Satsuki Nakai

An explanation for phonological word-final vowel shortening: Evidence from Tokyo Japanese

Abstract: This paper offers an account for the cross-linguistic prevalence of phonological word-final vowel shortening, in the face of phonetic final lengthening, also commonly observed across languages. Two contributing factors are hypothesized: (1) an overlap in the durational distributions of short and long vowel phonemes across positions in the utterance can lead to the misidentification of phonemic vowel length and (2) the direction of bias in such misidentification is determined by the distributional properties of the short and long vowel phonemes in the region of the durational overlap. Because short vowel phonemes are typically more frequent in occurrence and less variable in duration than long vowel phonemes, long vowel phonemes are more likely to be misidentified than short vowel phonemes. Results of production and perception studies in Tokyo Japanese support these hypotheses.

Satsuki Nakai: Queen Margaret University. E-mail: satsuki@ovod.net

1 Introduction

Final lengthening, i.e., phonetic lengthening of final elements of prosodic constituents such as the phrase and utterance, is a putative language universal found in many languages, including languages with phonemic vowel length contrasts such as Dinka (Remijsen and Gilley 2008), Estonian (Krull 1997), Finnish (Nakai et al. 2009), and Hungarian (Hockey and Fagyal 1999; White and Mády 2008). Curiously, in many languages with a phonemic vowel length contrast, word-final vowel length neutralization is also commonly observed towards the short vowel phoneme (see, e.g., Buckley 1998; Myers and Hansen 2007). Myers and Hansen (2007) ask how such length neutralization often arises in the face of cross-linguistically prevalent phonetic final lengthening. Why do word-final vowels tend to be neutralized towards the short vowel phoneme if they are often in a position to be lengthened phonetically? As a solution to the apparent paradox, Myers and Hansen (2007) propose that utterance-final devoicing, reported for
various languages, may be responsible for the cross-linguistic tendency for phonological word-final vowel shortening. Myers and Hansen’s (2007) proposal is based on their finding that the devoiced ending of a partially devoiced vowel is disregarded by the listener in phonemic length identification, producing perceptual shortening. This scenario may be a likely one for some languages that exhibit phonological word-final shortening, but not all, e.g., Tokyo Japanese, as we see below.

In this paper, I will present an alternative account to Myers and Hansen’s (2007) for the apparent paradox between the cross-linguistic tendency for phonetic final lengthening and phonological word-final vowel shortening. I will argue that the paradox is resolved by postulating two hypotheses:

**H1:** Short and long phonemes of similar durations across positions in the utterance can produce perceptual ambiguity for these phonemes.

**H2:** When phonemic length misidentification occurs due to distributional overlap in the durations of short and long phonemes, the listener is more likely to misidentify the phoneme whose frequency of occurrence is lower in the region of distributional overlap.\(^1\)

I will restrict the current investigation to the durational overlap arising from phonetic utterance-final lengthening, although H1 may potentially apply to durational overlap between short and long phonemes at other levels (e.g., position in the word). Possible effects of vowel centralization on phonemic length identification are regarded as part of H1, to the extent that the degree of vowel centralization and duration are correlated (see, e.g., Hirata and Tsukada 2009).

Crucially, H2 is not concerned with whether durational overlap arises from phonetic shortening or lengthening. The view taken here does not regard the durations of short and long phonemes in non-final position as more standard than their durations in final position, implicit in a view that associates phonetic final lengthening with the misidentification of the final short phoneme but not the non-final long phoneme. It is instead assumed that no context is inherently

---

\(^1\) The durational distributions of short and long phonemes can overlap also within the same position in the utterance, even under similar articulation rates. However, this probably is less common than an overlap across positions in the utterance; a durational overlap of short and long phonemes within the same position in the utterance is likely to be accompanied by a durational overlap of the two phonemes across positions in the utterance, but the reverse is not necessarily true. I assume that a durational overlap between short and long phonemes within the same context also produces perceptual ambiguity for their phonological length, and that H2 applies to their misidentification also.
more standard than others in the phoneme identification process, and that phonetic final lengthening can lead to the misidentification of both the final short phoneme and the non-final long phoneme, when phonetic final lengthening produces short and long phonemes of a similar duration across positions in the utterance. Below, I will justify the two hypotheses and illustrate how they predict phonological word-final vowel shortening.

### 1.1 Listener-driven account of sound change

Like Myers and Hansen’s (2007) account, the present account of phonological word-final vowel shortening can be classified as a listener-driven phonetic account of sound change, advocated most notably by Ohala (1981), with a number of followers (e.g., Beckman et al. 1992; Guion 1998; Jacewicz, Fox, and Salmons 2006). In this view, some sound changes result from ambiguity in the speech signal arising from phonetic conditioning factors such as coarticulation. When the listener misinterprets the ambiguous speech signal, a sound change occurs in the direction of the misinterpretation. For instance, the F2 of /u/ is raised by a neighboring /t/. This can make the F2 of /u/ similar to the F2 of /y/, which may cause the listener to misidentify the vowel in /ut/ as /y/. The listener would then produce the vowel as /y/, resulting in a sound change.

The precise ways in which articulatory and acoustic durations of speech segments are controlled differ depending on the purpose (e.g., accentual vs. final lengthening; see Edwards, Beckman, and Fletcher 1991; de Jong and Zawaydeh 1999). The precise implementation of long phonemes and phonetic final lengthening are therefore expected to differ articulatorily and acoustically. Yet, phonetic final lengthening can potentially produce short and long phonemes of a similar duration across different positions in the same utterance. This is especially likely for word-final short and long phonemes, as the word-final segment can undergo substantial phonetic final lengthening at the end of the phrase and utterance. Evidence suggests that mature, native listeners are likely to utilize contextual information available to them (e.g., other cues signaling the word’s position in the utterance) in phonemic length identification (Nooteboom and Doodeman 1980; see also Klatt and Cooper 1975; Verbrugge and Shankweiler 1977). However, this contextualization process is not error-free, according to the listener-based phonetic account of sound change. Thus, word-final short and long phonemes may sometimes be misidentified when their durations are similar across different positions in the utterance.

Hirata and Lambacher (2004) demonstrated that Japanese listeners can nearly always correctly identify short and long vowel phonemes of similar durations.
found in the same position of utterances produced at different articulation rates, provided that the test word is presented with some of the preceding or following speech material. Thus, similar durations of short and long phonemes across different articulation rates do not appear to pose a problem to the listener. This finding, however, does not entail that listeners can always correctly identify short and long phonemes of similar durations found across different positions in the utterance. Rather, one might expect such durational overlap between short and long phonemes to lead to phonemic length misidentification, given that phonemic length identification is dependent on the durations of other segments in the utterance, as found by Hirata and Lambacher (2004; see also Fujisaki, Nakamura, and Imoto 1975).

To give a simplified example, suppose that listeners can correctly identify a 100-ms vowel as phonemically short in a slow utterance, because other short vowel phonemes are ca. 100 ms and long vowel phonemes ca. 200 ms; and correctly identify a vowel of the same duration as phonemically long in a fast utterance, because other short vowel phonemes are ca. 50 ms and long vowel phonemes ca. 100 ms. Suppose also that, due to phonetic final lengthening, the utterance-final short vowel phoneme is 200 ms in the slow utterance and 100 ms in the fast utterance. At each articulation rate, the utterance-final short vowel phoneme is more similar in duration to other long vowel phonemes than short vowel phonemes, which could add noise to the phonemic length identification process.

Before moving on, it is perhaps useful to lay out the difference between the phonetic conditioning of duration and length neutralization. Under phonetic conditioning of duration, short and long phonemes both have shorter (or longer) durations in a certain (e.g., phrase-medial) context than in other (e.g., utterance-final) contexts, but the short phoneme is shorter in duration than the long phoneme in each context. In contrast, under length neutralization, when the process is complete, one of the length categories is lost, and only one phonemic length can occur. If the direction of length neutralization is towards the short phoneme, the listener would hear a short phoneme where there used to be a long phoneme, but this in itself is not a sufficient condition for regarding the phenomenon as length neutralization. If the listener-driven phonetic account of sound change is correct, the listeners can misidentify the phonological length of a phoneme in the absence of length neutralization, when the durations of short and long phonemes are similar across different contexts.

Operationally then, under length neutralization the sound in question is normally perceived as phonologically short (or long, depending on the direction of neutralization; more on this shortly), and its durational distribution has a single peak in the given context, forming a unimodal distribution. Under the phonetic
conditioning of duration, the long phoneme can be misidentified as phonologically short and vice versa, but the durational distributions of the short and long phonemes each have a separate peak in each context, together forming a bimodal distribution. I stress here the importance of comparing durational distributions of short and long phonemes context by context in distinguishing length neutralization from the phonetic conditioning of duration because the phonetic realizations of a phoneme, including duration, can be extremely variable depending on the context in which the phoneme occurs (e.g., Klatt 1976; Moon and Lindblom 1994; Fougeron and Keating 1997). Similar durations of short and long phonemes across different contexts may well cause length neutralization but does not constitute length neutralization itself.

1.2 The direction of phonemic length misidentification

Assuming that a similar duration of short and long phonemes resulting from phonetic final lengthening can produce phonemic length misidentification, I now discuss the likely direction of bias in such misidentification. H2 states that the direction of such misidentification is biased by the relative frequency of occurrence of the short and long phonemes in the region of durational overlap. This hypothesis falls out of a growing body of evidence that phonemic decisions are probabilistic and sensitive to the distributional properties of phonemes in the phonetic space (McMurray, Tanenhaus, and Aslin 2002; Clayards et al. 2008; Feldman, Griffiths, and Morgan 2009; Kirby 2010).

In many languages with vowel quantity oppositions, short vowel phonemes occur more frequently than long vowel phonemes, e.g., Czech (Kučera 1961), Finnish (Hakokari et al. 2005), Hungarian (Hockey and Fagyal 1999), Japanese (Hallé 1991), and possibly Swiss French (Grosjean et al. 2007). Across languages, short vowel phonemes are also typically less variable in duration than long vowel phonemes (Lehiste 1970: 36). Because of these properties, long vowel phonemes with ambiguous durations are more likely to be misidentified than their short counterparts, as illustrated below.

Figure 1a schematizes the frequency distributions of a more and a less frequent phoneme (Phonemes A and B) of an equal variance along an acoustic parameter, produced in a certain context. Phonemes A and B can be distinguished on the basis of their values along the specified acoustic parameter alone, as there is no distributional overlap between the two phonemes along the parameter. Figure 1b depicts the distributions of the same two phonemes when their values from another context are added to the distributions in Figure 1a. The two phonemes now partially overlap along the specified parameter, and are not
entirely distinguishable without information regarding the context in which the phonemes occurred. In Figure 1b the overlapping region of the parameter, or the range of ambiguous values, is divided into Regions 1 and 2. With no contextual information, the listeners are more likely to identify a value from Region 1 as Phoneme A than Phoneme B, since Phoneme A is more frequent than Phoneme B in this region. Likewise, the listeners are more likely to identify a value from Region 2 as Phoneme B than Phoneme A. As a result, the misidentification rate is higher for Phoneme B than for Phoneme A in Region 1; the reverse is true for Region 2. As can be gleaned from the figure, when the two sub-regions of overlap are combined, Phoneme A is more frequent than Phoneme B. Therefore, the misidentifi-
Phonological word-final vowel shortening

The direction of bias in misidentification is affected not only by the phonemes' overall frequency of occurrence but also by the spread of their distributions (Clayards et al. 2008). Figure 2 exemplifies how the change in the variance of Phonemes A and B would affect the misidentification of values in the region of distributional overlap. In Figures 2a and 2b, the overall frequency of Phoneme A is roughly twice as high as that of Phoneme B, as was the case in Figure 1b. As we just saw in Figure 1b, where two distributions had an equal variance, the

Fig. 2: (a) Schematic frequency distributions of a more frequent Phoneme A with a large variance and a less frequent Phoneme B with a small variance. (b) Schematic frequency distributions of a more frequent Phoneme A with a small variance and a less frequent Phoneme B with a large variance.
misidentification rate for the values in the region of overlap is predicted to be higher for the less frequent Phoneme B. If, however, the more frequent phoneme has a much greater variance than the less frequent phoneme, as in Figure 2a, the misidentification rate in the region of overlap is predicted to be higher for the more frequent Phoneme A, because Phoneme A is less frequent in the region of overlap. On the other hand, as a comparison of Figures 1b and 2b suggests, when the less frequent Phoneme B has a larger variance than the more frequent Phoneme A, the relative misidentification rate for the less frequent Phoneme B compared to the more frequent Phoneme A should be even higher than when the two phonemes have an equal variance.

Now, let us suppose that Figure 2b represents the durational distributions of a word-final short vowel phoneme (Phoneme A) and a word-final long vowel phoneme (Phoneme B). (Recall that, cross-linguistically, long vowel phonemes are less frequent in occurrence and more variable in duration.) If we also suppose that the overlap in the two phonemes’ distributions was caused by phonetic utterance-final lengthening, the region of distributional overlap would then largely consist of the word-final short vowel phoneme in utterance-final position and the word-final long vowel phoneme in non-utterance-final position. Mature, native listeners probably can use other information signaling the word’s position in the utterance to correctly identify the phonemic length of vowels in the region of distributional overlap. However, if this contextualization process is not constraining enough, the identification of phonemic length of vowels in the region of distributional overlap would be influenced by the frequency effect just described. Under the present scenario, the word-final long vowel phoneme in non-utterance-final position and the word-final short vowel phoneme in utterance-final position would both be susceptible to misidentification, but the former would be more prone to misidentification than the latter.

When viewed in this way, phonetic shortening and lengthening predict the same net direction of bias in phonemic length misidentification. If we suppose that the distributional overlap in Figure 2b resulted from phonetic shortening (e.g., arising from utterance-final devoicing), the region of overlap would now largely consist of the short vowel phoneme in a non-shortening context and the long vowel phoneme in a shortening context. The likelihood that a random value from the region of distributional overlap belongs to Phoneme A or Phoneme B is the same as when the overlap was assumed to have arisen from phonetic utterance-final lengthening. Thus, whether it arises from phonetic shortening or lengthening, the overall misidentification rate in the region of distributional overlap is predicted to be higher for the long vowel phoneme than for the short vowel phoneme. This bias in misidentification could, over time, result in word-final vowel length neutralization towards the short phoneme (see, e.g., Klein 1966).
1.3 Phonological word-final vowel shortening in Japanese

This paper looks at Tokyo Japanese (or Standard Japanese, hereafter Japanese) to investigate how the distributional properties of the durations of the short and long vowel phonemes in phrase-medial and utterance-final positions affect the identification of their phonemic length. Japanese has five short vowel phonemes (/i/, /e/, /a/, /o/, and /u/) and their long counterparts, which are transcribed as /iR/, /eR/, /aR/, /oR/, and /uR/ in the traditional analysis of Japanese phonology (Shibatani 1990). Phonetically, the five pairs of short-long vowel phonemes are often transcribed as [i]-[iː], [e]-[eː], [ɑ]-[ɑː], [o]-[oː], and [ɯ]-[ɯː], since the defining characteristic that distinguishes the members of each pair is duration (Vance 2008; Hirata and Tsukada 2009).

Like many other languages, Japanese has been reported to exhibit prosodic-constituent-final lengthening (Kaiki and Sagisaka 1992; Ueyama 1999; Fon 2002). Nevertheless, diachronic and synchronic phonological shortening of word-final long vowels have been reported for a group of Japanese words (Alfonso 1966; Sukegawa and Maekawa 1997; Sukegawa and Ogawara 1999; Kubozono 2002). These include:

– **tyo’o.tyo** (from *tyo’o.tyoo* ‘butterfly’: diachronic shortening; the apostrophe indicates a lexical pitch accent)
– **se’n.ta** (from *se’n.ta* ‘center’: synchronic shortening)
– **sen.se’** (from *sen.se’e* ‘teacher’: synchronic shortening)
– **kak. ko** (from *kak.koo* ‘appearance’: synchronic shortening; *kak.ko(o)* is lexically unaccented)

Myers and Hansen’s (2007) utterance-final devoicing account is unlikely to explain phonological word-final vowel shortening in Japanese, as the long vowel phonemes tend not to be devoiced in Japanese (Gordon 1998; Maekawa and Kikuchi 2005). The alternative account presented above, however, may explain phonological word-final vowel shortening in Japanese, as the Japanese short vowel phonemes are higher in frequency of occurrence than their long counterparts. For example, Hallé (1991) reports that the short and long vowel phonemes accounted for 89% and 11% of vowels in a spoken Japanese corpus, respectively. As can be seen in Table 1, the short member is more frequent than the long member for all five pairs of vowel phonemes, although the magnitude of the frequency difference varies among the pairs. Furthermore, the long vowel phonemes are generally more variable in duration than the short vowel phonemes in Japanese (Hirata and Tsukada 2009), as in other languages (Lehiste 1970: 36). Thus, vowels of ambiguous durations are more likely to be identified as phonemically short than long.
This is not the first paper that seeks a listener-driven phonetic account for phonological word-final vowel shortening in Japanese. Sukegawa and Ogawara (1999) attribute it to a greater shortening of the long vowel phonemes in the word-final compared to the word-initial syllable in fast speech. Kubozono (2002) attributes it to the shorter durations of the short and long vowel phonemes, and the smaller durational difference between the two vowel categories, in the word-final compared to the word-initial syllable, among other things. These phenomena can certainly contribute to phonological word-final vowel shortening, for they make a durational overlap between the word-final short and long vowel phonemes more likely across positions in the utterance, presumably increasing the likelihood of phonemic length misidentification. The view presented in this paper, however, differs somewhat from these (and probably many other) researchers’ in maintaining that not just phonetic shortening, but also phonetic lengthening can contribute to the emergence of phonological vowel shortening.

2 Production study

A production study was conducted to investigate whether utterance-final lengthening is likely to produce short and long vowel phonemes of similar durations across positions in the utterance. Specifically, Japanese speakers produced di-syllabic test words placed in phrase-medial vs. utterance-final position of tightly controlled test utterances.

2 Sukegawa and Ogawara (1999) do not necessarily regard the synchronic word-final vowel shortening as phonological. However, I take the view that at least some cases of synchronic shortening are becoming phonologized, as reflected in orthography and dictionary entries.

<table>
<thead>
<tr>
<th>Short vowel</th>
<th>Frequency</th>
<th>Long vowel</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>1110</td>
<td>/iː/</td>
<td>51</td>
</tr>
<tr>
<td>/e/</td>
<td>776</td>
<td>/eː/</td>
<td>135</td>
</tr>
<tr>
<td>/a/</td>
<td>1876</td>
<td>/aː/</td>
<td>20</td>
</tr>
<tr>
<td>/o/</td>
<td>1089</td>
<td>/oː/</td>
<td>288</td>
</tr>
<tr>
<td>/u/</td>
<td>735</td>
<td>/uː/</td>
<td>154</td>
</tr>
<tr>
<td>Total</td>
<td>5542</td>
<td>Total</td>
<td>648</td>
</tr>
</tbody>
</table>
2.1 Methods

2.1.1 Materials

Main test words were four disyllabic nonsense words with short and long vowel phonemes in all possible positions of the word and combinations (see Table 2). They constituted multiple minimal pairs differing in the phonemic length of a single vowel (e.g., /sasa/ vs. /sasaː/), which enabled direct comparisons of the durational distributions of the short and long vowel phonemes in different positions of the four word structures. Nonsense words were used because the vowel length distinction does not have a very high functional load in Japanese (Vance 2008), and it is difficult to find a minimal-pair quadruplet differing only in the phonemic length of the vowels. Using nonsense test words had an additional advantage of making the subsequent perception study free from lexical frequency effects (e.g., Savin 1963).

The consonant /s/ in the nonsense test words was chosen to allow for relatively straightforward acoustic segmentation (Turk, Nakai, and Sugahara 2006). For test vowels, the /a/-/aː/ pair was chosen because /a/ is less likely to be devoiced when flanked by voiceless consonants than other short vowel phonemes (see, e.g., Maekawa 2004). Additionally, the difference in the frequency of occurrence is the greatest between /a/ and /aː/ among the five short–long vowel pairs in Japanese (see Table 1 above). Therefore, the identification bias towards the short vowel phoneme in the later perception study, if any, is most likely to be evident for this pair.

The test words were produced with a lexical pitch accent on the first mora, the default pitch accent location for loanwords and nonsense words of the four word structures used here (Shinohara 2000). They were transcribed using the

<table>
<thead>
<tr>
<th>Test word</th>
<th>Structure</th>
<th>Phonemic transcription</th>
<th>Orthographic transcription in Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>sasa</td>
<td>CV.CV</td>
<td>/sasa/</td>
<td>ササ</td>
</tr>
<tr>
<td>sasaa</td>
<td>CV.CV:e</td>
<td>/sasaː/</td>
<td>ササー</td>
</tr>
<tr>
<td>saasa</td>
<td>CV:CV</td>
<td>/saːsa/</td>
<td>サーサ</td>
</tr>
<tr>
<td>saasaa</td>
<td>CV:CV:e</td>
<td>/saːsaː/</td>
<td>サーサー</td>
</tr>
</tbody>
</table>

Note. Lexically unaccented sasa is a real word.
katakana syllabary, in which the long vowel phonemes are indicated by the lengthening symbol —, e.g., ササー for sasaa (/sasaː/).

An additional set of test words was created from real words (see Table A1 in the Appendices). They were structurally comparable to the nonsense test words including the lexical pitch accent location, and were used to check if vowel durations in various contexts were comparable between nonsense-word and real-word test items. Words reported to exhibit synchronic phonological word-final vowel shortening were not used.

Both sets of test words were embedded in phrase-medial vs. utterance-final position of test utterances, as shown in (1) and (2) below. For each condition, all test words were embedded in basically the same utterance, differing in one word (the word denoting the question number; ichi ‘one’ in the example below). The question number was altered depending on the length of the test word, so that utterance length was equal in the number of syllables and moras, not just across phrase-medial and utterance-final conditions, but also across test words (14 syllables or 22 moras, including the test word). All of the question numbers (e.g., ichi) had a lexical pitch accent on the final syllable. Each test utterance was preceded by a precursor question designed to elicit phrasal accent on a word preceding the test word. This was done so that the test word would be de-accentuated and the durational difference between phrase-medial and utterance-final test vowels could be unambiguously attributed to the effect of position in the utterance.

(1) Phrase-medial condition

Precursor: Dare-ga saasaa tabun ichi-ban-tte itta-no?
‘Who said saasaa perhaps is (the answer to) No. 1?’

Test utterance: SENSEI-ga saasaa tabun ichi-ban-tte
teacher NOM saasaa perhaps one number that
itta-yo.
say PAST FP
‘TEACHER said saasaa probably is (the answer to) No. 1.’

(2) Utterance-final condition

Precursor: Saasaa-wa go-ban-no kotae-dakke?
‘Is saasaa the answer to No. 5?’

Test utterance: Toujou-sensei-ni kiitara ICHI-ban-ga
Toujou teacher DAT ask-if/when one number NOM
saasaa.

saasaa
‘According to Mr. Tojo, No. ONE is (the answer to) saasaa.’
2.1.2 Participants

Seven (4 female and 3 male) native speakers of Japanese participated in the study. They were either from Tokyo or a neighboring prefecture (Chiba and Kanagawa), areas associated with Tokyo Japanese. Their mean age was 32 years. None of the participants reported any history of speech or hearing problems. At the time of the recording they had lived in an English speaking country (or countries) for an average of 2.5 years. Consent to participate in the study was obtained following the British Psychological Society’s guidelines.

2.1.3 Procedure

Test utterances were produced in a sound-treated recording studio at the University of Edinburgh. The test utterances were visually presented to the participants on a computer screen, along with an appropriate precursor question. Each test utterance was produced after a pre-recorded precursor question read by a native speaker of Tokyo Japanese, as the auditory presentation of the precursor question facilitated the correct placement of phrasal accent in a pilot study. The recordings were digitized at a 48 kHz sampling rate and had 16-bit quantization.

Before the recording, the participants' familiarity with real-word test items and their pronunciation of nonsense-word test items were checked. Most participants reported that most of the real-word items were in their active vocabulary, and all could produce both real-word and nonsense-word test items with ease. A short practice session to familiarize the participants with the task followed. During the practice session, the participants were instructed to produce the test utterance in an appropriate manner as an answer to the precursor question (i.e., place phrasal accent on the specified word), and to produce each utterance without pausing in the middle.

Each participant produced each test utterance containing the real-word test item three times, and each utterance containing the nonsense-word test item twice, 160 utterances in total ([24 real words × 2 positions × 3 repetitions] + [4 nonsense words × 2 positions × 2 repetitions]). The utterances were first blocked by the type of test word (real-word vs. nonsense-word) and then by repetition. All participants produced utterances containing the real-word test items first, to

3 Some participants had heard of, but did not use: /misa/ (3 participants), /pata:/ (3 participants), /ku:pe/ (2 participants), /ma:ka:/ (1 participant). One participant had never heard of /ku:pe/. These items were included in the descriptive statistics in Table A1, as they did not appear to behave differently from the remaining real-word test items.
facilitate the production of utterances containing the nonsense-word test items. Within each block, utterances containing the same test word in phrase-medial vs. utterance-final position were paired and produced consecutively. The order of phrase-medial vs. utterance-final condition within each of these pairs, and the order of pairs containing different test words were random.

### 2.1.4 Acoustic measurements

To obtain the durations of test vowels, test words were acoustically segmented in Praat (Boersma and Weenink 2008) as follows:

First, consonant intervals were determined using supralaryngeal criteria. The interval for /s/ was the portion of the waveform with aperiodic noise in the frequency region characteristic of /s/, as determined by the waveform in conjunction with the spectrogram. The interval for nasal stops was the periodic portion of the waveform with a lower intensity and simpler waveform pattern than the surrounding vowels, which corresponded to nasal formants on the spectrogram. The interval for oral stops was from the cessation of the complex waveform pattern characteristic of the preceding vowel up to the stop release. Where there were multiple bursts at stop release, the peak of the last clear burst was regarded as the end of the stop interval.

The vowel onset was defined as the offset of the preceding consonant interval. The vowel offset was defined as the onset of the following consonant interval, where the vowel was followed by a consonant. The offset of utterance-final vowels was set to be the first peak of the last glottal pulse, where voicing ended abruptly (ca. 80% of utterance-final vowels). Where voicing ended gradually, the vowel offset was the point of the waveform corresponding to the end of (1) clear vowel formants on the spectrogram, or (2) voicing, whichever was earlier. There were several cases where the utterance-final vowel ended with ‘glottal squeak’, following creaky phonation (Redi and Shattuck-Hufnagel 2001). The glottal squeak was excluded from the vowel interval.

Two repetitions of each pair of phrase-medial and utterance-final test words (the second and third repetitions of the real-word items, and both repetitions of nonsense-word items) were acoustically segmented. When either a phrase-medial or utterance-final token of a real-word test item was judged unsegmentable due to, e.g., noise or unclear boundary between segments, both phrase-medial and utterance-final tokens from the same repetition were discarded. They were then replaced by another pair of tokens of the same word produced by the same participant during the first repetition, where possible. A pair of tokens of a real-word test item /seimei/ was also replaced by the first repetition because the vowel
in one token sounded like /ei/ rather than /eː/. No nonsense-word items were discarded. Altogether, 364 pairs of phrase-medial and utterance-final test words (728 words) were segmented (for a breakdown, see Table A1 in the Appendix).

2.2 Results

For each syllable of each test word in each position of the utterance, mean vowel duration was calculated (see Table A1 in the Appendix). Comparisons of the mean vowel durations for the real-word and nonsense-word test items suggested that the behaviors of short and long vowel phonemes in different contexts were similar between the two types of test words. Therefore, in the remainder of this section, I will focus on the durational distributions of vowels in the nonsense words, which allows direct comparisons of the durational distributions of short and long vowel phonemes in different positions of the four word structures.

Table 3 gives descriptive statistics of the duration of /a/ vs. /aː/ in the word-initial and word-final syllables of the nonsense words. The table shows that /a/ had a shorter mean duration than /aː/ in each position of the word and utterance, indicating the presence of two length categories. Figure 3a further shows that there was no overlap in the durational distribution of /a/ vs. /aː/ in the word-initial syllable, phrase-medially or utterance-finally. Figure 3b shows that the same was true for word-final /a/ vs. /aː/ in phrase-medial position, disregarding one outlier. There was some durational overlap between word-final /a/ and /aː/ in utterance-final position, but a majority (89%) of /a/ had a shorter duration than a majority (82%) of /aː/. Thus, in each position of the utterance, the durational distinction between the short and the long vowel category is considered to be present.

Table 3 and Figure 3 also suggest that vowel durations were generally longer in utterance-final than phrase-medial position for both the word-initial and word-final syllables. Thus, vowels in both syllables appear to have undergone

<table>
<thead>
<tr>
<th>Word position</th>
<th>Syllable position</th>
<th>/a/</th>
<th>/aː/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(ms)</td>
<td>(ms)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard deviation in parentheses</td>
<td></td>
</tr>
<tr>
<td>Phrase-medial</td>
<td>word-initial</td>
<td>58.4 (8.62)</td>
<td>134 (14.0)</td>
</tr>
<tr>
<td></td>
<td>word-final</td>
<td>59.7 (9.07)</td>
<td>112 (14.9)</td>
</tr>
<tr>
<td>Utterance-final</td>
<td>word-initial</td>
<td>67.1 (6.64)</td>
<td>156 (18.9)</td>
</tr>
<tr>
<td></td>
<td>word-final</td>
<td>91.5 (25.4)</td>
<td>184 (33.1)</td>
</tr>
</tbody>
</table>
utterance-final lengthening. Furthermore, consistent with Lehiste’s (1970: 36) observation about the durational variability of long phonemes, the duration of /aː/ appears more variable than the duration of /a/ across phrase-medial and utterance-final positions. This is at least partially due to the greater magnitude of utterance-final lengthening on /aː/ compared to /a/ in absolute duration. The mean difference between phrase-medial vs. utterance-final duration varied from 8 ms (sasa) to 37 ms (saasa) for /a/; and from 21 ms (saasa) to 72 ms (saasaa) for

Fig. 3: Phrase-medial vs. utterance-final duration of /a/ vs. /aː/ in the word-initial (a) and word-final (b) syllables in nonsense words. Each box represents the 25th–75th percentile of the distribution of vowel duration. Whiskers represent the entire distribution, excluding outliers (represented with circles). Horizontal lines inside the boxes represent median values.
Phonological word-final vowel shortening

/aː/. The magnitudes of utterance-final lengthening were comparable between /a/ and /aː/ in proportional terms, however; the phrase-medial vs. utterance-final durational difference varied from 14% (sasa) to 65% (saasa) for /a/, and from 16% (saasaa) to 66% (saasaa) for /aː/.

Despite utterance-final lengthening, there was no durational overlap between /a/ and /aː/ across phrase-medial and utterance-final positions in the word-initial syllable (see Figure 3a). In contrast, there was a considerable distributional overlap between word-final /aː/ in phrase-medial position and utterance-final /a/ (see Figure 3b). Although the median/mean duration was longer for word-final /aː/ in phrase-medial position than utterance-final /a/, the entire distribution of word-final /aː/ in phrase-medial position overlapped with the longer half of the utterance-final /a/. These short and long vowel phonemes in the region of distributional overlap cannot be differentiated from each other on the basis of absolute duration alone.

Two factors can be regarded as responsible for the above difference between the word-initial vs. word-final syllable. First, as can be seen in Table 4, the durations of /a/ and /aː/ were proportionally less distinct in the word-final syllable than in the word-initial syllable in each position in the utterance. Second, the proportional difference in phrase-medial vs. utterance-final duration was greater for vowels in the word-final syllable than in the word-initial syllable. That is, the effect of utterance-final lengthening was progressive. On average, word-final vowels were longer in utterance-final position than in phrase-medial position by 41% (sasa) to 66% (saasaa). The vowels in the word-initial syllable were longer in utterance-final position than in phrase-medial position by 14% (sasa) to 17% (saasa). Thus, the durationally less distinct, word-final /a/ and /aː/ were affected more by utterance-final lengthening, resulting in their substantial durational overlap across phrase-medial and utterance-final positions.

The above observations were subjected to statistical evaluation. A log transformation was performed on each participant’s mean duration of each of the two

<table>
<thead>
<tr>
<th>Word-initial syllable</th>
<th>Word-final syllable</th>
<th>Utterance-final</th>
</tr>
</thead>
<tbody>
<tr>
<td>sasa vs. saasa</td>
<td>sasaa vs. sasaasaa</td>
<td>sasa vs. saasa</td>
</tr>
<tr>
<td>2.3</td>
<td>1.9</td>
<td>2.3</td>
</tr>
<tr>
<td>saasa vs. saasaa</td>
<td>saasaa vs. saasaa</td>
<td>saasa vs. saasa</td>
</tr>
<tr>
<td>2.3</td>
<td>1.9</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 4: Durational ratios (VːV) between /a/ and /aː/ in nonsense words in various positions.
vowels in each of the four nonsense words in each position in the utterance. A repeated measures ANOVA was run on these values, with the following within-subject factors: Vowel Length (short vs. long), Syllable Position (word-initial vs. word-final), Other Vowel Length (short vs. long; e.g., sasa vs. sasaa, where the target vowel was in the word-initial syllable), and Word Position (phrase-medial vs. utterance-final). The effect of Vowel Length was significant, reflecting the overall longer durations of /aː/ compared to /a/: $F(1,6) = 218, p < .001$, partial $\eta^2 = .97$.

The effect of Word Position and its interaction with Syllable Position were both significant: $F(1,6) = 171, p < .001$, partial $\eta^2 = .97$; $F(1,6) = 19.7, p = .004$, partial $\eta^2 = .77$. Post-hoc Sidak tests indicated that vowels in both the word-initial and word-final syllables were longer in utterance-final than phrase-medial position at $p \leq .001$, although the magnitude of final lengthening was greater for the vowels in the word-final syllable. These results are consistent with the interpretation of progressive utterance-final lengthening. The Vowel Length × Word Position interaction was not significant ($F[1,6] < 1$), suggesting that the magnitudes of utterance-final lengthening on /a/ and /aː/ were proportionally comparable.

Furthermore, the Vowel Length × Syllable Position interaction was significant: $F(1,6) = 20.2, p = .004$, partial $\eta^2 = .77$. Post-hoc Sidak tests indicated that /a/ was shorter than /aː/ in both word-initial and word-final syllables at $p < .001$. Thus, the significant interaction is best interpreted to reflect the proportionally greater distinctiveness of /a/ and /aː/ in the word-initial than the word-final syllable. Finally, the effect of Other Vowel Length and its interaction with Syllable Position was significant: $F(1,6) = 6.33, p = .046$, partial $\eta^2 = .51$; $F(1,6) = 19.1, p = .005$, partial $\eta^2 = .76$. Post-hoc Sidak tests revealed that the duration of the vowel in the word-initial syllable was significantly shorter before /a/ (sasa and saasa) than before /aː/ (sasaa and saasaa) at $p = .001$, while word-final vowel durations were not affected by the phonological length of the vowel in the word-initial syllable. No other main effect or interaction was significant at $p < .05$.

The finding of utterance-final lengthening in the present data is consistent with Ueyama (1999) and Fon (2002), who report final lengthening in Japanese at

---

4 Because $\log(x) - \log(y) = \log(x/y)$, using log-transformed durations in the analysis is equivalent to comparing the proportional difference between the durations of /a/ and /aː/ in phrase-medial and utterance-final positions without collapsing the error terms across positions (Cardinal and Aitken 2006: 70). Here, comparing the proportional difference is more appropriate than the absolute difference, as the magnitudes of final lengthening on /a/ and /aː/ were more comparable in proportional terms.
the intonational phrase level and above, and at odds with Kaiki and Sagisaka (1992), who report sentence-final shortening (see Campbell [1992] for possible reasons for the discrepancy; see Hayashi [1982: 328] for a possible role of speaking style). The finding of progressive final lengthening is consistent with reports for several other languages (e.g., Dutch: Cambier-Langeveld 2000; English: White 2002; Byrd, Krivokapić, and Lee 2006; Turk and Shattuck-Hufnagel 2007; Estonian: Krull 1997; Finnish: Nakai et al. 2009; Guoyu and Putonghua: Fon 2002; Hebrew: Berkovits 1993), and contrasts with Fon’s (2002) finding of no final lengthening on the penultimate syllable at discourse disjunctures in Japanese. The small amount of final lengthening on the penultimate syllable in Japanese may have been obscured by factors such as segment types in Fon (2002), which used uncontrolled spontaneous speech material. Greater durational distinctiveness of the short and long vowel phonemes in the word-initial compared to the word-final syllable in Japanese was also reported by Kubozono (2002), and can be seen as a case of domain-initial strengthening, reported for several languages such as English, French, Korean, and Taiwanese (e.g., Fougeron and Keating 1997; Keating et al. 2003).

3 Perception study

Native speakers of Japanese were asked to identify the phonemic length of the vowels in the nonsense test words from the production study, in order to test the two hypotheses put forward in the introduction:

**H1:** Short and long phonemes of similar durations across positions in the utterance can produce perceptual ambiguity for these phonemes.

**H2:** When phonemic length misidentification occurs due to distributional overlap in the durations of short and long phonemes, the listener is more likely to misidentify the phoneme whose frequency of occurrence is lower in the region of distributional overlap.

---

5 Like the present study, Fon (2002) and Ueyama’s (1999) studies were both conducted in an English-speaking country. Therefore, the participants in those studies are likely to have been able to speak English, which probably was not the case with Kaiki and Sagisaka (1992). To check if utterance-final lengthening found in the present study is attributable to the participants’ knowledge of English, the participants were asked to produce the same utterances as those collected from monolingual Japanese speakers in a pilot study. The mean segmental durations of test words produced in phrase-medial vs. utterance-final position by the monolingual and bilingual groups were almost identical.
To assess these hypotheses, I will focus on the misidentification of word-final vowels, since the vowels in the word-initial syllable carried a lexical pitch accent that provided an additional acoustic cue to the phonemic length of the vowel, besides duration. Due to the lexical pitch accent, the long vowel phoneme – but not the short vowel phoneme – in the word-initial syllable was perceived with a pitch fall, which is known to affect Japanese listeners’ identification of the phonological length of a vowel (Nagano-Madsen 1992; Kinoshita, Behne, and Arai 2002). Given that the durational distributions of the short and long vowel phonemes in the word-initial syllable were distinct not just within but also across positions in the utterance (see Figure 3a), both H1 and the presence of a lexical pitch accent predict accurate phonemic length identification for vowels in the word-initial syllable, and the two effects cannot be easily distinguished in the present data.

For the word-final short and long vowel phonemes, the following predictions were made on the basis of the results of the production study. H1 predicts a low misidentification rate for word-final /a/ in phrase-medial position, whose duration hardly overlapped with the duration of word-final /aː/ in any position (see Figure 3b). By contrast, word-final /aː/ in phrase-medial position would be prone to misidentification, as its duration overlapped considerably with the duration of utterance-final /a/. If phonemic length judgment is perfectly constrained by the durational distributions of short and long vowel phonemes in an appropriate context, neither the short or long phoneme would be often misidentified in phrase-medial position, as their durations overlapped very little within this position. Although the mean durations of the short and long vowel phonemes were less distinct in the word-final than in the word-initial syllable, this by itself is unlikely to lead to a frequent misidentification of word-final vowels in phrase-medial position. In this position, 93% of word-final /a/ tokens were shorter than 70 ms, while 96% of the word-final /aː/ were longer than 90 ms (see Figure 3b); thus, a great majority of the short- vs. long-category members were separated by durations comfortably larger than the 75% just-noticeable difference of 7–10 ms estimated for a vowel sound by Fujisaki, Nakamura, and Imoto (1975) using Japanese listeners (at a mean mora duration of 200 ms). Regardless of the correctness of H1, some of the word-final /a/ and /aː/ in utterance-final position are likely to be prone to misidentification, assuming similar articulation rates of the test utterances, because of the distributional overlap between the short and long categories within utterance-final position (see Figure 3b).

If H2 is also correct, the overall misidentification rate is likely to be higher for word-final /aː/ than for word-final /a/ in the region of distributional overlap. This prediction is based on the estimated property of real-world durational distributions of these vowels given in Figure 4b, derived as follows. First, frequency dis-
Fig. 4: (a) Durational distributions of word-final /a/ and /aː/ in the production study, where the two phonemes occurred an equal number of times in each position in the utterance and word structure. Each phoneme’s phrase-medial and utterance-final durations in different word structures are aggregated. (b) Frequency-adjusted durational distributions of word-final /a/ and /aː/.

Distributions of vowel duration were tabulated for word-final /a/ and /aː/ in the nonsense words in the production study, where each type of vowel phoneme occurred equally often in each word structure and each position of the utterance. Figure 4a shows the resulting distribution of each phoneme, aggregated across word structures and positions in the utterance. Next, rough measures of real-world phrase-
medial and phrase/utterance-final occurrences of word-final /a/ and /aː/ were obtained for each word structure used in the production study (see Table 5). For this, a 50-minute spontaneous conversation between Japanese university students recorded by Adachi (2011) was used, as there appears to be no published data regarding the frequencies of occurrence of the Japanese short and long vowel phonemes in different positions in the utterance or word. Phrase-final and utterance-final vowels were grouped together because their durations are reported to be similar in Japanese (Ueyama 1999). Finally, frequency-adjusted distributional profiles of the durations of word-final /a/ and /aː/ were obtained as follows. The frequency of occurrence of each vowel category in each duration interval in Figure 4a was multiplied by an appropriate frequency value in Table 5, separately for each word structure and position in the utterance. The resulting values were aggregated for each vowel category and plotted in Figure 4b. (The distributional properties of word-final /a/ and /aː/ in Figures 4a and 4b differ not only in relative height but also shape, because these vowels did not occur equally often in phrase-medial vs. utterance-final position or in different word structures.) The figure uses a logarithmic scale on the y-axis, as frequency is thought to have effects at a roughly logarithmic scale (Feldman, Griffiths, and Morgan 2009).

In Figure 4a, where /a/ and /aː/ occurred an equal number of times in each position in the utterance and word structure, the region of distributional overlap comprises slightly fewer tokens of /a/ than /aː/ (37 vs. 40 tokens). As discussed in the introduction, H2 predicts that the misidentification rate of a phoneme in the region of distributional overlap inversely correlates with the phoneme’s probability density (the likelihood of a random value being that phoneme) in that region. Thus, when the real-world frequency difference between /a/ and /aː/ is not factored in, H2 would predict the misidentification rate in the region of overlap to be, if anything, higher for /a/. By contrast, in Figure 4b, which factored in the difference in the frequency of occurrence of word-final /a/ vs. /aː/ in a spontaneous conversation, the region of distributional overlap comprises far more tokens of

<p>| Table 5: Frequency of occurrence of word-final /a/ vs. /aː/ in phrase-medial vs. phrase/utterance-final position in a spontaneous conversation recorded by Adachi (2011). |
|----------------------------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Phoneme</th>
<th>Word structure</th>
<th>Phrase-medial</th>
<th>Phrase/utterance-final</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>CV.CV</td>
<td>182</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>CV.CVː</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>/aː/</td>
<td>CV.CVː</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>CV..CVː</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
On the basis of the frequency-adjusted durational distributions, therefore, H2 would predict a higher misidentification rate for word-final /a:/ than /a/ in the region of distributional overlap. Needless to say, random responding to the vowels in the region of distributional overlap should produce ca. 50% misidentification rate in the region of overlap for each vowel category.

Notice that, under the current scenario, H2 predicts an overall higher misidentification rate for /a:/ than /a/ in the region of distributional overlap, but not at the long end of the region. Figure 4b shows that, as the vowel duration increases from 60 to 140 ms, the difference in the estimated frequency of /a/ vs. /a:/ decreases, and the frequency of /a:/ becomes greater than that of /a/ in the 140–180 ms interval. Thus, as the vowel duration increases in the region of distributional overlap, the misidentification rate should generally decrease for /a:/ and increase for /a/, with a higher misidentification rate for /a/ than /a:/ at the longest duration range in the region, where only utterance-final /a/ and /a:/ are found.

### 3.1 Methods

#### 3.1.1 Materials

All 112 test utterances containing nonsense test words from the production study were used.

#### 3.1.2 Participants

A new group of 10 (5 female and 5 male) native speakers of Japanese participated in the perception study. They all grew up in Tokyo or neighboring prefectures (Chiba, Kanagawa and Saitama). Their mean age was 34 years.

#### 3.1.3 Procedure

The experiment was conducted via the internet. The participants listened to recorded test utterances over headphones and selected what the test word was from a choice of four options (sasa, saasa, sasaa and saasaa) given in the katakana syllabary. For each item, the whole utterance was presented to provide contextual information. The transcription of each utterance (excluding the test
word) was also provided in Japanese on the computer screen. The participants were instructed to choose an answer for every test word even if they were unsure. Each of the 112 test utterances was played once in a randomly determined order to half of the participants, and in the reverse order to the remaining half. The inter-stimulus interval was 5 seconds. Altogether 1120 responses (112 test words × 10 participants) were collected.

3.2 Results

Table 6 gives the breakdown of percentage responses to each type of test word in phrase-medial vs. utterance-final position in the perception study (all stimuli). The overall correct response rate was 86%, clearly above the chance level of 25%, taking account of the phonemic length of vowels in both the word-initial and word-final syllables. Despite the considerable durational overlap across phrase-medial and utterance-final positions, word-final /a/ and /aː/ were also identified correctly at a better-than-chance level. The lowest correct response rate was

Table 6: Contingency tables summarizing the results of the perception study (all stimuli). The columns represent word structures intended by the participants in the production study, and the rows represent responses in the perception study. The numbers are percentage responses for each column; the number of responses for each cell is given in brackets. The total number of responses for each column is 140 (14 tokens × 10 participants).

<table>
<thead>
<tr>
<th>Perceived word</th>
<th>Intended word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sasa</td>
</tr>
<tr>
<td>sasa</td>
<td>97.1 (136)</td>
</tr>
<tr>
<td>sasa</td>
<td>2.86 (4)</td>
</tr>
<tr>
<td>saasa</td>
<td>0 (0)</td>
</tr>
<tr>
<td>saasaa</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Perceived word</th>
<th>Intended word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sasa</td>
</tr>
<tr>
<td>sasa</td>
<td><strong>92.9</strong> (130)</td>
</tr>
<tr>
<td>sasa</td>
<td>7.1 (10)</td>
</tr>
<tr>
<td>saasa</td>
<td>0 (0)</td>
</tr>
<tr>
<td>saasaa</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
found for the word-final /aː/ in phrase-medial saasaa, whose phonemic length was correctly identified only 66% of the time (including the sasa response). This was above the chance level of 50% (considering the phonemic length of the word-final vowel only): $t = -2.4; p = .04; df = 9$. Recall that the durational distribution of the word-final /aː/ in phrase-medial position completely overlapped with the longer half of the durational distribution of the word-final /a/ in utterance-final position (see Figure 3b). Given this, the participants – more often than not – appear to have successfully used contextual information to identify the phonemic length of the word-final vowels, in line with the literature (Nooteboom and Doodeman 1980). This is evident in Figure 5, which plots the mean phrase-medial vs. utterance-final durations of the word-final /a/ and /aː/ at different misidentification rates. Where there are data points for both positions, the mean utterance-final duration is longer than the mean phrase-medial duration at each misidentification rate for both /a/ and /aː/, indicating the listeners’ expectations for longer vowel durations in utterance-final position.

At the same time, the participants were much more accurate in identifying the phonemic length of the vowels in the word-initial than word-final syllable, as predicted by H1 and the presence of a lexical pitch accent on the word-initial syllable. This is graphically presented in Figure 6a, which summarizes the information in Table 6. Importantly, consistent with the prediction of H1, in phrase-medial position the misidentification rate was much higher for word-final /aː/ than word-final /a/ (see Figure 6b). This difference was statistically significant, according to a paired $t$-test performed on individual participants’ square-root transformed percentage misidentification rates (Weisberg 2005: 179): $t = -8.71; p < 0.001; df = 9$.

H1 is also supported by further inspection of the data. Figure 7 presents misidentification rates for word-final /a/ and /aː/ in phrase-medial vs. utterance-final position, calculated for each of the 20-ms successive intervals along a vowel duration continuum using the following formula:

$$100 \times \frac{\text{number of erroneous responses within interval}}{\left(\text{number of stimulus within interval}\right) \times \left(\text{number of responses to each stimulus}\right)}.$$ 

Figure 7 shows that misidentification mostly occurred in the region of distributional overlap between the durations of word-final /a/ and /aː/ within and across positions in the utterance (i.e., 60–180 ms in Figures 4a and 4b). The figure also shows that, as vowel duration increased, the misidentification rate generally decreased for word-final /aː/ in each position of the utterance and increased for /a/ in utterance-final position, indicating that the listeners were not randomly
responding to the stimuli but were affected by the distributional properties of the two vowel categories. Notice that the discontinuity in the shift of misidentification rate for word-final /a:/ in phrase-medial vs. utterance-final position corroborates the earlier interpretation of the use of contextual information by the partici-

**Fig. 5:** Mean durations of word-final /a/ (a) and /a:/ (b) in phrase-medial vs. utterance-final position at different misidentification rates. The error bars represent the SE of the mean.
At the same time, word-final /aː/ in phrase-medial position was misidentified 50% of the time in the 80–100 ms interval and 19% of the time in the 100–120 ms interval. This was so, despite the fact that, in these intervals, word-final /aː/ in phrase-medial position overlapped in duration only with utterance-final /a/ and not with word-final /a/ in phrase-medial position (see Figure 3b). Similarly, utterance-final /a/ was misidentified 16% of the time in the 100–120 ms interval, even though no utterance-final /a:/ was observed in this interval. These
observations are consistent with H1 that an overlap in the durational distributions of the short and long vowel phonemes across positions in the utterance adds noise to phonemic length identification.

We now turn to the prediction of H2 that the overall misidentification rate in the region of durational overlap is higher for word-final /aː/ than for word-final /a/. In order to test this prediction, percentage misidentification rates in the region of durational overlap were calculated for word-final /a/ and /aː/ in each nonsense word (sasa, sasaa, saasa, saasaa) for each participant. Table 7 presents a summary. These misidentification rates were square-root transformed (Weisberg 2005: 179) and subjected to a repeated measures ANOVA with the following within-subject factors: V₁ Length (the vowel in the word-initial syllable; short vs.
long), V₂ Length (the word-final vowel; short vs. long) and Word Position (phrase-medial vs. utterance-final).

The effect of V₂ Length was significant (F[1,9] = 32.3, p < .001, partial \( \eta^2 = .78 \)), reflecting the overall higher misidentification rate for word-final /aː/ than word-final /a/ (25.1% vs. 7.0%), supporting H₂. The effect of Word Position was also significant (F[1,9] = 6.86, p = .028, partial \( \eta^2 = .43 \)), reflecting the overall higher misidentification rate for word-final vowels in utterance-final position ([12.6 + 26.3] ÷ 2 = 19.5%) than in phrase-medial position ([1.43 + 23.9] ÷ 2 = 12.7%), which is attributable to the low misidentification rate for word-final /a/ in phrase-medial position. Additionally, the V₂ Length × Word Position interaction was significant: F(1,9) = 7.09, p = .026, partial \( \eta^2 = .44 \). Post-hoc Sidak tests suggested that word-final /aː/ was misidentified significantly more often than word-final /a/ in phrase-medial position (p < .001), where vowel durations were relatively short, but not in utterance-final position (p = .15), where vowel durations were relatively long. This observation is consistent with the assumption of H₂ that listeners are sensitive to the distributional properties of the durations of short and long vowel phonemes (see Figure 4b).

Finally, the effect of V₁ Length and the V₁ Length × Word Position interaction were both significant: F(1,9) = 19.4, p = .002, partial \( \eta^2 = .68 \); F(1,9) = 5.85, p = .039, partial \( \eta^2 = .39 \). The significant effect of V₁ Length reflects the higher misidentification rate for the word-final vowel following the long V₁ (saasa and saasaa: [8.08 + 34.3] ÷ 2 = 21.2%) than the short V₁ (sasa and sasaː: [5.93 + 15.9] ÷ 2 = 10.9%). Post-hoc Sidak tests suggested that this difference was significant for utterance-final position (p = .001) but not for phrase-medial position (p = .11), which is attributable to the lack of misidentification for word-final /a/ in phrase-medial position saasa (see Table 7 above). No other main effect or interaction was significant at p < .05.

The relatively high misidentification rates for word-final vowels following the long V₁ (saasa and saasaa) compared to those following the short V₁ (sasa and sasaː) in some contexts render further support to H₁. Figure 8 plots the durational distribution of the word-final vowel in each nonsense word in phrase-medial vs. utterance-final position. As can be seen in the figure, the durational overlap between word-final /a/ and /aː/ within utterance-final position is greater between saasa and saasaa than between sasa and sasaː. Similarly, the durational overlap between utterance-final /a/ and word-final /aː/ in phrase-medial position is greater between saasa and saasaa than between sasa and sasaː. On the other hand, word-final /a/ in phrase-medial saasa did not overlap in duration with word-final /aː/ in phrase-medial or utterance-final saasaa. Thus, word-final /a/ and /aː/ were more often misidentified in saasa and saasaa than in sasa and sasaː, where the durations of the two vowel categories overlapped more between
sasaa than between sasa and sasaa within and across positions in the utterance, consistent with H1.

The high misidentification rate for word-final /aː/ in saasaa relative to sasa is consistent with Sukegawa and Ogawara’s (1999) observation of connected speech, where the word-final long vowel phoneme following a heavy syllable was perceived as phonologically short much more frequently than the word-final long

![Figure 8](image-url)
vowel phoneme following a light syllable. It is also compatible with the results in Hirata and Lambacher (2004), who used phrase-medial CV.CV, CVː.CV and CV.CVː words, and found low misidentification rates for the Japanese short and long vowel phonemes when presented with enough context (see Table 6 above). A greater overlap in the durational distributions of word-final short and long vowel phonemes following a heavy syllable as compared to a light syllable within and across positions in the utterance was also observed in real-word test words.6 (See Figure A1 in the appendix comparing the phrase-medial and utterance-final durations of word-final /o/ and /oː/ in -to and -too following a light vs. heavy syllable.)

These results can be seen as further support for the listener’s role in the emergence of phonological word-final shortening in Japanese, as Japanese word-final long vowels appear to be particularly prone to phonological shortening when they follow a heavy syllable (Alfonso 1966; Sukegawa and Ogawara 1999; Kubozono 2002). Conceivably, the high incidence of misidentification of the word-final long vowel phoneme following a heavy syllable contributes to this phenomenon. In light of the definition laid out in the introduction, phonemic length misidentification observed in the present study should, if anything, be regarded as a contributing factor towards length neutralization, not as evidence for it. In the current production data, the durational distributions of word-final /a/ and /aː/ had separate peaks forming a bimodal distribution in each position of the utterance, indicating the presence of the phonemic length distinction. In perception, the phonological lengths of these vowels were correctly identified at a better-than-chance level, also indicating the presence of the length distinction.

4 Summary and conclusions

This paper offered a listener-driven phonetic account for the cross-linguistic tendency for phonological word-final vowel shortening, in the face of phonetic

6 The long duration of the word-final vowel in some tokens of utterance-final saasa is consistent with a previous report regarding the stretchability of word-final, odd-numbered moras when they are not paired with another mora (Teranishi 1980, as cited in Poser 1990), and may have bearing on the bimoraic foot template hypothesized for Japanese, according to Poser (1990). The relatively short duration of the word-final vowel in saasa as compared to sasa may have bearing on the cross-linguistic tendency for the avoidance of adjacent heavy syllables in a word (*HH: see Anttila 1997). I will not speculate further on these issues, as they are beyond the scope of this paper, but note that these explanations per se would not account for why word-final (and not other) vowels following a heavy syllable are susceptible to phonological shortening (and not lengthening) in Japanese.
prosodic-constituent-final lengthening, also commonly observed across languages. While Myers and Hansen’s (2007) utterance-final devoicing account may explain phonological word-final vowel shortening in some languages, phonological word-final vowel shortening can also occur in languages that exhibit phonetic final lengthening and little utterance-final devoicing on the long vowel phoneme. It was hypothesized that phonetic final lengthening can contribute to the emergence of phonological word-final vowel shortening because (1) short and long vowel phonemes of similar durations across positions in the utterance can lead to the misidentification of phonological length of these vowels, and (2) such vowels are likely to be identified more often as phonologically short than long because of the low frequency of occurrence and the large durational variability typical of the long vowel phoneme. The present account of phonological word-final vowel shortening thus integrates the listener-driven phonetic account of sound change (e.g., Ohala 1981) with a growing body of evidence that phonemic identification is sensitive to the distributional properties of phonemes in the phonetic space (McMurray et al. 2002; Clayards et al. 2008; Feldman et al. 2009; Kirby 2010).

As supporting evidence for the above hypotheses, the results of Japanese production and perception studies that dealt with short and long vowel phonemes in disyllabic nonsense words placed in phrase-medial vs. utterance-final position were presented. In the production study, substantial final lengthening was observed on both the short and long vowel phonemes in utterance-final position. As a result, the longer half of the durational distribution of the word-final short vowel phoneme in utterance-final position completely overlapped with the durational distribution of the word-final long vowel phoneme in phrase-medial position. Likely due to this distributional overlap, in a subsequent perception study, the word-final long vowel phoneme in phrase-medial position was significantly more often misidentified than the word-final short vowel phoneme in phrase-medial position, whose durational distribution overlapped very little with the durational distribution of the long vowel phoneme in any position.

Furthermore, the pattern of shift in misidentification rate for word-final short and long vowel phonemes was consistent with the hypothesis that phonemic length misidentification is affected by the probability density of these phonemes in the phonetic space, producing an overall high misidentification rate for the word-final long vowel phoneme in the region of durational overlap. Moreover, the word-final short and long vowel phonemes generally overlapped more in duration and were misidentified more often after a heavy syllable than a light syllable, i.e., in word structures reported to be prone to phonological word-final vowel shortening in Japanese (Alfonso 1966; Sukeygawa and Ogawara 1999; Kubozono...
Together, these results are indicative of the listener’s role in the emergence of phonological word-final vowel shortening in Japanese.

The observed pattern of misidentification is hard to explain in a view that sees vowel durations in a fixed (e.g., phrase-medial) position as a point of reference used in phonemic length judgment. If we regard the phrase-medial durations as a reference, the misidentification of the utterance-final short vowel phoneme could be attributed to its longer duration than the reference, while the misidentification of the utterance-final long vowel phoneme could be regarded as a hypercorrective interpretation of its long duration (Ohala 1993). However, we would run into problems explaining the misidentification of the phrase-medial long vowel phoneme. Attributing this to phrase-medial shortening would raise a question about why the phoneme’s phrase-medial duration serves as a reference in some cases and not in others. The observed pattern of misidentification instead sits well with the hypothesis that distributional overlap in the durations of short and long phonemes not just within but also across positions in the utterance produces perceptual ambiguity for the phonemes in the region of distributional overlap.

While evidence presented in this paper does not include devoiced utterance-final vowels, phonological word-final vowel shortening can be seen to arise from essentially the same mechanism, whether it is caused by utterance-final devoicing or utterance-final lengthening, as discussed in the introduction. Both could potentially produce word-final short and long vowel phonemes of similar durations across positions in the utterance, making them prone to misidentification. The misidentification of such vowels is likely to be higher for the long vowel phoneme across languages because the long vowel phoneme is typically less frequent in occurrence and more variable in duration. This, in turn, could result in the cross-linguistic tendency for length neutralization towards short of word-final vowels.

It may be that other types of phonemic mergers towards the cross-linguistically less marked phoneme (Schane 1973: 115) are also more immediately the consequence of the pattern of distributional overlap between the relevant phonemes in the phonetic space, where the merger is triggered by listener misinterpretation (see Hume [2004] for a similar take on markedness; Chang, Plauché, and Ohala [2001] for an account of markedness in sound change at the auditory rather than the phonemic-decision level). If cross-linguistically marked phonemes are also low in frequency of occurrence in languages with marked phonemes (Prince and Smolensky 1997), the marked phonemes are more likely to also be low in frequency of occurrence where they overlap in distribution with other phonemes in the phonetic space. If so, the predicted direction of bias in the misinterpretation of an ambiguous phoneme based on the pattern of
distributional overlap is likely to be the same as the prediction based on cross-linguistic markedness.

A potential advantage of the present listener-based account of phonological word-final shortening over abstract accounts based on the markedness principle (e.g., Prieto-Vives, as cited in Myers and Hansen 2007) is that the present account might be able to straightforwardly explain why some languages with a vowel length opposition undergo word-final vowel length neutralization (e.g., Choctaw; see Buckley 1998), while others remain resistant to it (e.g., Northern Finnish; Suomi and Ylitalo 2011, personal communication). While this mostly remains an empirical question, the Northern Finnish data in Nakai, Kunnari, et al. (2009) and Nakai, Turk, et al. (2012), at least, appear consistent with the present account. That study found little overlap in the durational distributions of the voiced portion of word-final short (single) and long (double) vowel categories across phrase-medial and utterance-final positions, which could explain why the vowel length contrast is well maintained in word-final position in this variety of Finnish.

As a final note, the frequency effect assumed to bias the identification of phonological vowel length in this paper is at the phoneme level, but frequency effects at other levels such as phonological features, words, and phonotactics are likely to interact with the frequencies of phonemes in biasing the direction of misidentification (Savin 1963; Elman and McClelland 1988; Connine, Blasko, and Titone 1993; Connine, Titone, and Wang 1993; Pitt and McQueen 1998; Moreton and Amano 1999; Hay, Pierrehumbert, and Beckman 2003). A full understanding of the working of frequency effects would require detailed frequency analysis of spoken language at multiple levels (feature, phoneme, phonotactics, word, type, token) as well as perception experiments that manipulate these factors systematically. Furthermore, the distributional overlap of the durations of short and long phonemes can occur at various levels that were not studied in this paper. In the case of Japanese, for example, factors such as variations in vowel duration arising from pitch accent location (Mori and Erickson 2008) may also play a role in vowel length neutralization. Whatever the source of variation may be, however, where vowel length neutralization arises from perceptual factors, the distributional properties of the durations of the short and long phonemes make it likely that the direction of neutralization is towards the short vowel phoneme.

Acknowledgments: This work was funded by a Leverhulme research grant (F/00158/AF) and a British Academy small research grant (SG-39167). I am very grateful to Joe Toscano, an anonymous reviewer, and Mark Hasegawa-Johnson for their help in improving this paper. I am also grateful to Alice Turk, Mariko Sugahara, Mitsuhioko Ota, James Kirby, John Ohala, and James Scobbie for their
support at various stages of the study, and Minoru Ohtsuki and Ayumi Marushima for their help in finding participants for the perception study. Thanks also go to Kenneth de Jong and Jonathan Harrington for their comments on earlier versions of this paper.

References


Suomi, Kari, & Riikka Ylitalo. 2011, March 8. E-mail correspondence.
Appendix

Table A1: Mean test vowel durations (in ms). Standard deviations are given in brackets. 
$N$ indicates the number of measured pairs.

<table>
<thead>
<tr>
<th>Test word</th>
<th>$N$</th>
<th>Phrase-medial</th>
<th>Utterance-final</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$V_1$</td>
<td>$V_2$</td>
</tr>
<tr>
<td>/sasa/ nonsense word</td>
<td>14</td>
<td>56 (10)</td>
<td>63 (9)</td>
</tr>
<tr>
<td>/misa/ female name</td>
<td>14</td>
<td>34 (8)</td>
<td>59 (10)</td>
</tr>
<tr>
<td>/naka/ ‘inside’</td>
<td>12</td>
<td>59 (6)</td>
<td>62 (9)</td>
</tr>
<tr>
<td>/peke/ ‘(an) x’</td>
<td>8</td>
<td>68 (4)</td>
<td>67 (6)</td>
</tr>
<tr>
<td>/soto/ ‘outside’</td>
<td>13</td>
<td>50 (7)</td>
<td>57 (9)</td>
</tr>
<tr>
<td>/tate/ ‘lengthwise’</td>
<td>12</td>
<td>68 (11)</td>
<td>61 (12)</td>
</tr>
<tr>
<td>/toni/ ‘wealth’</td>
<td>10</td>
<td>69 (6)</td>
<td>58 (10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$V_1$</td>
<td>$V_2$</td>
</tr>
<tr>
<td>/sasaː/ nonsense word</td>
<td>14</td>
<td>61 (7)</td>
<td>116 (17)</td>
</tr>
<tr>
<td>/nise:/ ‘second generation’</td>
<td>12</td>
<td>59 (8)</td>
<td>100 (9)</td>
</tr>
<tr>
<td>/pata:/ ‘putter’</td>
<td>14</td>
<td>77 (7)</td>
<td>130 (9)</td>
</tr>
<tr>
<td>/sato:/ family name</td>
<td>14</td>
<td>67 (7)</td>
<td>120 (8)</td>
</tr>
<tr>
<td>/soni:/ ‘Sony’</td>
<td>12</td>
<td>79 (4)</td>
<td>107 (10)</td>
</tr>
<tr>
<td>/soso:/ ‘blunder’</td>
<td>13</td>
<td>55 (4)</td>
<td>117 (15)</td>
</tr>
<tr>
<td>/tena:/ ‘tenor’</td>
<td>14</td>
<td>85 (14)</td>
<td>129 (22)</td>
</tr>
<tr>
<td></td>
<td>$V_1$</td>
<td>$V_2$</td>
<td>$V_{11}$</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>/sa:sa/ nonsense word</td>
<td>14</td>
<td>126 (10)</td>
<td>57 (9)</td>
</tr>
<tr>
<td>/ka:to/ ‘cart’</td>
<td>14</td>
<td>146 (10)</td>
<td>57 (12)</td>
</tr>
<tr>
<td>/ku:pe/ car name</td>
<td>13</td>
<td>125 (9)</td>
<td>69 (10)</td>
</tr>
<tr>
<td>/pa:ma/ ‘permanent wave’</td>
<td>14</td>
<td>146 (8)</td>
<td>65 (7)</td>
</tr>
<tr>
<td>/se:to/ ‘pupil’</td>
<td>13</td>
<td>104 (7)</td>
<td>55 (10)</td>
</tr>
<tr>
<td>/so:mu/ ‘general affairs’</td>
<td>10</td>
<td>115 (9)</td>
<td>42 (7)</td>
</tr>
<tr>
<td>/to:ta/ ‘selection’</td>
<td>14</td>
<td>126 (6)</td>
<td>54 (9)</td>
</tr>
<tr>
<td>/sa:sa:/ nonsense word</td>
<td>14</td>
<td>142 (14)</td>
<td>109 (11)</td>
</tr>
<tr>
<td>/ko:ko:/ ‘filial duty’</td>
<td>14</td>
<td>131 (4)</td>
<td>113 (8)</td>
</tr>
<tr>
<td>/ma:ka:/ ‘marker’</td>
<td>14</td>
<td>130 (7)</td>
<td>117 (9)</td>
</tr>
<tr>
<td>/se:me:/ ‘full name’</td>
<td>13</td>
<td>123 (8)</td>
<td>100 (8)</td>
</tr>
<tr>
<td>/su:pa:/ ‘supermarket’</td>
<td>14</td>
<td>97 (12)</td>
<td>117 (5)</td>
</tr>
<tr>
<td>/ta:ki:/ ‘turkey’</td>
<td>13</td>
<td>146 (8)</td>
<td>113 (7)</td>
</tr>
<tr>
<td>/to:to:/ ‘finally’</td>
<td>14</td>
<td>133 (7)</td>
<td>111 (8)</td>
</tr>
</tbody>
</table>
Fig. A1: Phrase-medial vs. utterance-final durations of word-final /o/ and /oː/ in -to and -too in real words, following a light (a) vs. heavy syllable (b). Each box represents the 25th–75th percentile of the distribution of vowel duration, and the whiskers represent the entire distribution, excluding outliers (represented with circles). Horizontal lines inside the boxes represent median values.