FRENCH /y/-/u/ CONTRAST IN JAPANESE LEARNERS WITH / WITHOUT ULTRASOUND FEEDBACK: VOWELS, NON-WORDS AND WORDS

Claire Pillot-Loiseau[1], Takeki Kamiyama[1, 2], Tanja Kocjančič Antolík[1]

[1]Laboratoire de Phonétique et Phonologie (LPP) UMR 7018 CNRS / University of Sorbonne Nouvelle - Paris 3
[2]Linguistique Empirique: Cognition, Société et Langage (LECSeL), EA 1569, University of Paris 8, France
claire.pillot@univ-paris3.fr, takeki.kamiyama@univ-paris8.fr, tkocjancic@googlemail.com

ABSTRACT

This study investigated French /y/ and /u/ produced by one native speaker (NS) and seven Japanese-speaking learners; four of them received lessons on /u/ with ultrasound visual feedback of the tongue shape and position. Acoustic measurements of 10 repetitions of six words (coronal context) are compared to those of isolated /y/ and /u/, and of the non-words /tyty/ and /tutu/, both before and after the ultrasound lessons, in terms of i) Euclidean distance (ED) between /y/-/u/; ii) the distance F2-F3 of /y/; and iii) the distance F1-F2 of /u/. The isolated vowels were closest to NS's, followed by monosyllabic words, disyllabic ones, and then non-words. The values of all learners are significantly different from NS's at the first recording, but those of the learners having received ultrasound lessons approached NS's after the lessons.

Keywords: L2 vowel learning, French, Japanese, Ultrasound as visual feedback, acoustics.

1. INTRODUCTION AND GOAL

It has been widely shown that the vowel system of the first language (L1) has a significant impact on the acquisition of the vowels of second languages (L2). For example, front rounded vowels are known to cause difficulty to speakers of languages without such vowels ([8], [16], [22], *inter alii*).

In present-day Parisian French, the high front rounded /y/ and the high back /u/ are phonemically contrasted. Achieving this contrast is thus of great importance for learners of L2 French and often poses a challenge to learners without these vowels in L1. One such group are Japanese learners of French. The native speakers of Tokyo Japanese, on one hand, tend to produce the French /u/ with a higher F2 than French speakers, resulting in sound that is perceived as /ø/ by native listeners of French [12]. This is related to the nature of the high non-front /u/ in Tokyo Japanese, commonly transcribed [ɯ]. From the acoustic point of view, it shows a higher F2 (> 1000 Hz) than the French /u/ and from the articulatory point of view, the tongue is less far back and the lips are less rounded than in French /u/ ([6],

[25]). On the other hand, /y/ has been observed to be realized sometimes as a diphthong [ju], by Japanese learners [13], [15], in which F2 lowers progressively during the vowel [12], whereas native speakers of French produce this vowel with close F2/F3 around 1900 Hz [10], [11], [17], [26], *inter alii*). These tendencies lead to the difficulty that Japanese-speaking learners face in realizing the phonemic contrast /y/-/u/ [18] [21].

Ultrasound tongue imaging as a visual feedback tool has been so far successfully applied in the remediation of speech sounds in clinical populations [4] as well as in second language teaching and learning [27], including the case of English /r/ and /l/ acquired by Japanese-speaking learners [24].

Concerning the French vowels /y/-/u/, it has been shown that using ultrasound as a visual feedback improves the perceptual evaluation by native listeners [14], and enhances the contrast /y/-/u/ especially by enabling a fully back articulation, which is difficult for teachers to demonstrate and for learners to realize otherwise. Improvements have been reported for isolated vowels and non-words [19] in 4 Japanese learners of French as a foreign language (FFL) versus 3 Japanese FFL learners who did not receive such feedback. The goal of the present paper is to explore whether the same is true for words and to address the following questions: does vowel production change over time as a result of using ultrasound feedback? Does it depend on the utterance type, number of syllables in a word, and voicing of the preceding consonants? What are the differences between speakers?

2. METHOD

Seven female Japanese FFL learners (intermediate level) took part in a 12-week conventional pronunciation training in language laboratory. Four participants in experimental group (EXP), EXP1 (33 years old, B2 in the Common European Framework of Reference for Languages), EXP2 (31, B1) EXP3 (28, B1) and EXP4 (29, B2), received three additional 45-minute lessons using ultrasound visual feedback, with an interval of 1 to 2 weeks, whereas three participants in the control group (CTR), CTR1 (29 years old, B1 in CEFR), CTR2 (33 years old)

and CTR3 (33, B1), did not. They were all born and raised in Tokyo and the surrounding areas, except CTR2 from Fukuoka and EXP4 from Kobe. They started learning French as adults, and had been living in France for at least two years at the time of the study (except CTR1 only 5 months). One French native speaker (NAT, 42) participated as a reference for comparison.

To highlight the differences in the articulation of [y] and [u], the ultrasound and acoustic data were recorded simultaneously, with a headset stabilization for the probe [1, 2]. All seven Japanese speakers were recorded one week before their 12-week course (REC1) and in the 12th week of the course (REC2). Between these two recordings, three ultrasound lessons were administered to the EXPs, who were recorded again (REC3) two months after REC2. NAT was recorded only once.

The goal of the ultrasound lessons was to enable the learners to monitor the lingual position of [y] and [u] in real time. The training began with a repetition task, where participants repeated items after a French native speaker with ultrasound image feedback, then moved on to practicing the vowels [y] and [u] in isolation, syllables, words and sentences. The vowels were presented first in facilitating, then neutral, and finally in difficult phonetic contexts, as classified by Callamand (e.g. [ʁ_ʁ] is facilitating and [t_t] is difficult for [u] [7]). The details of lessons were adjusted to the abilities and preferences of each participant. Due to the limited number of hours, the vowel [u], which had been reported to be more difficult for Japanese-speaking learners to acquire [12], was mainly practiced.

The list of recorded items consisted of ten repetitions (two sets of five) 1) [y], [u], [a], [i] and for the seven Japanese speakers, the Japanese [ɯ] in isolation; 2) [y] and [u] in alternation; 3) disyllabic non-words C_1VC_2V (V = /y, u/ and $C_1 = C_2 = $ /p, t, k/), and 4) 28 words (14 minimal and quasi-minimal pairs) with /y/, and /u/ in various phonetic contexts. In order to obtain comparable phonetic contexts, only the words where /y/ and /u/ are preceded by dental stops are analyzed here and compared to non-words with the same consonantal contexts (/tyty/, /tutu/) and to the two vowels in isolation (*e.g.* /u/-/tutu/ - redoux /ʁədu/). The high front position of dental stops was selected because the tongue has to travel all the way from the coronal constriction to the back articulation required for [u], which is much more challenging than [u] in isolation. Following this condition, six words were selected: "doux, redoux, rendu, tu, tout, du". They were recorded as part of a word list but because they form (quasi-)minimal pairs (du /dy/ – doux /du/; rendu /ʁɑ̃dy/ –

redoux /ʁədu/; tu /ty/ – tout /tu/); they were not produced one after another.

The acoustic signals, recorded at 22050 Hz, 16 bits, were subjected to the analysis of the first three formants, manually measured approximately at 25%, 50% and 75% of vowel duration, using Praat [5]. This allowed measuring the Euclidean distance (ED) in order to quantify the degree of realization of the contrast, and to calculate the distance between F2 and F3 for [y] and F2-F1 for [u] [11]. French focal vowels [23] are known to be characterized by a distance of less than 800Hz [11] between the neighbouring formants: F1 and F2 for back vowels (/u o ɔ (ɑ)/), F2 and F3 for /y/ [26], which contributes to a higher concentration of energy, hence perceptual saliency.

Analyses were carried out using the R statistical computing software [20], and its lme4 (linear mixed-effects models) package [3]. Data from each speaker was analysed separately with two linear mixed-effects models, both having F1-F2, F2-F3 distances and ED as dependant variables. The first model addressed the relationship between dependant variables and two fixed effects: type of utterance (isolated vowel, non-word or word) and recording session (REC1, REC2, REC3). The second model was applied only to the subset of data containing words and addressed the relationship between dependant variables and three fixed effects: number of syllables in the word (one or two), voicing (voice, voiceless) and recording session. Repetition number of each item was entered into both models as random effects.
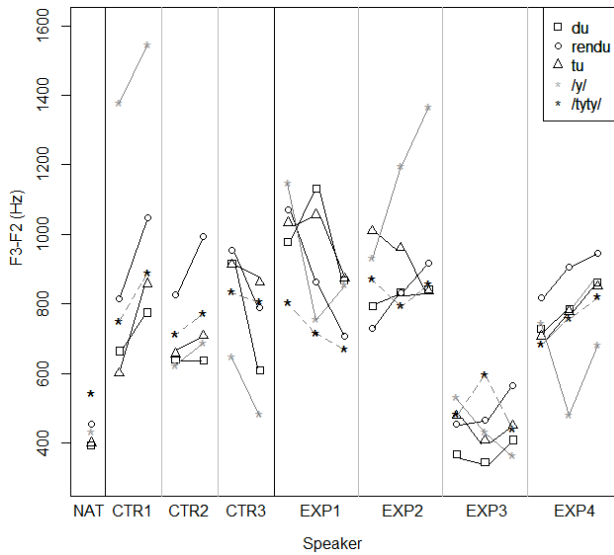
3. RESULTS

3.1. F2-F3 DISTANCE FOR /y/

NAT shows the smallest distance between F2 and F3 for words "du" (392Hz, SD 71Hz), "tu" (401Hz, SD 92Hz), followed by isolated /y/ (mean=434Hz, standard deviation SD=132Hz), the word "rendu" (453Hz, SD 163Hz), and finally by non-word /tyty/ (546Hz, SD 194Hz; average of 1st and 2nd /y/ (figure 1: SD is not represented for the sake of visibility).

Apart from EXP3, who shows an F2-F3 distance value close to that of the native (<600 Hz) in all recordings, the distance F2-F3 varies between 482 and 1546 Hz for the other learners (EXP+CTR). The values of non-words /tyty/ are less variable and cluster around 800Hz, whereas they are more variable and mostly higher for words. Only EXP1 and CTR3 show improvement over time, with EXP3 remaining similar to NAT, and other speakers showing even greater F2-F3 distances values at REC2 or REC3 than initially.
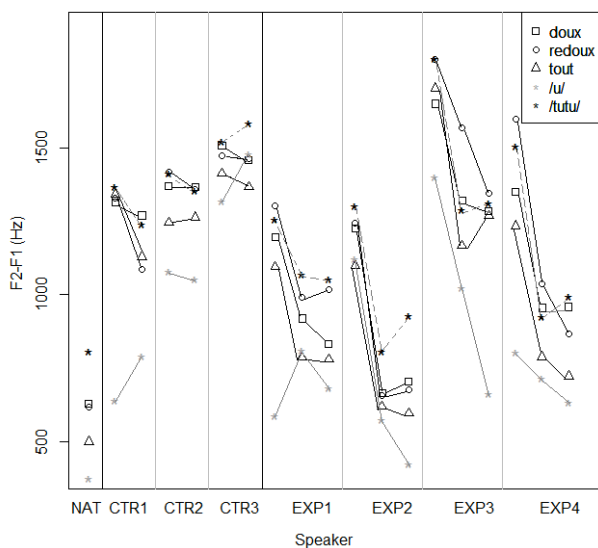
Figure 1: Mean (N = 30; 60 for non-words) F2-F3 distance for each speaker for the isolated /y/ (grey star), non-word /tyty/ (black star) and words "du" (square), "rendu" (circle) and "tu" (triangle). For each speaker, the lines connect REC1, REC2 (EXP and CTR) and REC3 (EXP).

3.2. F1-F2 DISTANCE FOR /U/

The isolated /u/ produced by NAT shows the smallest distance between F1 and F2 (376Hz, SD 95Hz), followed by "tout" (500Hz, SD 77Hz), "redoux" (617Hz, SD 145Hz) and "doux" (628Hz, SD 71Hz), and finally non-word /tutu/ (809Hz, SD 133Hz, figure 2).
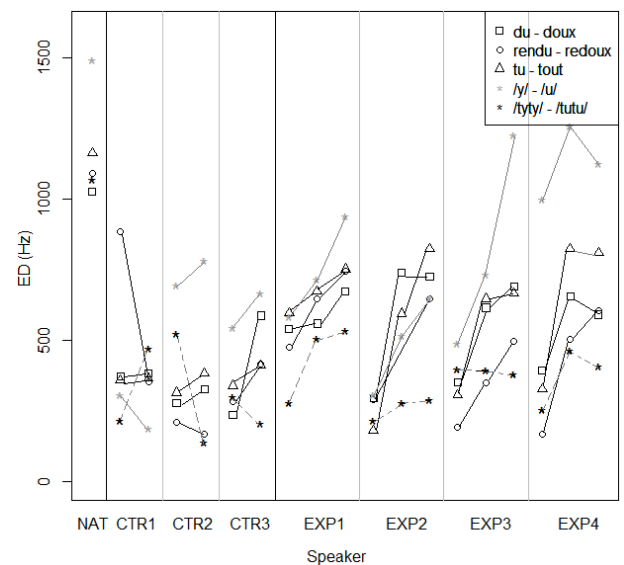
Figure 2: Mean (N = 30; 60 for non-words) F1-F2 distance for each speaker for the isolated /u/ (grey star), non-word /tutu/ (black star) and words "doux" (square), "redoux" (circle) and "tout" (triangle).

Similarly, all learners show the smallest distance F1-F2 for /u/ in isolation and the greatest for /tutu/ in most of recordings. Most EXP speakers have a decrease in F1-F2 distance between REC1 and REC2 of about 300-550Hz, while the greatest decrease for CTR speakers is 150Hz with some speakers increasing the values. These values are lower for "tout" than "doux", then than "redoux" for the most speakers and recordings, and almost the same for the rest.

3.3. EUCLIDEAN DISTANCE (ED)

Figure 3: Mean (N = 30; 60 for non-words) ED for each speaker for isolated /y/-/u/ (grey star), the non-words /tyty/-/tutu/ (black star), and the words "du-doux" (square), "rendu-redoux" (circle) and "tu-tout" (triangle).

NAT's ED is the greatest for isolated /y/-/u/ (1494Hz, SD 158Hz), followed by the words "tu-tout" (1163Hz, SD 137Hz), "rendu-redoux" (1091Hz, SD 238Hz), the non-words /tyty/-/tutu/ (1070 Hz, SD 150Hz), and finally "du-doux" (1026Hz, SD 93Hz). Most ED values of EXPs and CTRs are about half of NAT's. However, the increase from REC1 to REC2 is greater for EXPs speakers than for CTRs who also show decrease for some utterance types.

Statistical results (first linear model) show that for all speakers, the utterance type has a significant effect on all three measures. Similarly, recording session has a significant effect, except for CTR1's ED and EXP3's F2-F3 distance: the measures change significantly over time for almost all items. The number of syllables has a significant effect on most measures and speakers (second linear model). The most notable exception is EXP2: whether a word is either monosyllabic or disyllabic has no

effect on the three measures. Also, voicing has significant effects on F1-F2 distance: concerning F2-F3 distance, it has significant effects only for NAT, CTR3, EXP2 and EXP3 ($p<0.05$). Concerning ED, voicing has significant effects only for NAT, CTR2, EXP1 and EXP2. Finally, according to this second model, the recording session has significant effects for all learners (especially for CTR's, including "negative progress", see figures 1 to 3), except CTR2 (F1-F2 distance) and EXP2 (F2-F3 distance).

4. DISCUSSION

Firstly, our results illustrate an improved production of /u/ for EXP. At REC2, /u/ improves more (smaller F1/F2 distance) for EXPs than CTRs. Moreover, further improvement was observed two months after the ultrasound lessons (REC3) especially for EXPs 3 and 4, suggesting that the learners remembered the articulatory configuration necessary for /u/. The improvement of /y/ (smaller F2-F3 distance) was observed to a smaller extent across the speakers: it was notable only for EXP1 and CTR3. These results are most likely linked to a greater focus placed on /u/ during the ultrasound lessons where this vowel was practised more extensively in different contexts and with different utterance types than /y/, which is diphthongized with a lowering F2 in many cases throughout the recording sessions (lower mean F2 and greater F2-F3 distance). We can conclude that focused ultrasound feedback helps our FFL learners to improve lingual articulation of French vowels with a difficult articulatory target to attain.

Secondly, our results show significance effects of the utterance type, number of syllables and voicing. There is a notable increase in ED, which mirrors a decrease in F1-F2 distance and probably a decrease of F2 of /u/, particularly in EXPs 2, 3 and 4. The non-words do not seem to favour the production of the contrast between /y/ (in a facilitating context) and /u/ (in a difficult coronal context) among our learners, unlike monosyllabic words with the same difficult context for /u/, and the isolated vowels. This difference in favour of the words compared to non-words, may be due to our analysis of the latter: the vowel in the first syllable of the non-word, usually shorter, with a less focal quality [10], is included in the analysis. This factor could explain more variable results, smaller ED, and greater F1-F2 and F2-F3 distances for non-words than words. However, such differences about non-words are not reported in [19]. In addition, performance is enhanced when the stimulus is processed auditorily / psychoacoustically (*i.e.* without linguistic analysis, like in non-words) rather than linguistically (when the learner's attention is more focused on the content, like in familiar words [8]). Regarding the number of syllables, the disyllabic words "rendu" and "redoux" are generally the most difficult for the learners to produce (difficulty of production of the first syllable beginning with [ʁ], a difficult consonant for Japanese FFL learners). Regarding voicing, "tout" is produced with a smaller F1-F2 distance of /u/ than "doux" for all speakers; "du" is produced with a more focal /y/ than "tu" for most learners (smaller F2-F3 distance).

Greater ED for isolated vowels suggests that the articulation is more accurate in isolation because of a longer duration [9]. Marushima *et al.* [18] obtained ED values of 206Hz between the words "bulle" and "boule", corresponding to the lowest values of our learners' data, probably due to the labial context, which is not a difficult context neither for /u/ or /y/. They also obtained a F2-F3 distance mean of 924Hz in their most advanced Japanese FFL learners, values close to the mean of most of our learners. Our data corroborate these findings concerning greater F1-F2 distance than NAT, due to a higher F2 (more anterior lingual position for /u/ produced by Japanese learners). The values are significantly different for isolated vowels, non-words and words. There is therefore a need to work on different items in different consonantal and prosodic contexts during the pronunciation lessons.

Thirdly, our results suggest that each of the learners seems to have her own strategy to achieve the production of /y/ and /u/ in different utterance types. This means that we need to adapt the pronunciation training activities to the strategies of each learner and to his/her linguistic profile.

Finally, our following research perspectives are: analyses of F2 stability in function of time, analyses of the 22 other words with other phonetic contexts, quantitatively analysis of tongue contours for all utterances types, and perceptual evaluation of the same learners' productions of non-words and words by native French-speaking judges.

5. CONCLUSION

The production of vowels with challenging articulatory targets changes over time with the ultrasound feedback. It depends on the utterance type: isolated vowels are the closest to the native speaker's, and non-words the least. The disyllabic words are more difficult to produce than monosyllabic ones. Voicing positively affects only /y/, but with differences between speakers. Acquisition of /u/ is difficult for Japanese learners, both in perception and production; the articulatory approach adopted here suggests the possibility of production training improving perception [24].

7. REFERENCES

[1] Articulate Instruments Ltd. 2008. Ultrasound Stabilisation Headset Users Manual: Revision 1.4. Edinburgh: Articulate Instruments Ltd.

[2] Articulate Instruments Ltd. 2012. Articulate Assistant Advanced User Guide: Version 2.14. Edinburgh: Articulate Instruments Ltd.

[3] Bates D., Maechler M., Bolker B., Walker S. 2013. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-4. http://CRAN.R-project.org/package=lme4.

[4] Bernhardt, B.M., Bacsfalvi, P., Adler-Bock, M., Shimizu, R., Cheney, A., Giesbrecht, N., O'Connell, M., Sirianni, J., Radanov, B. 2008. Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada, *Clinical Linguistics & Phonetics* 22(2), 149–162.

[5] Boersma, P., Weenink, D. 2010. *Praat: doing phonetics by computer* [Computer program]. Version 5.1.26, http://www.praat.org/

[6] Bothorel, A., Simon, P., Wioland, F., Zerling, J. P. 1986. *Cinéradiographie des voyelles et consonnes du français*. Strasbourg: Publications de l'Institut de Phonétique de Strasbourg.

[7] Callamand, M. 1981. *Méthodologie de l'enseignement de la prononciation : organisation de la matière phonique du français et correction phonétique*. Paris: Création Loisirs Enseignement International.

[8] Flege, J. E. 1987. The production of new and similar phones in a foreign language: evidence for the effect of equivalence classification. *J. of Phonetics* 15, 47-65.

[9] Gendrot, C., Adda-Decker, M. 2005. Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. *Proc. Interspeech 2005*, Lisbon, 2453-2456.

[10] Gendrot, C., Adda-Decker, M., Vaissière, J. 2008. Les voyelles /i/ et /y/ du français: aspects quantiques et variations formantiques. *Proc. Journées d'Étude sur la Parole* Avignon, 205-208.

[11] Georgeton, L., Paillereau, N., Landron, S., Gao, J., Kamiyama, T. 2012. Analyse formantique des voyelles orales du français en contexte isolé : à la recherche d'une référence pour les apprenants de FLE. *Proc. Journées d'Étude sur la Parole -TALN-RECITAL* Grenoble, 145–152.

[12] Kamiyama, T., Vaissière, J. 2009. Perception and production of French close and close-mid rounded vowels by Japanese-speaking learners. *Acquisition et interaction en langue étrangère Aile... Lia* 2, 9-41.

[13] Kamiyama, T., Detey, S., Kawaguchi, Y. Les japonophones. In: S. Detey, I. Racine, Y. Kawaguchi, J. Eychenne (ed.), *La prononciation du français dans le monde : du natif à l'apprenant*. Submitted.

[14] Kamiyama, T., Pillot-Loiseau, C., Kocjančič-Antolík, T. 2014. Opposition /y/-/u/ chez sept apprenantes japonophones avec et sans aide de l'ultrason lingual: 2. identification perceptive des voyelles isolées par 16 francophones natifs, *Proc. Journées d'Étude sur la Parole*, Le Mans, 649-657.

[15] Lauret, B. 2007. *Enseigner la prononciation du français : questions et outils*. Paris: Hachette.

[16] Levy, E. S., Law II, F. F. 2010. Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *J. Acoust. Soc. Am.* 128(3), 1290-1305.

[17] Liénard, J.-S. 1977. *Les processus de la communication parlée : introduction à l'analyse et la synthèse de la parole*. Paris: Masson.

[18] Marushima, N., Detey, S., Kawaguchi, Y. 2010. Caractéristiques phonétiques des voyelles orales arrondies du français chez des apprenants japonophones. *Flambeau* 36, 53-72.

[19] Pillot-Loiseau, C., Kocjančič-Antolík, T., Kamiyama, T. 2014. Opposition /y/-/u/ chez 7 apprenantes japonophones avec et sans l'aide de l'ultrason lingual: 1. analyse acoustique comparative voyelles isolées / logatomes, *Proc. Journées d'Étude sur la Parole,* Le Mans, 202-210.

[20] R Core Team. 2012. *R: A language and environment for statistical computing*. Viena: R Foundation for Statistical Computing.

[21] Racine, I., Detey, S., Kawaguchi, Y. 2012. Les voyelles /y-u/ dans IPFC : évaluation perceptive de productions natives, hispanophones et japonophones. *Proc. Journées d'Étude sur la Parole -TALN-RECITAL* Grenoble, 385-392.

[22] Rochet, B. L. 1995. Perception and production of second-language speech sounds by adults. In: Strange, W. (ed), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press, 379-410.

[23] Schwartz, J.-L., Boë, L.-J., Vallée, N., Abry, C. 1997. The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics* 25(3), 255-286.

[24] Tateishi, M., Winters, S. 2013. Does ultrasound training lead to improved perception of a non-native sound contrast? Evidence from Japanese learners of English, *Proc. 2013 annual conference of the Canadian Linguistic Association*, Victoria, 1-15.

[25] Uemura, Y., Takada, M. (Kokuritsu Kokugo Kenkyûjo [National Institute for the Japanese Language]) 1990. *Nihongo no boin, shiin, onsetu: chouon undou no jikken-onseigakuteki kenkyû* [Vowels, consonants and syllables in Japanese: an experimental phonetic study on articulatory gestures]. Tokyo: Shûei shuppan.

[26] Vaissière, J. 2011. On the acoustic and perceptual characterization of reference vowels in a cross-language perspective. *Proc. 17th ICPhS* Hong Kong, 52–59.

[27] Wilson, I. 2014. Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology* 35, 285–289.