

Why is L1 not easy to hear?

Shinobu Mizuguchi & Koichi Tateishi*

Abstract. We naively believe that L1 is easier to hear than L2. Generally, this belief is correct, but not always. Japanese contrastive focus is more challenging to identify than English focus even for L1 speakers. To account for why Japanese is hard to perceive, we first conducted production and perception experiments, to understand linguistic mechanisms. We found that Japanese lacks a part of focus effects and is an acoustically weak language contra previous studies. English, on the other hand, is an acoustically strong language and uses the F0 feature as a focus cue. We then conducted an fMRI experiment to see whether or not linguistic mechanisms for them are implemented in the brain. We found that we employ different neural networks to process English and Japanese; the right dorsolateral frontal cortex is activated to process Japanese CF, but not English CF. Japanese is a pitch language and requires processing both lexical accents and pitch contours. English, on the other hand, needs to process lexical accent only, and it activates left superior temporal gyrus, insular, and supramarginal regions, but not right dorsolateral frontal cortex. We conclude that processing burdens lead to perception difficulty, even for L1 Japanese speakers.

Keywords. focus; perception of prosody; Japanese; English; fMRI

1. Introduction. Speakers produce, and listeners perceive certain parts of an utterance as more or less prominent. To highlight prominence in an utterance, prosody – part of the acoustic information – plays an important role. Previous cross-linguistic studies on prosody (cf. Cole et al. 2010, 2019, Bauman and Winter 2018, among others) have found that prosodic cues vary among individuals and languages. Lee et al. (2015) conducted a cross-linguistic perception experiment of contrastive focus, using ten-digit numbers of the form XXX-XXX-XXXX, where one number in the series was with focus under a Question-Answer sequence as in (1). Giving them only the response, we asked participants to identify which number was focused.

- (1) A: Is Mary’s number 215-418-5623?
 B: No, the number is 215-417-5623.

Table 1 shows that focused digits, as compared with digits without a focus, exhibit significantly higher duration, intensity, and pitch in Mandarin Chinese and American English but not in Seoul Korean and Tokyo Japanese, and that languages can adopt different strategies when the speakers communicate the location of contrastive focus.

* Many thanks go to the experiment participants at Kobe University, Kobe College, and the University of Leuven. We would also like to express our gratitude to the audience of our presentation at the General Meeting of LSA 2020 in New Orleans. We most deeply thank Tim Mahrt for his analysis and discussion of our perception experiment. This study is partly supported by JSPS (the Japan Society for Promotion of Science) grant #15K02480 given to Shinobu Mizuguchi. Authors: Shinobu Mizuguchi, Kobe University (mizuguti@kobe-u.ac.jp) & Koichi Tateishi, Kobe College (tateishi@mail.kobe-c.ac.jp).

	American English	Mandarin Chinese	Seoul Korean	Tokyo Japanese
Duration	0.95	1.19	0.13	0.10
Intensity	1.25	0.36	0.24	-0.24
Pitch	2.96	3.13	0.62	0.60

Table 1: Median z-score values of focused digits (adapted from Lee et al. 2015)

Lee et al. 2015 claim that languages with strong acoustic cues show high accuracy in prominence perception: 97.3% for American English and 94.9% for Mandarin Chinese. However, for languages with weaker prosodic marking, low accuracy was obtained in the perception study: 41.6% for Seoul Korean.

This paper considers prominence perception and processing in Japanese. Japanese is a mora-timed pitch language. Acoustically, it has Accented and Unaccented words that are lexically determined. Prosodically, its utterance has a downstepping pitch contour, called ‘downtrend’ (cf. McCawley 1968). We replicated Lee et al.’s perception experiment on Japanese and English by Japanese (L1)/English (L2) (J/E, henceforth) late-onset bilinguals. Our finding was that the identification ratio of Japanese contrastive focus (86.2%) was worse than that of American English (98.6%). This fact suggests that L1 is more difficult to perceive than L2.

The finding of ours is contra the previous work on L1/L2 perception. Gandour et al. 2003 investigated the perception of sentence focus in Mandarin Chinese (L1) and English (L2). They recruited ten late-onset, medium proficiency Chinese (L1)/English (L2) (C/E, henceforth) bilinguals, and used sentence-pairs with two potential locations of sentence focus (initial and final) both in Chinese and English, recorded by a male speaker of Mandarin and English. The focus identification ratio is lower in English (86.2%) than in Chinese (95.8%). They further considered whether the neural substrates are shared or segregated in multilingualism. The whole-brain cluster analysis revealed extensive overlapping activation between Chinese and English stimuli in frontal, parietal and temporal areas. However, C/E bilinguals exhibited significantly higher bilateral activity in the anterior insula (aINS) ($F(1,9) = 12.15, p < 0.01$) and the anterior Superior Frontal Sulcus (aSFS) ($F(1,9) = 18.54, p < 0.005$) when presented with English stimuli as compared to Chinese stimuli. Gandour et al. 2003 suggest that the activity in the anterior insula is graded in response to the task difficulty in L2 and that L2 requires more extensive cortical resources than L1.

Gandour et al. 2003 claim that L2 is more difficult to perceive, while our data show that L1 is more complicated than L2. In this paper, we would like to consider why Japanese is not easy to hear even to L1 speakers, based on the linguistic and neuro-biological experiments. The structure of this paper is as follows; Section 2 is on an acoustic study on the focus perception in Japanese. Section 3 is on an fMRI study on the neural implementation of Japanese and English focus processing. Discussion and conclusion follow in Section 4 and Section 5.

2. Perception experiments of contrastive focus. We replicated Lee et al. 2015’s perception experiment of the contrastive focus of Japanese with J/E late-onset bilinguals. We also conducted a perception experiment of the contrastive focus of English with J/E bilinguals, to see the differences or similarities in perception between these two languages. We predicted that Japanese would be low in identification accuracy since Japanese has weaker acoustic cues than English (cf. Table 1).

2.1. METHOD AND MATERIALS. The method was the same as the perception experiment by Lee et al. 2015. We used Japanese tokens as in (2) and English tokens as in (1) as our materials,

recorded by a male speaker of Tokyo Japanese and a male speaker of Midwest American English.

- (2) A: Yamada-san-no bango-wa 215-418-5623-desu-ka?
 Yamada-Mr/Ms-GEN number-TOP 215-418-5623-copula-Q
 ‘Is Yamada’s number 215-418-5623?’
 B: Ie. 215-418-6623-desu.
 no -copula
 ‘No. (It) is 215-418-6623.’

Giving them only the response portion, we asked participants to identify which number was focused. They heard 30 utterances in each language twice, presented in random order. Since we conducted the two experiments on separate occasions, we had two groups of participants: one for the Japanese experiment (M2, F20, mean age 20.45, SD=0.87) and the other for the English experiment (M13, F5, mean age 20.5, SD=1). They reported no auditory difficulties, and we do not believe that different groups of participants have affected our results. We conducted the experiments in a quiet room with the audio stimuli projected from a room speaker.

2.2. RESULTS. Table 2 shows the results; the mean contrastive focus identification ratio was 86.2% in Japanese and 98.6% in English.

Task	Target word position (% accuracy)										median
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	
Japanese	100	88	91	98	92	92	65	91	82	58	86.2
English	100	100	100	100	100	98	100	98	96	94	98.6

Table 2: Accuracy in words bearing contrastive focus by J/E

Our prediction is borne out; Japanese contrastive focus is harder to identify than English contrastive focus, even though Japanese is the L1 of the participants.

We further conducted supplementary perception experiments with non-native speakers of Japanese, to see whether or not Japanese contrastive focus is difficult to hear with L3 speakers. We recruited 18 Chinese (L1)/Japanese (L3) (C/J, henceforth) late-onset bilinguals (M5, F13, mean age 25.59, SD=2.67) at a Japanese university, and 9 Dutch (L1) /Japanese (L3) (D/J, henceforth) late-onset bilinguals (M4, F5, mean age 24.44, SD=5.05) at a Belgian university. The materials and the procedure were the same as the perception experiment with J/E participants. Table 3 shows the result.

Participants	Target word position (% accuracy)										median
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	
C/J	100	98	91	91	96	98	77	95	70	89	90.5
D/J	100	96	81	100	93	100	44	85	63	93	85.5

Table 3: Accuracy in words bearing contrastive focus by C/J and D/J

Our results show that the difficulty of focus perception of Japanese is not a matter of native language, but that it is a matter of the Japanese language. Below we would like to discuss why Japanese is not easy to hear from an acoustic point of view.

2.3. DISCUSSION.

2.3.1. ACOUSTIC ANALYSIS OF CONTRASTIVE FOCUS. We have observed the higher accuracy of contrastive focus identification in English than Japanese (cf. Table 2). Lee et al. 2015 claim that languages with strong acoustic cues show high accuracy in the prominence perception (cf. Table 1). Figure 1 shows the feature values by position in the digit sequences used in our experiment. The F0 maximum and the Intensity correspond with the focus position in English, but not in Japanese.

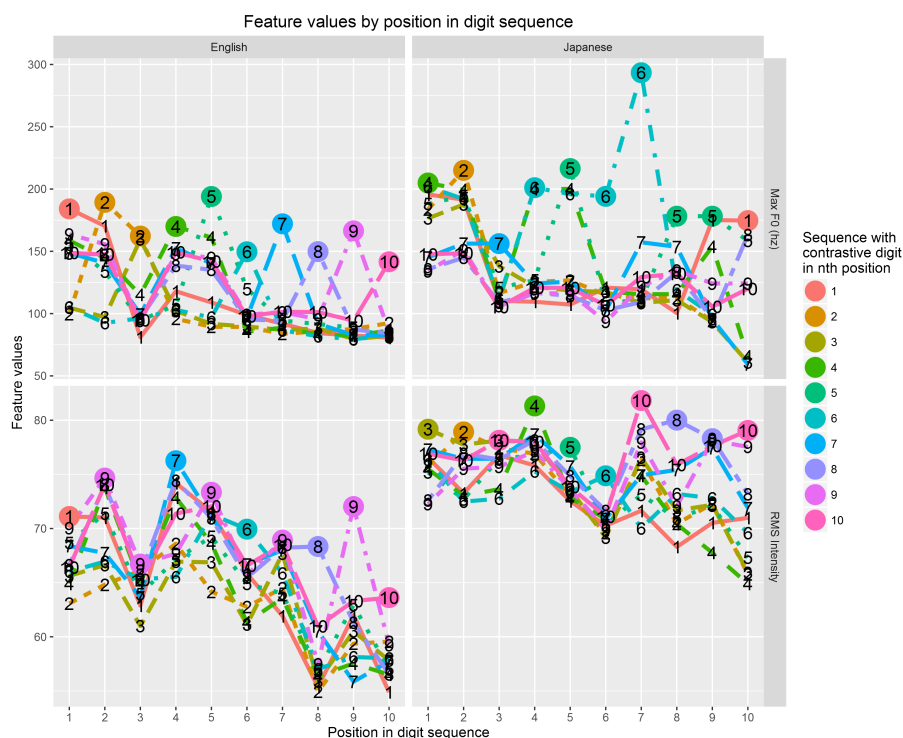


Figure 1. F0 maximum (above) and Intensity (below) by position in digit sequence

The patterns in the figure explain why English contrastive focus is easy to hear. Its acoustic cues correspond with the position of the foci, which means that the English F0 feature is ‘segmental’ and shows which segment is highlighted directly by F0. However, this is not the case with Japanese; the Japanese F0 feature does not highlight the focused segment. Below we will consider why F0 does not function as a focus cue in Japanese.

2.3.2. ACOUSTIC PROPERTIES OF JAPANESE FOCUS. Japanese is a pitch language and has Accented (A) and Unaccented (U) words, which are lexically determined, as in (3). Japanese

realizes its accent by a falling H*-L bi-tonal contour.¹

- (3) a. i'noti (H*LL) 'life' (Initial-Accented)
 b. koko'ro (LH*L) 'heart' (Medial-Accented)
 c. atama' (LHH*(-L)) 'head' (Final-Accented)
 d. sakana (LHH) 'fish' (Unaccented)

The Japanese literature on focus prosody (cf. Pierrehumbert and Beckmann 1986, Ishihara 2003, 2016, Kubozono 2007, among many others) agrees that focus on A words boosts the pitch on the accent peak and lowers the one on the following tone (cf. Figure 2).

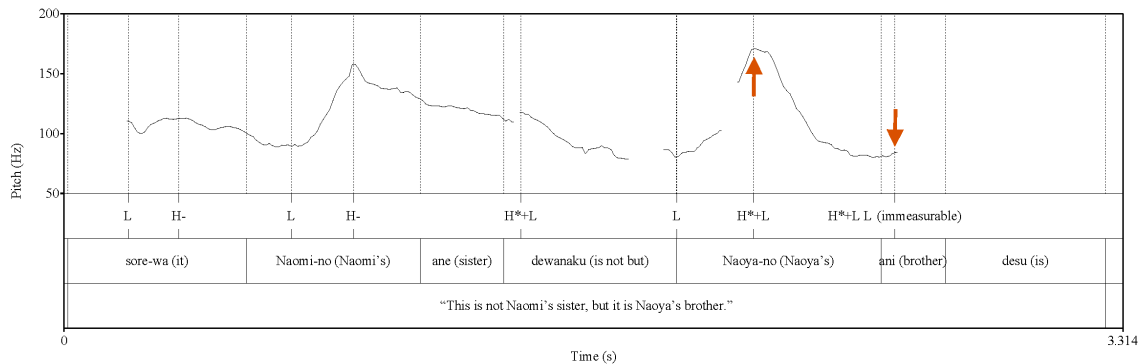


Figure 2. Sentence pitch contour, with contrastive focus on *Na'oya* and compression on *a'ni* 'brother' (from Ishihara 2016)

It is not clear what type of cue makes U words focused in the literature yet, so we conducted a production and perception experiment on two-noun sequences (cf. Mizuguchi and Tateishi 2018). We predicted that i) focus on both A and U words would be boosted, and ii) post-focal lowering would be observed in focus on A words only, due to the absence of lexical fall in U words. We also predicted that iii) A words would be easier to perceive than U words since the former is with more acoustic cues.

The materials of our production experiment consist of two nouns W1W2 of similar segmental characteristics, covering the four possible accentual sequences UU (e.g. *amai ume* 'sweet plum'), UA (e.g. *amai u'ni* 'sweet urchin'), AU (e.g. *ao 'i ume* 'blue plum'), and AA (e.g. *ao 'i u'ni* 'blue urchin'). With a [+Focus] condition, one stimuli set comes in 12 conditions: UU, U[+F]U, UU[+F], UA, U[+F]A, UA[+F], AU, A[+F]U, AU[+F], AA, A[+F]A, and AA[+F]. Two different stimuli sets were used, mixed with 24 fillers, in a pseudorandomized order. We re-recruited 8 Tokyo-Japanese speakers (F2, M6, we abandoned the data of 3 speakers due to their inconsistent use of lexical accents). We asked them to produce an answer to wh-questions like (4).

¹ For the representation of pitch accents, we use the method frequently used in theoretical studies of Japanese accent. The accent mark (') means that the immediately preceding mora is accented.

(4) a. Contrastive Focus (CF)

Q: Kore-wa aoi u'ni desu-ka?
 this-TOP blue sea-urchin COPULA-Q
 'Is this blue sea-urchin?'

A: Iie. Ao'i ume desu.
 no blue plum COPULA
 'No. (It's) blue plum.'

b. Broad Focus (BF)

Q: Kore-wa nani desu-ka?
 this-TOP what COPULA-Q
 'What is this?'

A: Ao'i ume desu.
 Blue plum COPULA
 '(It's) blue plum.'

Our materials were 48 in total, and we presented them using Microsoft PowerPoint to the participants. The sound were recorded at 44.1Kh, 16bits directly to a computer.

For the analysis, we follow Ishihara 2016 and take the normalized F0-means of the six measurement points: the 1st F0-minimum, F0-maximum, the 2nd F0-minimum of W1 (L1-1, H1, L1-2), along with those of W2(L2-1, H2, L2-2).

Figure 3 shows the result. Contra our predictions, only a few Focus effects were observed; F0-rise (the upward arrow in Figure 3 (right)) is statistically significant only in AU[+F] and UA[+F] (W=7, p=0.026 on Mann-Whitney Test for both positions). We find statistically significant Post-focal Fall (the downward arrow in Figure 3 (left)) in A[+F]U only (W=41, p=0.038 on Mann-Whitney Test).

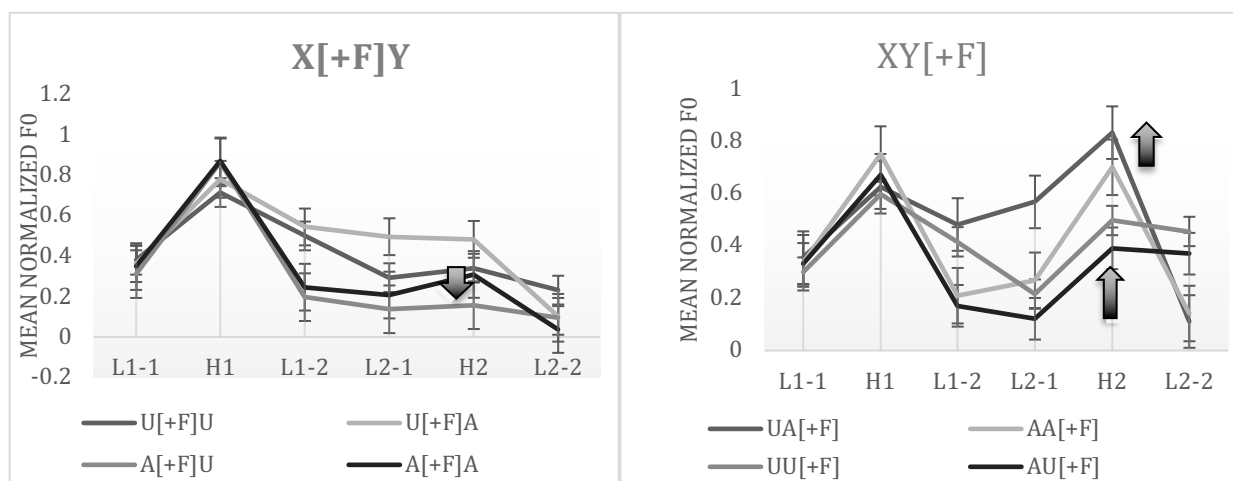


Figure 3. Mean normalized F0 of W1[+F]W2 (left) and W1W2[+F](right)

Figure 3 shows that the acoustic cues of contrastive focus in Japanese are weak, and they vary depending on the context, which leads us to predict that focus perception is not easy in Japanese.

We conducted a perception experiment of contrastive focus (CF) in Japanese, using the materials recorded in our production experiment. The subjects were 23 L1 Japanese (F12, M11, mean age 19.65). The task was to mark the words which they thought were focused. For the broad focus (BF), we instructed the participants to mark “The whole sentence” as focused.

Table 4 shows the results; the focus identification ratio varies, depending on the contexts, and it appears that the participants perceived focus on A words without difficulty at first glance, but, as observed in Figure 4, that is the case only in the context preceding or following U words.

	Max(%)	Min(%)	Mean(%)	SD
CF on A words	91.4	68.5	76.8	16.1
CF on U words	76.1	57.6	65.8	4.6
BF	42.1	15.3	36.8	5.4

Table 4. The focus identification ratio in W1W2 conditions

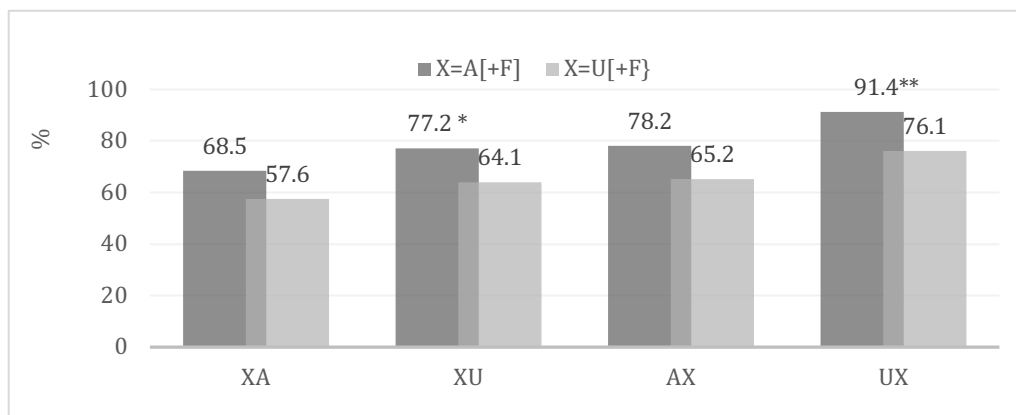


Figure 4. Contrastive focus identification ratio in %
(* and ** indicate significant difference $p < 0.05$ on T-test)

Our predictions i)-iii) were not borne out; contra predictions i) and ii), only a few cases of F0-rise and post-focal lowering were statistically significant. Furthermore, the focus on A-words was perceived more than the focus on U-words in restricted contexts only.

Our experimental data leads us to claim that focus effects are not always realized as claimed in previous studies (cf. Pierrehumbert and Beckman 1988, Kubozono 2007, Ishihara 2016, among many others), and actual utterances may lack them. ‘Segmental’ F0 rise in English as in Figure 1 is not expected in Japanese. We will see why Japanese lacks focus effects below.

2.3.3. SUPRASEGMENTAL PROPERTIES OF JAPANESE. In addition to lexical accents, Japanese has a suprasegmental downstepping pitch contour, called ‘downtrend’(cf. McCawley 1968), as observed in Figure 5.

Due to downtrend, the initial H in an Intonational Phrase (IP, henceforth) is with the max F0. In Figure 3 (left), in the context of W1[+F]W2, W1[+F] shows the max F0. The locus of the max F0 is due to the initiation of downtrend, not due to the focus effect. In the context of W1W2[+F] in Figure 3 (right), W2[+F] shows the max F0 only in the context of UA[+F]. In other

contexts, i.e., AA[+F], UU[+F], and AU[+F], the F0 of W2[+F] is lower than that of W1, and we hardly observe the focus effect.

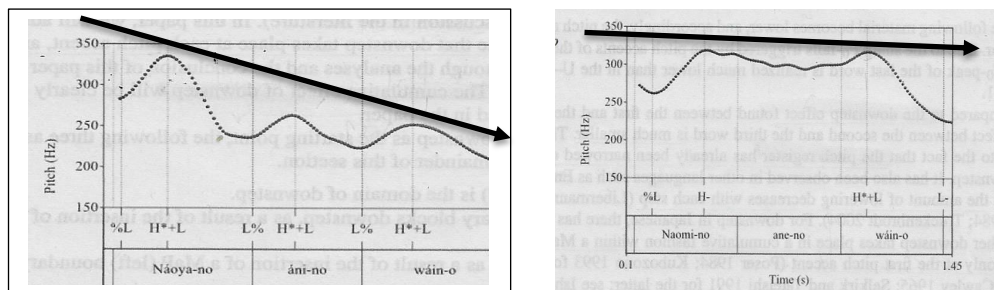


Fig.5. F0-contour of AAA sequence *Na'oya-no a'ni-o -no wa'in-o* ‘Naoya’s brother’s wine-ACC’ (left) and UUA sequence *Naomi-no ane-no wa'in-o* ‘Naomi’s sister’s wine-ACC’ (right) (adapted from Ishihara 2016: 1391)

The lack of focus effect leads to weaker identification of focus; Figure 4 shows that the focus identification ratio is higher in UA[+F] (91.4%) than other contexts of AA[+F] (78.2%), UU[+F] (76.1%) and AU[+F] (65.2%).

We conclude that contra English, F0 is not effective for focus identification in Japanese, and this is why max F0 does not correspond to the digit position in Japanese in Figure 1. Back to the perception experiment in Section 2.1, the 10-digit phone numbers XXX-XXX-XXXX of our Japanese experiment materials are with the prosodic structure (5), due to the accent rules of Japanese,

$$(5) \quad [IP_1[UAA] IP_2[UAA] IP_3[[UA][UA]]]$$

Table 2 and Table 3 show that the focus identification ratio varies among positions, and Figure 1 illustrates that the F0 cues do not correspond with the focus position. Nevertheless, a look at IPs leads to a different story; the focus identification ratio is almost the same among the three IPs, as given in Table 5.

Task	Target IP position (% accuracy)		
	IP1	IP2	IP3
Japanese	98	98	97

Table 5: Accuracy in IPs bearing contrastive focus by J/E

In this section, we have argued that J/E bilinguals identify the contrastive focus of English, their L2, with less difficulty than Japanese, their L1, and that the perception of contrastive focus is affected by the features of focus cues, not by the L1/L2 distinction. Figure 1, Tables 2, 3, and 5 lead us to claim that participants perceive the contrastive focus in English on the ‘segmental’ level and that the contrastive focus in Japanese is processed on the IP level. Japanese is a pitch language with lexically determined accents as well as prosodic downtrend. To process Japanese, we need to perceive the lexical accent on the segmental level and the downtrend on the prosodic level. In a word, we have to deal with two types of acoustic information. If we are on

the right track, different substrates are employed to process segmental and prosodic acoustic features. We will conduct an fMRI experiment to see whether or not such perceptual mechanisms are implemented in the brain.

3. fMRI experiment. Methods and materials are as follows.

3.1. METHODS AND MATERIALS.

Participants: We recruited 22 right-handed late-onset J/E bilinguals (M11, F11, mean age 26.7, SD=11.1, average PBT TOEFL score=595, SD=55.7). They reported no auditory disorders nor a history of neurological or psychiatric disorders. All participants gave written consent before participation in the experiment. The experiment was approved by the Ethics Committee of Kobe University.

Stimuli: We used 80 English and Japanese broad focus and contrastive focus 10-digit numbers (20X2 (broad/contrastive) X2 (Japanese/English)), which are the same as the stimuli we used in perception experiments in Section 2.

Procedure: Before the experiment, participants sat for a practice session and received the instruction about the distinction between broad and contrastive focus in Question-Answer dialogues, like (1) and (2).

Task: The task was to judge whether or not the audio stimuli contained contrastive focus, and the “Yes” and “No” responses were provided by pressing the buttons with the index and the middle fingers of the left hand.

fMRI paradigm: We used a 3T MRI scanner (Siemens, Prisma), Presentation (Neuro Behavioral System), and an MRI-compatible electrostatic headphone in our experiment. We employed a block design of the average trial duration of 4 seconds and a response interval of 2 seconds. We had two experimental blocks (J/E order for eleven participants, E/J order for the other eleven). Data processing was carried out by Matlab 9.3 and analyzed by SPM 12 (Mathworks). Data were realigned to the first volume, normalized into standard stereotactic space (voxel size 3X3 mm, template provided by the Montreal Neurological Institute).

3.2. BEHAVIORAL RESULTS. Table 6 shows the results.

	English (L2)	Japanese (L1)
Identification Ratio		
Broad Focus	95.19% (SD=0.05)	99.09% (SD=0.01)
Contrastive Focus	90.06% (SD=0.11)	80.02% (SD=0.19)
Response Time	451.6 ms	467.3 ms

Table 6. Focus identification ratio and response time

The identification ratio of contrastive focus is significantly lower in Japanese than the one in English on T-test: $t(38) = -1.78, p=0.042$. The response time is shorter in English than in Japanese, but the difference is not statistically significant.

3.3. fMRI RESULTS. Figure 6 shows the activated brain regions when contrastive focus (CF) is processed in Japanese (above) and English (below). Table 7 gives each value of the activated regions to process broad focus (BF) and contrastive focus (CF) in Japanese. We find a substantial

overlap in the bilateral Superior Temporal Gyrus (STG), Insula (INS) and pre-Motor, and left Supramarginal Gyrus (SMG). Right Dorsolateral Frontal Cortex (DLFC) is involved in processing Japanese CF. The contrast of CF > BF in Japanese reveals the involvement of left Primary Motor and preMotor.

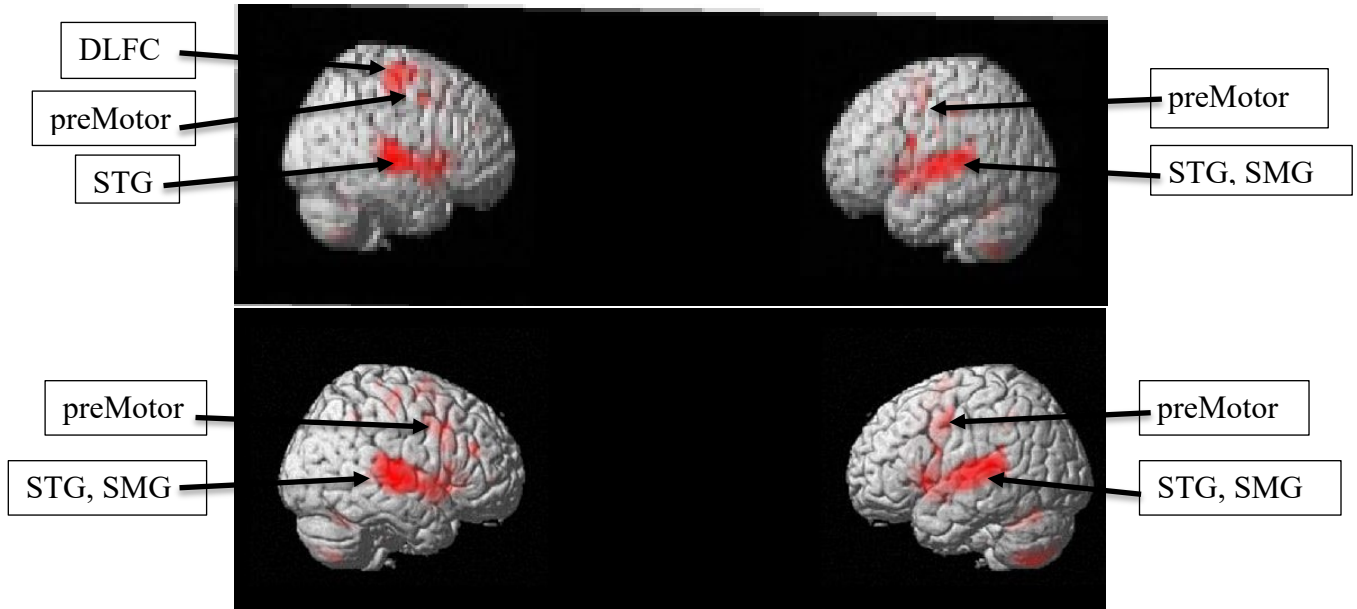


Figure 6. Activated regions to process Japanese CF (above) and English CF (below)

Region	BA	k	z-value	p-value	MNI coordinates		
					x	y	z
Japanese BF							
L STG	22	5879	7.33	0.000	-52	-36	10
	22		7.17		-58	-36	-2
R STG	22		7.19		52	-8	-4
L INS	13		7.15		-30	24	0
R INS	13		7.61		34	22	-4
R PrimAuditory	41	6034	7.74	0.000	54	-22	4
L preSMA	8	2324	6.92	0.000	-4	16	44
L preMotor	6		6.46		-6	2	58
	6	95	5.59	0.000	-30	-2	48
R preMotor	6		6.43		8	6	64
	6	227	5.60	0.000	56	4	44
R Primary Motor	4	737	6.06	0.000	38	-22	54
L SMG	40	321	5.68	0.000	-48	-34	46
	40		5.23		-56	-20	44
	40		4.86		-40	-44	44
L Ventral Posterior Cingulate Cortex	23	68	5.28	0.000	-4	-22	30
R Thalamus		114	5.25	0.000	10	-20	4

Japanese CF

R PrimAuditory	41	4820	7.61	0.000	54	-20	2
			7.36		56	-10	-2
R STG	22		7.34		58	0	-2
L STG	22		7.40		-52	-34	10
	22		7.18		-58	-20	0
L preMotor	6	2620	7.03	0.000	-6	4	58
	6	183	5.74	0.000	-32	-2	48
	6		5.10		-52	-2	44
R preMotor	6		6.09		30	-16	64
	6	156	5.63	0.000	54	4	42
L preSMA	8		6.71		-4	16	42
L SMG	40	319	5.82	0.000	-50	-28	44
L Primary Motor	4		5.43		-58	-16	26
L Primary Sensory	1		5.26		-38	-34	42
R DLFC	9	60	5.78	0.000	32	40	26

Japanese CF > Japanese BF

L Primary Motor	4	8	4.94	0.006	-38	-22	60
L preMotor	6	2	4.83	0.022	-42	-16	60

Table 7. BF/CF Activated regions and the contrast of CF > BF in Japanese

Note: Peak voxels in clusters are in boldface. P-values refer to p-cluster < 0.05, FEW-corrected.

Abbreviations: BA, Brodmann area; L, left hemisphere; R, right hemisphere; k, cluster size

Table 8 gives each value of the activated regions to process broad focus (BF) and contrastive focus (CF) in English. The contrast of CF > BF in English reveals the involvement of left STG and preMotor, as well as bilateral Primary Auditory Area.

Region	BA	k	z-value	p-value	MNI coordinates		
					x	y	z
English BF							
R STG	22	20730	7.84	0.000	62	-22	2
L STG	22		7.65		-52	-34	10
R INS	13		7.57		34	26	-2
R preSMA	8	3169	7.44	0.000	2	20	42
L preMotor	6		6.94		-4	10	50
R SMG	40	851	6.83	0.000	38	-48	44
L SMG	40	1021	6.82	0.000	-36	-52	48
	40		6.51		-54	-38	44
R Somatosensory Association Cortex	7	100	5.76	0.000	10	-66	46
L	7		5.16		-4	-72	46
L Primary Visual	17	271	5.63	0.000	-10	-72	8
R	17		5.38		14	-72	16

English CF								
L STG	22	16031	Inf	0.000	-52	-32	10	
R	22		7.81		62	-20	2	
R preSMA	8	2903	7.28	0.000	10	20	40	
L	8		5.45		-52	12	42	
L preMotor	6	689	6.31	0.000	-44	4	38	
	6		6.83		-6	10	48	
	6		6.45		-6	4	60	
R	6		7.71		52	-4	-4	
	6	78	5.28		34	-8	56	
L SMG	40	228	5.51	0.000	-38	-46	42	
	40		5.04		-46	-38	42	
English CF > BF								
L STG	22	52	5.24	0.000	-54	-38	10	
L preMotor	6	19	5.26	0.001	-52	-4	42	
R Prime Auditory	41	8	4.93	0.006	58	-6	-2	

Table 8. BF/CF Activated regions and the contrast of CF > BF in English

Note: Peak voxels in clusters are in boldface. P-values refer to p-cluster < 0.05, FEW-corrected.

Abbreviations: BA, Brodmann area; L, left hemisphere; R, right hemisphere; k, cluster size

Table 9 shows the contrast between English CF > BF and Japanese CF > BF, and we can see left STG and bilateral Prime Auditory regions are activated more in English than in Japanese.

Region	BA	k	z-value	p-value	MNI coordinates		
					x	y	z
(English CF > BF) - (Japanese CF > BF)							
L STG	22	410	5.49	0.000	-54	-14	0
	22		5.47	0.000	-60	-28	6
L Prime Auditory	41		5.35	0.000	-56	-20	6
R	41	45	5.04	0.000	60	-18	4

Table 9. Contrast between English CF>BF and Japanese CF>BF

4. Discussion. Our research question was why Japanese CF is hard to perceive even to L1 speakers. To answer this question, we conducted linguistic production and perception experiments to understand language mechanisms. We further conducted an fMRI experiment to see how language mechanisms reside in the brain.

Our findings are (i) the identification ratio of focus perception is better in English and Japanese, both with our acoustic experiments and our fMRI experiment (cf. Tables 2, 3, and 6). (ii) English uses F0 features to perceive CF, but Japanese does not (cf. Figure 1). In other words, English employs ‘segmental’ features as a focus cue, but Japanese does not have such segmental focus cues. It relies on suprasegmental cues, instead. (iii) English and Japanese activate bilateral STG and preMotor and left SMG strongly to process CF (cf. Figure 6, Tables 7, 8), and (iv) only Japanese requires right DLFC to activate for CF processing (cf. Figure 6, Table 7).

Table 9 shows that left STG and bilateral Prime Auditory regions are more activated to process English focus. Traditionally these regions are considered to be typically employed to process acoustic features of a language (cf. Gandour 2007, Perrone et al. 2010, among many others). Japanese, on the other hand, does not activate these regions substantially but activates motor areas (BA4 and 6) more for CF processing (cf. Table 7). In a word, English needs acoustic features to be processed for CF perception, while Japanese does not.

This difference in the neural network between English and Japanese accounts for the acoustic F0 feature illustrated in Figure 1. We can claim that English recruits STG and Prime Auditory regions to process F0 features. Perrone et al. 2010 consider the perception of BF and CF in French and report that bilateral IFC and SMG, left STG, MTG, preMotor, and INS are recruited for BF processing. The lack of significant activation in these areas in Japanese (cf. Table 7) means that Japanese does not employ F0 features for CF processing, and this supports our claim that Japanese CF lacks Focus effects, discussed in Section 2.3.

Another difference in the neural network between English and Japanese is the activation of right DLFC in Japanese. This area is responsible for intonation processing (cf. Chien et al. 2020) and pitch memory (cf. Schaal et al. 2017). Since our experiment materials are all statements and the intonation is the same, we believe that pitch, not intonation, is involved in the activation of right DLFC in our study.

Green and Abutalebi (2013) explore the nature of control process in the neural network in bilingual speakers, and propose the Adaptive Control Hypothesis: they claim that language control processes adapt to the recurrent demand placed on them by the interactional context and changes a parameter or parameters about the way it works. Figure 7 provides a schematic description of the neural structures and their connections that they associate with language control processes.

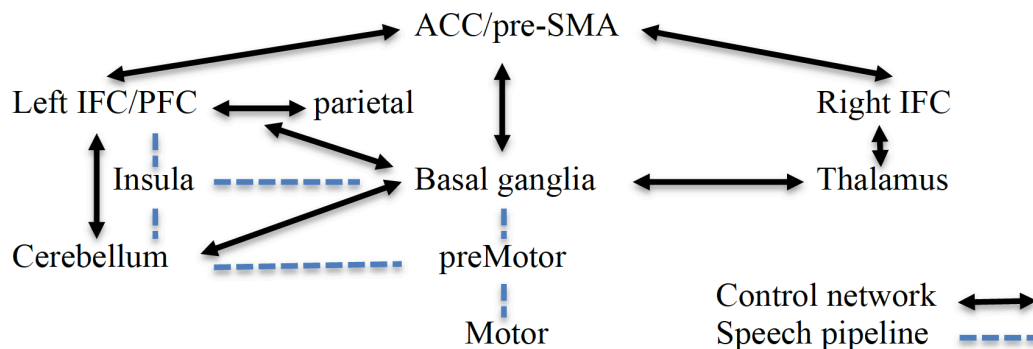


Figure 7. Language control network and speech production regions (from Green and Abutalebi 2013) Abbreviations: ACC, anterior cingulate cortex; pre-SMA, presupplementary motor area; IFC, inferior frontal cortex; PFC, prefrontal cortex)

They identify ACC and pre-SMA with conflict monitoring, and pre-SMA initiates speech in language switching. They associate left PFC with the control of interference, parietal cortices with the maintenance of task presentation. Basal ganglia switch languages.

Applying Figure 7 to our study, we can claim that J/E bilinguals activate motor areas to switch languages and change their parameters in their neural network. Tables 7 - 9 illustrate that left STG and INS are employed to process both English and Japanese, and but Japanese need to activate right DLPC and Thalamus. In a word, J/E bilinguals switch parameters, depending on

the language they use.

Gandour et al. 2007 report that C/E bilinguals recruit left STG, SMG, INS to process both Chinese (L1) and English (L2). We understand that Chinese and English use the same parameter in the left regions in Figure 7. They report that the L2 task is tougher, and the graded activity in the anterior INS is responsible for task difficulty in English (L2).

J/E bilinguals employ a different parameter when they switch languages and activate both the left and the right brain regions to process Japanese. The graded activity in the right DLPC leads to task difficulty in Japanese, which is why Japanese is difficult to hear, even to L1 speakers.

If we are on the right track, tone languages like Chinese and pitch languages like Japanese employ different parameters in the neural network, even though they both use pitch. Another puzzle is why right DLPC is activated to process CF in Japanese. Japanese is a pitch language, and if right DLPC were responsible for processing pitch, it should be activated to perceive BF, too. It is a puzzle why right DLPC is involved in the Japanese neural parameter only when we process CF. We will leave this puzzle for our future study.

5. Conclusion. Japanese contrastive focus is significantly harder to identify than English focus even for L1 speakers. To account for why Japanese is hard to perceive, we conducted acoustic experiments and an fMRI experiment. We found that Japanese lacks Focus effects contra previous studies and is an acoustically weak language. We also found that we employ different neural networks to process English and Japanese; right dorsolateral frontal cortex activates to process Japanese, but not English, CF. Japanese is a pitch language and requires to process both lexical accent and suprasegmental pitch contour. English, on the other hand, needs to process lexical accent only, and it activates left superior temporal gyrus, insular and supramarginal regions, but not right dorsolateral frontal cortex. We conclude that processing burdens lead to perception difficulty, even for L1 speakers.

References

- Chien, Pei-Ju, Angela D. Friederici, Gesa Hartwigsen & Daniela Sammler. 2020. Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Human Brain Mapping*. 1–17. <https://doi.org/10.1002/hbm.24916>.
- Cole, Jennifer, Yun-S. Mo & S-D Beak. 2010. The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes* 25. 1141–1177. <https://doi.org/10.1515/labphone.2010.022>.
- Cole, Jennifer, José I. Hualde, Caroline L. Smith, Christopher Eager, Timothy Mahrt & Ricardo Napoleão de Souza. 2019. Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics* 75. 113–147. <https://doi.org/10.1016/j.wocn.2019.05.002>.
- Bauman, Stephan and Bodo Winter. 2018. What makes a word prominent? : Predicting untrained German listeners' perceptual judgement. *Journal of Phonetics* 70. 20–38. <https://doi.org/10.1016/j.wocn.2018.05.004>.
- Gaab, Nadine, Christian Gaser, Tino Azehle, Lutz Jancke & Gottfried Schlaug. 2003. Functional anatomy of pitch memory: An fMRI study with sparse temporal sampling. *NeuroImage* 19. 1417–1426. [https://doi.org/10.1016/S1053-8119\(03\)00225-6](https://doi.org/10.1016/S1053-8119(03)00225-6).
- Gandour, Jackson, Yunxia Tong, Thomas Talavage, Donald Wong, Mario Dzemidzic, Yisheng Xu, Xiaojian Li & Mark Lowe. 2007. Neural basis of first and second language processing

- of sentence-level linguistic prosody. *Human Brain Mapping* 28. 94-108.
<https://doi.org/10.1002/hbm.20255>.
- Green, David and Jubin Abutalebi. 2013. Language control in bilinguals: The adaptive control hypothesis. *Studies of Cognitive Psychology*. 25 (5). 515–530.
<https://doi.org/10.1080/20445911.2013.796377>.
- Ishihara, Shin'ichiro. 2003. *Intonation and interface conditions*. Cambridge, Mass.: MIT dissertation.
- Ishihara, Shin'ichiro. 2016. Japanese downstep revisited. *Natural Language & Linguistic Theory* 34. 1389-1443. <https://doi.org/10.1007/s11049-015-9322-8>.
- Kubozono, Haruo. 2007. Focus and intonation in Japanese: Does focus trigger pitch reset? *Workshop on Prosody Syntax, and Information Structure (WPSI)* 2. 1-27.
- Lee, Yong-Cheol, BeiWang, Sisi Chen, Martine Adda-Decker, Angélique Amelot, Satoshi Sambu & Mark Lieberman. 2015. A crosslinguistic study of prosodic focus. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 4754–4758. <https://doi.org/10.1109/ICASSP.2015.7178873>.
- McCawley, James. 1968. *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.
- Mizuguchi, Shinobu & Koichi Tateishi. 2018. Focus prosody in Japanese reconsidered. *Proceedings of the Linguistic Society of America (PLSA)* 3.28. 1-15.
<https://doi.org/10.3765/plsa.v3il.4291>.
- Perrone, Marcela, Marion Dohen, Hélène Lœvenbruck, Marc Sato, Cédric Pichat & Monica Baci. 2013. Neural correlates of the perception of contrastive prosodic focus in French: A functional magnetic resonance imaging study. *Human Brain Mapping* 34. 2574-2591.
<https://doi.org/10.1002/hbm.22090>.
- Pierrehumbert, Janet & Mary Beckmann. 1988. *The Japanese Tone Structure*. Cambridge, MA: The MIT Press.
- Sato, Marc, Pascale Tremblay & Vincent L. Gracco. 2009. A mediating role of the premotor cortex in phoneme segmentation. *Brain & Language* 2009. 1-9.
<https://doi.org/10.1016/j.bandl.2009.03.002>.
- Schaal, Nora K., Marina Kretschmer, Ariane Keitel, Vanessa Krause, Jasmin Pfeifer & Bettina Pollok. 2017. The significance of the right dorsolateral prefrontal cortex for pitch memory in non-musicians depends on baseline pitch memory abilities. *Frontiers in Neuroscience* 11. 1–9.
<https://doi.org/10.3389/fnins.2017.00677>.