Journal of Phonetics and Audiology

Tomita, J Phonet and Audiol 2016, 2:1
DOI: 10.4172/2471-9455.1000112

Research Article Open Access

Virtual Space of English Consonants: Shorter Distance Produced by Japanese Learners of English

Kaoru Tomita*

Faculty of Literature and Social Sciences, Yamagata University, 1-4-12 Kojirakawa-machi, Yamagata 990-8560, Japan

Abstract

This study investigates how Japanese learners of English pronounce two consonants, /s/ and /S/, or /b/ and /v/, of English minimal-paired words whose corresponding words are English-based loanwords in Japanese and written in katakana. Frequency of spectral peak, duration, and intensity of these consonants produced by six Japanese learners of English and six native English speakers are measured with acoustic equipment. Among these phonetic features, significant differences in frequency of spectral peak between /s/ and /S/ are observed. This holds true for both native English speakers and Japanese learners of English. There are also significant differences in duration between /b/ and /v/, and intensity between /S/ and /s/, or /b/ and /v/. A hypothesis that distance between the values of these features for each paired consonants tends to be smaller for the Japanese learners of English than for the native English speakers is also verified. Implications for further research are briefly discussed.

Keywords: Consonant; Consonantal distance; Duration; Formant; Intensity

Introduction

Japanese and English have different phonological systems, which produce differences in timing, such as a stress-timed language or a mora-timed language. That might also cause differences in length, manner and even position of consonant articulation. A standard textbook of Japanese-English phonetics and phonology states that there is not a big difference in number of English and Japanese consonants, and broadly the former is uttered with stronger breath than the latter.

Manners and positions of English and Japanese consonants presented in Table 1, however, invite several questions as to those explanations of Japanese and English phonological features.

It is then expected that these different phonological systems would affect Japanese learners of English at some learning stages. For example, one may ask to what extent Japanese speakers' /S/ in *she* is different from English speakers' one in terms of acoustic properties such as frequency of noise region. Transferring from one's native language plays an important role in learning a foreign language [1]:

- Foreign accents are not the result of just "missing the mark" in random ways. To the contrary, careful inspection shows that the deviations between the goal and what is achieved are systematic; and can usually be attributed to the phonology, including the phonological rules, of one's native language. The phenomenon of mispronunciations in a second language in ways attributable to the phonology of the first language is called transfer.

The transfer from a native language, for example, the one from English-based loan words in Japanese to English, is not studied empirically [2]:

- The argument that loanwords in Japanese are of (great) detriment to learners of English has been generated from observations of errors and evidenced with anecdotes. These studies have focused on perceived interference in pronunciation and word meaning. As with Contrastive Analysis itself, there is little empirical evidence presented, only descriptions of gross, superficial features of produced language.

Consonantal transferring would be observed in its phonetic features, such as duration and intensity, and they can be analyzed with values of a concentration of acoustic energy, the formant.

Consonants are easy to describe in articulatory terms whereas vowels are easier to describe in acoustic terms [3]. They are more difficult than vowels to be measured with a single category [4]. They, however, can be dealt with based on the same concepts [3]:

- This use of different dimensions in the description of consonants and vowels suggests that the articulation of these two classes of sounds has little in common. However, the articulatory description of both consonants and vowels is largely based on location of constriction ("place of articulation" in consonants, "frontness" in vowels) and degree of constriction ("manner of articulation" in consonants, "height" in vowels).

Formant values would be used to describe consonantal transferring observed in utterances by Japanese learners of English.

Consonants are generally classified by voicing/unvoicing, place of articulation and manner of articulation. Consonants, /s/ and /S/, for example, are both unvoicing sounds produced with a stream of air directed at the upper teeth, which creates noisy turbulent flow. Only their place of articulation is different. The sound, /s/, is made by touching the tip or blade of the tongue to a location just forward of the alveolar ridge. The sound, /S/, is made by touching the blade of the tongue to a location just behind the alveolar ridge.

With looking at the spectral noise region, we can tell, /s/ from /S/: there is the importance of a high-frequency noise region for /s/ and a low-frequency noise region for /S/. This noise region has some ranges: The energy for /s/ is largely above 4,000 Hz and that for /S/ begins lower at around 2,500 Hz [5]. Identification of [s] appeares to depend on energy peaks at about 5000 and 8000 Hz, whereas identification of [S] is related to a peak at about 2500 Hz [4]. The difference between frequency spectral peak of [s] and [S], in which [S] typically exhibits a

***Corresponding author:** Kaoru Tomita , Faculty of Literature and Social Sciences, Yamagata University, 1-4-12 Kojirakawa-machi, Yamagata 990-8560, Japan, Tel: (023) 628-4793; Fax: (023) 628-4793; E-mail: kaoru@human.kj.yamagata-u.ac.jp

Received January 29, 2016; **Accepted** February 25, 2016; **Published** March 01, 2015

Citation: Tomita K (2016) Virtual Space of English Consonants: Shorter Distance Produced by Japanese Learners of English. J Phonet and Audiol 2: 112. doi:10.4172/2471-9455.1000112

Citation: Tomita K (2016) Virtual Space of English Consonants: Shorter Distance Produced by Japanese Learners of English. J Phonet and Audiol 2: 112. doi:10.4172/2471-9455.1000112

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glotal
Plosive	p b		t d	*t d*			k g		
Affricate				t^s	***tΣ dZ***				
Nasal	m		n	*n*			*N*	N	
Flap					}				
Fricative		*f v*	*T Δ*	s z	*Σ **Z***				h
Approximant				¨		j	w		
Lateral approximant				*l*					

Table 1: Consonants which has the same manner and position of articulation for English and Japanese are presented in small letters and those whose manner or position differs in big letters, with italic for English and non-italic for Japanese.

mid-frequency spectral peak around 2,500-3,500 Hz and alveolar [s] is produced with a shorter anterior cavity than [S] and therefore display a primary spectral peak at higher frequencies, ranging from 4,000 to 7,000 Hz [3].

A consonant, /v/, is also a voicing sound formed by touching the lower lip to the upper teeth with a tight constriction which is made so that air passing through the constriction flows turbulently, making a hissing noise. A consonant, /b/, is defined to be a voicing sound formed by two lips with the airflow through the mouth is momentarily closed off.

Consonants are also discussed in groups that are distinctive in their articulatory and acoustic properties: plosives, affricates, nasals, fricatives and approximants. Spectral change is a vital part of creating characteristic timbre of consonants [6]. They are always coproduced with a vowel, and are acoustically characterized by a period of silence (corresponding to the vocal tract closure), followed by a sudden broadband burst of energy (as the vocal tract constriction is released), followed by formant transitions that typically last about 50 msec.

Languages traditionally classified as stress-timed have low %V (percent of duration occupied by vowels) and high ΔC values (consonantal interval variability), while languages traditionally classified as syllable timed or mora timed have high %V and low ΔC values [6]. This would affect the duration of consonants in stress timed language when they are produced by learners whose native language is syllable or mora timed.

With looking at the duration of noise segments, we can tell, /b/ from /v/. It is reported that when stops, affricates and fricatives are compared in an equivalent context, the fricatives generally have the longest noise segments. The interval from release of a consonant constriction to the onset of voicing is larger in fricatives than in stops [7]. In a study of the noise segment durations for stops, affricates, and fricatives in the languages of Mandarin, Czech, and German, the following durational boundaries are identified: 62 to 78 msec for the stop-affricate boundary, and 132 to 133 msec for the affricate-fricative boundary [8-10].

On the basis of these phonological features, it is estimated that there would be a significant difference in the values of their production of frequency of noise region, duration and intensity between paired English consonants, /S/ and /s/, or /b/ and /v/ produced by both native English speakers and Japanese learners of English. Besides, for minimal-paired consonants, such as /s/ and /S/, Japanese learners of English would produce /s/s that display noise region at lower frequencies, and /S/s that display noise region at higher frequencies than those produced by native speakers of English. For minimal-paired consonants, such as /v/ and /b/, Japanese learners of English would produce /v/s that hold shorter duration, and /b/s that hold longer duration than those produced by native speakers of English. For both of these minimal-paired consonants, Japanese learners of English would produce /S/s in lower intensity and /s/s in higher intensity or /b/s in lower intensity and /v/s in higher intensity than those produced by native speakers of English

On the basis of these predictions, it is hypothesized that distance between the locations that the blade of the tongue touches for producing /s/ and /S/, which can be called virtual consonant space, would smaller for Japanese learners of English than for native English speakers. Also the distance between the duration of /v/ and /b/, or the distance between the intensity of /S/ and /s/, or /b/ and /v/, which can be also called virtual consonant space, would smaller for Japanese learners of English than for native English speakers.

Experiment

Method

Subjects: Three female speakers of American English (hereafter FE1, FE2, FE3), three male speakers of American English (hereafter ME1, ME2, ME3) and four female Japanese learners of English (FJ1, FJ2, FJ3, FJ4) and two male Japanese learners of English (MJ1, MJ2) participated in the experiment. The native English speakers aged 23 or 20 came from Oklahoma, U.S.A. as exchange students with one year term. The Japanese learners of English who were from Northern Japan were college students and their ages were 21 or 20 years. On the basis of the TOEIC® scores, they were regarded as intermediate-level learners of college English.

Stimuli: The stimuli consisted of a pair of monosyllabic or disyllabic words. They were (1) veering /vI«¨IN/, (2) beer /bI«¨/, (3) view /vju:/, (4) boom /bu:m/, (5) virtual /vÎ:]u«l/, (6) bargain /bA:]gIn/, (7) seafood /si:]u:d/, (8) shifting /SIftIN/ (9) suit /su:]/, (10) shoot /Su:]/, (11) son / sÃn/, and (12) shutting /SÃtIN/. All these words except "veering" and "son" were used as English-based loanwords in Japanese. They were called in such ways as /bi±u/, /bju:/, /bu:mu/, /ba:tSa±u/, /ba:gen/, /si:-u:do/, /siutingu/, /su:tsu/, /sju:to/, /sjaÂtingu/. Each stimulus item was printed in a carrier phrase, *I said* "_"., on a 21 cm × 30 cm sheet of paper.

Procedure

Each speaker was presented with test sheets and asked to clearly produce test items ten times. He/she was instructed to pronounce each word as clearly as possible. This instruction was important particularly for native English speakers because elision of consonants in conversational speech was not uncommon.

Recordings were made of 1440 items (12 speakers × 12 items × 10 times per subject) and they were recorded in a sound treated room using a Sony unidirectional dynamic microphone (F-V640) and a Marantz solid state recorder (PMD670). The microphone was positioned at a lip-to-mouth distance of approximately five cm. Recordings lasted approximately one hour for each subject.

Acoustic measurements

The speech samples were analyzed using the Praat speech analyzing

Citation: Tomita K (2016) Virtual Space of English Consonants: Shorter Distance Produced by Japanese Learners of English. J Phonet and Audiol 2: 112. doi:10.4172/2471-9455.1000112

software (http: www.praat.org). The sampling rate was 44.1 kHz with a 16 bit resolution. For each consonant, the mean frequency of noise region, the mean duration and the mean intensity values were used for analyses.

Analyses

Frequency of noise region, duration and intensity were compared between subjects. An analysis of variance (ANOVA) with repeated measures was a basic tool used for mean comparisons.

Results

The frequency ranges from 1129 to 10890 Hz. The durations ranges from 13 to 149 msec. The intensity ranges from 36 to 79 dB.

Frequency of noise regions for /s/ and /]/ comparison

The spectral noise region is measured with acquiring frequency of the highest spectrum by observing spectral slice. Mean frequencies of the spectral peak in hertz, that are observed in consonants, /s/ or /S/, of the experimental words produced by each subject is shown in Tables 2 and 3. A 2 × 7 ANOVA with two consonants and seven female speakers and the 2 × 5 ANOVA with two consonants and five male speakers are performed to examine the speaker effect and the consonantal quality effect. As predicted, comparison of /s/ and /S/ shows that the frequency of noise region of /s/ is significantly higher than that of /S/ for all native English speakers and Japanese learners of English except MJ2's seafood-shifting.

Duration of segments for /b/ and /v/ comparison

The duration is measured with observing spectrogram and acquiring range of energy spread fairly evenly. Mean durations in milliseconds of consonants, /v/ or /b/, produced by each subject is shown in Tables 4 and 5. A 2 × 7 ANOVA with two consonants and seven female speakers

/s/-/∑/	FE1	FE2	FE3	FJ1	FJ2	FJ3	FJ4	Mean	F-value[a]	p-value	Comparison
Seafood	7634	7994	10890	7288	2562	6560	5665				
Shifting	4095	3714	1129	3609	3304	4724	5358				
Mean	5864	5854	6009	5448	2931	5642	5511	6941	27.35	<0.001	FE3>FE2, FE1, FJ1, FJ3, FJ4>FJ2
F-value[b]	115.64	114.81	610.88	14.00	3.93	9.61	4.16	3704	18.12	<0.001	FJ4, FJ3>FE1, FE2, FJ1, FJ2,>FE3
p-value	0.001	0.001	0.001	0.001	0.06	0.06	0.05				
Suit	5572	6629	8635	5489	4545	6556	6185				
Shoot	4132	3639	3753	3698	3629	4201	3972				
Mean	4847	5134	6194	4593	4087	5378	5078	6230	5.76	<0.001	FE3>FE2, FJ3, FJ4, FE1, FJ1
F-value[b]	6.53	56.45	24.75	14.25	30.01	15.94	266.20	3860	6.65	<0.001	FJ3, FE1, FJ4>FE3, FJ1, FE2, FJ2
p-value	0.02	0.001	0.001	0.001	0.001	0.001	0.001				
Son	7192	7694	9599	8424	6293	6932	5050				
Shutting	4189	3501	3861	3762	4368	4568	6588				
Mean	5690	5597	6730	6093	5330	5750	5819	7312	18.85	<0.001	FE3, FJ1>FE2, FE1, FJ3, FJ2,>FJ4
F-value[b]	467.78	91.69	545.07	137.71	11.21	41.18	24.64	4405	38.78	<0.001	FJ4>FJ3, FJ2, FE1, FE3, FJ1>FE2
p-value	0.001	0.001	0.001	0.001	0.004	0.001	0.001				

Table 2: Mean frequency of spectral peak for the female speakers [Hz]. [a] The degrees of freedom are all 6 and 63. [b] The degrees of freedom are all 1 and 18.

/s/-/∑/	ME1	ME2	ME3	MJ1	MJ2	Mean	F-value[a]	p-value	Comparison
Seafood	8067	5603	6998	6670	4661				
Shifting	4639	2781	4291	3513	4667				
Mean	6353	4192	5644	5091	4664	6400	13.83	<0.001	ME1, ME3>MJ>ME2, MJ2
F-value[b]	33.86	291.80	226.61	35.74	40.49	3978	42.91	<0.001	ME1, ME3>MJ1>MJ2>ME2
p-value	0.001	0.001	0.001	0.001	NS				
Suit	6042	5516	5225	6881	4687				
Shoot	4068	2541	3180	3196	3801				
Mean	5055	4028	4202	5038	4244	5670	10.42	<0.001	MJ1, ME1>ME2, ME3, MJ2
F-value[b]	20.67	224.68	32.55	279.56	40.846	3357	19.90	<0.001	ME1, MJ2>MJ1, ME2>ME2
p-value	0.001	0.001	0.001	0.001	0.001				
Son	6576	5830	5716	7733	5623				
Shutting	4313	2700	4222	3629	4133				
Mean	5444	4265	4969	5681	4878	5695	27.34	<0.001	MJ1>ME1>ME2, ME3, MJ2
F-value[b]	69.66	222.50	49.60	318.44	87.65	3799	42.91	<0.001	MJ1>ME1, ME3, MJ2>ME2
p-value	0.001	0.001	0.001	0.001	0.001				

Table 3: Mean frequency of spectral peak for the male speakers [Hz]. [a] The degrees of freedom are all 4 and 45. [b] The degrees of freedom are all 1 and 18. *NS* not significant.

/v/-/b/	FE1	FE2	FE3	FJ1	FJ2	FJ3	FJ4	Mean	F-value[a]	p-value	Comparison
Veering	149	47	97	52	32	68	36				
Beer	24	17	44	38	53	45	37	68	12.49	<0.001	FE1>FE3>FJ3, FJ1, FE2, FJ3, FJ2
Mean	86	32	70	45	42	56	36	37	75.23	<0.001	FJ2, FJ3, FE3, FJ1, FJ4, FE1>FE2
F-value[b]	695.71	80.73	94.13	6.87	17.81	5.58	0.09				
p-value	0.001	0.001	0.001	0.01	0.001	0.03	NS				
View	82	30	52	59	57	85	32				
Boom	29	26	48	52	59	60	51	56	18.28	<0.001	FJ3, FE1>FJ1, FJ2, FE3>FJ4, FE2
Mean	55	28	50	55	58	72	41	46	13.98	<0.001	FJ2, FJ2, FJ1, FJ4, FE3>FE1, FE2
F-value[b]	24.59	3.24	0.41	2.41	0.32	12.08	11.74				
p-value	0.001	NS	NS	NS	NS	0.003	0.003				
Virtual	98	26	50	61	57	61	42				
Bargain	48	29	61	48	55	63	35	56	16.92	<0.001	FE1>FJ1, FK3, FJ2, FE3>FJ4, FE2
Mean	73	27	55	54	56	62	38	48	4.79	<0.001	FJ3, FE3, FJ1, FJ2, FE1, FJ4>FE2
F-value[b]	11.07	0.68	8.25	1.49	0.06	0.06	3.64				
p-value	0.004	NS	0.010	NS	NS	NS	NS				

Table 4: Mean duration for the female speakers [msec]. [a] The degrees of freedom are all 6 and 63. [b] The degrees of freedom are all 1 and 18. *NS* not significant.

Citation: Tomita K (2016) Virtual Space of English Consonants: Shorter Distance Produced by Japanese Learners of English. J Phonet and Audiol 2: 112. doi:10.4172/2471-9455.1000112

/v/-/b/	ME1	ME2	ME3	MJ1	MJ2	Mean	F-value[a]	p-value	Comparison
Veering	30	71	44	25	57	45	2.12	NS	MJ2, ME3>MJ1, ME2>ME1
Beer	13	23	38	29	41	28	22.44	<0.001	
Mean	21	47	41	27	34				
F-value[b]	61.28	84.16	1.68	2.91	0.52				
p-value	0.001	0.001	NS	NS	NS				
View	26	33	50	35	41	37	6.66	<0.001	ME3, MJ2>MJ1, ME2, ME1
Boom	20	26	40	39	30	31	11.54	<0.001	ME3, MJ1>MJ2, ME2, ME1
Mean	23	29	45	37	35				
F-value[b]	3.46	2.80	4.84	0.56	4.48				
p-value	NS	NS	0.04	NS	0.04				
Virtual	24	29	49	38	35	35	24.92	<0.001	ME3>MJ1, MJ2>ME2, ME1
Bargain	14	39	43	48	33	35	18.59	<0.001	MJ1, ME3, ME2>MJ2>ME1
Mean	19	34	46	43	34				
F-value[b]	19.72	12.18	1.07	6.20	0.40				
p-value	0.001	0.003	NS	0.023	NS				

Table 5: Mean duration for the male speakers [msec]. [a] The degrees of freedom are all 4 and 45. [b] The degrees of freedom are all 1 and 18. *NS* not significant.

and the 2 × 5 ANOVA with two consonants and five male speakers are performed to examine the speaker effect and the consonantal quality effect. As predicted, there are significant differences between duration of /v/ and /b/ for both native English speakers and Japanese learners of English. Comparison of /v/ and /b/ shows that the duration of /v/ was significantly longer than that of /b/ for native English speakers except FE2's *view-boom*, *virtual-bargain*, FE3's *view-boom*, ME1's *view-boom*, ME2's *view-boom*, ME3's *veering-beer, virtual-bargain*. The comparison of /v/ and /b/ also showed that the former was significantly longer than the latter for Japanese learners of English except FJ1's *view-boom, virtual-bargain*, FJ2's *view-boom*, *virtual-bargain*, FJ3's *virtual-bargain*, FJ4's *veering-beer*, *virtual-bargain,* MJ1's *veering-beer, view-boom*, MJ2's *veering-beer* and *virtual-bargain*.

Intensity of segments for /S/ and /s/ or /b/ and /v/ comparison

The intensity is measured with observing spectrogram and acquiring energy values. Mean intensities in decibels of consonants, /S/ and /s/, or /b/ and /v/, produced by each subject is shown in Tables 6 and 7. A 2 × 7 ANOVA with two consonants and seven female speakers and the 2 × 5 ANOVA with two consonants and five male speakers are performed to examine the speaker effect and the consonantal effect. As is expected, comparison of /S/ and /s/ shows that the intensity of /S/ is significantly higher than that of /s/ and comparison of /b/ and /v/ shows that the intensity of /b/ is significantly higher than that of /v/ for both native English speakers and Japanese learners of English. Comparison of /S/ and /s/ shows that the intensity of /S/ is significantly higher than that of /s/ for native English speakers except ME1's *shoot-suit*, *shutting-son*, ME3's *shifting-seafood*, *shutting-son*. Comparison of /b/ and /v/ shows that intensity of /b/ is significantly higher than that of /v/ for native English speakers except FE3's *bargain-virtual*, ME2's *bargain-virtual*, and ME3's *beer-veering*. Comparison of /S/ and /s/ or /b/ and /v/ produced by Japanese learners of English shows that among 18 cases of /S/ and /s/ comparison, six ones do not show a significant difference, and among 18 cases of /b/ and /v/ comparison, 12 ones do not show a significant difference.

Discussion

This study examines precise descriptions of English consonant

/Σ/-/s/	FE1	FE2	FE3	FJ1	FJ2	FJ3	FJ4	Mean	F-value[a]	p-value	Comparison
Shifting	56	49	54	54	67	40	47	52	85.49	<0.001	FJ2>FE1, FE3, FJ1>FJ4, FE2>FJ3
Seafood	56	42	47	43	73	38	46	43	115.40	<0.001	FJ2>FE1>FE3, FJ1, FJ4>FE2, FJ3
Mean	56	45	50	48	70	39	46				
F-value[b]	0.23	72.90	46.97	102.93	22.19	0.52	0.52				
p-value	NS	0.001	0.001	0.001	0.001	NS	NS				
Shoot	61	51	51	56	48	43	46	50	39.82	<0.001	FE1>FJ1>FE2, FE3>FJ2, FJ4>FJ3
Suit	59	44	47	49	58	37	45	48	36.08	<0.001	FE1, FJ2>FJ1, FE3, FJ4, FE2>FJ2
Mean	60	47	49	52	53	40	45				
F-value[b]	7.83	22.25	15.17	16.33	13.87	26.45	0.01				
p-value	0.010	0.001	0.001	0.001	0.002	0.001	NS				
Shutting	60	47	53	54	42	42	46	48	46.65	<0.001	FE1>FJ1, FE3>FE2, FJ4>FJ2, FJ3
Son	57	40	47	45	56	36	44	46	46.96	<0.001	FE1, FJ2>FE3, FJ1, FJ4>FE2, FJ3
Mean	58	43	50	49	49	39	45				
F-value[b]	8.32	39.03	32.54	42.43	58.31	25.79	1.03				
p-value	0.010	0.001	0.001	0.001	0.001	0.001	NS				
/b/-/v/	FE1	FE2	FE3	FJ1	FJ2	FJ3	FJ4	Mean	F-value[a]	p-value	Comparison
Beer	73	53	59	72	72	57	62	64	90.89	<0.001	FE1, FJ1, FJ2>FJ4, FE3>FJ3>FE2
Veering	57	48	57	70	70	53	68	60	118.61	<0.001	FJ1, FJ2, FJ4>FE1, FE3>FJ3>FE2
Mean	65	50	58	71	71	55	65				
F-value[b]	138.44	17.66	13.88	5.86	2.43	7.48	20.48				
p-value	0.001	0.001	0.002	0.026	NS	0.014	0.001				
Boom	77	59	66	73	76	54	64	67	71.57	<0.001	FE1, FJ2, FJ1>FE3, FJ4>FE2, FJ3
View	58	50	57	71	70	45	64	59	134.83	<0.001	FJ1, FJ2>FJ4>FE1, FE3>FE2>FJ2
Mean	67	54	61	72	73	49	64				
F-value[b]	436.50	50.87	183.50	2.76	31.22	43.70	0.18				
p-value	0.001	0.001	0.001	NS	0.001	0.001	NS				
Bargain	70	57	62	65	74	55	68	64	17.49	<0001	FJ2, FE1, FJ4>FJ1, FE3>FE2, FJ3
Virtual	56	53	59	68	73	52	67	61	38.67	<0.001	FJ2, FJ1>FJ4>FE3, FE1>FE2, FJ3
Mean	63	55	60	66	73	53	67				
F-value[b]	22.12	15.69	1.39	0.63	1.18	1.18	0.042				
p-value	0.001	0.001	NS	NS	NS	NS	NS				

Table 6: Mean intensity for the female speakers [dB]. [a] The degrees of freedom are all 6 and 63. [b] The degrees of freedom are all 1 and 18. *NS* not significant.

Citation: Tomita K (2016) Virtual Space of English Consonants: Shorter Distance Produced by Japanese Learners of English. J Phonet and Audiol 2: 112. doi:10.4172/2471-9455.1000112

/s/-/Σ/	ME1	ME2	ME3	MJ1	MJ2	Mean	F-value[a]	p-value	Comparison
Shifting	52	48	68	48	53				
Seafood	49	44	65	40	54	53	78.94	<0.001	ME3>MJ2, ME1>ME2, MJ1
Mean	50	46	66	44	53	58	89.58	<0.001	ME3>MJ2>ME1>ME2.MJ1
F-value[b]	4.49	18.76	3.71	22.37	1.81				
p-value	0.04	0.001	NS	0.001	NS				
Shoot	56	51	62	49	53				
Suit	54	45	56	40	50	54	36.88	<0.001	ME3>ME1, MJ2>ME2, MJ1
Mean	55	48	59	44	51	49	23.43	<0.001	ME3, ME1>MJ2, ME2>MJ1
F-value[b]	2.07	38.71	40.09	14.04	2.81				
p-value	NS	0.001	0.001	0.001	NS				
Shutting	54	48	66	45	46				
Son	52	45	66	40	46	51	99.22	<0.001	ME3>ME1, MJ2>ME2, MJ1
Mean	53	46	66	42	46	49	90.36	<0.001	ME3>ME1>MJ2, ME2>MJ1
F-value[b]	1.88	10.96	0.00	11.83	44.11				
p-value	NS	0.004	NS	0.003	0.001				
/b/-/v/	ME1	ME2	ME3	MJ1	MJ2	Mean	F-value[a]	p-value	Comparison
Beer	76	58	71	60	74				
Veering	62	48	69	58	70	67	89.97		ME1, MJ2>ME3>MJ1, ME2
Mean F-value[b]	69	53	70	59	72	61	50.25		MJ2, ME3>ME1, MJ1>ME2
p-value	67.63	42.80	2.20	2.20	8.32				
Boom	0.001	0.001	NS	NS	0.010			<0.001	
View	79	61	79	58	65			<0.001	
Mean	60	55	67	57	61				
F-value[b]	69	58	73	57	63	68	65.31		ME1, ME3>MJ2, ME2>MJ1
p-value	432.01	4.79	38.10	0.78	1.72	60	13.22		ME3>MJ2, ME1, MJ1>ME2
Bargain	0.001	0.001	0.001	NS	NS			<0.001	
Virtual	76	59	79	61	67			<0.001	
Mean	67	57	75	62	66				
F-value[b]	71	58	77	61	66	68	69.95		ME3, ME1>MJ2>MJ1, ME2
p-value	44.41	2.99	7.66	2.59	0.32	65	41.87		ME3>ME1, MJ2>MJ1>ME2
	0.001	NS	0.013	NS	NS				

Table 7: Mean intensity of spectral peak for the male speakers [dB]. [a] The degrees of freedom are all 4 and 45. [b] The degrees of freedom are all 1 and 18. *NS* not significant

qualities produced by Japanese learners of English with using four consonants, /s/, /]/, /b/, /v/, and their frequency of spectral peak, duration and intensity values are measured. There are significant differences between the frequency of spectral peak of /s/ and /S/ produced by both English native speakers and Japanese learners of English. There are significant differences between the duration of /b/ and /v/ produced by both of them, and there are also significant differences between the intensity of /S/ and /s/, and /b/ and /v/ produced by both of them. However, the number of the cases that do not show a significant difference is not the same for English native speakers and Japanese learners of English. There are much more cases for Japanese learners of English that do not show a significant difference between the paired consonants, /s/ and /S/, or /b/ and /v/, than for English native speakers.

English distinguishes /b/ and /v/ but Japanese does not have /v/. It has /b/ only. English distinguishes /s/ and /S/ but Japanese does not distinguish /s/ and /S/ either. It has /s/ only. As for the Japanese /s/, it is palatalized in some contexts [11]:

- Palatalization of consonants before /i/ is regular in Japanese, and its effect is especially notable with /s/, /z/, /t/, /d/, /n/, /h/. This can threaten intelligibility when transferred to English consonants preceding /iù/ and /I/.

From this difference in these two languages, it is expected that the virtual distance pictured by the frequency of spectral peak of the constituent consonants, the duration in them or intensity values in them is expected to be different for the native English speakers and the Japanese learners of English.

Shorter consonant-distance hypotheses

Distance in the frequency of spectral peak of a pair of consonants, /s/ and /S/ is compared for each pair of words, which occurs in the phonological contexts of __ /I/, __ /u/ or __ //. The mean difference of distance in frequencies of spectral peak for the consonants in paired words is shown in Table 8. A 1 × 7 ANOVA with one paired distance of consonants and seven female speakers and the 1 × 5 ANOVA with one paired distance of consonants and five male speakers are performed to examine the speaker effect.

Distance in the duration of a pair of consonants, /b/ and /v/ is compared for each pair of words, which occurs in the phonological contexts of __ /I/, __ /u/ or __ /A/. The mean difference of distance in frequencies of spectral peak for the consonants in paired words is shown in Table 8. A 1 × 7 ANOVA with one paired distance of consonants and seven female speakers and the 1 × 5 ANOVA with one paired distance of consonants and five male speakers are performed to examine the speaker effect.

Distance in the intensity of a pair of consonants, /s/ and /S/ or /b/ and /v/ is compared for each pair of words, which occurs in the phonological contexts of __ /I/, __ /u/ or __ /A/. The mean difference of distance in frequencies of spectral peak for the consonants in paired words is shown in Tables 8-11. A 1 × 7 ANOVA with one paired distance of consonants and seven female speakers and the 1 × 5 ANOVA with one paired distance of consonants and five male speakers are performed to examine the speaker effect (Tables 12 and 13).

Virtual distance of /s/ and /S/ measured in frequency of spectral

/s/-/Σ/	FE1	FE2	FE3	FJ1	FJ2	FJ3	FJ4	Mean	F-value[a]	p-value	Comparison
Seafood-Shifting	3539	4280	6884	3679	-7422	1836	3065	2265	25.54	<0.001	FE3, FE2, FJ1>FE1, FJ4>FJ3>FJ2
Suit-shoot	1439	2989	4881	1791	9162	2355	2213	3547	5.78	<0.001	FJ2>FE3, FE2>FJ3, FJ4, FJ1, FE1
Son-shutting	3003	4193	5738	4662	1925	2364	-1538	2477	40.56	<0.001	FE3, FJ1>FE2, FE1>FJ3, FJ2>FJ4

Table 8: Mean distance of frequency of spectral peak between a pair of consonants for female speakers [Hz]. The degrees of freedom are all 6 and 63.

/s/-/Σ/	ME1	ME2	ME3	MJ1	MJ2	Mean	F-value[a]	p-value	Comparison
Seafood-Shifting	3428	2821	2706	3156	693	2560	12.15	<0.001	ME1, MJ1, ME2, ME3>MJ2
Suit-Shoot	1974	2975	2045	3684	886	2312	10.99	<0.001	MJ1, ME2>ME3, ME1>MJ2
Son-shutting	2263	3129	1493	4104	1490	2495	31.52	<0.001	MJ1>ME2>ME1, ME3, MJ2

Table 9: Mean distance of frequency of spectral peak between a pair of consonants for male speakers [Hz]. The degrees of freedom are all 4 and 45.

/s/-/Σ/	FE1	FE2	FE3	FJ1	FJ2	FJ3	FJ4	Mean	F-value[a]	p-value	Comparison
Veering-beer	1245	298	533	143	-208	236	-10	319	68.20	<0.001	FE1>FE3, FE2>FJ3>FJ1, FJ4>FJ2
View-boom	532	31	34	75	-23	254	-195	101	13.88	<0.001	FE1>FJ3, FJ1, FE3, FE2>FJ2, FJ4
Virtual-bargain	502	-28	-112	132	23	-16	69	81	15.65	<0.001	FE1>FJ1, FJ4, FJ2, FJ3, FE2>FE3

Table 10: Mean distance of duration between a pair of consonants for female speakers [msec]. The degrees of freedom are all 6 and 63.

/s/-/Σ/	ME1	ME2	ME3	MJ1	MJ2	Mean	F-value[a]	p-value	Comparison
Veering-beer	164	485	58	-45	-74	117	37.38	<0.001	ME2>ME1, ME3>MJ1, MJ2
View-boom	55	69	100	-35	112	60	2.35	NS	
Virtual-bargain	95	-99	56	-100	16	-6	6.19	<0.001	ME1, ME3, MJ2>ME2, MJ1

Table 11: Mean distance of duration between a pair of consonants for male speakers [msec]. The degrees of freedom are all 4 and 45. *NS* not significant.

/S/-/s/	FE1	FE2	FE3	FJ1	FJ2	FJ3	FJ4	Mean	F-value[a]	p-value	Comparison
Shifting-Seafood	6	72	73	110	-65	16	19	33	15.31	<0.001	FJ1, FE3, FE2>FJ4, FJ3, FE1>FJ2
Shoot-suit	22	69	42	67	-97	58	2	23	15.45	<0.001	FE2, FJ1, FJ3, FE3, FE1>FJ4>FJ2
Shutting-Son	35	70	61	94	-145	56	22	27	35.76	<0.001	FJ1, FE2, FE3, FJ3>FE1, FJ4>FJ2
Beer-veering	156	51	18	24	18	38	-60	35	36.39	<0.001	FE1>FE2, FJ3, FJ1, FE3, FJ2>FJ4
Boom-view	187	91	96	23	56	93	-9	76	27.33	<0.001	FE1>FE3, FJ3, FE2>FJ2, FJ1>FJ4
Bargain-virtual	146	37	2	-29	8	36	2	28	6.47	<0.001	FE1>FE2, FJ3, FJ2, FE3, FJ4, FJ1

Table 12: Mean distance of intensity between a pair of consonants for female speakers [dB]. The degrees of freedom are all 6 and 63.

/s/-/Σ/	ME1	ME2	ME3	MJ1	MJ2	Mean	F-value[a]	p-value	Comparison
Shifting-Seafood	30	42	28	87	-16	34	8.14	<0.001	MJ1, ME2>ME1, ME3, MJ2
Shoot-suit	18	61	60	6	34	51	3.71	<0.011	MJ1, ME2, ME3, MJ2>ME1
Shutting-son	20	36	0	50	88	38	6.28	<0.001	MJ2, MJ1, ME2, ME1, ME3
Beer-veering	143	101	25	18	39	65	14.23	<0.001	ME1, ME2>MJ2, ME3, MJ1
Boom-view	185	64	117	13	39	83	16.85	<0.001	ME1>ME3, ME2>MJ2, MJ1
Bargain-virtual	89	19	39	-17	12	28	7.96	<0.001	ME1, ME3>ME2, MJ2, MJ1

Table 13: Mean distance of intensity between a pair of consonants for male speakers [dB]. The degrees of freedom are all 4 and 45.

peak produced by the native English speakers and the Japanese learners of English are presented in Figures 1 and 2.

Among three pairs of words measured by frequency of spectral peak, one pair, that holds /s/ or /S/ before /i/ present much shorter distance for the Japanese learners of English than the native English speakers. This phenomenon is explained with referring to an effect of regular palatalization of /s/ before /i/ in Japanese. The palatalization of /k/ before /i/ in English is simulated [12]. *Seafood* as *katakana* version is pronounced as [si]uùdo] and this Japanese [s^j] is similar to English /S/. *Katakana* letters are used for putting loanwords in Japanese words and that would affect the sound of these words in English produced by Japanese learners of English. Different *katakana* letters, サ[sa] and シャ[sja] or ス[su] and シュ[sju], are used for the pairs of /s/ or /S/ before /A/, and /s/ or /S/ before /u/, whereas the same letter, シ[s^ji] is used for the pair of /s/ or /S/ before /i/.

Virtual distance of /v/ and /b/ measured in duration produced by

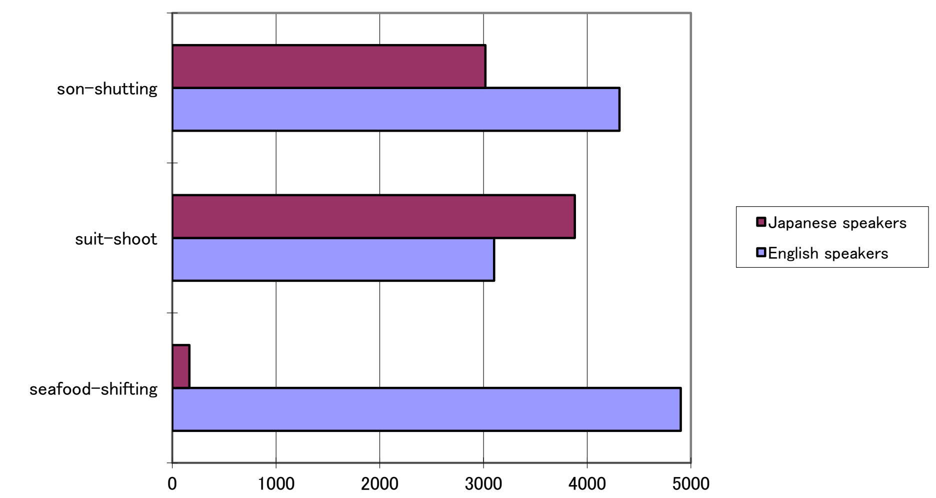

Figure 1: Distances of /s/ and /Σ/ in paired-words, 1: seafood-shifting, 2: suit-shoot, 3: son-shutting, measured by frequency of spectral peak of female speakers [Hz].

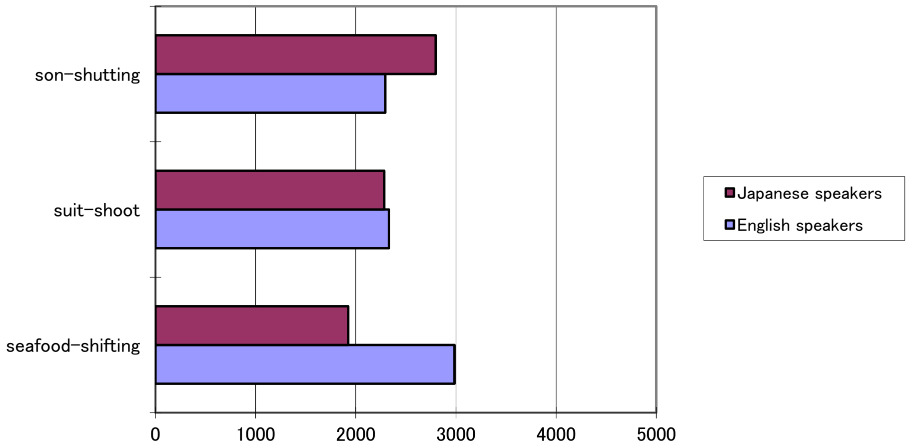

Figure 2: Distances of /s/ and /Σ/ in paired-words, 1: seafood-shifting, 2: suit-shoot, 3: son-shutting, measured by frequency of spectral peak of male speakers [Hz].

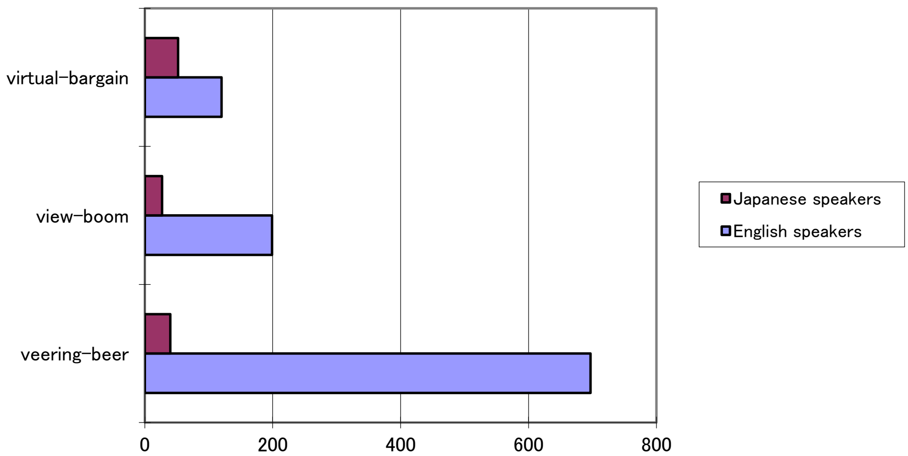

Figure 3: Distances of /v/ and /b/ in 1: veering-beer, 2: view-boom, 3: virtual-bargain, measured by duration of female speakers [msec].

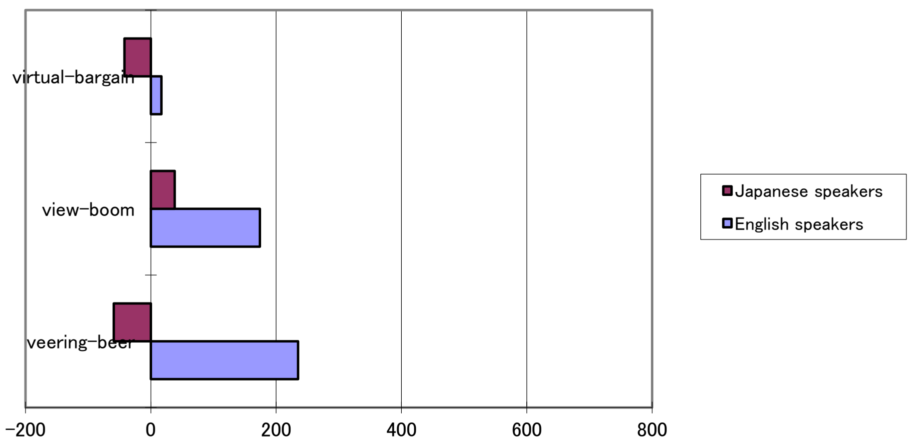

Figure 4: Distances of /v/ and /b/ in 1: veering-beer, 2: view-boom, 3: virtual-bargain, measured by duration of male speakers [msec].

native English speakers and Japanese learners of English are presented in Figures 3 and 4.

The distance of all the pairs of word measured by duration presents shorter one for Japanese learners of English than for native English speakers.

Virtual distance of /S/ and /s/, or /b/ and /v/ measured in intensity produced by native English speakers and Japanese learners of English is presented in Figures 5 and 6.

All the pairs of words measured in intensity present shorter distance for Japanese learners of English than for native English speakers. Among

Citation: Tomita K (2016) Virtual Space of English Consonants: Shorter Distance Produced by Japanese Learners of English. J Phonet and Audiol 2: 112. doi:10.4172/2471-9455.1000112

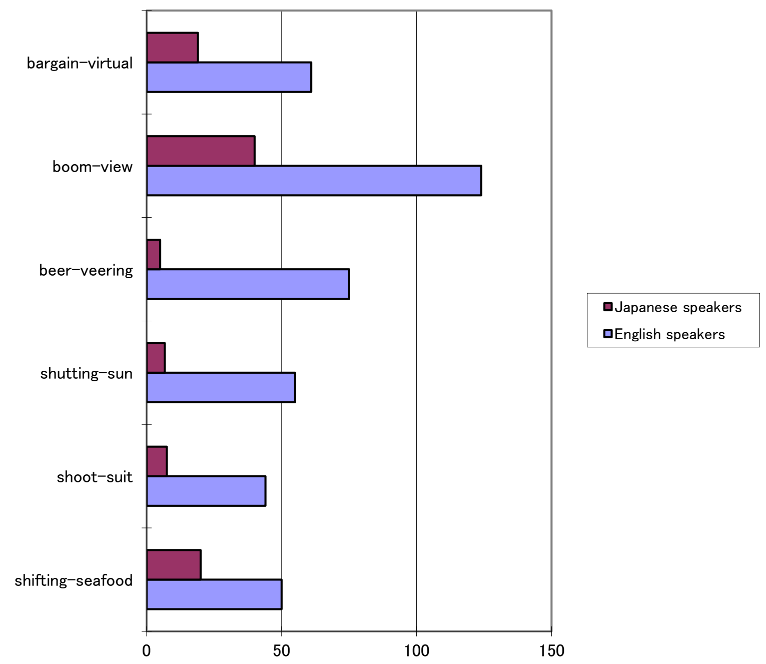

Figure 5: Distances of /Σ/ and /s/ in paired-words, 1: shifting-seafood, 2: shoot-suit, 3: shutting-son, and /b/ and /v/ in 4: beer-veering, 5: boom-view, 6: bargain-virtual, measured by intensity of female speaker [dB].

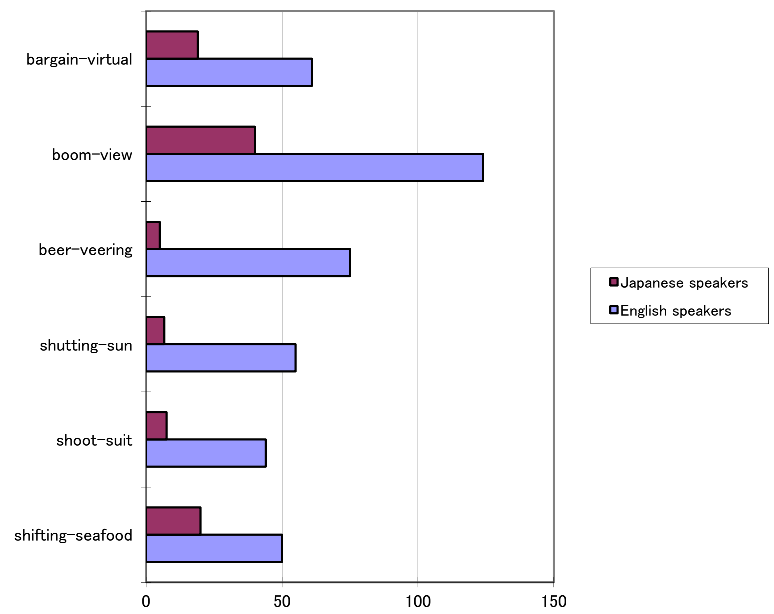

Figure 6: Distances of /Σ/ and /s/ in paired-words, 1: shifting-seafood, 2: shoot-suit, 3: shutting-son, and /b/ and /v/ in 4: beer-veering, 5: boom-view, 6: bargain-virtual, measured by intensity of male speaker [dB].

three types of phonological features used for the norms of analyses in this study, intensity presentes the largest differences between the native English speakers and the Japanese learners of English. These differences, which are reflected on the virtual space of English consonants as is called in this study, would be easier to grasp with visualized three-dimensions. Of course, the units used for each norm show different phonetic variations. Mean distances of all the paired words in each norm, frequency, duration and intensity are added and their means are calculated. They are put into three dimensions: frequency (4105 Hz for native English speakers and 2354 Hz for Japanese learners of English), duration (2537 msec. for native English speakers and 2335 msec. for Japanese learners of English), and intensity (474 dB for native English speakers and 18 dB for Japanese learners of English). Virtual space of English consonants in a cubic ellipse is calculated by multiplying the frequency, the duration, the intensity and 4/3π. The one produced by native English speakers is 20634 × 10⁶ Hz msec dB and the one produced by Japanese learners of English is 413 × 10⁶ Hz msec dB. Simple multiplication may not reveal the precise values of virtual space but these two numbers at least show that there is a big difference between what the native English speakers and the Japanese learners of English can deal with.

This study presents comparison of phonetic features, such as frequency of spectral peak, duration and intensity in the production of paired consonants, /S/ and /s/, or /b/ and /v/ between native English speakers and Japanese learners of English in the phonological contexts of _ /i/, _ /u/ or _ /A/. One of the points that is not hypothesized but newly observed in the results is that, among three types of phonological contexts, that of /i/ seems to affect the precedent consonants very much and produce the clear differences between the native English speakers and the Japanese learners of English. This might be because that the front part of vowel space for /i/ is very narrow. It would be a good way to focus on the phonological context of /i/ and observe the different pronunciation of consonants by the native English speakers and the Japanese learners of English.

Further research

The present study takes only a first step in broader research on Japanese learners' consonant qualities in English. As such, many problems and questions remain for future investigations. Some of them are mentioned here.

First of all, the vowel sound of /i/s or /I/s is used to form the phonological contexts in this study, but it would be better to select the either one. The same is true for the context of /A/. This study includes /A/, /Ã/ and /Î/ to arrange words of minimal paired consonants in the phonological context, but it would also better to select the one type from the vowel contexts of /A/, /Ã/ or /Î/.

Gestures required to contrast manner of articulation produced at some places are different from those at other places [13]. How to measure the phonetic features of each consonant in these different places precisely is an important issue. Researches focusing on a more dynamic cue are introduced and recommended [3]:

- Subsequent research on invariance focused on a more *dynamic* cue, namely the change in distribution of high-frequency energy as compared to low-frequency energy between consonant onset and the onset of the following vowel. This approach was a refinement of the earlier research in that it still captured the basic notion that bilabials are characterized by a relative predominance of energy in the lower frequencies while alveolars showed a predominance of energy in the higher frequencies. Using this dynamic criterion, 91 percent of the labial, dental, and alveolar plosives in English, French, and Malayalam were correctly classified.

From this point of view, the measurement of consonants, such as /b/ and /v/ is better to consider their dynamic change like the following characteristics described [3]:

- Since plosives are produced with a complete constriction followed by a release, the change in energy from plosive to vowel is relatively large, certainly larger than that for approximants, which have only a moderate constriction.

As for the distances, not only real distances produced by speakers with pronouncing two different consonants, but also so-called perceptual distances should be taken into consideration. Speakers know the perceptual distances between two phonological elements, and based on this knowledge, they attempt to minimize the perceptual disparity between two corresponding elements in phonology [14].

Furthermore, there are several items that need to be clarified. A first question to ask may be whether and how consonant quality as found in this study contributes to the putatively low intelligibility noted for spoken English words produced by Japanese learners of English. Although some educationists support the idea of lingua franca core, that produces so-called pronunciation norm for non-native speakers [15], others including the authors of this study still think these norms are not adequate to apply for language learning in classrooms. A second question of interest may be which English consonants are difficult for Japanese (and other language) speakers to acquire and why. A third question which the author finds interesting involves the possible variability in consonant quality within speakers.

Japanese speakers of this study distinguish paired consonants fairly well. It is, however, too early to conclude that Japanese learners of English acquire English sounds very well. The virtual space of consonants produced by the Japanese learners of English is much smaller than the one by the native English speakers. This means although the former distinguishes paired sounds very well, the degree of discrimination is less for them than the ones produced by the latter.

Acknowledgment

The authors wish to thank Emily Goodwill, Thomas Green Jr., Amber Numamoto, Aaron Molinas, Zoe Nieves, Ahren Kerwood, Risa Endo, Hitomi Hori, Keiko Sakai, Shunetsu Arai, Misato Kameyama, Tatsuhiro Higuchi for their active participation in the language experiment.

References

1. Hayes B (2009) Introductory Phonology. MA: Wiley-Blackwell.
2. Daulton FE (2007) Japan's Built-in Lexicon of English-based Loanwords. Toronto: Multilingual Matters LTD.
3. Reetz H, Jongman A (2009) Phonetics: Transcription, Production, Acoustics, and Perception. Oxford: Blackwell Publishing.
4. Kent RD, Read C (1992) The Acoustic Analysis of Speech. California: Singular Publishing Groups Inc.
5. Crystal D (2010) The Cambridge Encyclopedia of Language. 3rd edition Cambridge CUP.
6. Patel AD (2008) Music, Language, and the Brain. Oxford: Oxford University Press.
7. Shaw JA, Davidson L (2011) Perceptual similarity in input-output mappings: A computational/experimental study of non-native speech production. Lingua 121: 1344-1358.
8. Klatt DH (1975) Voice onset time, frication, and aspiration in word-initial consonant clusters. J Speech Hear Res 18: 686-706.
9. Klatt DH (1976) Linguistic uses of segmental duration in English: acoustic and perceptual evidence. J Acoust Soc Am 59: 1208-1221.
10. Shinn P (1984) A cross-language investigation of the stop, affricate and fricative manner of articulation. Providence RI.
11. Walker R (2011) Teaching the Pronunciation of English as a Lingua Franca. Oxford: Oxfored University Press.
12. Morley RL (2014) Implications of an exemplar-theoretic model of phoneme genesis: a velar palatalization case study. Lang Speech 57: 3-41.
13. Jong KJ, Hao YC, Park H (2009) Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. Journal of Phonetics 37: 357-373.
14. Kawahara S, Shinohara K (2009) The role of psychoacoustic similarity in Japanese puns: A corpus study. Journal of Linguistics 45: 111-138.
15. Jenkins J (2002) A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. Applied Linguistics 23: 83-103.