Remarks on Japanese /w/

Kikuo Maekawa
National Institute for Japanese Language and Linguistics

1 Introduction*

Articulatory phonetics is the basis of all phonetic sciences, including phonology. It is known, however, that articulatory phonetic descriptions of speech often involve uncertainty due to the subjective nature of observation. Since 2017, the author has been involved in a research project, "Reconstruction of articulatory phonetics by means of real-time MRI." The main goal of this project is to design and construct a novel database of the articulatory movements of the Japanese consonants and vowels, thereby providing a firm quantitative basis for the articulatory phonetics of Japanese (前川他2018, 浅井他 2018, 能田他2019, Maekawa 2019, Saito, Yurong & Maeawa 2019, Takemoto et al. 2019, 前川2019, Asai, Kikuchi & Maekawa 2019, Maekawa submitted). The issues of Japanese phonetics so far analyzed in this project include consonant palatalization, the place of articulation of the utterance-final moraic nasal /N/, and the place and manner of Japanese /r/.

In the current study, the articulatory characteristics of Japanese /w/ will be analyzed: hereafter, the symbol /w/ stands for the consonant that appears in the Japanese mora written with the kana わ or ワ, unless specified otherwise. In the literature of Japanese linguistics, the /w/ sound is often transcribed with the IPA (or seemingly IPA) symbols [w] (i.e., a voiced bilabial-velar double articulation approximant) or [ɰ] (a voiced velar approximant), but it is often difficult to know precisely the intentions with which these symbols are used by the authors.

Most studies of the Japanese language use [w] without explanation, like Shibatani (1990). In the illustration of Japanese in the *Handbook of the International Phonetic Association* (IPA 1999),[1] the symbol [w] appears in the column for voiced velar approximant (which is usually occupied by the [ɰ] symbol in the IPA consonant chart) without explanation. In this case, the readers will thus be left with at least two competing interpretations: either there is no bilabial constriction, or there is bilabial constriction but its degree is secondary to the primary dorso-velar constriction.

On the other hand, Vance (2008) stated clearly why he used the symbol [ɰ] as follows.

> The semivowel in wa is a back or velar semivowel, since it's produced like the high back Japanese vowel /u/. The semivowel involves the same kind of lip activity as the vowel: compression (not rounding) in careful pronunciation that weakens or disappears in ordinary conversation. (pp. 89f)

This view of /w/ is similar to that of Kawakami (川上1977), who wrote:

> The starting point of [wa] is the sound of the vowel [ɯ] with a little more narrowing of the lips. (p. 55, translation mine)

Lastly, there are scholars like Labrune who state that Japanese /w/ is somewhere between the [w] and [ɰ] of IPA.

> La semi-consonne labio-vélaire /w/ est légèrement moins arrondie que son équivalent anglais (dans *way*) ou français (dans *oiseau*). Sa réalisation phonétique se situe entre celle des symboles [ɰ] et [w] de l'API. (Labrune 2006, p. 106)

Clearly, there is lack of consensus regarding the relative importance of the bilabial and dorso-velar articulations in Japanese /w/. In the rest of this paper, the /w/ will be analyzed on the basis of the evidence of real-time MRI (rtMRI) data of its articulatory movements with the aim of clarifying the relative importance of bilabial

* Work supported by the KAKENHI grants 17H02339 and 19K21641 to the present author and the budget of the Center for Corpus Development of NINJAL.
[1] The Japanese illustration was written by late Professor Hideho Okada of Waseda University.

and velar articulations that might be involved in the production of the /w/.

2 Data

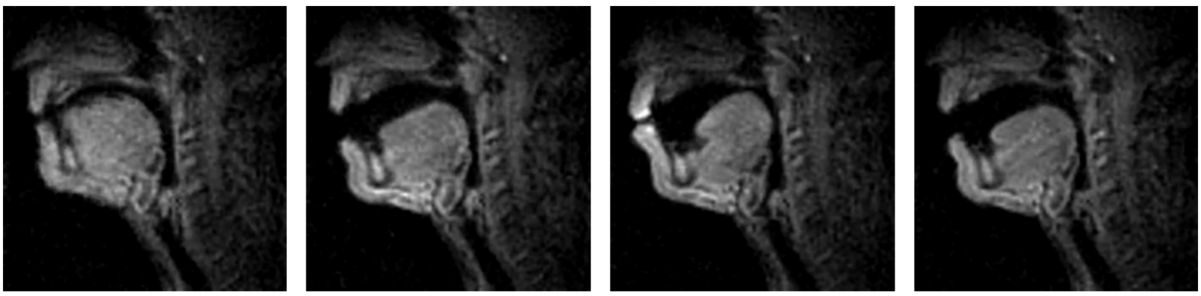
2.1 *Real-time MRI data* rtMRI is a technique for continuous recordings of the shape of the whole vocal tract (See Mohammad et al. 1997, Masaki et al. 1999, Ramanarayanan et al. 2014, Lingala et al. 2016, among many others). This new technique makes it possible to record the movements of the whole vocal tract with a time resolution of 10–100 frames per second. In addition, because the rtMRI does not involve x-ray exposure, it is possible to acquire large samples from a single subject.

In this paper, two kinds of rtMRI data will be analyzed. The main source of the data is a database named "The real-time MRI database of Japanese speech sounds" (rtMRIDB, for short) that is currently under construction by the research group of the current author (Maekawa submitted). The data acquisition of the rtMRIDB is performed at the Brain Activity Imaging Center of the ATR Promotion Laboratories (ATR-BAIC) in Kyoto, Japan.[2] The other source is the "real-time MRI IPA charts" database developed by the SPAN group of the University of Southern California (Toutios et al., 2016), which will be abbreviated SPAN-IPA hereafter.[3]

The movies in the rtMRIDB have high spatial resolution (i.e., 256 × 256 pixels, with one pixel corresponding to 1 mm, and image thickness of 10 mm), but their temporal resolution is limited (i.e., about 14 frames per second or fps). On the other hand, the samples of SPAN group are characterized by very high frame acquisition rate (i.e., about 89 fps), but the quality of the acquired image tends to show deterioration in the peripheral areas. Compare the images shown in Figures 1 and 2 below.

RP C V

Figure 1. Frames of a Japanese /wa/ corresponding to three measurement timings: rest position (RP), the /w/ (C), and the /a/ (V). Samples from the rtMRIDB.



RP V0 C V

Figure 2. Frames of IPA [w] corresponding to four measurement timings: rest position (RP), preceding /a/ vowel (V0), the [w] (C), and the following /a/ vowel (V). Samples from the SPAN-IPA.[1]

The materials analyzed below differ depending on the database. First, Japanese moras /wa/ ワ, /Fa/ ファ, and

[2] The author thanks Drs. Shinobu Masaki and Yasuhiro Shimada of ATR-BAIC for their help and advice on data acquisition.
[3] The data analyzed below are downloadable from https://sail.usc.edu/span/rtmri_ipa/. I thank the people of the University of Southern California who made these data publicly available.

/u/ ウ uttered in isolation were extracted from the rtMRIDB. Each was uttered one time without a carrier sentence by 11 native speakers of Standard Japanese (seven male and four female speakers, ages in the range 29–66 at the time of data acquisition). Second, data for the nonsense sequence /awa/ uttered by four trained phoneticians (two males and two females, each sequence uttered once) were extracted from the SPAN-IPA.

2.2 *Measurement*

2.2.1 *Timing* The aim of the measurements reported below is to quantify the labial and velar articulations in the production of the Japanese /w/. To this end, three time points in the rtMRI data were chosen as the timings of measurement. At the first point, the utterance has not started yet (with respect to the acoustic waveform) and the vocal tract is in breathing condition (as marked by lowered soft palate and neutral tongue position). This timing is referred to as rest position (RP). Second, the timing of a typical /w/ articulation is chosen and referred to as W. It is the frame in which the distance between the upper and lower lips is the smallest, and the distance between the highest point of the tongue (the v2 point, see the next section) and the palate is also the smallest. Lastly, the timing of typical /a/ articulation after the /w/ is chosen as the timing A. This timing is characterized by the most backward location of the tongue as a whole and the lowest location of the v2 point. Figure 1 compares the snapshots of the frames corresponding to these timings.

For SPAN-IPA, four timings are chosen for the measurement: the timing of rest position (RP), the timing of typical preceding /a/ vowel (A0), the timing of /w/ (C), and the timing of following /a/ vowel (A). Frames corresponding to these timings are compared in Figure 2.

2.2.2. . *Measurement points* At each timing thus determined, the *x*- and *y*-values of nine different measurement points are recorded as in Figure 3. Points UL and LL represent the positions of the extremal parts of the upper and lower lips, respectively. Points v1, v2, v3, and v4 are located on the tongue contour.

Point v1 is the apex of the tongue, v2 is the highest point of the contour, v3 is the most backward point of the tongue contour, and v4 is the edge of tongue blade. Point PL is located where the distance between v2 and the palate is the smallest. Lastly, ANS and PNS are the locations of the anterior and posterior nasal spines. These two points are used in the normalization for the skull size (see next section).

As mentioned earlier, one frame of rtMRI consists of 256 × 256 pixels, and 1 pixel corresponds to 1 mm. The origin of the *x*- and *y*-axes is at the upper left corner of the frame; larger values of *x* and *y* mean respectively that the point is further back and further down. In the rest of this paper, the values of UL, LL, v2, and PL will be used in analysis.

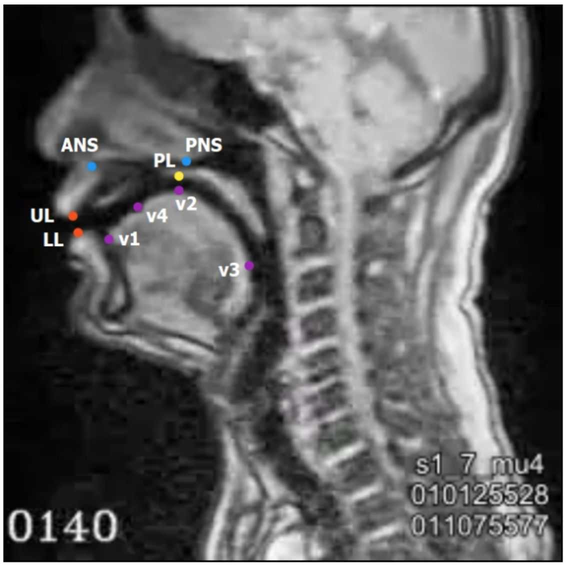


Figure 3. The nine measurement points.

2.2.3 *Normalization* The size of the human vocal tract differs considerably from subject to subject. This difference needs to be removed and normalized prior to analysis. In this study, the original data points were normalized as follows. First, the origin of the coordinate system is moved to the ANS. Second, the new coordinates are rotated around the new origin by an angle −Θ, where Θ is the angle between the original *x*-axis and the line connecting the ANS and PNS. Finally, both the *x*- and *y*-values of the points are divided by a constant D, which is the distance between the ANS and PNS. For an evaluation of this simple normalization technique, see Maekawa (submitted).

3 Analysis and discussion

3.1 *Analysis of Japanese /wa/* First, the changes in the distance between the two lips and the distance between the highest point of the tongue and the palate are analyzed. The distance between the upper lip (point UL) and lower lip (LL) is computed and referred to as the lip distance. Its distribution is shown in the left-hand panel of Figure 4. The distance between the points v2 and PL is referred to as the palate distance; its distribution is shown in the right-hand panel of Figure 4. The primary aim of Figure 4 is a comparison between the rest position (RP) and /w/ (C); the distribution of /a/ (V) is shown for reference.

A paired *t*-test was applied to the difference between the RP and C. The lip distance is significantly smaller in /w/ than in the rest position ($p < 0.001$, $t = 4.5724$, $df = 11$), while the palate distance is not significantly different between the rest position and /w/ ($p = 0.554$, $t = -0.61$, $df = 11$). These results suggest that Japanese /w/ is accompanied by a clear lip narrowing gesture, but the tongue raising gesture for dorso-velar constriction is not clear or is absent.

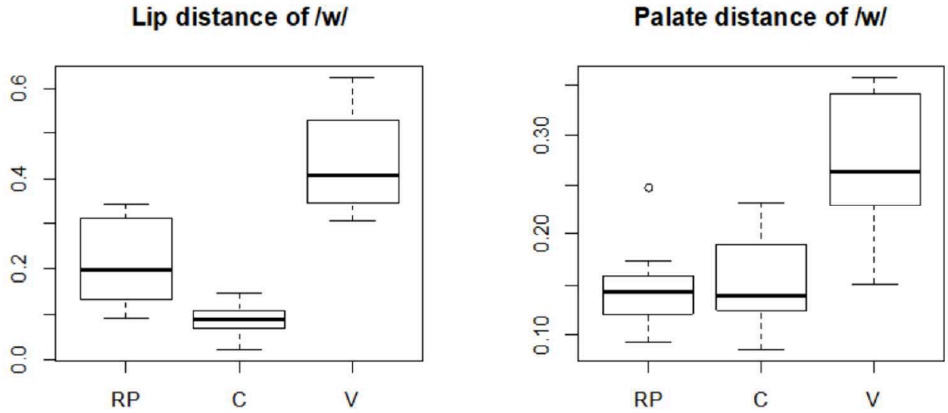


Figure 4. Box-whisker plot of the lip and palate distances of Japanese /wa/.

RP C V

Figure 5. Frames of Japanese /Fa/ corresponding to three measurement timings: rest position (RP), the consonant (C), and the following vowel (V). The same speaker as in Figure 1.

3.2 *Comparison with /Fa/* In the previous sub-section, it was stated that the tongue raising gesture for the dorso-velar constriction was not clearly observed in Japanese /w/. If there is no such gesture in the /w/, the articulation of /w/ is expected to be very similar, if not identical, to that of Japanese [ɸ], a voiceless bilabial fricative. It is of great interest to know whether there is any difference between the vocal tract articulations of /w/ and /F/. Figure 5 shows the rtMRI data of typical Japanese /Fa/ uttered by the same subject as in Figure 1.

As the first step of analysis, samples of /Fa/ are analyzed in the same way as in the analysis of /w/. The results are shown in Figure 6, where the C stands for the timing of [ɸ]. A paired *t*-test is applied to the difference between the RP and F. The lip distance is significantly smaller in /F/ than in the rest position ($p < 0.01$, $t = 4.3695$, $df = 10$), while the palate distance is not significantly different between the rest position and the /w/ ($p < 0.085$, $t = -1.914$, $df = 10$). These results suggest that Japanese /F/ is accompanied by lip narrowing gesture but not by tongue rising gesture, as expected.

The second step of analysis is a comparison of the lip and palate distances between the /F/ and /w/. Figure 7 shows the results. A paired *t*-test reveals that none of the /F/-/w/ differences are significant ($p < 0.462$, $t = -0.765$, $df = 10$ for the lip distance, and $p < 0.102$, $t = -1.803$, $df = 10$ for the palate distance).

These two-step analyses reveal that the articulations of Japanese /w/ and /F/ are quite similar.

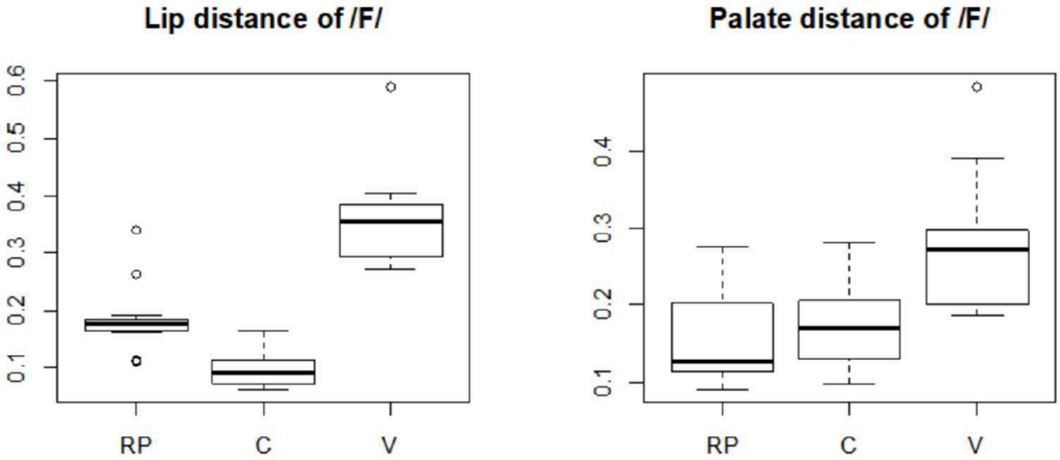


Figure 6. Box-whisker plots of the lip and palate distances of Japanese /Fa/.

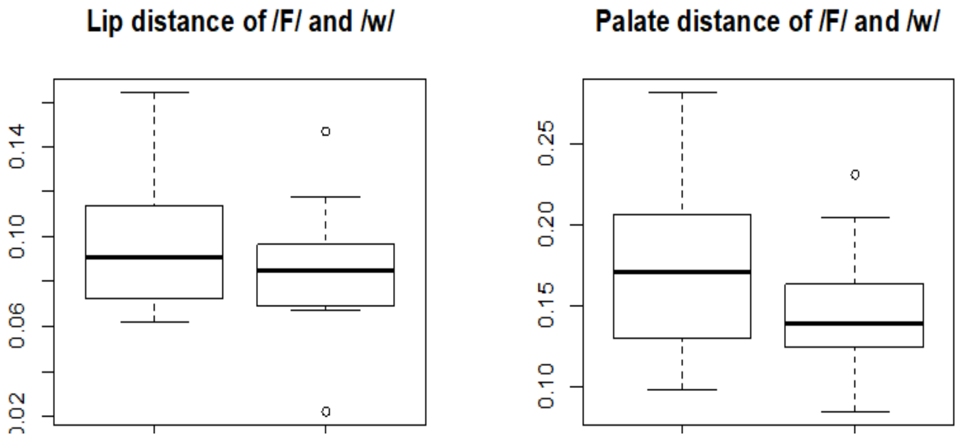
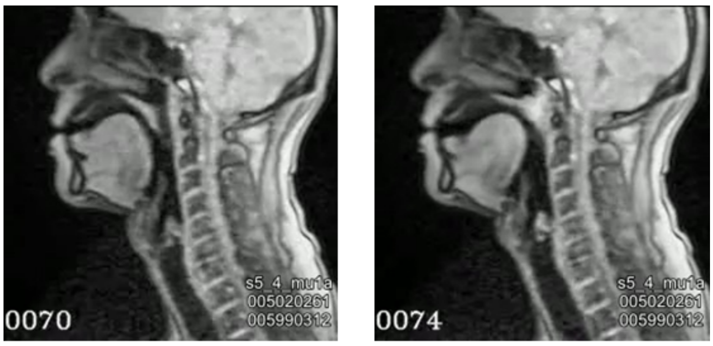


Figure 7. Comparison between the Japanese /w/ and /F/.



RP V

Figure 8. Frames of Japanese /u/ corresponding to two measurement timings, rest position (RP) and target vowel (V). The same speaker as in Figure 1.

3.3 *Comparison with /u/* Next, the assumption of Kawakami and Vance that /w/ is /u/ with compression of the lips is examined. If this assumption is correct, then on the one hand, the lip distance is expected to be smaller for /w/ than for /u/, and on the other, the palate distance should be similar between /u/ and /w/. Figure 8 shows rtMRI data of Japanese /u/ uttered by the same subject as in Figures 1 and 5.

Figure 9 compares the lip and palate distances between /u/ (timing V of Figure 8) and /w/ (timing C of Figure 1). The results of a paired *t*-test reveal that the differences in lip distance are not significant (t = 1.011, df = 10, p

= 0.336), while those of the palate distance are significant ($t = -3.470$, $df = 10$, $p = 0.006$).

These results completely contradict the assumptions above. First, the degree of labial constriction is the same in /u/ and /w/. Second, the dorso-velar constriction differs significantly between /u/ and /w/. Third, it is to be noted that /w/ has a larger palate distance than /u/. Given that the constriction is expected to be narrower in approximants than in vowels, /w/ can hardly be classified as an approximant so far as the dorso-velar articulation is concerned.

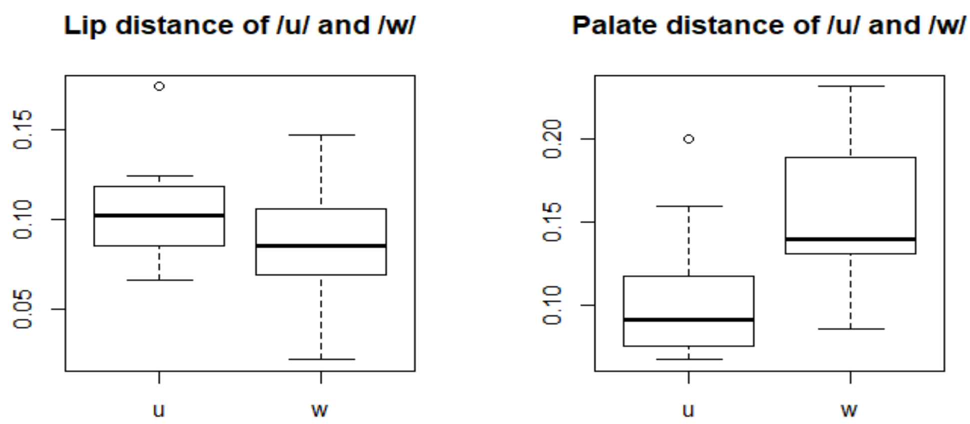


Figure 9. Comparison between Japanese /u/ and /w/.

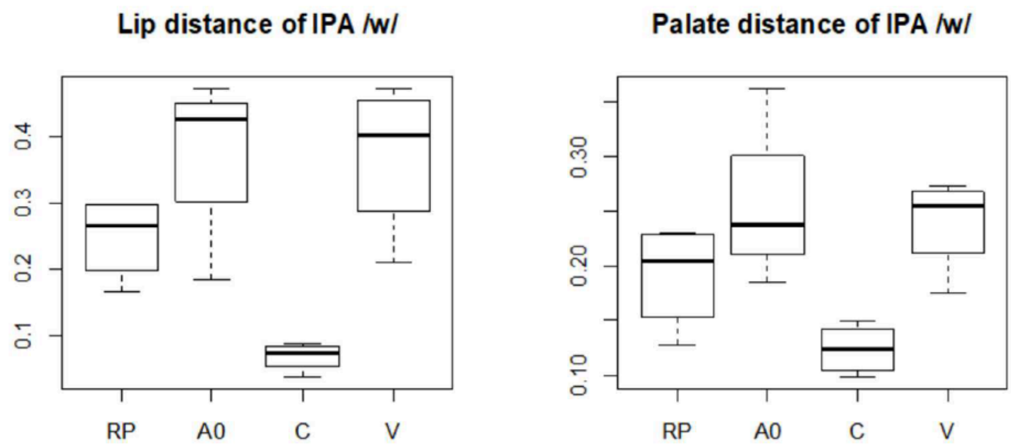


Figure 10. Box-whisker plot of the lip and palate distances of IPA [w].

3.4 *Comparison with IPA [w]* Lastly, the Japanese /w/ is compared to the IPA [w] sound. Because IPA [w] is uttered in the context /a_a/, the lip distance and palate distance are measured at four timings corresponding to the rest position (RP), the first vowel (V0), the IPA [w] (C), and the second vowel (V). Sample frames of IPA [w] have already been shown in Figure 2.

The results are shown in Figure 10. As for the lip distance (the left-hand panel), it is significantly smaller for /w/ (C) than in the rest position (RP) ($p < 0.02$, $t = 4.583$, $df = 3$) as in Japanese /w/, but the palate distance (right-hand panel) is much smaller at timing C than at RP unlike the Japanese /w/. Although the difference is not significant ($p < 0.058$, $t = 2.9976$, $df = 3$), this is probably due to too few samples ($N = 4$).

These results suggest that IPA [w] is characterized by both lip narrowing and tongue raising, in contrast with Japanese /w/ (cf. Figures 1 and 4).

4 Concluding remarks

The results reported above suggest strongly the interpretation that Japanese /w/ is not a consonant with double articulation. Its primary (and presumably only) place of articulation is (bi)labial, rather than velar. It seems to be the case that Japanese /w/ is a voiced approximant counterpart of Japanese [ɸ], at least from a purely phonetic and articulatory point of view.[4]

The results of the current study show that it is misleading to use the IPA symbol [w] for Japanese /w/. Note also that it is even more misleading to use [ɰ] instead, because there is little evidence to interpret Japanese /w/ as a velar consonant. If for some reason one wishes to use this symbol, it should be accompanied by a diacritic of labialization, [ɰʷ]. Even with this modification, however, this symbol fails to indicate that it is the labial articulation that plays the primary role in the production of /w/. The best solution then is probably to create a new IPA symbol for the category of voiced bilabial approximant, which is lacking in the present consonant chart of IPA.[5]

References

Asai, Takuya, Hideaki Kikuchi & Kikuo Maekawa. (2019) Motion detection of articulatory movement with paralinguistic information using real-time MRI movie. *Proc. Oriental COCOSDA 2019*, Cebu City, Philippine, paper 33.

International Phonetic Association. (1999) *Handbook of the International Phonetic Association: A Guide to the Use of International Phonetic Alphabet.* Cambridge University Press.

Labrune, Laurence. (2006) *LA PHONOLOGIE DU JAPONAIS* (Collection Linguistique Publiée par la Société de Linguistique de Paris XC). Peters Leuven-Paris.

Lingala, Sajan G., Brad P. Sutton, Marc Miquel & Krishna S. Nayak. (2016) Recommendations for real-time speech MRI. *Journal of Magnetic Resonance Imaging* 43, 28-44.

Maekawa, Kikuo. (2019) A real-time MRI study of Japanese moraic nasal in utterance-final position. *Proc. ICPhS 2019*, Melbourne, 1987-1991.

Maekawa, Kikuo. (submitted) A real-time MRI study of utterance-final moraic nasal.

Mohammad, M., E. Moore, J. N. Carter, C. H. Shadle & S. J. Gunn. (1997) Using MRI to image the moving vocal tract during speech. *Proceedings EUROSPEECH 2019*, 2027-2030.

Masaki, Shinobu, Mark K. Tiede, Kiyoshi Honda, Yasuhiro Shimada, Ichiro Fujimoto, Yuji Nakamura & Noburo Ninomiya. (1999) MRI-based speech production study using a synchronized sampling method. *Journal of the Acoustical Society of Japan (E)* 20(5), 375-379.

Mohammad, M., E. Moore, J. N. Carter, C. H. Shadle & S. J. Gunn. (1997) Using MRI to image the moving vocal tract during speech. *Proceedings EUROSPEECH '97*, Rhodes, Greece. (http://www.isca-speech.org/archive)

Ramanarayanan, Vikram, Louis Goldstein, Dani Byrd & S. Narayanan Shrikanth. (2014) An investigation of articulatory setting using real-time magnetic resonance imaging. *Journal of the Acoustical Society of America* 134 (1), 510-519.

Saito, Yoshio, Yurong & Kikuo Maekawa. (2019) An investigation into modern Mongolian vowel harmony using real-time Magnetic Resonance Imaging. *Proceedings ICPhS 2019*, Melbourne, 1431-1434.

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

Takemoto, Hironori, Tsubasa Goto, Yuya Hagihara, Sayaka Hamanaka, Tatsuya Kitamura, Yukiko Nota & Kikuo Maekawa. (2019) Speech organ contour extraction using real-time MRI and machine learning method. *Proceedings INTERSPEECH 2019*, 904-908.

Toutios, Asterios, Sajan Goud Lingala, Colin Vaz, Jangwon Kim, John Esling, Patricia Keating, Matthew Gordon, Dani Byrd, Louis Goldstein, Krishna Nayak & Shrikanth Narayanan. (2016) Illustrating the production of the International Phonetic Alphabet sounds using fast real-time magnetic resonance imaging. *Proc. INTERSPEECH 2016*, San Francisco, 2428-2432.

Vance, Timothy J. (2008) *The Sounds of Japanese*. Cambridge: Cambridge University Press.

浅井拓也・菊池英明・前川喜久雄 (2018)「調音運動動画アノテーションシステムの開発と応用」日本音声学会第32回全国大会予稿集, 201-206.

川上蓁 (1977)『日本語音声概説』東京:桜楓社.

能田由紀子・籠宮隆之・前川喜久雄 (2019)「磁気センサシステムをもちいた計測による座位・仰臥位・腹臥位における舌運動の差異の検討」日本音響学会 2019 年春季研究発表会講演論文集, 823-824.

前川喜久雄 (2019)「日本語ラ行子音の調音：リアルタイム MRI による観察」日本音声学会第 33 回全国大会予稿集, 98-103.

[4] This conclusion provides a natural explanation for the historical change of word-medial /h/ ([ɸ]) into /w/ (as voiced bilabial approximant) known as *hagyoo-tenko*.

[5] In the past, the Phonetic Society of Japan (PSJ) has recommended a symbol that looked like Greek letter omega ("ω") for the transcription of Standard Japanese /w/, but this symbol was used without a definition or explanation. Readers can find this symbol in the "Transcription of Standard Japanese" printed in the inside page of the front cover of the *Bulletin of the PSJ* (Nos. 51–87) published in the years 1938–1955.

前川喜久雄・能田由紀子・北村達也・竹本浩典・石本祐一 (2018)「日本語撥音の調音音声学的記述の精緻化：rtMRI データによる試み」日本音響学会 2018 年春季研究発表会講演論文集, 1247-1248.