音声研究　第14巻第１号
2010(平成22)年４月
60–75頁

Journal of the Phonetic Society
of Japan, Vol.14 No.1
April 2010, pp.60–75

Cross-linguistic Production and Perception of Japanese- and Dutch-accented English

Mirjam Broersma*, Makiko Aoyagi** and Andrea Weber***

日本語とオランダ語の訛りの入った英語をめぐる言語相互の生成と知覚

ミリアム・ブルーズマ *，青柳真紀子 **，アンドレア・ウェバー ***

要旨：日本語話者とオランダ語話者が英単語を発話する際に見られる各言語特有の訛り（例：*act* をそれぞれ /'akto/ と /ɛkt/，*move* を /'mu:bu/ and /mu:f/ と発音）について，まず，それぞれの話者が両言語訛りで読み上げた音声の音響的比較を行った。その結果，両言語話者間に，母語訛りの音声だけでなく非母語訛りの音声にも，母語のリズム特性に起因すると思われる時間制御の違いが見られた。次に，それらの音声特徴と，日本語・オランダ語話者について行った英語語彙認識の聴覚プライム視覚語彙認識実験（Weber, Broersma and Aoyagi, 投稿中）の結果との関係を分析した。その結果，オランダ語話者にとっては /ɛkt/ のように /æ>ɛ/ の訛りが強いほど *act* の語彙認識が容易になるという例を除いては，両言語話者とも，母語訛りと非母語訛りの音声のいずれを聴いた場合でも，音響的に英語音声に近づくほど語彙認識が容易になることが分かった。

Key words: foreign-accented speech, speech production, L2 listening, spoken-word recognition, Dutch, Japanese

1. Introduction

Moobu or *moof*, which word is more easily recognizable as the English word *move*? We presented Japanese listeners with English words with a typical Japanese accent (Thompson 2001), pronouncing *act* as /'akto/, *move* as /'mu:bu/, and *indeed* as /ɪn'di:do/. We also presented them with English words with a typical Dutch accent (Tops, Dekeyser, Devriendt and Geukens 2001), pronouncing *act* as /ɛkt/, *move* as /mu:f/, and *indeed* as /ɪn'di:t/. Where /æ/ was replaced with /a/ in the Japanese-accented form, it became /ɛ/ in the Dutch-accented version, and where final voiced obstruents underwent vowel paragoge and final /v/ became /b/ in the Japanese-accented version, final obstruents were devoiced in the Dutch-accented version.

The two versions of the accented words were recorded both by a Japanese speaker and by a Dutch speaker. Thus, each speaker produced the English words in their own accent, but also in an accent they were unfamiliar with. We then presented Japanese and Dutch listeners with the recorded stimuli, and we hypothesized that the two listener groups should differ in the ease with which they recognized the accented words. We predicted that both experience with an accent but also perceptual confusability with the correct English pronunciation would influence spoken-word recognition.

Japanese speakers are commonly found to add a vowel after final obstruents in English, to retain the common Japanese CV syllable structure (Tanaka 2009),

Fig. 1　Illustration of vowel epenthesis in written form: *grand* misspelled as *grando*; Maruyama Park, Kyoto.

* Research Scientist, Donders Centre for Cognition, Radboud University Nijmegen（ラトバウト大学ナイメーヘン，ドンデルス認知センター研究員）
** Associate Professor, Department of English, Faculty of Foreign Languages, Dokkyo University（獨協大学外国語学部英語学科准教授）
*** Research Group Leader, Max Planck Institute for Psycholinguistics（マックスプランク心理言語学研究所研究グループ長）

such that *indeed* /ɪn'di:d/ becomes *indeedo* /ɪn'di:do/, and *grand* becomes *grando*, sometimes even in written form, as illustrated in Fig. 1. Japanese has no labiodental fricatives, and Japanese speakers often replace /v/ with /b/ in English (Thompson 2001), pronouncing *move* /mu:v/ as *mooboo* /'mu:bu/. Further, Japanese has only five vowels, including /a/ but not /æ/, and Japanese learners tend to perceive and produce English /æ/ as /a/ (Nishi and Kewley-Port 2007), such that *happy* /'hæpɪ/ is pronounced as *hAppy* /'hapɪ/. Dutch has voiced and voiceless obstruents (/b/, /p/, /d/, /t/, /k/, /v/, /f/, /z/, /s/), but due to final devoicing only voiceless obstruents occur at the end of words in isolation. Whereas Dutch listeners can distinguish English voicing contrasts in final position accurately in phonetic categorization tasks (Broersma 2005, 2008), they do not use the voicing distinction accurately for word recognition; thus, they recognize /mu:f/ pronounced by an English speaker as the English word *move* (Broersma and Cutler 2008). Dutch speakers often pronounce final voiced obstruents as voiceless in English (Tops et al. 2001), e.g., *indeed* pronounced as *indeet* /ɪn'di:t/ and *move* as *moof* /mu:f/. Dutch has 19 monophthongs and 3 diphthongs (Gussenhoven 1999) but lacks /æ/; Dutch learners commonly produce and perceive English /æ/ as /ɛ/ (which is present in Dutch) (Tops et al. 2001). Thus, Dutch learners pronounce *happy* as *heppy* /'hɛpɪ/.

Experience with one's own accent should make it easier for Japanese listeners to recognize words with a typical Japanese accent, and for Dutch listeners to recognize words with a typical Dutch accent. For Dutch listeners, the substitutions in words with a typical Japanese accent should perceptually clearly stand out, and deviate from the English norm. Based on the phonology of Dutch (Gussenhoven 1999), the contrast between /æ/ and /a/, as well as between /v/ and /b/ should be easy to distinguish for Dutch listeners, and the structural change due to the addition of a word-final vowel should be perceived as a clear mismatch with the English pronunciation too. Words with a typical Dutch accent, on the other hand, might not be noticeably different from the English norm for Japanese listeners. The difference between English /æ/ and Dutch /ɛ/ might go unnoticed as the Japanese phoneme inventory (Shibatani 1990) does not include the vowels /æ/ and /ɛ/ (indeed, Japanese listeners frequently confuse the English /æ/ and /ɛ/; Nishi and Kewley-Port 2007), the devoicing of final /b/, /d/, and /v/ might go unnoticed as Japanese has no word final obstruents, and Japanese listeners have difficulty distinguishing /v/ and /f/ in general (Takata and Nabelek 1990).

The Japanese-accented English speech should thus lead to fast and accurate word recognition for Japanese listeners only, whereas the Dutch-accented speech should be recognized well both by Dutch and by Japanese listeners. Indeed, the results reported in Weber, Broersma and Aoyagi (submitted) showed exactly this pattern. In a series of Cross-Modal Priming experiments, hearing a word in one's native accent facilitated the subsequent recognition of the same word in written form in comparison with recognition of that word following an unrelated prime word (Fig. 2). Significantly faster RTs following accented prime words indicate that the listeners had interpreted the auditory prime word correctly. For the Japanese listeners, hearing a Dutch-accented word also facilitated recognition of the written word, but for Dutch learners, hearing a Japanese-accented word did not lead to facilitation. Thus, Dutch-accented words were well recognizable for both groups of listeners, and Japanese-accented words only for the Japanese listeners.

Surprisingly, the results of the listening experiments were the same for the recordings of the two speakers. It did not matter, for Japanese or for Dutch listeners, whether they heard the Japanese or the Dutch speaker producing the two accent forms; the pattern of results stayed the same. Yet, the two speakers sounded clearly distinct. Japanese listeners (for example those present at the *International Phonetics and Phonology Forum 2009* in Kobe, Japan) reported that the Japanese-accented speech produced by the Dutch speaker, although easily understandable, was clearly not spoken by a Japanese speaker, and Dutch listeners reported the same for the Dutch-accented speech produced by the Japanese speaker. Both speakers were, of course, very

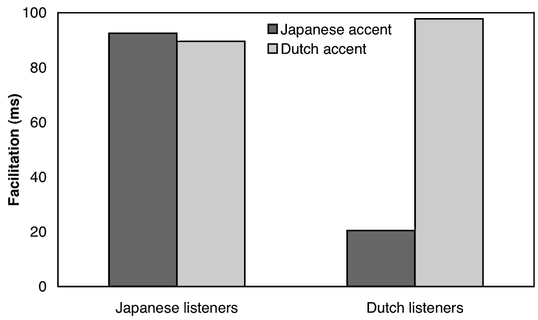

Fig. 2 Facilitation effects in the results of Weber, Broersma and Aoyagi (submitted), calculated as the difference between reaction times of correct responses after a Japanese- or Dutch-accented word and those after an unrelated word.

familiar with their native language accent, but not familiar with the accent of the other language they were instructed to produce. Although they could produce accented speech that was, on a segmental level, convincing enough to lead to accurate word recognition in the listening experiments, their pronunciation was not intended to be more than a rough approximation of the other-language accented speech. This rough approximation of the accent of another language was apparently sufficient to significantly facilitate word recognition for listeners who came from that language background, but in this paper we want to do a more finely tuned analysis of the influence of speaker background on word recognition.

For this purpose we acoustically analyze the speech materials in more detail and assess the effect of subtle acoustic differences on word recognition performance. First, we investigate the differences between the Japanese and the Dutch speakers' pronunciation of Japanese- and Dutch-accented English. As Japanese is mora-timed, whereas Dutch is stress-timed (e.g., Vroomen, Van Zon and De Gelder 1996), durational differences between the two speakers' productions might be expected, in particular in stressed syllables. Further, in the case of vowel epenthesis, as Dutch has long vowels but not short vowels in open syllables, the Dutch speaker might produce longer final vowels than the Japanese speaker. In order to assess these possible differences, the two speakers were compared on a number of acoustic measures.

Next, for both speakers, the phonetic details of the Japanese- and Dutch-accented stimuli were compared. All stimuli of course contained the required segments, but acoustic details might have varied between the Japanese- and Dutch-accented stimuli. Acoustic measurements were analyzed to ascertain that there were no unforeseen differences in the pronunciation of the Japanese-accented and Dutch-accented stimuli at a subsegmental level.

Finally, we relate the production with the perception side of the story by asking how listeners responded to the variation in the stimuli, within speakers and within stimulus types. It seems plausible that, when listening to the other-language accented speech, a pronunciation that approaches the English norm more must be easier to understand than a strongly accented pronunciation unfamiliar to the listener. When listening to their native language accent, on the other hand, two options seem possible: listeners might find it easier to recognize words pronounced with a strong accent than words more in line with the correct English pronunciation, because they are very familiar with their own accent; alternatively, they might still benefit from a more correct, English-like pronunciation, because it approximates the norms of the second language.

2. Materials

Four different types of stimuli were selected, with 12 mono- and disyllabic English words per type. Words of the first type contained the vowel /æ/ in canonical form (e.g., *act* /ækt/), which was replaced by /a/ in the Japanese-accented form, and depending on the final consonant, the vowel /u/, /o/, or /i/ was appended (e.g., *Acto* /'akto/). In the Dutch-accented form the vowel /æ/ was replaced by /ɛ/ (*ect* /ɛkt/). Words of the second type also contained the vowel /æ/, but ended in a vowel, /r/, or /n/, after which Japanese speakers do not typically add a vowel (Tanaka 2009); thus, *happy* /'hæpɪ/ became *hAppy* (/'hapɪ/) in Japanese-accented form, and *heppy* (/'hɛpɪ/) in Dutch-accented form. Words of the third type ended with a /v/ in canonical form (e.g., *move* /mu:v/). In Japanese-accented form the final /v/ was replaced with /b/ and the vowel /u/ was added (*mooboo* /'mu:bu/); in Dutch-accented form, the final /v/ was replaced with /f/ (*moof* /mu:f/). Words of the fourth type ended with a /d/ in canonical form (e.g., *indeed* /ɪn'di:d/); the vowel /o/ was added in the Japanese-accented form (*indeedo* /ɪn'di:do/), and final /d/ was replaced by /t/ in the Dutch-accented form (*indeet* /ɪn'di:t/).

One female native speaker of Japanese and one female native speaker of Dutch, students at Dokkyo University and Radboud University Nijmegen, respectively, recorded all Japanese- and Dutch-accented words multiple times, in clear citation style, in a soundproof booth with a Sennheiser microphone. Speakers had a moderate accent in English and a basic knowledge of phonetics. For both speakers, native language accented items were presented in regular English spelling. Other-language accented items were presented both in regular and in modified spelling; the speaker was asked to produce the modified spelling, and consulted the regular spelling in case of uncertainty. One token per speaker of each item was selected by the authors and excised from the recording using the speech editor Praat. For more details about the construction of the materials, see Weber et al. (submitted). After having recorded the four types of items, the speakers were asked to read aloud the 11 American English vowels in a hVbə context three times, following the procedure described in Strange et al. (2007), to determine the speakers' vowel space in English. Note that even though the hVbɔ nonword items were visually preceded by two English example words (e.g., *creep* and *steep* for recording *heeba*), the speakers' productions were not necessarily modeled on correct American English pronunciation (i.e., the speakers did not hear the correct AE pronunciation immediately before they produced the hVbɔ nonwords, neither did we correct their pronunciation).

3. Acoustic Measurements

First, for the hVbɔ nonwords, F1 and F2 were measured at temporal midpoint of the first vowel. Averaging over the three tokens of each vowel, both speakers' English vowel charts were drawn. Fig. 3a shows that for the Japanese speaker, all English vowels were centered around five locations, presumably corresponding to the five vowels of Japanese. For the Dutch speaker (Fig. 3b), the vowel space approached the English distribution better. However, as expected, the Dutch speaker did not differentiate between English /ɛ/ and /æ/.

Next, averages of acoustic measurements per speaker,

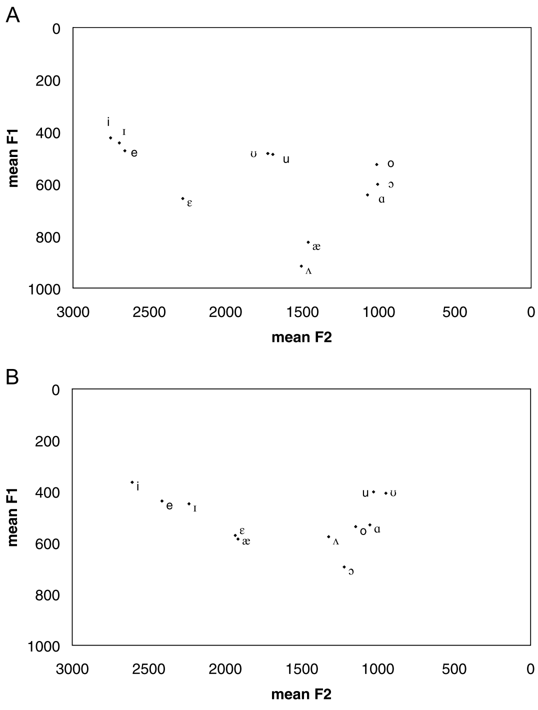

Fig. 3 F1/F2 (Hz) plots of 11 English vowels.
A: Japanese speaker; B: Dutch speaker.

accent, and stimulus type are presented in Table 1. For items with /a/ or /ɛ/ (i.e., *Acto, ect, hAppy, heppy*), the duration of /a/ or /ɛ/ was measured (hence: Vowel Duration), as well as F1, F2, and F3 at the temporal midpoint of that vowel. The ratio of Vowel Duration divided by the total stimulus duration was calculated (Relative Vowel Duration). For items of the *mooboo*, *moof*, *indeedo*, and *indeet* type, the duration of the vowel preceding the last consonant was measured (Vowel Duration), and Relative Vowel Duration was calculated as above. For items with vowel epenthesis (i.e., *Acto, mooboo, indeedo*), the duration of the final vowel, and F1, F2, and F3 at final vowel midpoint were measured (hence: Final Vowel Duration, Final Vowel F1, etc.), and the ratio of Vowel Duration divided by Final Vowel Duration was calculated (Relative Duration Two Vowels). For items with a final /f/ (i.e., *moof*), the duration of the final fricative was measured (Fricative Duration), and Fricative Duration divided by total stimulus duration was calculated (Relative Fricative Duration). For items of the *mooboo* type, the duration of the closure preceding the final stop (Closure Duration) and Vowel Duration plus Closure Duration was determined (Vowel Plus Closure Duration). For *indeedo* and *indeet* type items, only Vowel Plus Closure Duration was determined (as for some of those items, the last stop was preceded by /n/, e.g., *behindo*, such that Closure Duration could not be determined separately). For *mooboo*, *indeedo*, and *indeet* type items, Vowel Plus Closure Duration divided by total stimulus duration was calculated (Relative Vowel Plus Closure Duration). For those same items, the duration of the closure period without voicing was measured, and divided by, first, Closure Duration (Proportion Voiceless Closure; note that this could be determined even for stimuli with /n/ preceding the last stop, as the duration of closure without voicing was always 0 in those cases) and, second, total stimulus duration (Relative Voiceless Closure Duration); F1 at offset of the vowel before the last stop was determined (Closure F1), the duration of the release burst of the last stop was measured (Burst Duration), and Burst Duration divided by total stimulus duration was calculated (Relative Burst Duration).

4. Results and Discussion

4.1 Speaker differences

Acoustic measures were analyzed to assess whether the two speakers differed in the pronunciation of the stimuli. For each stimulus type, the acoustic measurements that could be meaningfully compared across

Table 1　Averages of acoustic measures (durations: ms; formants: Hz; relative measures: ratios).

	Japanese speaker		Dutch speaker	
	Japanese accent	**Dutch accent**	**Japanese accent**	**Dutch accent**
Acto – ect				
Vowel Duration	111.0	105.3	95.1	156.3
Relative Vowel Duration	.152	.144	.141	.228
F1	817.4	686.2	918.0	735.9
F2	1623.2	2251.6	1981.2	2009.0
F3	3090.8	3151.0	3639.3	2883.4
Final Vowel Duration	149.1	—	238.5	—
Final Vowel F1	482.6	—	335.1	—
Final Vowel F2	1772.8	—	1387.1	—
Final Vowel F3	2776.8	—	2538.3	—
Relative Duration Two Vowels	.755	—	.403	—
hAppy – heppy				
Vowel Duration	120.6	108.2	111.3	127.6
Relative Vowel Duration	.180	.150	.168	.178
F1	842.3	721.5	988.5	706.4
F2	1577.4	2190.2	2279.5	2008.7
F3	3047.5	3166.6	3754.0	2850.8
mooboo – moof				
Vowel Duration	215.2	212.2	152.4	302.1
Relative Vowel Duration	.321	.299	.231	.418
Closure Duration	44.4	—	32.2	—
Vowel Plus Closure Duration	259.5	—	184.6	—
Burst Duration	41.0	—	60.1	—
Closure F1	262.5	—	262.1	—
Proportion Voiceless Closure	.117	—	.153	—
Relative Vowel Plus Closure Duration	.386	—	.280	—
Relative Voiceless Closure Duration	.006	—	.010	—
Relative Burst Duration	.063	—	.088	—
Final Vowel Duration	173.9	—	263.7	—
Final Vowel F1	460.6	—	330.3	—
Final Vowel F2	1751.3	—	960.4	—
Final Vowel F3	2695.2	—	2451.4	—
Relative Duration Two Vowels	1.313	—	.607	—
Fricative Duration	—	304.2	—	228.1
Relative Fricative Duration	—	.426	—	.319

Table 1 (continued)

	Japanese speaker		Dutch speaker	
	Japanese accent	**Dutch accent**	**Japanese accent**	**Dutch accent**
indeedo – indeet				
Vowel Duration	235.2	206.1	157.2	285.8
Relative Vowel Duration	.312	.256	.224	.385
Vowel Plus Closure Duration	273.2	298.8	177.5	337.1
Burst Duration	38.0	124.4	30.1	77.6
Closure F1	537.1	486.6	263.7	299.5
Proportion Voiceless Closure	.018	.750	0	.302
Relative Vowel Plus Closure Duration	.362	.372	.253	.454
Relative Voiceless Closure Duration	.001	.089	0	.019
Relative Burst Duration	.051	.156	.044	.104
Final Vowel Duration	135.7	—	228.5	—
Final Vowel F1	547.0	—	486.3	—
Final Vowel F2	1183.5	—	1166.7	—
Final Vowel F3	2641.3	—	2514.5	—
Relative Duration Two Vowels	1.779	—	.705	—

speakers (including relative durational cues and proportions, but not absolute durations and spectral cues) were analyzed in a multiple regression analysis (Method Stepwise), with Speaker as dependent variable. When two variables were correlated too strongly (with r>.9), only the variable that correlated with the dependent variable most strongly was used in the regression analysis. Table 2 shows which variables were considered for analysis ('possible variables'), which ones were actually used (the variables that were excluded due to high collinearity are indicated with '-' in the second column), which predictors were included in the final regression models, and their relative importance in the final regression models (with a larger Beta value, either positive or negative, indicating greater importance).

The multiple regression analyses indicate which of the acoustic measures distinguish best between the two speakers, taking into account that the acoustic measures are often correlated with one another. If an acoustic measure is not included in the final regression model, that does not imply that the measure does not differ for the two speakers, but only that other cues differentiate the two speakers better.

For Japanese-accented items of the *Acto* type, the predictors were Relative Vowel Duration and Relative Duration Two Vowels. Both were included in the final regression model, with the latter being the most important predictor. Relative Duration Two Vowels was larger for the Japanese than for the Dutch speaker; thus, as predicted, the Dutch speaker had a relatively longer final vowel than the Japanese speaker. The role of Relative Vowel Duration is more difficult to interpret. Table 1 shows that this ratio was larger for the Japanese than for the Dutch speaker; the Beta weight (Table 2), however, indicates an effect into the opposite direction. This suggests that the predictor functions as a suppressor variable. Indeed, Relative Vowel Duration has a higher correlation with the other predictor, Relative Duration Two Vowels (r=.444), than with the dependent variable Speaker (r=.191). We therefore assume that Relative Vowel Duration contributes to the regression model by increasing the predictive power of the other predictor, and not by its own relation with the dependent variable Speaker.

For Japanese-accented items of the *mooboo* and *indeedo* type, a regression model was formed with Relative Duration Two Vowels, with a higher value for the Japanese than for the Dutch speaker. This indicates that the Japanese speaker produced relatively shorter final vowels than the Dutch speaker, as predicted, and

Table 2 Speaker differences. Multiple regression models for different item types; possible variables, variables included in the regression analysis, variables included in the final regression model, Beta weight (+ sign indicates positive and - sign negative correlation with the Japanese speaker) and *p* value.

Possible variables	Variables included in analysis	Variables included in model	Beta	*p*<
Japanese accent: *Acto* *F*(2, 23)=54.3, *p*<.001; *adjusted* R^2: .823				
Relative Vowel Duration	Y	Y	–.253	.05
Relative Duration Two Vowels	Y	Y	.999	.001
Japanese accent: *hAppy*: –				
Relative Vowel Duration	Y	–		
Japanese accent: *mooboo & indeedo* *F*(1, 47)=56.0, *p*<.001; *adjusted* R^2: .539				
Relative Vowel Duration	–	–		
Proportion Voiceless Closure	–	–		
Relative Vowel Plus Closure Duration	Y	–		
Relative Voiceless Closure Duration	Y	–		
Relative Burst Duration	Y	–		
Relative Duration Two Vowels	Y	Y	.741	.001
Dutch accent: *ect & heppy* *F*(1, 47)=12.7, *p*<.001; *adjusted* R^2: .200				
Relative Vowel Duration	Y	Y	–.465	.001
Dutch accent: *moof* *F*(2, 23)=17.2, *p*<.001; *adjusted* R^2: .584				
Relative Vowel Duration	Y	Y	–.516	.001
Relative Fricative Duration	Y	Y	.562	.001
Dutch accent: *indeet* *F*(1, 23)=38.3, *p*<.001; *adjusted* R^2: .618				
Relative Vowel Duration	Y	Y	–.797	.001
Proportion Voiceless Closure	–	–		
Relative Vowel Plus Closure Duration	–	–		
Relative Voiceless Closure Duration	Y	–		
Relative Burst Duration	Y	–		

similar to the *Acto* type items.

For Dutch-accented items of the *ect* and *heppy* type, only Relative Vowel Duration could be compared between speakers. A regression model was formed with this predictor, showing that the Japanese speaker produced /ɛ/ vowels with a shorter relative duration than the Dutch speaker did.

For Dutch-accented /f/-final stimuli (*moof*), a regression model was formed with the predictors Relative Fricative Duration and Relative Vowel Duration. Relative Fricative Duration was longer and Relative Vowel Duration shorter for the Japanese speaker than for the Dutch speaker.

For Dutch-accented /t/-final items (*indeet*), as for the

/f/-final items, Relative Vowel Duration was longer for the Dutch than for the Japanese speaker.

These analyses show that the speakers differed on several details of the acoustic make-up of the stimuli. In the Japanese-accented vowel epenthesis case, as expected, the Dutch speaker produced relatively longer final vowels than the Japanese speaker did, most likely because Dutch has long but not short vowels in open syllables. For Dutch-accented items, the Japanese speaker produced a shorter /ɛ/, shorter final /t/-preceding vowels, and shorter final /f/-preceding vowels than the Dutch speaker did. Crucially, those vowels occurred (with three exceptions) in stressed syllables. As Dutch is a stress-timed language, where word stress is expressed by (among other things) vowel duration, whereas Japanese is mora-timed, this difference in stressed vowel duration is likely to be due to the speakers' native language rhythmic structure.

The instructions that the speakers received might have contributed to the speaker differences. For the Dutch but not for the Japanese speaker, the suffixed vowel was spelled out in the written stimuli, which may have induced the Dutch speaker to produce longer vowels. For the Dutch-accented items, the Japanese but not the Dutch speaker was explicitly instructed to produce /ɛ/, /f/ and /t/; the Japanese speaker produced a more /ɛ/-like vowel than the Dutch speaker did (in English /ɛ/ is shorter than /æ/; Flege, Bohn and Jang 1997), a more voiceless /f/ (with both longer fricatives and shorter preceding vowels; Watson 1983) and a more clearly voiceless final /t/ (with shorter preceding vowels; Watson 1983).

Thus, due to interference from the speakers' native language phonology, and possibly also the instructions they received, timing differences were found between the two speakers' renditions of both accents.

4.2 Accent differences

To assess how the pronunciation of the two accent types varied, acoustic measurements were compared for the Japanese-accented and Dutch-accented version of the stimuli, for the Japanese and Dutch speakers separately. For each stimulus type, the acoustic measurements that could be meaningfully compared between the Dutch-accented and Japanese-accented version of each word were analyzed in a multiple regression analysis (as above). The independent variable was Accent; the multiple regression analyses thus indicate which of the acoustic measures distinguish between the two accent types best. Tables 3 and 4 show for each speaker which predictors were used, and which ones were included in the final regression models, and their importance in those models.

For *Acto-ect* type items, for the Japanese speaker, F2 was higher for Dutch-accented /ɛ/ than for Japanese-accented /a/, as to be expected (compare Ladefoged 1999, Shibatani 1990). For the Dutch speaker, F1 was higher for /a/ than for /ɛ/, again as to be expected, and Vowel Duration of Dutch-accented /ɛ/ was longer than that of Japanese-accented /a/, in line with the common finding that the duration of vowels is generally longer in shorter words (e.g., Salverda, Dahan and McQueen 2003).

For *hAppy-heppy* type items, for the Japanese speaker, again, F2 was higher for Dutch-accented /ɛ/ than for Japanese-accented /a/. For the Dutch speaker F1, again, and F3 were higher for /a/ than for /ɛ/.

For *mooboo-moof* type items, for the Dutch speaker, Vowel Duration was longer in the Dutch-accented *moof* items than in the Japanese-accented *mooboo* items, similar to the finding for the *Acto-ect* type items, and in line with the commonly found pattern of longer vowel duration in shorter words.

For *indeedo-indeet* type items, for the Japanese speaker, Proportion Voiceless Closure was included in the final regression model, with a higher ratio for Dutch-accented /t/ final items than for Japanese-accented /do/ final items. Thus, there was more voicing during the closure before a voiced Japanese-accented stop than before a voiceless Dutch-accented stop, as to be expected. For the Dutch speaker, Vowel Plus Closure Duration was longer for the Dutch-accented than for the Japanese-accented items. This is again likely to be due to the difference in vowel duration in longer versus shorter words, as for the *Acto-ect* and *mooboo-moof* type items.

Thus, the differences between the Japanese- and Dutch-accented stimuli are all in line with the expectations. They include spectral differences for the /a/-/ɛ/ distinction for both speakers, a difference in voicing during closure preceding /d/ versus /t/ for the Japanese speaker, and differences in Vowel Duration when the Japanese-accented stimuli were one syllable longer than the Dutch-accented stimuli for the Dutch speaker. All differences between the Japanese- and Dutch-accented stimuli were in the right direction, corresponding to the segmental structure of the items the speakers were asked to produce, and no unforeseen differences were found between the Japanese-accented and the Dutch-accented stimuli.

Table 3 Japanese speaker, accent differences. Multiple regression models for different item types; possible variables, variables included in the regression analysis, variables included in the final regression model, Beta weight (+ sign indicates positive and – sign negative correlation with the Japanese accent) and *p* value.

Possible variables	Variables included in analysis	Variables included in model	Beta	*p*<
Acto – ect $F(1, 23)$=222.1, p<.001; *adjusted* R^2: .906				
Vowel Duration	Y	–		
Relative Vowel Duration	Y	–		
F1	Y	–		
F2	Y	Y	–.954	.001
F3	Y	–		
hAppy – heppy $F(1, 23)$=156.4, p<.001; *adjusted* R^2: .871				
Vowel Duration	Y	–		
Relative Vowel Duration	Y	–		
F1	Y	–		
F2	Y	Y	–.936	.001
F3	Y	–		
mooboo – moof				
Vowel Duration	Y	–		
Relative Vowel Duration	Y	–		
indeedo – indeet $F(1, 23)$=30.8, p<.001; *adjusted* R^2: .565				
Vowel Duration	Y	–		
Relative Vowel Duration	Y	–		
Vowel Plus Closure Duration	Y	–		
Closure F1	Y	–		
Proportion Voiceless Closure	Y	Y	–.764	.001
Relative Vowel Plus Closure Duration	Y	–		
Relative Voiceless Closure Duration	–	–		
Relative Burst Duration	–	–		

4.3 Perceptual relevance

To what extent did the variation in the pronunciation affect Japanese and Dutch listeners' recognition of the Japanese- and Dutch-accented words? We investigate whether a more English-like, less strongly accented, pronunciation facilitated recognition, both when listeners heard their native language accented speech, and when they heard the other-language accented speech. To assess how the pronunciation of the two accent types affected word recognition, reaction times (RTs) of correct responses following accented words in the Cross-Modal Priming listening experiments (Weber et al. submitted) were related to the newly obtained acoustic measurements. Shorter lexical decision times

Table 4 Dutch speaker, accent differences. Multiple regression models for different item types; possible variables, variables included in the regression analysis, variables included in the final regression model, Beta weight (+ sign indicates positive and – sign negative correlation with the Japanese accent) and *p* value.

Possible variables	Variables included in analysis	Variables Included in model	Beta	*p*<
Acto – ect *F* (2, 23)=19.1, *p*<.001; *adjusted* R^2: .612				
Vowel Duration	Y	Y	–.403	.05
Relative Vowel Duration	Y	–		
F1	Y	Y	.528	.01
F2	Y	–		
F3	Y	–		
hAppy – heppy *F*(2, 23)=44.6, *p*<.001; *adjusted* R^2: .791				
Vowel Duration	Y	–		
Relative Vowel Duration	Y	–		
F1	Y	Y	.521	.01
F2	Y	–		
F3	Y	Y	.417	.05
mooboo – moof *F*(1, 23)=32.3, *p*<.001; *adjusted* R^2: .567				
Vowel Duration	Y	Y	–.771	.001
Relative Vowel Duration	–	–		
indeedo – indeet *F* (1, 23)=68.4, *p*<.001; *adjusted* R^2: .746				
Vowel Duration	–	–		
Relative Vowel Duration	–	–		
Vowel Plus Closure Duration	Y	Y	–.870	.001
Closure F1	Y	–		
Proportion Voiceless Closure	Y	–		
Relative Vowel Plus Closure Duration	–	–		
Relative Voiceless Closure Duration	–	–		
Relative Burst Duration	–	–		

generally imply faster word recognition. Note that lexical decisions were made to written stimuli that were presented immediately at offset of the auditory stimuli, and RTs were also measured from offset of the auditory stimuli. To investigate the effect of variation in production on the perception of the stimuli, multiple regression analyses were done with acoustic measurements (as above), separately for each accent, stimulus type, and speaker. The independent variables were the RTs of the Japanese listeners' and the Dutch listeners' correct responses. (Note that very few errors were made in the experiments, such that RTs of correct responses and not error rates are the appropriate measure.) Table 5 shows which predictors were used, and which ones were included in the final regression models, and their importance in those models. Table 5 also indicates which combinations of speakers and listeners were tested; for each item type, one group of Japanese listeners was exposed to either the Japanese or the Dutch speaker, and two groups of Dutch listeners to one speaker each.

For Japanese-accented *Acto* type items, for the Japanese speaker and Dutch listeners, F2 and Final Vowel F3 were included in the final regression model. For those listeners, a higher F2 and a lower Final Vowel F3 were associated with shorter RTs. The English vowel /æ/ is more fronted, with a higher F2, than the Japanese /a/ (Ladefoged 1999, Shibatani 1990). Thus, if F2 was closer to English /æ/, Dutch listeners recognized the words faster than when the vowel was clearly less fronted than the English vowel. Here, similarity to the correct English pronunciation aided the Dutch listeners' recognition of the stimuli.

For *mooboo* type items, for the Dutch speaker and Dutch listeners, three predictors were included in the final regression model; in order of importance: Burst Duration, Closure Duration, and Proportion Voiceless Closure. Responses were faster when the release burst and the closure were shorter, both of which are associated with more voiced stops (Watson 1983). Thus, for Dutch listeners, if the /b/ was more clearly voiced, recognition of the words was easier. For Proportion Voiceless Closure, on the other hand, a higher proportion, indicating relatively less voicing, was associated with shorter RTs; however, this was a less important predictor in the regression model, and may have served as a suppressor variable for Closure Duration; its correlation with that predictor was stronger than with the dependent variable (with r=.364 and r=.110, respectively).

For *indeedo* type items, for the Japanese speaker and Japanese listeners, Closure F1 and Proportion Voiceless Closure (in order of importance) were included in the final regression model. A lower Closure F1 led to shorter RTs. Lower Closure F1 is associated with more voicing of the following stop (Watson 1983), and led to faster word recognition for the Japanese listeners. Like for *mooboo* type items, Proportion Voiceless Closure is difficult to interpret and seems to have served as a suppressor variable. Like for *mooboo* type items, less voicing during closure was associated with shorter RTs, but the correlation of this predictor with the other predictor in the model was stronger than with the dependent variable, with r=.605 and r=−.002, respectively. Next, for the Japanese speaker and the Dutch listeners, like for the Japanese listeners, lower Closure F1 led to shorter RTs. Further, lower Final Vowel F3 values led to shorter RTs. Thus, for both Japanese and Dutch listeners, a more clearly voiced final /d/, in terms of Closure F1, led to easier word recognition.

For Dutch-accented *ect* type items, for the Dutch speaker and Dutch listeners, lower F1 was related to shorter RTs. The Dutch /ɛ/ is higher, with lower F1, than the English /æ/. For the Dutch listeners, vowels with F1 values more like that of the Dutch vowel led to faster word recognition than vowels with an F1 more like that of the English target vowel. Here, similarity to the Dutch vowel, and dissimilarity from the correct English pronunciation, facilitated word recognition for the Dutch listeners.

For *heppy* type items, for the Japanese speaker and Japanese listeners, F3 was included in the final regression model, with lower F3 values leading to faster recognition.

For *indeet* type items, for the Japanese speaker and Japanese listeners, Relative Voiceless Closure Duration was included in the final regression model, with shorter durations correlated with shorter RTs. The duration of the closure itself could not be determined for all stimuli, but as both a shorter closure duration and a relatively short duration of voiceless section during closure is associated with more voiced stops (Watson 1983), it can be taken that shorter Relative Voiceless Closure Duration indicates more voiced-like stops. More voiced-like stops thus led to faster word recognition for the Japanese listeners. Next, for the Japanese speaker and Dutch listeners, longer Relative Vowel Duration led to shorter RTs. As long preceding vowels are associated with voiced stops (Peterson and Lehiste 1960), more voicing led to faster word recognition for the Dutch listeners too, similar to the Japanese listeners. For the Dutch speaker and Dutch listeners, two predictors were included in the final regression model: Relative Vowel Plus Closure Duration, and Closure F1 (in order of importance). With respect to the former predictor, longer Relative Vowel Plus Closure Duration led to shorter RTs. Longer vowels are associated with more voiced stops; longer closure durations, on the other hand, are associated with more voiceless stops. The effect of vowel duration and closure duration could not be separated here, but it seems likely that the much longer vowel duration annihilates the effect of closure duration. Thus, we interpret this predictor to show that more voiced-like stops led to faster word recognition. The effect of the second predictor goes in the same direction, with lower Closure F1, indicating more voicing of the final stop, associated with shorter RTs.

Table 5 Perceptual effects. Multiple regression models for different item types; possible variables, variables included in the regression analysis, variables included in the final regression model, Beta weight (+ sign indicates positive and – sign negative correlation with reaction times of correct responses) and *p* value. For each item type, three combinations of speakers and listeners are presented. In the second and third column, the three values refer to those combinations.

Possible variables	Variables included in analysis	Variables included in model	Beta	$p<$
Japanese accent: *Acto* Japanese speaker, Dutch listeners: $F(2, 11)=13.9$, $p<.01$; *adjusted* R^2: .701 Dutch speaker, Japanese listeners: – Dutch speaker, Dutch listeners: –				
Vowel Duration	Y Y Y	– – –		
Relative Vowel Duration	Y Y Y	– – –		
F1	Y Y Y	– – –		
F2	Y Y Y	Y – –	–.579	.01
F3	Y Y Y	– – –		
Final Vowel Duration	Y Y Y	– – –		
Final Vowel F1	Y Y Y	– – –		
Final Vowel F2	Y Y Y	– – –		
Final Vowel F3	Y Y Y	Y – –	.546	.01
Relative Duration Two Vowels	Y Y Y	– – –		
Japanese accent: *hAppy* Japanese speaker, Japanese listeners: – Japanese speaker, Dutch listeners: – Dutch speaker, Dutch listeners: –				
Vowel Duration	Y Y Y	– – –		
Relative Vowel Duration	Y Y Y	– – –		
F1	Y Y Y	– – –		
F2	Y Y Y	– – –		
F3	Y Y Y	– – –		
Japanese accent: *mooboo* Japanese speaker, Dutch listeners: – Dutch speaker, Japanese listeners: – Dutch speaker, Dutch listeners: $F(3, 11)=19.3$, $p<.001$; *adjusted* R^2: .833				
Vowel Duration	– – –	– – –		
Relative Vowel Duration	– – Y	– – –		
Closure Duration	Y Y Y	– – Y	.796	.001
Vowel Plus Closure Duration	Y – –	– – –		
Burst Duration	– – Y	– – Y	.834	.001
Closure F1	Y Y Y	– – –		
Proportion Voiceless Closure	– – Y	– – Y	–.320	.05

Table 5 (continued)

Possible variables	Variables included in analysis	Variables included in model	Beta	*p*<
Relative Vowel Plus Closure Duration	Y – –	– – –		
Relative Voiceless Closure Duration	Y Y –	– – –		
Relative Burst Duration	Y Y –	– – –		
Final Vowel Duration	Y Y Y	– – –		
Final Vowel F1	Y Y Y	– – –		
Final Vowel F2	Y Y Y	– – –		
Final Vowel F3	Y Y Y	– – –		
Relative Duration Two Vowels	Y Y –	– – –		
Japanese accent: *indeedo* Japanese speaker, Japanese listeners: $F(2, 11)=9.4$, $p<.01$; *adjusted* R^2: .605 Japanese speaker, Dutch listeners: $F(2, 11)=10.5$, $p<.01$; *adjusted* R^2: .634 Dutch speaker, Dutch listeners: –				
Vowel Duration	– – Y	– – –		
Relative Vowel Duration	Y Y Y	– – –		
Vowel Plus Closure Duration	Y Y Y	– – –		
Burst Duration	Y Y Y	– – –		
Closure F1	Y Y Y	Y Y –	1.033 1.052	.01 .001
Proportion Voiceless Closure	Y Y Y	Y – –	–.627	.05
Relative Vowel Plus Closure Duration	– – Y	– – –		
Relative Voiceless Closure Duration	– – Y	– – –		
Relative Burst Duration	Y Y Y	– – –		
Final Vowel Duration	Y Y Y	– – –		
Final Vowel F1	Y Y Y	– – –		
Final Vowel F2	Y Y Y	– – –		
Final Vowel F3	Y Y Y	– Y –	.603	.05
Relative Duration Two Vowels	Y Y Y	– – –		
Dutch accent: *ect* Japanese speaker, Dutch listeners: – Dutch speaker, Japanese listeners: – Dutch speaker, Dutch listeners: $F(1, 11)=5.6$, $p<.05$; *adjusted* R^2: .295				
Vowel Duration	Y Y Y	– – –		
Relative Vowel Duration	Y Y Y	– – –		
F1	Y Y Y	– – Y	.599	.05
F2	Y Y Y	– – –		
F3	Y Y Y	– – –		

Table 5 (continued)

Possible variables	Variables included in analysis	Variables included in model	Beta	*p*<
Dutch accent: *heppy* Japanese speaker, Japanese listeners: $F(1, 11)=5.4$, $p<.05$; *adjusted* R^2: .287 Japanese speaker, Dutch listeners: – Dutch speaker, Dutch listeners: –				
Vowel Duration	Y Y Y	– – –		
Relative Vowel Duration	Y Y Y	– – –		
F1	Y Y Y	– – –		
F2	Y Y Y	– – –		
F3	Y Y Y	Y – –	.593	.05
Dutch accent: *moof* Japanese speaker, Dutch listeners: – Dutch speaker, Japanese listeners: – Dutch speaker, Dutch listeners: –				
Vowel Duration	Y Y Y	– – –		
Relative Vowel Duration	Y Y Y	– – –		
Fricative Duration	Y Y Y	– – –		
Relative Fricative Duration	Y Y Y	– – –		
Dutch accent: *indeet* Japanese speaker, Japanese listeners: $F(1, 11)=6.5$, $p<.05$; *adjusted* R^2: .334 Japanese speaker, Dutch listeners: $F(1, 11)=13.4$, $p<.01$; *adjusted* R^2: .530 Dutch speaker, Dutch listeners: $F(2, 11)=23.8$, $p<.001$; *adjusted* R^2: .806				
Vowel Duration	Y Y –	– – –		
Relative Vowel Duration	Y Y –	– Y –	–.757	.01
Vowel Plus Closure Duration	Y Y Y	– – –		
Burst Duration	Y Y Y	– – –		
Closure F1	Y Y Y	– – Y	.428	.05
Proportion Voiceless Closure	– – Y	– – –		
Relative Vowel Plus Closure Duration	Y Y Y	– – Y	–.861	.001
Relative Voiceless Closure Duration	Y Y –	Y – –	.628	.05
Relative Burst Duration	Y Y Y	– – –		

Thus, for both speakers and both groups of listeners, perceptual cues associated with more voicing of the final stop were found to be helpful for recognition of *indeet* type words.

5. General Discussion

In this paper, we provided a detailed assessment of the acoustic characteristics of English words with segmental substitutions typical for a Dutch and a Japanese accent, produced both by a Japanese and a Dutch speaker. First, we found several differences between the speakers in the way they pronounced the accents. In Dutch-accented speech, the Japanese speaker produced shorter vowels in stressed syllables than the Dutch speaker did, and in Japanese-accented speech the Dutch speaker produced longer suffixed vowels than the Japanese speaker did. Both differences are most likely due to native language interference; because of the rhythmic structure of Japanese, the Japanese speaker did not lengthen stressed vowels in the same way as the Dutch speaker did, and as Dutch does not have short vowels in open syllables, the Dutch speaker lengthened suffixed vowels compared to the Japanese speaker.

For both speakers, the differences between words with a typical Japanese accent and words with a typical Dutch accent were analyzed, and found to be in line with the phonological forms the speakers were asked to produce. There were no unexpected differences between the acoustic details of the Japanese-accented and the Dutch-accented materials, confirming the validity of the materials for the speech perception experiments reported in Weber et al. (submitted).

Finally, the results of the listening experiments reported in Weber et al. (submitted) were further analyzed, to assess how listeners responded to variation in the pronunciation of accented stimuli. In most cases, listeners found it easier to recognize words when they approximated the correct English pronunciation more. Interestingly, this was found not only for listeners hearing the other-language accented speech, but also for listeners hearing their native-language accented speech.

For the Japanese-accented vowels (e.g., *Acto, hAppy*), Dutch listeners recognized words faster when the frontness of the vowel was more like that of the English /æ/ than like that of Japanese /a/. Similarly, for all items with a final consonant manipulation, for both listener groups, more voicing of the final consonant was helpful for word recognition. This was the case not only for Dutch-accented speech (e.g., *indeet*), where the final voiced consonants were pronounced as voiceless, but also for Japanese-accented speech (e.g., *mooboo, indeedo*), where the final consonants were pronounced as voiced, and with a vowel added to it. Thus, for *mooboo* type items, Dutch listeners recognized the words more readily when the /b/ was more clearly voiced, and for *indeedo* type items, both Japanese and Dutch listeners recognized words more readily when the last /d/ was more clearly voiced. Similarly, for Dutch-accented *indeet* type items, voiceless final stops that contained more characteristics of voicing led to faster word recognition, both for Japanese and Dutch listeners. For the Dutch-accented items, this suggests that a greater similarity to the English norm was helpful for recognition, for Japanese listeners, and even for Dutch listeners, who were familiar with final devoicing in Dutch-accented speech. For the Japanese-accented items, it suggests that even within the category of voiced stops, clearer voicing is still helpful, and that both groups of listeners were sensitive to such subtle differences in pronunciation.

There was one exception to this general pattern. For Dutch-accented vowels, Dutch listeners did not find it easier to recognize words when the vowel approached the English standard more. To the contrary, Dutch listeners recognized words faster when they contained vowels that were more similar in height to the Dutch /ɛ/ than to the English target /æ/. Here, deviation from the English norm, and similarity to the vowel that Dutch speakers typically substitute the English vowel with, facilitated word recognition for Dutch listeners. This suggests that the Dutch listeners might lack the awareness that the replacement of /æ/ by /ɛ/ in Dutch-accented speech does not conform to the English target pronunciation, whereas they might be more aware of that where the devoicing of final consonants in Dutch-accented speech is concerned, like the Japanese listeners might be more aware of the anomaly of the substitutions made in Japanese-accented speech.

Note that the listeners' sensitivity to subtle differences in pronunciation did not depend on the speaker. Perception was affected by the acoustic details of the accented speech regardless of whether the speaker had the same native language background as the listeners or not.

The results reported in Weber et al. (submitted) showed that Dutch-accented speech was easy to understand for Japanese listeners, whereas Japanese-accented speech was not so easy to understand for Dutch listeners. The results presented in the current paper show that there is more to be said about the perception of Japanese- and Dutch-accented speech by Japanese and Dutch listeners. Within the global pattern that Japanese-accented English words are easy to recognize for Japanese but not for Dutch listeners, and Dutch-accented words are easy to recognize for both Dutch and Japanese listeners, the exact pronunciation of those words turns out to matter. Dutch listeners recognize Dutch-accented words better when the target vowel /æ/ is pronounced in a way more similar to Dutch /ɛ/; in all other cases: the more the pronunciation approaches that of the English target form, the better, both for listeners who are unfamiliar with the accent and for those who share the speaker's native language.

Acknowledgments

This research was supported by a VENI grant from the Netherlands Organisation for Scientific Research (NWO) awarded to the first author, and by the Minerva W2 program of the Max Planck Society to the third author. We thank Sammie Tarenskeen for performing the acoustic measurements.

References

Broersma, M. (2005) "Perception of familiar contrasts in unfamiliar positions," *Journal of the Acoustical Society of America* 117, 3890–3901.

Broersma, M. (2008) "Flexible cue use in nonnative phonetic categorization," *Journal of the Acoustical Society of America* 124, 712–715.

Broersma, M. and Cutler, A. (2008) "Phantom word activation in L2," *System: An International Journal of Educational Technology and Applied Linguistics* 36, 22–34.

Flege, J.E., Bohn, O.S. and Jang, S. (1997) "Effects of experience on non-native speakers' production and perception of English vowels," *Journal of Phonetics* 25, 437–470.

Gussenhoven, C. (1999) "Dutch." In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (pp.74–77), Cambridge: University Press.

Ladefoged, P. (1999) "American English," In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (pp.41–44), Cambridge: University Press.

Nishi, K. and Kewley-Port, D. (2007) "Training Japanese listeners to perceive American English vowels: Influence of training sets," *Journal of Speech, Language and Hearing Research* 50, 1496–1509.

Peterson, G.E. and Lehiste, I. (1960) "Duration of syllable nuclei in English," *Journal of the Acoustical Society of America* 32, 693–703.

Salverda, A.P., Dahan, D. and McQueen, J.M. (2003) "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension," *Cognition* 90, 51–89.

Shibatani, M. (1990) *The Languages of Japan*. Cambridge: University Press.

Strange, W., Weber, A., Levy, E.S., Shafiro, V., Hisagi, M. and Nishi, K. (2007) "Acoustic variability within and across German, French, and American English vowels: Phonetic context effects," *Journal of the Acoustical Society of America* 122, 1111–1129.

Takata, Y. and Nabelek, A.K. (1990) "English consonant recognition in noise and in reverberation by Japanese and American listeners," *Journal of the Acoustical Society of America* 88, 663–666.

Tanaka, S. (2009) *Nitizyou gengo ni hisomu on-housoku no sekai* [Phonological laws and patterns hidden in our language]. Tokyo: Kaitakusha Publisher.

Thompson, I. (2001) "Japanese speakers." In M. Swan and B. Smith (Eds.), *Learner English: A Teacher's Guide to Interference and Other Problems* (pp.296–309) Cambridge: University Press.

Tops, G.A.J., Dekeyser, X., Devriendt, B. and Geukens, S. (2001) "Dutch speakers." In M. Swan and B. Smith (Eds.), *Learner English: A Teacher's Guide to Interference and Other Problems* (pp.1–20). Cambridge: University Press.

Vroomen, J., Van Zon, M. and De Gelder, B. (1996) "Cues to speech segmentation: Evidence from juncture misperception and word spotting," *Memory and Cognition* 24, 744–755.

Watson, I. (1983) "Cues to the voicing contrast: A survey," *Cambridge Papers in Phonetics and Experimental Linguistics* 2, 1–34.

Weber, A., Broersma, M. and Aoyagi, M. (submitted) "Spoken-word recognition in foreign-accented speech by L2 listeners."

(Received Dec. 31, 2009, Accepted May 12, 2010)