

The Phonetics of Emphatic Vowel Lengthening in English

Aaron Braver^a, Natalie Dresher^b, and Shigeto Kawahara^c

^aTexas Tech University, ^bRutgers University, and ^cKeio University

1 Introduction

It is well established that languages make use of phonetic duration to signal a lexical contrast (e.g., Japanese [no] ‘field’ with a short vowel contrasts with [noo] ‘brain’ with a long vowel). Languages can also use phonetic duration to signal a different kind of meaning: emphasis. For example, English speakers can say *She’s soooooo cool* to emphasize the degree of their belief. While lexical short/long contrasts are well represented in the literature, this second use of duration—emphatic lengthening—has received less attention. A primary difference between these two uses of phonetic duration is that lexical contrasts tend to be binary distinctions, whereas emphatic lengthening can make use of more fine-grained distinctions.

Early studies of emphatic lengthening, primarily in Japanese, are based on impressionistic observations and lack any quantitative or phonetic analysis (Aizawa, 1985; Higuchi & Haraguchi, 2006; though see Nasu, 1999 and Kawahara, 2006, 2013 for judgment experiments). More recent studies have examined the phonetic properties of this pragmatic lengthening, finding that at least some Japanese speakers can make six-way durational contrasts in both vowels (Kawahara & Braver, 2013) and consonants (Kawahara & Braver, 2014) to express different degrees of emphasis. In these studies, Japanese speakers were asked to read words with different levels of emphasis—expressed by way of orthographic gemination and vowel length marks—ranging from level 0 emphasis (i.e., no emphasis) to level 5 emphasis. Illustrative figures are given in Figures 1 and 2, which show that Japanese speakers can make up to six-way durational contrasts—much finer-grained distinctions than found in the binary length contrasts typical of lexical short/long distinctions.

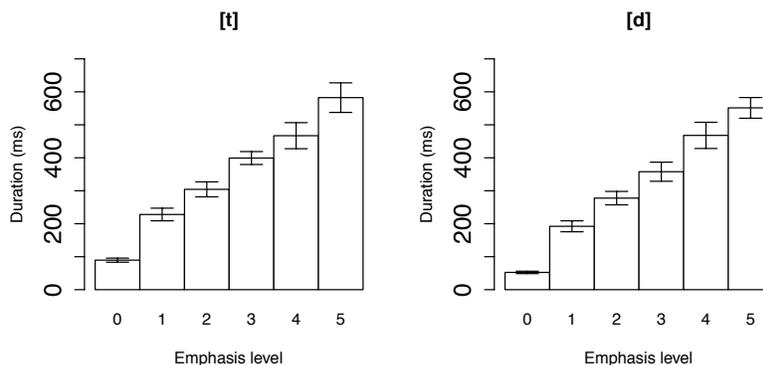


Figure 1: This Japanese speaker makes a six-way consonant duration distinction (for [t, d]) to express degrees of emphasis. Taken from Kawahara & Braver (2014).

One question that remains is whether native language plays a role in a speaker’s ability to produce such fine-grained durational distinctions. In other words, are Japanese speakers able to make these distinctions, at least in part, because their native language makes use of a (lexical) duration contrast? Or, alternatively, do speakers in general possess this ability, regardless of native language? Further, Japanese speakers make their largest duration distinction between the no emphasis and level 1 emphasis conditions. Should we expect the same result from, e.g., English speakers, who lack a lexical length contrast? To address these questions, the current study investigates whether speakers of English—a language which does not exploit durational

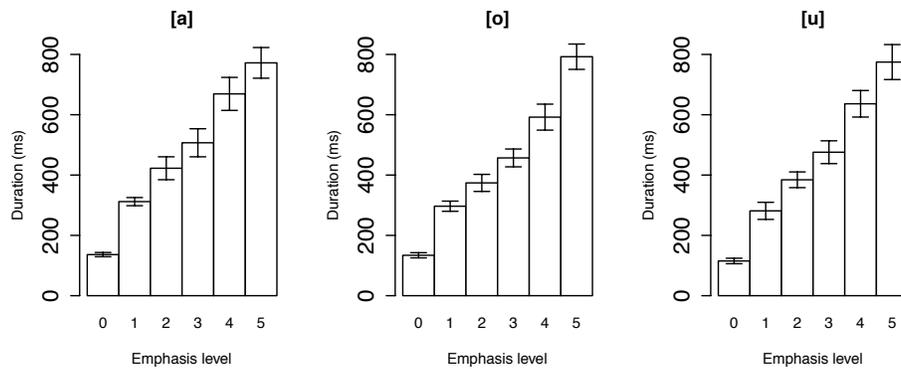


Figure 2: This Japanese speaker makes a six-way vowel duration distinction (for [a, o, u]) to express degrees of emphasis. Taken from Kawahara & Braver (2013).

differences for lexical contrasts—can make similar durational distinctions.

A related remaining question is why there is such an overwhelming cross-linguistic preference for binary vowel length distinctions. Vowel duration contrasts tend, cross-linguistically, to be binary. There are a few rare typological exceptions to this claim such as Estonian, in which this contrast can be ternary (Prince, 1980), but in general the distribution of such contrasts is restricted by various prosodic and morphological factors (see Ladefoged & Maddieson, 1996; Lehiste, 1970; Prince, 1980 for discussion). Additional possible exceptions noted by Ladefoged & Maddieson (1996) include Mixe (Hoogshagen, 1959; cf. Jany, 2006, 2007) and Yavapai (Thomas & Shaterian, 1990). In any case, the vast majority of vowel duration contrasts are binary, and in those languages where ternary contrasts exist, such contrasts are prosodically and/or morphologically restricted. We know of no convincing cases of four-way or greater lexical duration contrasts.¹ Why should this be? One potential answer to this question is that three-way durational contrasts may be difficult to produce or perceive. The experiment presented here shows that speakers can, in fact, produce fine-grained durational distinctions, suggesting that restrictions on production are not (solely) responsible for the preference for binary duration contrasts.

2 Experiment

In this experiment, we examined the extent to which English speakers can produce fine-grained durational distinctions to express pragmatic emphasis. If they are able to do so, it would suggest that the ability to produce these distinctions does not depend on speaking a language with a binary lexical length contrast.

2.1 Method

2.1.1 Stimuli The target words for this experiment were seven English intensifier adverbs which are used to show emphasis in casual speech.² Each adverb was placed into a frame sentence, as in the first column in Table 1. Each sentence was then modified to create 5 levels of emphasis by orthographically lengthening the target adverb, as in the remaining columns in Table 1. For example, the target adverb *so* was placed into the frame sentence *That guy is so creepy* (no emphasis), and five additional sentences were created, replacing *so*

¹ Four-way durational contrasts may appear to exist in cases where two phonological contrasts interact. For example, vowels tend to be longer before voiced stops than before voiceless stops (Chen, 1970), and this lengthening effect may interact with a phonemic vowel length contrast to yield a four-way durational distinction (e.g., VT < VD < VVT < VVD). However, we never observe a single durational lexical contrast that is realized as a four-way durational distinction.

² The use of the adverb *mad* is generally restricted to the greater New York area. All participants in this study were from New Jersey, where this use of *mad* is acceptable.

with *soo*, *sooo*, *sooooo*, *soooooo*, and *sooooooo* to represent emphasis levels 1–5.

2.1.2 Speakers Eight native speakers of New Jersey English participated in this production study. Participation was limited to female speakers since the emphatic lengthening of adverbs is considered by some to be characteristic of feminine speech.

2.1.3 Recording All stimuli (7 adverbs \times 6 levels of emphasis = 42 stimuli) were randomized by Superlab (Cedrus Corporation, 2010) and were visually presented to speakers for oral production. Speakers produced all stimuli 10 times, with order of stimuli randomized within each repetition. Recordings were performed in a sound-attenuated booth using an AT44040 cardioid capacitor microphone with a pop filter, amplified through an ART TubeMP microphone pre-amplifier and JVC RX-554V receiver. The speech was digitized as WAV files at a sampling rate of 44.1kHz using Audacity (Audacity Team, 2008).

2.1.4 Analysis Acoustic analysis was performed using Praat (Boersma & Weenink, 2009). For adverbs with a monophthong that is emphatically lengthened (e.g. *mad*), the duration of that vowel was measured. For diphthong target vowels (e.g., *way*), the duration of the entire diphthong was measured due to the difficulty of placing a boundary between the main vowel and the following offglide. For the adverbs *very* and *really*, since boundaries between vowels and [ɹ, ɪ] are difficult to determine, duration was measured by calculating the duration of the entire word and subtracting the duration of the initial consonant (e.g., very). Sample waveforms and spectrograms are given in Figures 3a and 3b to illustrate our segmentation procedure.

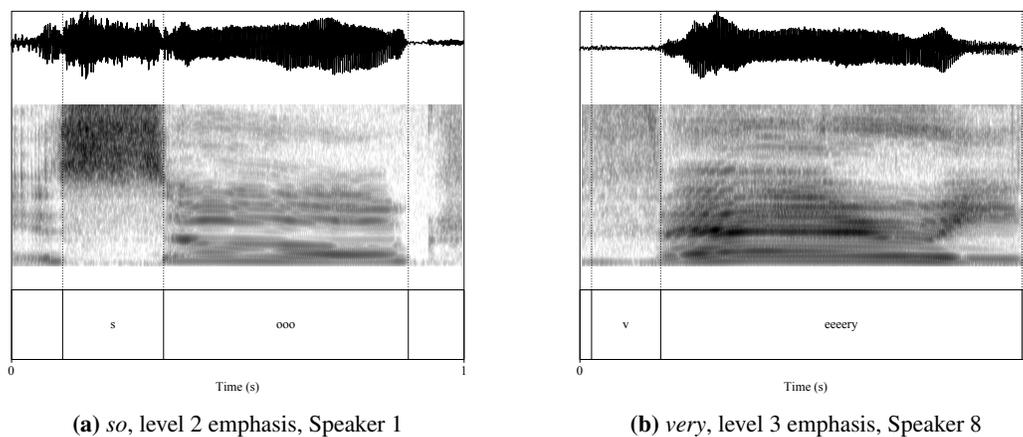


Figure 3: Waveforms and spectrograms illustrating our segmentation procedure. Time scales both 1000 ms.

Following previous work on emphatic lengthening, we provide the Pearson correlation (r) as a measure of correlation between emphasis level and duration (Kawahara & Braver, 2013, 2014). In this calculation, the non-emphatic condition was excluded because the relationship between this condition and the emphatic conditions is non-linear, as we will discuss below. We also ran a regression analysis on the emphatic conditions to assess how many milliseconds speakers increase their vowel duration per emphasis level.

Finally, in the appendix, we provide non-paired t -tests for each speaker comparing each successive pair of emphasis levels (e.g., level 0 vs. level 1, level 1 vs. level 2. . .). In order to avoid Type I error we Bonferroni-adjust our significance level to $\alpha = .01$ (each speaker has 5 comparisons; 0.5/5).

2.2 Results

2.2.1 Individual patterns Since there is non-trivial inter-speaker variation, again similar to Kawahara & Braver (2013), we discuss the behavior of each speaker in turn; summary and comparison of all speakers is provided in §2.2.2. We start with speakers who show the clearest distinctions among different levels of emphasis.

Three speakers—Speakers 1, 5, and 6—show a clear six-way durational contrast (illustrated in Figures 4, 5, and 6). Beginning with Speaker 1, we observe (a) that there is a fairly substantial increase in duration

No emphasis	Level 1	Level 2	Level 3	Level 4	Level 5
That bag is mad expensive.	... maad...	... maaad...	... maaaaad...	... maaaaaad...	... maaaaaaad...
That model is very tall.	... veery...	... veecery...	... veeceryy...	... veeceryyy...	... veeceryyyy...
That baby is too cute.	... tooo...	... tooooo...	... toooooo...	... tooooooo...	... toooooooo...
That guy is so creepy.	... soo...	... soooo...	... sooooo...	... soooooo...	... sooooooo...
That band is way cool.	... waay...	... waaay...	... waaaay...	... waaaaay...	... waaaaaay...
That joke is super funny.	... suuper...	... suuuper...	... suuuuper...	... suuuuuper...	... suuuuuuper...
That lecture was really boring.	... reaaally...	... reaaaally...	... reaaaaally...	... reaaaaaally...	... reaaaaaaally...

Table 1: Stimuli used in the production experiment

between the non-emphatic condition and the first level of emphasis (just as with Japanese speakers), and (b) there is a steady increase in duration from one emphasis level to the next. The r -value for Speaker 1, assessing correlation between level of emphasis and duration, is $r = .75$, which is statistically significant ($p < .001$). The regression analysis shows a best-fitting coefficient of 92 ms, suggesting that for Speaker 1, each additional level of emphasis corresponds to approximately 92 ms of additional duration. Non-paired t -tests show that comparisons between each successive emphasis level are significant at the level of $p < .001$ (see Table 3 in the appendix).

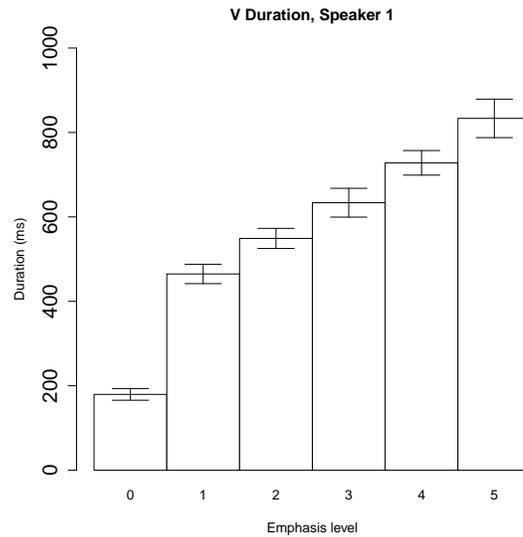


Figure 4: Results for Speaker 1. $r = .75$, regression coefficient = 92 ms. (Error bars here and throughout represent 95% CI.)

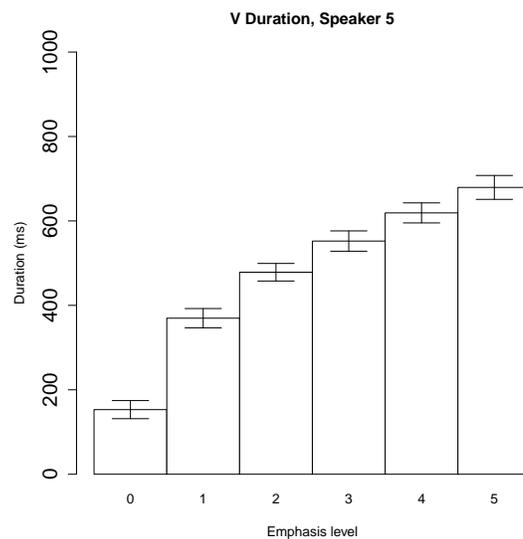


Figure 5: Results for Speaker 5. $r = .73$, regression coefficient = 76 ms.

Speakers 5 and 6 perform almost as well as Speaker 1, illustrated in Figures 5 and 6, respectively. Like Speaker 1, they have a large gap between the no-emphasis condition and the first level of emphasis, and duration increases along with emphasis level. The Pearson r -values for the emphatic conditions are $r = 0.73$

(Speaker 5) and $r = 0.72$ (Speaker 6), which are both significant at the $p < .001$ level. The regression coefficients are 76 ms and 65 ms, respectively. We note that in Figure 6, the error bars for the level 4 and level 5 conditions overlap, suggesting that while Speaker 5 clearly differentiates 6 levels of emphasis, Speaker 6 only produces 5 levels of emphasis. This is reflected in the results of non-paired t -tests between each successive emphasis level—for Speaker 5, every comparison are significant to at least the $p < 0.01$ level, whereas for Speaker 6 the comparison between levels 4 and 5 does not rise to significance (see Table 3 in the appendix).

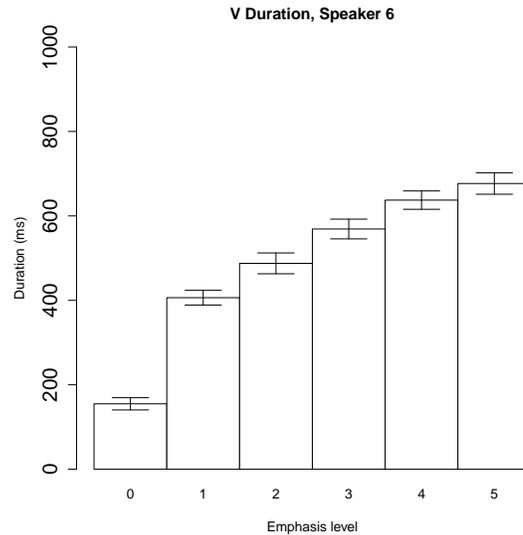


Figure 6: Results for Speaker 6. $r = .72$, regression coefficient = 69 ms.

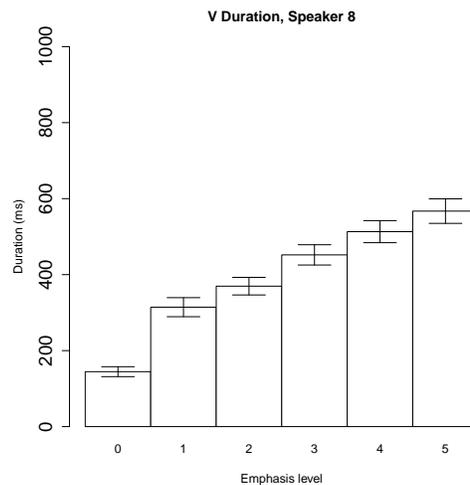


Figure 7: Results for Speaker 8. $r = .63$, regression coefficient = 65 ms.

Turning now to Speaker 8, illustrated in Figure 7, we again see a steady increase in duration as emphasis level increases. This speaker's Pearson r -value is $r = .63$ ($p < .001$), and her regression coefficient is 65 ms. Speaker 8 has a large jump in duration between the no emphasis and level 1 emphasis conditions. Like Speaker 6, Speaker 8 has a relatively small difference in duration between emphasis levels 4 and 5 (in Figure 7, the error bars for these two conditions overlap). This situation is again reflected in the non-paired t -tests (Table 3 in the appendix)—all comparisons are significant to at least the $p < .01$ level except for the

comparison between levels 4 and 5.

Speaker 3 and Speaker 2 have the next highest r -values, with $r = .44$ and $r = .42$ (both $p < .001$), respectively. The regression analysis indicates that for Speaker 3, each emphasis level results in an additional 60 ms of duration, and 52 ms for Speaker 2. Both of these speakers show relatively little change in duration between emphasis levels 2, 3, and 4 (the error bars overlap in these conditions for Speaker 3 in Figure 8 and for Speaker 2 in Figure 9). The paired t -tests (Table 3 in the appendix) support this observation—for both speakers all comparisons are significant to at least $p < .01$ except for the comparisons between levels 2–3 and 3–4.

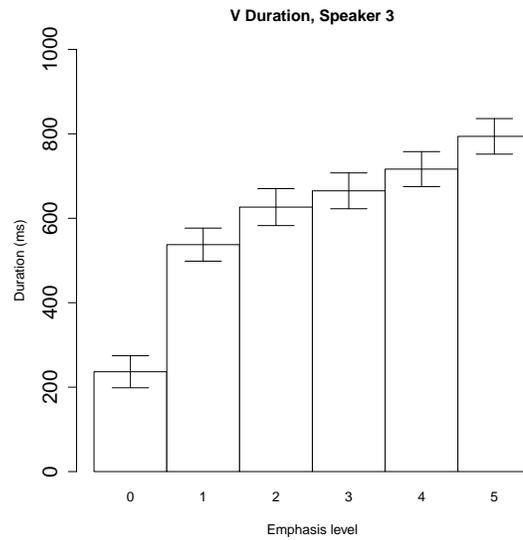


Figure 8: Results for Speaker 3. $r = .44$, regression coefficient = 60 ms.

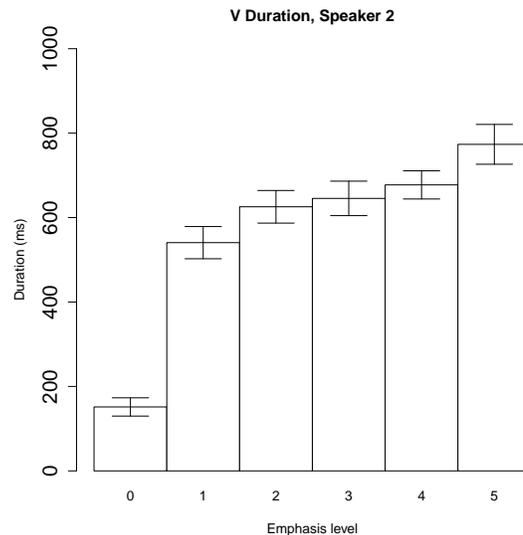


Figure 9: Results for Speaker 2. $r = .42$, regression coefficient = 52 ms.

Finally, Speakers 4 and 7 both show a positive correlation between emphasis level and duration ($r = .38$ for Speaker 4 and $r = .35$ for Speaker 7; both $p < 0.001$). However, as can be seen in Figures 10 and 11 respectively, the duration differences between each level of emphasis are quite small—error bars overlap

between almost every pair of conditions. This is reflected in the relatively small regression coefficients for these speakers: 22 ms and 26 ms, respectively. The biggest duration difference for both speakers is between the no emphasis condition and level 1 emphasis—they clearly differentiate *no* emphasis from *some* emphasis, but more fine-grained degrees of emphasis are not clearly reflected in the duration measurements. The non-paired *t*-tests (Table 3 in the appendix) show a similar pattern: Speaker 4 shows a significant difference between emphasis levels 0 and 1 ($p < .001$), but no other comparisons show significance. For Speaker 7, the comparisons between levels 0 vs. 1, and 1 vs. 2 are significant ($p < .001$ and $p < .01$, respectively), but no other comparisons are significantly different.

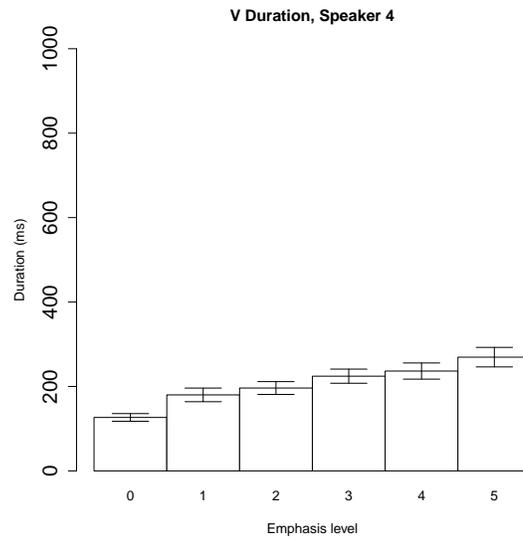


Figure 10: Results for Speaker 4. $r = .38$, regression coefficient = 22 ms.

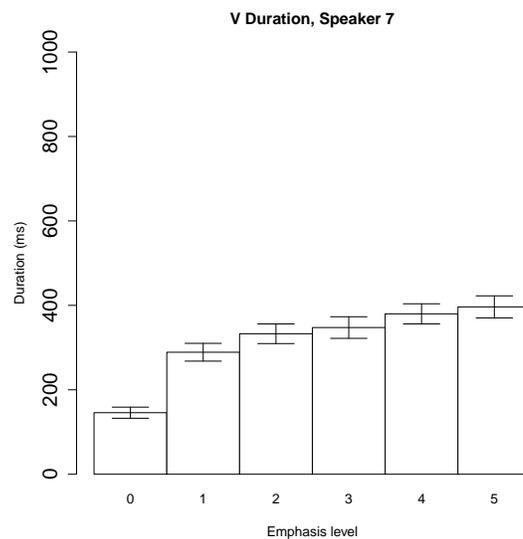


Figure 11: Results for Speaker 7. $r = .35$, regression coefficient = 26 ms.

2.2.2 Summary Table 2 gives a summary of each speaker’s data. For each speaker, we provide an r -value, regression coefficient, and as a measure of speakers’ duration range, the maximum token duration.

Table 3 in the appendix provides the results of t -tests on comparisons between each pair of consecutive emphasis levels for each speaker.

Speaker	r	Coefficient (ms)	Max duration (ms)
1	0.75	92	1265
5	0.73	76	1020
6	0.72	69	962
8	0.63	65	876
3	0.44	60	1347
2	0.42	52	1427
4	0.38	22	702
7	0.35	26	803

Table 2: Speakers' r -values, regression coefficients, and maximum vowel/rhyme duration.

The regression analysis shows that all speakers have a significant ($p < .001$) positive correlation between duration and emphasis levels 1–5. Inter-speaker variability is also evident: correlations range from $r = 0.75$ (Speaker 1) to $r = 0.35$ (Speaker 7). All speakers show a large jump in duration from the no emphasis condition to the level 1 emphasis condition, with smaller duration gaps between other emphasis levels.

Speakers 1, 5, and 6 show a clear six-way durational contrast without much overlap in error bars (see Figures 4, 5, and 6). While other speakers do not show these distinctions quite as clearly, they do show a mostly steady linear increase in duration as emphasis level increases. Overall there are no evident significant reversals—i.e., no speaker produces shorter durations as emphasis level increases. Speaker 7, and to a lesser extent Speaker 4, show an (almost) binary contrast between the no emphasis condition on the one hand, and all other emphasis conditions on the other—error bars overlap between all conditions except no emphasis and level 1 emphasis.

In Table 2 we note that there is a weak correlation between r -value and overall maximum duration: Speakers 1 and 5 have high maximum durations and high r -values, whereas Speakers 4 and 7 have low r -values and maximum durations. This association is not perfect—Speakers 3 and 2 have the highest maximum durations, but relatively low r -values.

3 Discussion

3.1 Summary The current study provides a first experimental description of the emphatic vowel lengthening process in English. In spite of some inter-speaker variability, several speakers produce a six-way durational distinction. Among those speakers who fail to produce such a fine-grained distinction, all except Speakers 4 and 7 produce distinctions more fine-grained than simply binary emphasis/no emphasis. A general implication of this study is that English speakers possess the articulatory control to produce duration distinctions more fine-grained than short/long.

3.2 Effect of a native binary vowel length contrast A question posed in the introduction is whether native language plays a role in the ability to produce fine-grained duration distinctions. Evidence from emphatic lengthening in Japanese suggests that at least some Japanese speakers are able to make 6-way duration contrasts (Kawahara & Braver, 2013, 2014). These speakers have a potential advantage over speakers of English, since Japanese makes use of length in a lexical short/long contrast. English does not have a contrastive length distinction, yet the speakers in the experiment presented here perform similarly to speakers of Japanese—they are able to produce fine-grained durational distinctions. We argue that a native binary vowel length contrast, therefore, is not necessary for the production of ternary or greater length distinctions—a claim which should ideally be tested in languages beyond Japanese and English.

3.3 Why are lexical contrasts binary, and other further questions In light of the results of this experiment, a question that arises is why natural languages are generally restricted to two-way length contrasts. We noted the possibility in the introduction that binary contrasts are preferred due to restrictions on speakers' ability to produce or perceive more fine-grained distinctions. Given that both English speakers

(in this experiment), and Japanese speakers (Kawahara & Braver, 2013, 2014) can produce up to six-way duration distinctions, we argue that the restriction does not lie in production.

A more tenable solution is that three-way or greater durational contrasts are difficult to perceive without ambiguity (see, e.g., Flemming, 1995, 2001, 2004; Liljencrants & Lindblom, 1972; Lindblom, 1986; see especially Engstrand & Krull, 1994; Podesva, 2000; Kawahara, 2012 for the grammatical imperatives on perceptual dispersion in durational contrasts). This perception-restriction hypothesis is currently being explored in a forthcoming experiment in which English speakers are asked to determine the degree of emphasis of tokens from the experiment presented above.

3.4 *Against the ‘counting hypothesis’* A possible alternative analysis of our results suggests that speakers do not show a linguistic ability to produce fine-grained duration distinctions *per se*, but rather are simply counting the number of vowels in the orthographic representation they were shown. For example, a speaker could count that *soooo* has four (orthographic) vowels, and could then count four beats during production. We argue that this is not the strategy actually employed by the speakers in this study.

All speakers in this experiment (and indeed, all speakers in both Kawahara & Braver, 2013 and Kawahara & Braver, 2014) show a large duration jump from the non-emphatic condition to level 1 emphasis—almost universally a larger duration gap than between any two other conditions. If speakers were simply counting, we should expect a uniformly linear correlation between duration and emphasis level. In fact, we see that speakers treat the distinction between level 0 emphasis and level 1 emphasis differently than they treat other distinctions.

We suggest that English speakers, like Japanese speakers, overall make a binary distinction between non-emphatic and emphatic vowels, and that within the category of emphatic vowels, speakers vary their productions to express further degrees of emphasis. This categorical difference between emphasis levels 0 and 1 should therefore be expected to be qualitatively different from the more fine-grained distinctions between levels 1 and 5 which do not cross a category boundary.

Appendix

Speaker	Comparison	mean diff. (ms.)	t(df)	p
1	level 0 vs. level 1	285.27	$t(87.32) = 21.32$	$p < .001$
	level 1 vs. level 2	84.17	$t(107.96) = 5.09$	$p < .001$
	level 2 vs. level 3	84.53	$t(86.69) = 4.08$	$p < .001$
	level 3 vs. level 4	94.46	$t(96.22) = 4.23$	$p < .001$
	level 4 vs. level 5	105.35	$t(91.80) = 3.90$	$p < .001$
2	level 0 vs. level 1	389.16	$t(96.52) = 17.75$	$p < .001$
	level 1 vs. level 2	84.92	$t(123.00) = 3.13$	$p < .01$
	level 2 vs. level 3	19.88	$t(123.63) = 0.71$	<i>n.s.</i>
	level 3 vs. level 4	32.00	$t(117.45) = 1.21$	<i>n.s.</i>
	level 4 vs. level 5	96.04	$t(110.51) = 3.32$	$p < .01$
3	level 0 vs. level 1	301.08	$t(135.96) = 11.02$	$p < .001$
	level 1 vs. level 2	88.91	$t(135.13) = 3.02$	$p < .01$
	level 2 vs. level 3	38.54	$t(136.80) = 1.26$	<i>n.s.</i>
	level 3 vs. level 4	51.45	$t(137.92) = 1.73$	<i>n.s.</i>
	level 4 vs. level 5	77.55	$t(137.96) = 2.62$	$p < .01$
4	level 0 vs. level 1	53.44	$t(107.48) = 5.75$	$p < .001$
	level 1 vs. level 2	16.26	$t(136.36) = 1.46$	<i>n.s.</i>
	level 2 vs. level 3	28.07	$t(134.37) = 2.48$	<i>n.s.</i>
	level 3 vs. level 4	12.13	$t(126.63) = 0.95$	<i>n.s.</i>
	level 4 vs. level 5	32.93	$t(128.08) = 2.19$	<i>n.s.</i>
5	level 0 vs. level 1	216.56	$t(133.69) = 13.79$	$p < .001$
	level 1 vs. level 2	108.88	$t(135.62) = 7.03$	$p < .001$
	level 2 vs. level 3	73.84	$t(135.21) = 4.63$	$p < .001$
	level 3 vs. level 4	66.69	$t(137.97) = 3.94$	$p < .001$
	level 4 vs. level 5	60.34	$t(132.56) = 3.26$	$p < .01$
6	level 0 vs. level 1	251.38	$t(133.48) = 22.13$	$p < .001$
	level 1 vs. level 2	81.30	$t(124.15) = 5.37$	$p < .001$
	level 2 vs. level 3	81.37	$t(137.60) = 4.77$	$p < .001$
	level 3 vs. level 4	68.42	$t(137.26) = 4.28$	$p < .001$
	level 4 vs. level 5	39.34	$t(134.72) = 2.35$	<i>n.s.</i>
7	level 0 vs. level 1	143.19	$t(115.54) = 11.5$	$p < .001$
	level 1 vs. level 2	43.76	$t(136.67) = 2.78$	$p < .01$
	level 2 vs. level 3	14.70	$t(136.75) = 0.85$	<i>n.s.</i>
	level 3 vs. level 4	32.47	$t(137.32) = 1.85$	<i>n.s.</i>
	level 4 vs. level 5	16.49	$t(136.85) = 0.93$	<i>n.s.</i>
8	level 0 vs. level 1	169.74	$t(103.80) = 11.98$	$p < .001$
	level 1 vs. level 2	55.32	$t(137.14) = 3.23$	$p < .01$
	level 2 vs. level 3	82.30	$t(135.34) = 4.65$	$p < .001$
	level 3 vs. level 4	61.13	$t(137.03) = 3.10$	$p < .01$
	level 4 vs. level 5	54.10	$t(136.46) = 2.49$	<i>n.s.</i>

Table 3: Non-paired *t*-tests between each emphasis level for all speakers, showing the effect of emphasis level on duration. $\alpha = .01$ after Bonferroni adjustment (each speaker has five comparisons; $0.05/5$).

References

- Aizawa, Yoshiko (1985). Intensification by so-called “choked sounds”—long consonants—in Japanese. *The study of sounds 21*, The Phonetic Society of Japan, 313–324.
- Audacity Team (2008). Audacity v. 1.2.6. Computer program, URL <http://audacity.sourceforge.net>.
- Boersma, Paul & David Weenink (2009). Praat: Doing phonetics by computer. Computer program, URL <http://www.praat.org>.
- Cedrus Corporation (2010). Superlab v. 4.5. Computer program.
- Chen, Matthew (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22:129–159.
- Engstrand, Olle & Diana Krull (1994). Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion. *Phonetica* 51, 80–91.
- Flemming, Edward (1995). *Auditory Representations in Phonology*. Doctoral dissertation, UCLA.
- Flemming, E. (2001). *Auditory Representations in Phonology*. Garland Press, New York.
- Flemming, Edward (2004). Contrast and perceptual distinctiveness. Hayes, B., R. Kirchner & D. Steriade (eds.), *Phonetically Based Phonology*, Cambridge University Press, Cambridge, 232–276.
- Higuchi, Marii & Shosuke Haraguchi (2006). Final lengthening in Japanese. *On-in Kenkyu [Phonological Studies]* 9, 9–16.
- Hoogshagen, Searle (1959). Three contrastive vowel lengths in Mixe. *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft* 12, 111–115.
- Jany, Carmen (2006). Vowel length and phonation contrasts in Chuxnabán Mixe. *Santa Barbara Papers in Linguistics 18: Proceedings from the 9th Annual Workshop on Native American Language*.
- Jany, Carmen (2007). Phonemic vs. phonetic correlates of vowel length in Chuxnabán Mixe. *Proceedings of Berkeley Linguistic Society 33S: Languages of Mexico and Central America*, Berkeley Linguistics Society, Berkeley.
- Kawahara, Shigeto (2006). Mimetic gemination in Japanese: A challenge for Evolutionary Phonology. *Theoretical Linguistics* 32:1, 411–424.
- Kawahara, Shigeto (2012). Amplitude changes facilitate categorization and discrimination of length contrasts. *IEICE Technical Report*, The Institute of Electronics, Information, and Communication Engineers, vol. 112, 67–72.
- Kawahara, Shigeto (2013). Emphatic gemination in Japanese mimetic words: A wug-test with auditory stimuli. *Language Sciences* 40, 24–35.
- Kawahara, Shigeto & Aaron Braver (2013). The phonetics of emphatic vowel lengthening in Japanese. *Open Journal of Modern Linguistics* 3:2, 141–148.
- Kawahara, Shigeto & Aaron Braver (2014). Durational properties of emphatic consonants in Japanese. *Journal of the International Phonetic Association* 44:3.
- Ladefoged, P. & I. Maddieson (1996). *The Sounds of the World's Languages*. Blackwell Publishers, Oxford.
- Lehiste, Ilse (1970). *Suprasegmentals*. MIT Press, Cambridge.
- Liljencrants, Johan & Björn Lindblom (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48, 839–862.
- Lindblom, Björn (1986). Phonetic universals in vowel systems. Ohala, John & Jeri Jaeger (eds.), *Experimental Phonology*, Academic Press, Orlando, 13–44.
- Nasu, Akio (1999). Chouhukukei onomatope no kyouchou keitai to yuuhousei [Emphatic forms of reduplicative mimetics and markedness]. *Nihongo/Nihon Bunka Kenkyuu [Japan/Japanese Culture]* 9, 13–25.
- Podesva, Robert (2000). Constraints on geminates in Burmese and Saleyarese. Bilerey-Mosier, Roger & Brook Danielle Lillehaugen (eds.), *Proceedings of West Coast Conference on Formal Linguistics*, Cascadilla Press, Somerville, vol. 19, 343–356.
- Prince, Alan (1980). A metrical theory for Estonian quantity. *Linguistic Inquiry* 11, 511–562.
- Thomas, Kimberly D. & Alan Shaterian (1990). Vowel length and pitch in Yavapai. Langdon, Margaret (ed.), *Papers from the 1990 Hokan-Penutian Languages Workshop*, SIU, Carbondale, 144–153.