for LL1 and LL2 in Appendix B), indicating that, on the whole, the boundaries between short and long consonants in the two languages are very close. Although we cannot directly compare our results with those of Aoyama (2001) due to many differences in the experimental procedure, the mean $\theta_{LL[FIN],TL[FIN]}$ (108 ms, Section 3.2.2) is also very close to the one we estimated from Aoyama's data (106 ms, in Section 1.1).[6] The mean $\theta_{LL[JPN],TL[JPN]}$ observed in the current data (114 ms) is also quite close (107 ms).

Even more striking across the languages is the degree to which the details of the results coincide across the two groups of listeners. The most compelling and consistent aspect of these results is a persistent shift in categorization boundary between stimuli made from original geminate productions and ones made from original singleton productions (Section 3.3.2). These differences indicate quantity information in the signal besides the closure duration, which the listeners integrate into their categorization behavior. Most importantly, these covariates are apparently produced by both language groups in a way that is available for both listener groups. That is, Finnish productions contain covariate information that is used by Japanese listeners, and Japanese productions contain information that is used by Finnish listeners. This effect of Base is the most consistent effect on threshold across the board. What this highlights is that pervasive and complex patterns in production need not be entirely idiosyncratic to particular language groups, but can be profitably transferred from language to language.

[6] In Aoyama (2001), the target consonants are nasals (/n/ and /nn/) and the range of the closure duration is larger in the top ends (69–177 ms for Finnish and 59–167 ms for Japanese), with a geminate (/nn/) token used as the base. Also, the participants only listened to the stimuli from their own language.

4.2. Language-specific aspects in quantity categorization

This is not to say, however, that all covariate patterns are simply low-level and universally available. One language-specific pattern in the current experiment is the effect of anti-compensation in Japanese. The current stimuli reflected the previously often-noted and obvious effect of vowel lengthening preceding a target geminate in Japanese, and this pattern was not obvious in the Finnish productions (Section 1.1). Perceptually, this difference appears to have given rise to an interaction effect with the Japanese listeners where the Base effect is larger for the Japanese listeners hearing Japanese productions than all the other three cases (Section 3.3.2). Here, where anti-compensation is in the signal, boundaries are shifted to a larger degree than for the Finnish listeners. On the converse side, Base-induced shifting is also smaller in degree for the Japanese listeners hearing Finnish speech, which does not have the anti-compensation pattern (Table 2).

The perception experiment using non-speech stimuli in Kingston et al. (2009) showed that a longer preceding vowel analog (i.e., an anti-compensating pattern) enhances "long" responses for all the 4 language groups studied (Japanese, Norwegian, Italian and English). According to the authors, this could arise from general auditory character of listeners adding the durations of the consonant and vowel intervals. Their experiment using speech stimuli, on the other hand, found that only Japanese and English listeners exhibit enhancement of "long" perception in the same direction as for the non-speech stimuli. In the present case, Japanese stimuli contain an anti-compensating pattern in duration of the preceding vowel, which should integrate with closure duration into a contrastive perceptual property for quantity perception. The results suggest, however, only Japanese listeners seem to exploit this anti-compensating pattern for quantity categorization. This is presumably because the pattern of anti-compensation in segmental duration is robust and consistent only in Japanese productions; thus it constitutes part of the phonetic knowledge for only Japanese listeners.

Furthermore, the current experiment also shows that for Japanese listeners, Finnish stimuli generally sound more like geminates than Japanese stimuli do, whereas for Finnish listeners, smaller but significant Talker Language effects in the opposite direction are found (Section 3.2.2).

Also another result pointing to the language-specific nature of quantity categorization is the fact that the slopes of the categorization functions are steeper when Finnish listeners hear stimuli from their native language, showing perceptual difficulty for the stimuli from the Japanese speaker, while Japanese listeners do not exhibit such difficulty to Finnish stimuli (Sections 3.2.1 and 4.3). These patterns suggest that the Japanese listeners were able to integrate aspects of the Finnish stimuli into their quantity categorization functions better than were the Finnish listeners for the Japanese stimuli.

4.3. The effect of word prosody on quantity categorization

The current study also examined the role of language-specific effects of word prosody – word-initial syllable structure – on the covariates, with a view toward determining whether such prosodic effects, by affecting these covariates, might create patterns which get inappropriately appropriated by listeners from the other language for the quantity categorization (Section 3.4.2). The overall effect of such prosodic differences was to create a range of small boundary shifts in various directions. However, unexpectedly, these shifts occurred with listeners listening to stimuli made from their *own* language. It seems likely that such shifts indicate, in addition to the prosodic patterns, the existence of covariates in the stimuli which encourage singleton or geminate categorization. Troublesome for this conclusion, however, are two facts. First, the threshold shifts are not in the same direction for the Japanese and Finnish listeners. That is to say, if these are the type of covariates which give rise to the large Base effects, they are not ones which are expected by both listener groups and must be language-specific.

Second, especially the Finnish listeners exhibit shifts in the same direction for both Finnish and Japanese stimuli, suggesting that some of the shifts are partially due to the listeners. During previous presentations of the current results and during the process of review a number of suggestions for various of these effects have been discussed, including possible effects of footing in Finnish placing a foot-boundary before the target in the *manta*- and *mata*-condition (c.f., production work on Finnish in Suomi, 2005, 2007; Suomi et al., 2003, and parallel work on other languages in Cho & Keating, 2001, 2009), the possibility that Finnish listeners might be more likely to parse the five-syllable stimuli as comprising a compound with a morphological break before the target, language specific articulatory declination in longer words (c.f production work on other languages by Krakow, 1999; Krakow, Bell-Berti, & Wang, 1995; Vayra & Fowler, 1987) and the role of word initial accents in the Finnish productions (c.f. production and perception work on Japanese by Nagano-Madsen, 2007).

Regardless of what accounts for each of the effects, the overall point is clear, namely that many properties of the stimuli can create subtle effects on categorization threshold, and these can do so in a language-specific manner. In the current data, these shifts are numerous, but are much smaller than the overall Base effect which is strongly evident across the languages.

There is, however, one readily interpretable effect of Word Prosody in the current experiment, and that concerns language-specific effects on categorization slopes. Here, the very clear effect of Word Prosody is not that of providing information which gets integrated into the quantity contrast, but is rather that of interfering with the categorization. Specifically, the condition which creates the largest differences in the vowel duration preceding the closure in Japanese creates problems for Finnish listeners (Section 3.4.1). While stimuli produced in other WP conditions had ('half-long') vowels of roughly the same duration as those in the Japanese stimuli, the longer preceding vowels in the *manta*-condition apparently reduced the steepness of the Finnish categorization function. That is, *not* having the shorter vowel in the *manta*-condition stimuli produced by the Japanese talker is particularly disruptive for Finnish listeners. An anonymous reviewer has even suggested that the vowel preceding the target consonant in Japanese stimuli may actually be long enough to be perceived as “long”.[7] This could add further difficulty in quantity categorization for Finnish listeners in this condition. These results suggest that prosodic variation can cause problems for listeners from other languages, though not necessarily by misleading the listeners into integrating it into the quantity contrast.

5. Concluding remarks

The current study examined quantity perception in Finnish and Japanese listeners in a fully-crossed perception study. The results highlight both the similarity and difference between the two languages. First, a pervasive across-the-board Base effect indicates that the two languages share substantial acoustic covariates to closure duration that are relevant to quantity categorization. This points out that, even though it may be useful to isolate a phonetic parameter such as closure duration as a primary indicator of a contrast, this parameter occurs in a plausible context with a larger pattern of lawfully-governed variation in the signal. Such patterns of co-variation are appreciable even for listeners confronted with it in the context of a novel language. Thus, while phonetic variation is rich and complex, this does not mean that it presents itself as a puzzle for listeners in a second language environment.

On the other hand, Finnish listeners show evidence that stimuli from Japanese are more difficult to categorize, while Japanese listeners do not exhibit such difficulty. This is remarkable, considering that the segments used for the stimuli are not ones which previous descriptions of Japanese and Finnish would generally describe as being different. So, despite the fact that languages can be typologically and phonologically very similar, and that phonetic variation arises from articulatory constraints and the auditory character of humans in general (Kingston & Diehl, 1994), there is yet a rich repository of phonetic variability which is language specific, and does present itself as a problem for the listeners confronted with a new language. Some of these attributes are obviously detectable in language descriptions, such as the variation in phonologically short vowels in Finnish (half-long vs. short); however, various small but consistent shifts in threshold due to Word Prosody suggest that there are many more that remain unaccounted for.

This, of course, suggests that more work on the specific nature of these covariates would be called for, perhaps through specific re-synthesis approaches. The current study used this approach, manipulating closure duration. However another point of the current study should be noted; ultimately, determining what phonetic variability is important for a phonological contrast needs to take place in the context of the natural variability in speech. The current study has suggested many possible further topics in natural variation to be studied, especially concerning the variety of threshold shifting effects found. While this may have produced too much of interest for our comfort, the use of natural speech is crucial for finding these phenomena.

Finally, concerning real speakers confronted with a new language, our results concerning the slopes of the categorization functions identify certain cross-language effects that may pose problems in second language learning, and these can easily go unnoticed in experiments using single language base words, or different language bases as in Aoyama (2001). Prosodic patterns which create local conditions which differ widely between the two languages (such as the *manta*-condition here) are liable to be a particular problem for learners. Also generally, Finnish and Japanese listeners may widely deviate in where to put singleton/geminate boundary for words with different prosodic contexts, which could give rise to perceptual errors in the L2. However, the current results are not all bad news for L2 learners. One major point not to be missed is that a language learner comes to the second language with excellent experience with the rich variation that surrounds the quantity contrast, and this variation is often informative and helpful even for a language different from one's native languages.

Acknowledgments

We thank Reijo Aulanko, Marjut Mäenpää and Seppo Kittilä at University of Helsinki for providing support for running subjects in Finland, Donna Erickson, Takuya Oomae and Yosuke Igarashi for providing support in Japan, and Jussi Ylikoski and Keiko Kuriyama for references. We also acknowledge the audience at 2010 Meeting of LSA, members at Riken Brain and Language Lab and three anonymous reviewers for helpful comments on this research, and Isabelle Darcy's help with using DMDX was invaluable. We also thank the Finlandia Foundation Fellowship Award 2009 for financial support for experiments in Finland. Finally, we thank three anonymous reviewers, Holger Mitterer, and Taehong Cho for many helpful comments on previous versions of this work.

Appendix A

See Table A1.

Appendix B

See Table B1.

[7] Lehtonen's (1970: 116–7) data show that consonants are shorter after long vowels than before short ones. Given this, it is possible that perception of a long vowel made Finnish listeners biased toward geminate perception in this condition.

Table A1
Summary of posterior distributions of the predictors involving β (Eq. (3) in Section 2.4), including the mean, lower and upper bounds of 95% HDI, and proportion of the posterior smaller or larger than 0.

	Mean	95% HDI		Posterior > 0	Posterior < 0
		Lower	Upper		
Japanese listeners (LL1)					
β_0	0.171	0.147	0.197		
β_{TL} (Japanese)	0.001	−0.006	0.009	63.7	36.3
β_B (/p/)	0.003	−0.004	0.011	79.7	20.3
β_{WP1} (mata)	−0.003	−0.013	0.006	24.3	75.7
β_{WP2} (manta)	−0.002	−0.012	0.009	36.5	63.5
$\beta_{TL\times B}$ (JPN × /p/)	0.005	−0.003	0.013	90.2	9.8
$\beta_{TL\times WP1}$ (JPN × mata)	0.002	−0.008	0.011	65.9	34.1
$\beta_{TL\times WP2}$ (JPN × manta)	−0.007	−0.017	0.003	91.0	9.0
Finnish listeners (LL2)					
β_0	0.167	0.146	0.188		
β_{TL} (Japanese)	−0.030	−0.038	−0.023	0.0	100.0
β_B (/p/)	−0.005	−0.002	0.013	92.3	7.7
β_{WP1} (mata)	0.009	−0.0004	0.018	96.8	3.2
β_{WP2} (manta)	−0.020	−1.200	0.800	100.0	0.0
$\beta_{TL\times B}$ (JPN × /p/)	0.001	−0.006	0.009	36.6	63.4
$\beta_{TL\times WP1}$ (JPN × mata)	0.009	−0.001	0.018	96.1	3.9
$\beta_{TL\times WP2}$ (JPN × manta)	−0.016	−0.024	−0.006	0.0	100.0

Table B1
Summary of the posterior distributions of the predictors involving θ (Eq. (4) in Section 2.4).

	Mean	95% HDI		Posterior > 0	Posterior < 0
		Lower	Upper		
Japanese listeners (LL1)					
θ_0	107.0	102.5	111.4		
θ_{TL} (Japanese)	7.4	6.4	8.4	100.0	0.0
θ_B (/p/)	6.3	5.3	7.2	100.0	0.0
θ_{WP1} (mata)	−2.3	−3.4	−1.3	0.0	100.0
θ_{WP2} (manta)	2.5	1.4	3.5	100.0	0.0
$\theta_{TL}\times B$ (JPN × /p/)	1.6	0.6	2.6	99.9	0.1
$\theta_{TL}\times WP1$ (JPN × mata)	1.0	−0.04	2.0	96.4	3.6
$\theta_{TL}\times WP2$ (JPN × manta)	1.3	0.3	2.4	99.0	1.0
Finnish listeners (LL2)					
θ_0	106.4	103.6	109.4		
θ_{TL} (Japanese)	−1.7	−2.6	−0.8	0.0	100.0
θ_B (/p/)	5.9	4.9	6.8	100.0	0.0
θ_{WP1} (mata)	5.1	5.1	6.1	100.0	0.0
θ_{WP2} (manta)	−0.2	−1.2	0.8	100.0	0.0
$\theta_{TL\times B}$ (JPN × /p/)	0.0	−0.9	0.9	49.1	50.9
$\theta_{TL\times WP1}$ (JPN × mata)	−1.9	−2.8	−0.9	0.0	100.0
$\theta_{TL\times WP2}$ (JPN × manta)	0.9	0.0	2.0	96.8	3.2

References

Akpanglo-Narty, J. (1982). *On fricative phones and phonemes: Measuring the phonetic differences within and between languags*. UCLA working papers in phonetics, 55, pp. 1–181.
Amano, S., & Hirata, Y. (2010). Perception and production boundaries between single and geminate stops in Japanese. *Journal of the Acoustical Society of America*, *128*(4), 2049–2058.
Aoyama, K. (2001). *A psycholinguistic perspective on Finnish and Japanese prosody*. Dordrecht: Kluwer Publishers.
Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris Publications.
Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer. (Version 5.1.08). From ⟨http://www.praat.org/⟩ Retrieved 20.06.09.
Campbell, N. (1999). A study of Japanese speech timing from the syllable perspective. *Journal of Phonetic Society of Japan*, *3*(2), 29–39.
Caramazza, A., Yeni Komshian, G. H., Zurif, E. B., & Cabone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French English bilinguals. *Journal of the Acoustical Society of America*, *54*, 421–428.
Cho, T., & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*, 155–190.
Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*, 466–485.
Doty, S. C., Idemaru, K., & Guion, S. (2007). Singleton and geminate stop in Finnish—Acoustic correlates. In *Proceedings of the 8th annual conference of the international speech communication association* (pp. 2737–2740). Antwerp, Belgium.
Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, *40*, 280–288.
Ferrari Disner, S. (1977). *Vowels in Germanic languages*. UCLA working papers in phonetics, 40, pp. 1–79.
Ferrrari Disner, S. (1983). *Vowel quality: The relation betwen universal and langauge-specific factors*. UCLA working papers in phonetics, 58, pp. 1–158.
Forster, K., & Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–124.
Ham, W. H. (2001). *Phonetic and phonological aspects of geminate timing*. New York: Routledge.
Hirata, Y., & Whiton, J. (2005). Effects of speaking rate on the single/geminate stop distinction in Japanese. *Journal of the Acoustical Society of America*, *118*(3), 1647–1660.

Idemaru, K., & Guion, S. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of International Phonetic Association*, *38*(2), 167–186.

Idemaru, K., & Guion, S. (2010). Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica*, *67*, 25–46.

Isei-Jaakkola, T. (2010). Durational variability of vowel quantity boundary for Japanese, Finnish and Czech Speakers in perception. *Paper presented at Speech Prosody*, 2010.

Järvikivi, J., Aalto, D., Aulanko, R., & Vainio, M. (2007). Perception of vowel length: Tonality cues categorization even in a quantity language. In *Proceedings of 16th international congress of phonetic sciences* (pp. 693–696). Saarbrücken.

Järvikivi, J., Vainio, M., & Aalto, D. (2010). Real-rime correlates of phonological quantity reveal unity of tonal and non-tonal languages. *PLoS One*, *5*(9), e12603, http://dx.doi.org/10.1371/journal.pone.0012603.

Jo'o, H. (1977). Gendai nihongo no on'in (Phonology of contemporary Japanese). In S. Ohno, & T. Shibata (Eds.), *Iwanami Kooza Nihongo 5: On'in* (pp. 107–145). Tokyo: Iwanami Shoten.

Kato, Hiroaki, Makiko, Muto, Minoru, Tsuzaki, & Yoshinori, Sagisaka (2003). Relationship between control precision and perceptual sensitivity to segmental duration. In *Proceedings of 15th international congress of phonetic sciences* (pp. 2043–2046).

Keating, P. A., Mikos, M. J., & Ganong, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acousticsal Society of America*, *70*, 1261–1271.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419–454.

Kingston, J., Kawahara, S., Chambless, D., Mash, D., & Brenner-Alsop, E. (2009). Contextual effects on the perception of duration. *Journal of Phonetics*, *37*, 297–320.

Krakow, R. A. (1999). Physiological organization of syllables: a review. *Journal of Phonetics*, *27*, 23–54.

Krakow, R. A., Bell-Berti, F., & Wang, E. Q. (1995). Supralaryngeal declination: Evidence from the velum. In F. Bell-Berti, & L. J. Rafael (Eds.), *Producing speech: Contemporary issues*. New York: AIP Press.

Kruschke, J. K. (2010a). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 658–676.

Kruschke, J. K. (2010b). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*(7), 293–300.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*(3), 299–312, http://dx.doi.org/10.1177/1745691611406925.

Kruschke, J. K. (2013). Bayesian estimation supersedes the *t* test. *Journal of Experimental Psychology: General*, *142*(2), 573–603, http://dx.doi.org/10.1037/a0029146.

Kruschke, J. K. (2015). *Doing Bayesian data analysis. A tutorial with R, JAGS, and Stan* ((2nd ed.). Waltham, MA: Academic Press.

Lehtonen, J. (1970). *Aspects of quantity in standard Finnish*. Jyväskylä: Jyväskylä University Press.

Local, J., & Simpson, A. (1999). Phonetic implementations of geminate in Malayalam nouns. In *Proceedings of the 14th international congress of phonetic sciences* (pp. 595–598).

Maddieson, I. (1980). *Palato-alveolar affricates in several languages*. UCLA working papers in phonetics, 51, pp. 120–126.

Maddieson, I. (1985). *Phonetic cues to syllabification*. UCLA working papers in phonetics, 59, pp. 85–101.

Moreton, E., & Amano, S. (1999). Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies. *Paper presented at Eurospeech '99*.

Myers, S., & Hansen, B. B. (2005). The origin of vowel-length neutralization in vocoid sequences: Evidence from Finnish speakers. *Phonology*, *22*, 317–344.

Nagano-Madsen, Y. (1992). *Mora and prosodic coordination: A phonetic study of Japanese, Eskimo and Yoruba*. Lund, Sweden: Lund University Press.

Pickett, E. R., Blumstein, S. E., & Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate contrast in Italian. *Phonetica*, *56*(3–4), 135–157.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003)*. Vienna, Austria. ISSN 1609-395X.

Port, R., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, *81*(5), 1574–1585.

R Development Core Team (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ⟨http://www.R-project.org⟩.

Research Institute for the Languages of Finland. (2007). Dialect distribution map (Finnish). URL: ⟨http://www.kotus.fi/index.phtml?l=en&s=368⟩.

Richardson, U., Leppänen, P., Leiwo, M., & Lyytinen, H. (2003). Speech perception of infants with high familial risk for dyslexia differ at the age of 6 months. *Developmental Neuropsychology*, *23*(3), 385–397.

O'Dell, M. (2003). *Intrinsic timing and quantity in Finnish* (Doctoral dissertation). Tampere, Finland: University of Tampere.

Ofuka, E., Mori, Y., & Kiritani, S. (2005). Sokuon no chikaku ni taisuru senkoo kouzoku boinchoo no eikyoo (Perception of a Japanese geminate stop: The effect of the duration of the preceding/following vowel). *Journal of Phonetic Society of Japan*, *9*(2), 59–65.

Pols, L. C. W., van der Kamp, L. J. T., & Plomp, R. (1969). Perceptual and physical space for vowel sounds. *Journal of the Acoustical Society of America*, *46*, 458–467.

Strange, W., Bohn, O.-S., Trent, S.A, & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, *115*, 1791–1807.

Suomi, K. (2005). Temporal conspiracies for a tonal end: Segmental durations and accentual *f*0 movement in a quantity language. *Journal of Phonetics*, *33*, 291–309.

Suomi, K. (2007). On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics*, *35*, 40–55.

Suomi, K., Toivonen, J., & Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics*, *31*, 113–138.

Terbeek, D. (1977). *A cross-language multidimensional scaling study of vowel perception*. UCLA working papers in phonetics, 37, pp. 1–271.

Terbeek, D., & Harshman, R. (1971). *Cross-language difference the perception of natural vowel sounds*. UCLA working papers in phonetics, 19, pp. 26–38.

Vainio, M., Järvikivi, J., Aalto, D., & Suni, A. (2010). Phonetic tone signals phonological quantity and word structure. *Journal of the Acoustical Society of America*, *128*(3), 1313–1321.

Vayra, M., & Fowler, C. A. (1987). Declination of supralaryngeal gestures in spoken Italian. *Phonetica*, *49*, 48–60.

Ylitalo, R. (2009). *The realization of prominence in three varieties of standard Finnish* (Doctoral dissertation). Oulu, Finland: University of Oulu.