RELATIVE ROLES OF THREE SUPRASEGMENTAL PARAMETERS IN PERCEIVED DEGREES OF FOREIGN ACCENT IN JAPANESE

Yukari Hirata[1] and Hiroaki Kato[2]

[1] Dept. of East Asian Languages and Literatures, Colgate University
[2] Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology
[1] yhirata@colgate.edu [2] kato.hiroaki@nict.go.jp

ABSTRACT

This study examined perceived degrees of foreign accent using speech of second language learners and native speakers morphed to each other in terms of three suprasegmental parameters. Results indicated that duration and f0, but not intensity, are central to accuracy and accentedness of a simple Japanese sentence. The effects of these two parameters still existed even when we evaluate only 'accurate' utterances. Analysis on individual learners suggested that the relative roles of duration and f0 depended on the state of the learners original utterances.

The learners' stimuli modified and matched to the three parameters of native speakers were still rated as more accented than the original native stimuli. In addition, the native stimuli matched to all of the learners' acoustic parameters were still rated as less accented than the original learners' stimuli. These results indicate that segmental aspects of learners' speech significantly contribute to perceived accent as well.

Keywords: duration, fundamental frequency, intensity, morphing

1. INTRODUCTION

Many studies have examined the relative roles that different acoustic properties of speech play in listeners' judgments on intelligibility and foreign accents, comparing roles of, e.g., segments vs. suprasegmentals, or duration vs. fundamental frequency (f0) [2, 3, 5, 6]. These studies showed that various factors such as target languages, listeners and speakers' linguistic backgrounds, and kinds of stimulus manipulation affect listeners' judgments.

For Japanese, native Japanese (NJ) listeners gave 'less native-like' responses when duration was incorrect than f0 was in [2], but duration and f0 equally contributed to accentedness rating in [3]. The present study explored this issue further by examining the degrees to which duration, f0, and intensity contribute to perceived degrees of foreign accent when a simple Japanese sentence produced by intermediate learners of Japanese was modified to that of NJ speakers in those three acoustic dimensions by way of morphing [4]. First, which parameter and which combination of parameters improve native listeners' accent rating? Second, how does accent rating differ when the learners' sentence includes phonologically 'inaccurate' pronunciation vs. when it includes only 'accurate' pronunciation? Finally, does the accentedness reach the native level when these three dimensions were modified, or were there other acoustic dimensions such as segmentals that contribute to foreign accent in a short sentence?

2. METHOD

2.1. Participants

(a) 7 native English speakers (e-M1, e-M2, e-M3, e-F1, e-F2, e-F3, e-F4) ("M" for male; "F" for female): They had studied Japanese for two years in the U.S.A. (ages: 19-21).
(b) 2 NJ speakers (j-M1, j-F1): They were from [1], who read a variety of words embedded in a carrier sentence. A subset of their recording was taken as the native speaker base of the present experiment.
(c) 18 NJ listeners: Recruited in the Kansai region of Japan for 'accent' rating. They had no knowledge of phonetics or phonology, and had little contact with foreign accented speech.
(d) 4 NJ listeners: Recruited in the Kansai region of Japan for 'accuracy' judgments. They are trained in Japanese phonetics and phonology.

2.2. Stimuli

A stimulus base was /soko wa kako to jomimasu/ 'That part is read as *past*.') and the same sentence read with /kakko/ 'parenthesis' instead of /kako/ 'past'. Each participant in groups (a) and (b) above read each sentence twice.

Using a speech morpher, STRAIGHT [4], three acoustic parameters of duration, f0, and intensity in the learners' base stimuli (e) were modified to completely match with those of the NJ speaker (j) in the corresponding gender. Including the various combinations of the three acoustic parameters, eight

variations were created per base stimulus as a result of morphing: U (unmodified); D (duration modified); F (f0 modified); I (intensity modified); DF (both duration & f0 modified); DI (both duration & intensity modified); FI (both f0 & intensity modified); DFI (all three parameters modified).

Similarly, the three acoustic parameters of the two native speaker base stimuli (j) were modified to completely match with those of the gender-matched learners. Thus, j-M1 was morphed to e-M1, e-M2, and e-M3, and j-F1 morphed to e-F1, e-F2, e-F3, and e-F4.

As a result of the above morphing, a total of 448 stimuli were created:

SET 1: KAKO sentence

- 7 learners x 8 modifications x 2 repetitions* = 112
- 1 male NJ x 8 modifications x 3 learners to morph to x 2 reps = 48
- 1 female NJ x 8 modifications x 4 learners to morph to x 2 reps = 64

(* Two different tokens produced by a learner are each morphed to two different tokens of an NJ speaker.)

SET 2: KAKKO sentence included the identical stimulus structure as SET 1.

The stimuli were randomized within each set and broken into 6 blocks.

2.3. Procedure

2.3.1. Accent rating

Eighteen NJ listeners, participant group (c) in 2.1, rated accentedness of each stimulus on a Likert scale ranging from 1 (not at all native-like) to 7 (native-like). They were given each stimulus only once, and given 4 seconds to respond to each. Set 1 (/kako/ sentence) and Set 2 (/kakko/ sentence) were given separately with a break in between, and the sentence was written in Japanese on the answer sheet for each set so that they knew what the utterance should be. The order of the two sets was randomized across participants.

2.3.2. Accuracy judgments

Four NJ listeners, participant group (d) in 2.1, were given evaluation sheets in which the sentence was spelled out for each stimulus e.g., "[ ] so ko wa ka ko to jo mi ma su." They marked "○" or "x" in the square bracket for the overall correct or incorrect pronunciation, respectively. For incorrect responses, they marked the location and wrote what was heard. For example, /kako/ with HL (high-low) pitch accent could be heard as /kakkoo/, the pitch accent of HH, or vowel /o/ missing, and the initial word /soko/ with LH could be heard as HL or /soku/.

2.4. Analyses

For accent rating, mean scores for 8 stimulus modification types (U, D, F, I, DF, DI, FI, DFI) were calculated for e-data (including both kako & kakko sets) and separately for j-data. Paired-sample t-tests made the following 13 comparisons (with a criterion for significance, $p = 0.005/13 = 0.00385$):

- (1) U&D, (2) U&F, (3) U&I: to examine whether one-parameter modification improves the originals on accent rating
- (4) D&F: to see whether one parameter is more influential than the other [2, 3]
- (5) D&DI, (6) D&DF, (7) F&FI, (8) F&DF, (9) I&DI, (10) I&FI: to see whether an addition to one-parameter modification improves accent rating
- (11) DF&DFI, (12) DI&DFI, (13) FI&DFI: to see whether 3-parameter modification adds to 2-parameter modification

For accuracy judgments, the mean number of 'accurate' and 'inaccurate' judgments for each stimulus was calculated. In addition, further details were compiled for the stimulus inaccuracy due to (1) timing, (2) pitch, and (3) other, separately for the focus word /kako kakko/ and for other parts of the sentence.

In order to separate accuracy from the overall accentedness, final analyses were conducted with the tokens that were judged as 'accurate' by 3 or 4 evaluators. Mean accent scores of these accurate tokens were calculated for each of the 8 modification types (U, D, F, I …), and paired sample t-tests were performed.

3. RESULTS

3.1. Accent rating for all data

Fig. 1 shows the accent rating given to all stimuli by all listeners. The left panel shows e-data (e.g., U as the learners' original tokens and DFI as their three acoustic parameters morphed to the native speakers'). Paired sample t-tests (Table 1) showed that the mean scores were highest (most native-like) for DFI & DF, followed by DI=D > FI=F > I=U, showing significant roles of duration, f0, and the combination of the two, and an ignorable role of intensity.

The right panel of Fig. 1 shows j-data (e.g., U as the NJ speakers' original tokens and DFI as their three parameters morphed to the learners'). Paired sample t-tests showed the results identical to those of e-data (Table 1), except that the differences were

in the opposite direction: The mean scores were highest for U & I, followed by FI=F > DI=D > DFI=DF, showing that applying the learners' duration and f0 (but not intensity) to the NJ speakers' had a detrimental role in accent rating.

The mean score of e-DFI (4.59) was found to be still significantly lower than that of j-U (6.80) even though these two categories contained the three identical suprasegmental properties of NJs [$p<0.001$]. Similarly, e-U (2.38) was still significantly lower than j-DFI (3.39) [$p<0.001$]. These differences indicate that acoustic properties other than these suprasegmentals such as individual consonants and vowels contributed to the accentedness.

Figure 1: Mean and standard errors of accent rating (7=Native-like; 1=least native-like) for all of E-data (left) and J-data (right).

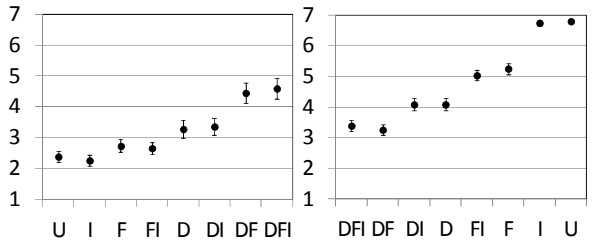
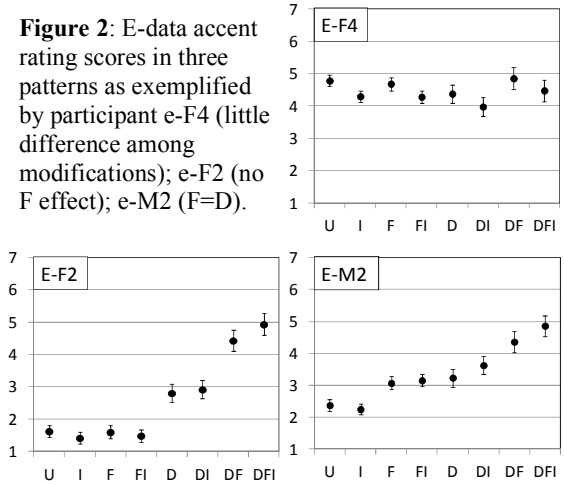


Table 1: E-data significant differences on accent ratings. Asterisks indicate $p<0.00385$; "ns" indicates $p>0.00385$. J-data showed identical results, though the direction of differences was opposite as in Fig. 1.

	Acoustic Parameter Modification						
	I	F	D	FI	DI	DF	DFI
U	ns	*	*				
I				*	*		
F			*	ns		*	
D					ns	*	
FI							*
DI							*
DF							ns

Figure 2: E-data accent rating scores in three patterns as exemplified by participant e-F4 (little difference among modifications); e-F2 (no F effect); e-M2 (F=D).



When we examined the individual learners separately, we identified three patterns (Fig. 2): (1) {e-F4} (top right) had the highest mean rating for her originals and none of the modifications improved the scores, (2) {e-F2, e-M1, e-M3} (bottom left) had little f0 but a large duration effect, and (3) {e-F1, e-F3, e-M2} (bottom right) showed the equal degree of duration and f0 effects. The key to understanding these results is shown in 3.2.

Table 2: E-tokens in % judged as accurate by 3 or 4 evaluators.

	Learners							Total
	e-F2	e-M1	e-M3	e-F1	e-F3	e-M2	e-F4	
U	0	0	0	0	0	25	75	14.3
I	0	0	0	0	0	25	75	14.3
F	0	0	0	50	75	50	75	35.7
D	0	100	100	0	0	25	100	46.4
FI	0	0	0	50	75	50	100	39.3
DI	0	100	100	0	50	25	100	53.6
DF	100	100	100	100	100	100	100	100.0
DFI	100	100	100	100	100	100	100	100.0
Total	25.0	50.0	50.0	37.5	50.0	50.0	90.6	

Table 3: J-tokens in % judged as accurate by 3 or 4 evaluators.

	Learners that J-tokens were modified to							Total
	e-F2	e-M1	e-M3	e-F1	e-F3	e-M2	e-F4	
U	100	100	100	100	100	100	100	100.0
I	100	100	100	100	100	100	100	100.0
F	25	100	100	0	0	25	100	50.0
D	0	0	0	50	100	50	100	42.9
FI	0	100	100	0	25	50	100	53.6
DI	0	0	0	50	100	50	100	42.9
DF	0	0	0	0	50	25	75	21.4
DFI	0	0	0	0	25	25	75	17.9
Total	28.1	50.0	50.0	37.5	62.5	53.1	93.8	

Table 4: Mean numbers of 'inaccurate' evaluations in terms of timing, pitch, and other properties of speech, given to the learners' original four tokens (4=all 4 evaluators judged all 4 tokens as 'accurate').

Property of inaccuracy	Inaccuracy location	Learners						
		e-F2	e-M1	e-M3	e-F1	e-F3	e-M2	e-F4
Timing	kako/kakko	4	4	3.5	1	0.75	2	3
	other words	1.5	0.25	0	0	0	0	0
Pitch	kako/kakko	1.75	0	0	3.75	3.25	2.5	0
	other words	0.25	0.25	0	3	1.25	2.5	0
Other	kako/kakko	0	0	0	0	0	0.5	0
(e.g., segments)	other words	0	0.5	0	0	0	0	0

3.2. Accuracy judgments

Tables 2 & 3 show the proportion of tokens judged as accurate by at least 3 out of the total 4 evaluators for e- and j-tokens, respectively. Table 4 shows which acoustic/phonetic properties of the learners' original tokens were judged as inaccurate. Table 4 indicates that learners e-F2, e-M1, and e-M3 were more inaccurate in their original timing than in pitch, whereas e-F1, e-F3, and e-M2 were more inaccurate

in their original pitch than in timing. This grouping corresponds well with the pattern of results in Fig. 2.

3.3. Accent rating on accurate tokens

As a final analysis, accent scores of the tokens judged as accurate by at least 3 out of the 4 evaluators (Tables 2 & 3) were compared in Fig. 3. This analysis excluded the tokens of e-U, e-I, j-DFI, and j-DF because more than 75% of the tokens were judged as inaccurate (Tables 2 & 3), thus no longer representing those categories well. Paired sample t-tests (Tables 5 & 6) showed effects of duration and f0 manipulations similar to the results of all data (3.1), particularly for j-data. This suggests that, even after separating accuracy from accent rating, there was a significant amount of contribution of duration and f0. The effects were smaller for e-data. In particular, there was no significant difference between D and F, which contrasted with the overall results in 3.1 where they showed a greater effect of duration than f0 modification.

Figure 3: Mean and standard errors of accent rating (7=most native-like; 1=least native-like) for tokens that 3 or 4 evaluators judged as accurate. E-data (left) and J-data (right).

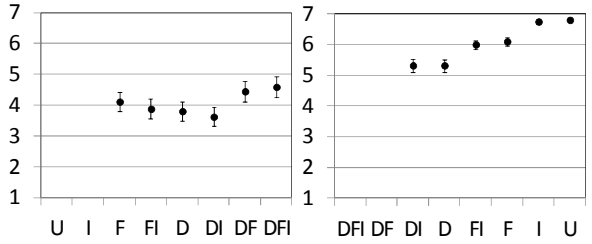


Table 5: Significant differences on accent ratings for E-data identified as accurate by 3 or 4 evaluators. * p<0.00385; ns p>0.00385

	Acoustic Parameter Modification						
	I	F	D	FI	DI	DF	DFI
U							
I							
F			ns	ns		ns	
D					ns	*	
FI							*
DI							*
DF							ns

Table 6: Significant differences on accent ratings for J-data identified as accurate by 3 or 4 evaluators. * p<0.00385; ns p>0.00385

	Acoustic Parameter Modification						
	I	F	D	FI	DI	DF	DFI
U	ns	*	*				
I				*	*		
F			*	ns			
D					ns		

4. DISCUSSION AND CONCLUSIONS

This study with learners' and native speakers' speech morphed to each other demonstrated that duration and f0, but not intensity, are central to the accuracy and accentedness of Japanese speech. Critical effects of these two parameters were present even when we evaluated only 'accurate' utterances. Effects of duration and f0 were clear not only in [3] where bilinguals imitated accented speech, but also in the present study with real learners of Japanese at an intermediate level.

This study also suggests that there were speech dimensions other than duration and f0 (such as segmentals) that contribute to accent rating significantly. The learners' stimuli modified to all of the NJ speakers' parameters of duration, f0, and intensity were still rated as more accented than the original NJ stimuli (section 3.1). There might have been a bias that signal manipulation/distortion itself might have contributed to this result. However, the NJ stimuli modified to all of the three learners' acoustic parameters were still rated as less accented than the original learners' stimuli (3.1), which shows that the learners' low scores were not simply a result of signal manipulation/distortion.

As for the relative extent to which duration and f0 played a role, a larger effect was found when the learners' duration was modified than their f0 was in the overall accentedness scores. However, analysis on individual learners suggested that the relative roles of the two parameters depended on the state of the learners original utterances. One group with inaccurate duration first had to get the duration modified in order for the effect of f0 to manifest, whereas another group with inaccurate f0 did not need correct f0 to obtain the benefit of duration modification. This primacy of duration is consistent with [2].

5. REFERENCES

[1] Hirata, Y., Whiton, J. 2005. Effects of speaking rate on the single/geminate stop distinction in Japanese. *J. Acoust. Soc. Am.* 118(3), 1647–1660.

[2] Ishihara, S. et al. 2011. What constitutes "good pronunciation" from L2 Japanese learners' and native speakers' perspectives? A perception study. *Electronic J. Foreign Language Teaching* 8(1), 277–290.

[3] Kato, S. et al. 2012. Effects of learners' language transfer on native listeners' evaluation of the prosodic naturalness of Japanese words. *Sp. Prosody,* 198–201.

[4] Kawahara, H. et al. 1999. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous frequency-based F0 extraction. *Sp. Comm.* 27(3-4), 187–207.

[5] Tajima, K. et al. 1997. Effects of temporal correction on intelligibility of foreign-accented English. *J. Phonet.* 25, 1-24.
[6] Winters, S., O'Brien, M. G. 2013. Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Sp. Comm.* 55, 486–507.