Perception of Contrastive Emphasis by American English and Japanese Listeners

Donna Erickson

Department of International Cultural Studies
Gifu City Women's College, Gifu, Japan
erickson@gifu-cwc.ac.jp

Abstract

Acoustic and articulatory measurements were made of contrastively emphasized digits in dialogs read by two American English speakers. The averaged duration, F1, F2-F1 pattern, tongue dorsum and jaw positions were significantly different for the emphasized *vs*. unemphasized digits for both speakers, but only one of the speakers showed a significant difference in peak F0. For both American and Japanese listeners, the digits best perceived as emphasized were those produced with lower jaw and more tongue dorsum movement in the direction of the phonological specification of the vowel, with acoustic correlates of longer duration and more peripheral formant frequencies, and increased F0. The results suggest that the phonetic correlates of contrastive emphasis/contrastive focus are very similar in the two languages, even though the languages have different rhythm and accent typologies.

1. Introduction

Languages vary according to their prosodic typologies—the types of accents (tone, pitch, stress accents) and types of rhythm (stress-, syllable-, mora-timed) [1]. Exploring how the native language of a listener influences perception of prosody of another language is a useful way to understand the phonetic characteristics of a language. For instance, [2] reported with regard to perception of prominence, amplitude cues overrode duration cues for American English listeners, while for Estonian listeners, duration cues were more important.

This paper examines the perception by American English and Japanese listeners of American English contrastive focus; specifically, contrastively emphasized digits in a read dialog. In terms of rhythm, English is said to be a stress-timed language (in which the syllable is the unit of stress) while Japanese is a mora-timed language. In terms of accent types, English is often said to be a stress-accent language, Japanese, a pitch-accent language (in which the weight of the syllable affects the placement of the pitch fall) [3]. In English, pitch accents are phrasally-assigned, can have a variety of phonetic shapes, and signal increased prominence; in Japanese, pitch accents are both lexically and phrasally-assigned [4] , but have only one phonetic shape (in standard Tokyo Japanese) and do not signal increased prominence [5].

Contrastive focus is used by speakers of languages to maximally differentiate a word in an utterance so the contrasting information will be more likely to be perceived by listeners. Contrastively focused words in both Japanese [5] and American English [6] are characterized by, among other things, expansion of pitch range and increased duration. In addition, both Japanese and American English show contrastive emphasis-related formant changes, such that emphasized low vowels become more compact and emphasized high vowels, more diffuse [7, 8]. These contrastive emphasis-induced formant changes were more robust in American English than in Japanese. In terms of articulation, contrastively emphasized American English vowels, regardless of vowel height, have significantly lowered jaw position, whereas for Japanese, only low- and mid- contrastively focused vowels show substantially increased jaw opening. For both English and Japanese, the tongue dorsum tends to move in the direction of the phonological specification of the vowel. A recent study by [9] with contrastive emphasis in spontaneous dialog showed that words well-perceived by American listeners as emphasized have larger jaw opening than those poorly perceived, but that the pitch accent associated with the emphasized word differs both within and across speakers.

Given the similarities and differences of prosodic characteristics in Japanese and English, we expect to find differences in the perception of contrastive emphasis by native listeners of the two languages, which may provide a window to a better understanding of the phonetic cues underlying contrastive emphasis.

2. Methods

Articulatory and acoustic data were recorded at the University of Wisconsin X-Ray Microbeam Facilities, Madison, Wisconsin from 2 American English male speakers (Midwest dialect, Wisconsin.) Question-answer sentences, like "Is it 9 5 9 Pine Street? Yes, it's 9 5 9 Pine Street" or "Is it 9 9 9 Pine Street? No, it's 9 FIVE 9 Pine Street" were randomized (10-13 repetitions) and read from a monitor screen, with the digit to be emphasized in capital letters. The target digit was either a 5 or a 9, and appeared in either initial, middle or final position of the 3 digit street address. Articulatory measurements (jaw x-y and tongue dorsum (T3) x-y positions) and LPC-Cepstrum method formant extraction using a MATLAB-based program (written by J. Dang) were made at the time of maximum jaw opening during the digit. Peak F0 and vowel duration measurements were made using WaveSurfer (www.speech.kth.se). The method of collecting and analyzing the articulatory data is reported in more detail [8].

The answer parts of only "no-sentences" with intended contrastive emphasis were presented auditorily by headphones to American (Midwest dialect, South Dakota) and Japanese (Gifu) college student listeners using a Macintosh computer and Psyscope Software. The Japanese listeners had at least six years of English education. For speaker 1, there were 76 utterances which were presented to 15 American listeners and 21 Japanese listeners; for speaker 2, there were 88 utterances presented to 14 American listeners and 13 Japanese listeners. The listeners were asked to indicate which of the three digits was "emphasized/stronger," by typing either 1, 2, or 3. Each sentence was presented twice with a gap of 1 second; the test

was self-paced. A short practice test of 6 utterances preceded the test.

3. Results

For the articulatory and acoustic analysis, we examined only the emphasized and unemphasized middle-5 digit for S1 and S2, as shown in Table 1 below. Generally, for both speakers, emphasized digits have higher F0, higher F1, lower F2, more compact F1-F2 pattern, longer duration, more back and down tongue dorsum (T3), and lower jaw (J) position. A t-test was done to test whether the values of the emphasized *vs*. unemphasized digits were significantly different. Asterisks indicate a significance of p<.01, and ! indicates p<.05. For S1 the difference in horizontal tongue dorsum and jaw positions for emphasized *vs*. unemphasized digits are not significant, and for S2, both peak F0 and F2 are not significantly different.

Table 1. *Averaged articulatory and acoustic values for middle emphasized 5-digits. "Ue" indicates "unemphasized," "E", emphasized. Negative x-values indicate more forward articulator positions, negative Jy, lower jaw positions, and positive T3y, lower tongue dorsum positions.*

S1	*F0	*F1	*F2	*F2-F1	*Dur	T3x	*T3y	Jx	*Jy
Ue	122	622	1498	876	0.15	-47	14	-1	-7
E	140	766	1359	593	0.25	-48	10	0	-9
S2	F0	*F1	F2	*F2-F1	*Dur	*T3x	*T3y	*Jx	!Jy
Ue	119	691	1409	718	0.15	-50	5	1	-6
E	124	818	1364	546	0.23	-52	4	2	-7

The results of the perception tests with Japanese and American English listeners are shown in Table 2 below.

Table 2. *Perception test results for middle digit emphasis for American and Japanese listeners.*

Speaker/Listeners	**Digit Position**		
S1	1	2	3
Am Eng listeners (N=15)	0%	100%	0%
Japanese listeners (N=21)	1%	97%	2%
S2			
Am Eng listeners (N=14)	2%	91%	7%
Japanese listeners (N=13)	8%	90%	2%

The acoustic values with perception scores for each of the middle digits are shown in Tables 3 and 4 below. JP indicates perception scores of Japanese listeners, AEP, American English listeners. Table 3 shows that for S1, American listeners perceived as emphasized the middle digits (intended to be emphasized) 100% of the time, and Japanese listeners, only the middle digits of the last 6 of the utterances were perceived 100% as emphasized.

Table 4 shows that perception of middle digit emphasis was more difficult for S2, for both American and Japanese listeners. For some of the utterances, the Japanese listeners performed better than the American listeners in identifying the middle digit as contrastively emphasized.

In order to show any underlying relationships among the acoustic and articulatory parameters and perception scores, a Pearson correlation analysis was done. For purposes of the correlation analysis, the middle-5 digits in the yes-utterances (with no intended contrastive emphasis) were assigned a 0-rating. For S1, both American English and Japanese listeners displayed a significant correlation (p<.01) between emphasis and F0, F1, F2, F2-F1, duration, and the vertical positions of tongue dorsum and jaw, while for S2, a significant correlation (p<.01) was seen for all acoustic measures, except F2 and F0; also, only for the horizontal positions of tongue dorsum and jaw, not the vertical positions, was there significant correlation.

Table 3. *S1 Acoustic values of middle-5 digit for S1. JP indicates perception results from Japanese listeners, AEP, from American English listeners.*

ID	JP	AEP	DUR	F0	F1	F2	F2-F1
72.1	91%	100%	0.22	136	711	1444	733
55.1	95%	100%	0.25	136	754	1316	562
58.2	95%	100%	0.25	140	796	1295	499
63.1	95%	100%	0.28	140	764	1337	573
70.1	95%	100%	0.27	138	785	1401	616
50.1	95%	100%	0.27	163	807	1316	509
54.1	95%	100%	0.27	126	796	1316	520
29.2	100%	100%	0.31	151	785	1327	542
51.1	100%	100%	0.21	144	754	1369	615
57.1	100%	100%	0.21	139	732	1337	605
69.1	100%	100%	0.25	138	785	1412	627
74.1	100%	100%	0.24	134	754	1316	562
75.1	100%	100%	0.23	132	732	1475	743

Table 4. *Acoustic values of middle-5 digit for S2. (Headings same as for Table 3.)*

ID	JP	AEP	DUR	F0	F1	F2	F2-F1
103.1	62%	71%	0.19	119	785	1369	584
101.1	85%	93%	0.22	128	849	1422	573
104.1	85%	93%	0.24	122	796	1401	605
105.1	85%	86%	0.25	131	860	1348	488
50.1	85%	93%	0.24	123	785	1327	541
66.1	85%	93%	0.20	131	764	1348	584
63.1	92%	100%	0.21	120	817	1412	594
102.1	100%	93%	0.23	123	807	1380	573
31.2	100%	93%	0.25	112	892	1348	456
52.1	100%	93%	0.21	126	807	1359	552
58.2	100%	93%	0.24	123	839	1295	456
68.1	100%	93%	0.23	127	817	1359	541

We also examined the F0 contours of the utterances. Fig. 1 shows the F0 contours for S1, all of whose middle digits were 100% of the time perceived as emphasized by American listeners. The pitch accents on the middle-5 digit were all H* or L+H*, except for the first one, which was L*+H. This one was least-well perceived by Japanese listeners (91%). The middle digits in the next six utterances, in which the peak F0 occurred at the end of the syllable, were perceived by Japanese listeners 95% of the time as emphasized. The peak F0 in these digits occurred at the end of the syllable. The middle digits in the final six utterances, in which peak F0

occurred at the beginning or middle of the syllable, were perceived 100% of the time as emphasized by Japanese listeners.

Fig. 2 shows the F0 contours for S2. The pitch accents were generally H*, except for the first one, which was L* pitch

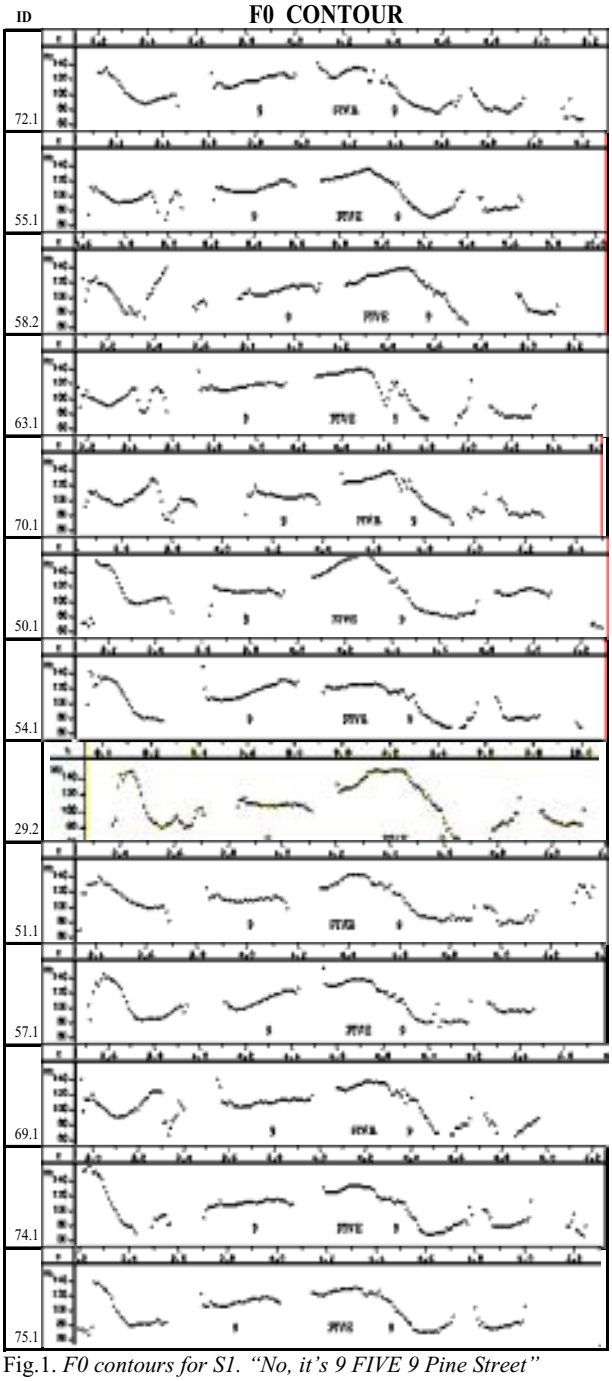

Fig.1. *F0 contours for S1. "No, it's 9 FIVE 9 Pine Street"*

accent. This one was least well-perceived by both American and Japanese listeners. The H* accents of speaker 2 compared to those of speaker 1 were phonetically different in terms of timing of the F0 peak. For speaker 1, F0 reached a peak toward the middle or end of the syllable (which is the more usual pattern for H* pitch accents); for speaker 2, the F0 peak generally occurred at the beginning of the syllable and continued downward.

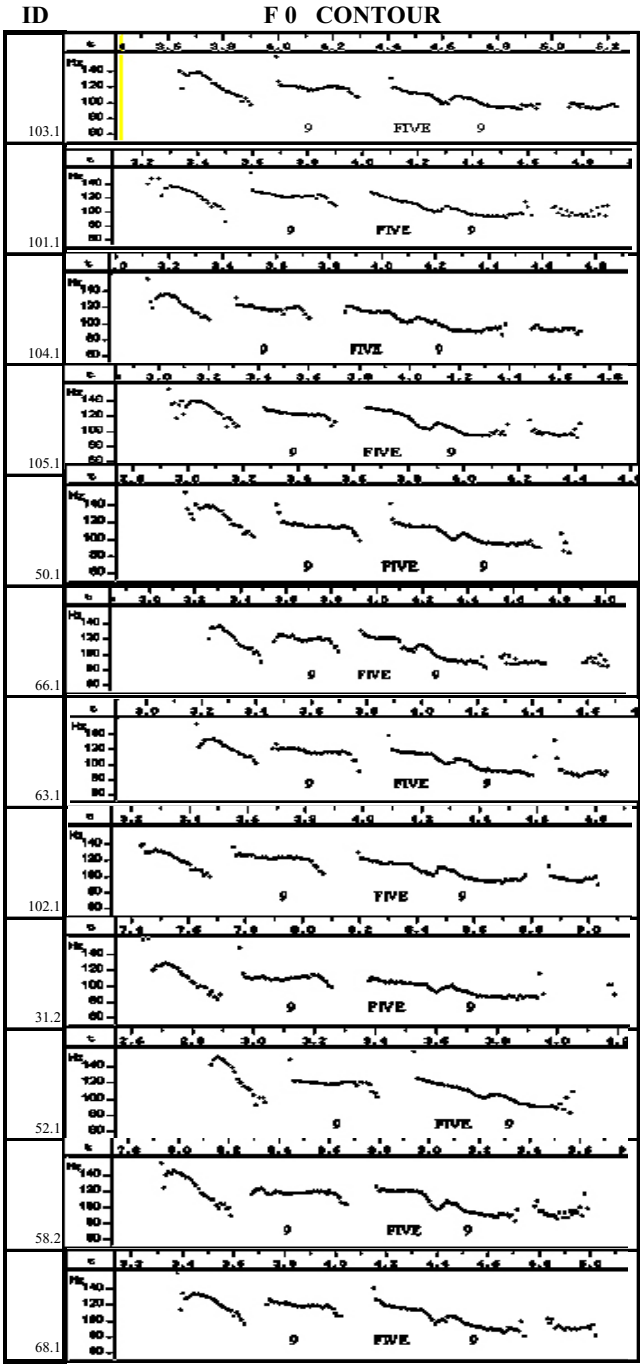

Fig.2. *F0 contours for S2. "No, it's 9 FIVE 9 Pine Street"*

4. Discussion

Generally, the middle 5-digits that were intended to be emphasized by the dialog paradigm were well perceived as emphasized by both American English and Japanese listeners. Analysis of the acoustic and articulatory measurements showed that the averaged duration, F1, F2-F1 pattern, tongue dorsum and jaw positions were significantly different for the

emphasized *vs*. the unemphasized digits for both speakers. A Pearson correlation analysis showed a significant correlation between perception of emphasis and the above measures by both Japanese and American listeners. These results show that those syllables produced with a lower jaw and more tongue dorsum movement (in the direction of the phonological specification of the vowel), and having acoustic correlates of longer duration and more peripheral formant frequencies, were well-perceived as contrastively emphasized by both American and Japanese listeners. In this sense, I suggest that an increase in syllable weight, in terms of the syllable's acoustic and articulatory make-up, signals emphasis to both American and Japanese listeners.

There were some differences in production of contrastive emphasis by the two speakers, which affected the perception by listeners. For speaker 1, peak F0 was significantly higher for emphasized *vs*. non-emphasized digits; moreover, there was a significant correlation between F0 and perception of emphasis by both groups of listeners. American listeners perceived these digits as emphasized 100% of the time; Japanese listeners, 97% of the time. It is interesting that the middle digits produced by speaker 1 that were not perceived 100% of the time as emphasized by Japanese listeners were those in which the F0 peak occurred at the end of the syllable. According to [5], it is not permissible in Japanese for a pitch accent to fall at the end of a heavy syllable, only at the middle or beginning. It may be that the timing of the pitch fall affected the perception of emphasis by a few of the Japanese listeners.

For speaker 2, however, the difference in averaged peak F0 was not significant for the emphasized *vs*. unemphasized digits, and there was no significant correlation between f0 and perception of emphasis. Moreover, both groups of listeners showed a poorer rate of perception of emphasis for speaker 2 than speaker 1—91% by American listeners and 90% for Japanese listeners. The phonetic shape of the H* pitch accent used by speaker 2 in terms of timing of the F0 peak was different from that of the H* pitch accent used by speaker 1. For speaker 1, the F0 peaked toward the middle or end of the syllable (which is the more usual pattern for H* pitch accents) whereas for speaker 2, the F0 peak generally occurred at the beginning of the syllable and continued downward. The utterance which was most poorly perceived as emphasized had L* pitch accent. The L* pitch accent and phonetic shape of the H* accent used by speaker 2 may be a dialectal variant not well used by Midwest dialect speakers of South Dakota. This needs to be explored further. It is curious that Japanese listeners tended to do better perceiving emphasis for certain of speaker 2's utterances than American listeners. For Japanese listeners, pitch accents *per se* do not signal contrastive emphasis/contrastive focus, rather it is the timing of the pitch fall within the syllable that is important. Because of this, it could be that the H* pitch accents with initial high F0 were more easily perceived as emphasized by Japanese than by some of the American listeners.

4. Conclusion

For both American and Japanese listeners the digits best perceived as emphasized were those produced with a lower jaw and more tongue dorsum movement in the direction of the phonological specification of the vowel, with acoustic correlates of longer duration and more peripheral formant frequencies, and accompanied by higher F0. Simply changing syllable weight, without increasing F0, was sufficient for both language listener groups to perceive emphasis, but with a less high rate of accuracy.

An interesting finding from this study is that that the phonetic shape of pitch accents also affected both sets of listeners in their perception of emphasis, but in different ways. For American listeners (from South Dakota) the best perceived shapes were ones in which the peak F0 occurred at the middle or end of the syllable, not at the beginning. For Japanese listeners, the digits best perceived as emphasized were ones in which F0 peak occurred at the beginning or middle of the syllable, not at the end.

An additional finding is that both American and Japanese listeners are sensitive to increased syllable weight and increased F0. It is interesting that the phonetic correlates of contrastive emphasis/contrastive focus are very similar in the two languages, even though the languages have different rhythm and accent typologies

Future work involves perception tests using more speakers, speakers from different American dialects (*e.g*., Wisconsin/Ohio) and from other languages. Experiments with synthetic speech to vary the acoustic parameters of formants, F0, duration, pitch accents, timing of F0 fall, *etc*., would be useful to better understand which cues are more salient to listeners from various language and dialect backgrounds for signaling contrastive emphasis/contrastive focus.

5. References

[1] Komatsu, M., 2003. Essay on acoustic correlates of prosodic typology. In *A New Century of Phonology and Phonological Theory, A Festschrift for Prof. Haraguchi*, T. Honma, M. Okazaki, T. Tabata & S. Tanaka (eds.) Tokyo: Kaitakusha, 492-507.

[2] Lehiste, I.; Fox, R.A., 1992. Perception of prominence by Estonian and English listeners. *Language and Speech* 35(4), 419-434.

[3] Kubozono, H., 1999. Mora and syllable. In *The Handbook of Japanese Lingusitics,* N. Tsujimura (ed.). Oxford: Blackwell, 31-61.

[4] Fujimura, O., 2003. Stress and tone revisited: Skeletal *vs*. melodic and lexical *vs*. phrasal. In *Cross-linguistic Studies of Tonal Phenomena,* S. Kaji (ed.). Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

[5] Maekawa, K., 1997. Effects of focus on duration and formant frequency in Japanese. In *Computing Prosody,* Sagisaka, Campbell, Higuchi (eds.).New York: Springer, 129-153.

[6] Cooper, W.E..; Eady, S.J.; Mueller, P.R., 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77(6), 2142-2156.

[7] Erickson, D.; Hashi, M.; Maekawa, K., 2000. Articulatory and acoustic correlates of prosodic contrasts: A comparative study of vowels in Japanese and English. *Acoustical Society of Japan, Spring Meeting,* 265-266.

[8] Erickson, D., 2002. Articulation of extreme formant patterns for emphasized vowels. *Phonetica* 59, 134-149.

[9] Menezes, C.; Erickson, D.; Fujimura, O., 2002. Contrastive emphasis: Comparison of pitch accents with syllable magnitudes. *Speech Prosody 2002, Aix-en-Provence,* 495-497.