A REAL-TIME MRI STUDY OF JAPANESE MORAIC NASAL IN UTTERANCE-FINAL POSITION

Kikuo Maekawa

National Institute for Japanese Language and Linguistics (NINJAL)
kikuo@ninjal.ac.jp

ABSTRACT

Moraic nasal of Japanese, often symbolized as /N/, is a nasal segment that has the status of an independent mora. It is widely acknowledged that the place of articulation of /N/ is determined by the assimilation to the following consonants; for example, /aNma/, /aNta/, and /aNka/ become [amma], [anta], and [aŋka] respectively. There is, however, a lack of consensus concerning the realization of /N/ in the utterance-final position. Places of articulation of utterance-final /N/ hitherto stipulated in the literatures include velar [ŋ], uvular [N], and nasalized vowels. A real-time MRI movie database was analyzed to solve this problem. Data of three male speakers revealed consistent results. The location of closure for the final /N/ is highly predictable by the membership of the immediately preceding vowel. Closure locations predicted by a generalized linear mixed-effect model regression analysis showed high correlation (between .887-.986) with the observed locations.

Keywords: Realtime MRI movie, moraic nasal, Japanese, utterance-final, coarticulation

1. INTRODUCTION

Moraic nasal of Japanese, often symbolized as /N/, is a nasal segment that is counted as an independent mora in Japanese phonology. It is hence an exception to the basic C_0V mora structure of the language. Additionally, it is acknowledged that the place of articulation for /N/ is completely underspecified and determined by assimilation to the following consonants; for example, /aNma/ 'massage', /aNta/ 'you', and /aNka/ 'cheap' become [amma], [anta], and [aŋka] respectively. Based on these characteristics, moraic nasal /N/ is often called a "special" mora. As for the treatment of /N/ in Japanese linguistics, see [1] among many others.

There is, however, a lack of consensus concerning the realization of /N/ in the utterance-final position, where nothing follows the /N/. Places of articulation of utterance-final /N/ hitherto stipulated in the literature include velar [ŋ], uvular [N], and nasalized vowels. See Saito [2], Vance [3] and Hashi [4] for a review of past works. As Hashi appropriately pointed out, the principal reason for the lack of consensus is the paucity of objective data on the articulation of final /N/. It is difficult, if not entirely impossible, to make objective observations of articulations that include the velum, uvula, tongue root, and pharynx wall using traditional devices of articulatory observation such as x-ray microbeam [5], EMA [6], and WAVE [7].

Observation of articulatory movements by a real-time magnetic resonance imaging movie (rtMRI) is a solution to this problem [8,9]. It provides a clear image of the whole vocal tract including the velum, uvula, tongue-root, and pharynx cavity with reasonable time-resolution (see below).

2. DATA

2.1. The database

In this paper, an rtMRI database, whose construction has been underway since 2017 was analyzed [10]. The recording of articulatory movements was conducted in the ATR Brain Activity Imaging Center in Kyoto using a 3T MRI system (Siemens MAGNETOM Prisma fit 3T). Realtime recording of articulatory movements in the midsagittal plane was made with FLASH sequence with acceleration factor 3. Spatial resolution was 256 x 256 pixels with a pixel size of 1 mm x 1 mm and slice thickness of 10 mm. Temporal reconstruction rate was 14 frames per second.

Currently, the database covers 9 subjects of Tokyo Japanese each providing speech of about one hour including a special portion devoted to moraic nasal. Because the annotation of the database is underway, samples of three male speakers of Tokyo Japanese were analysed in this paper. For each speaker, the special part of the database contained 62 instances of /N/, of which 38 and 24 were in word-medial and word-final positions respectively. Because the words were read in isolation, the word-final /N/ was realized utterance-finally. Note that samples of medial /N/ having [h] and [w] as following consonant were excluded. These samples will be analysed in a separate paper.

2.2. Measurement and normalization

Each instance of /N/ was characterized by five measurement points. In Fig. 1, the measurement

points were shown on a snapshot of the MRI movie. Points c1 and c2 (shown by an empty circle) stand respectively for the beginning and end of consonantal closure for /N/, here dorso-velar closure. Points v1, v2, and v3 (circle with cross) stand respectively for the location of the tip of the tongue, the highest point of the tongue, and, the most backward point of the tongue. These points provide a rough overall tongue shape during the /N/ articulation. The same v1-v3 measurement was also conducted for the vowel that immediately preceded the /N/. The x- and y-coordinate values of these points are represented as (c1x, c1y) or (v2x, v2y) in the rest of the paper.

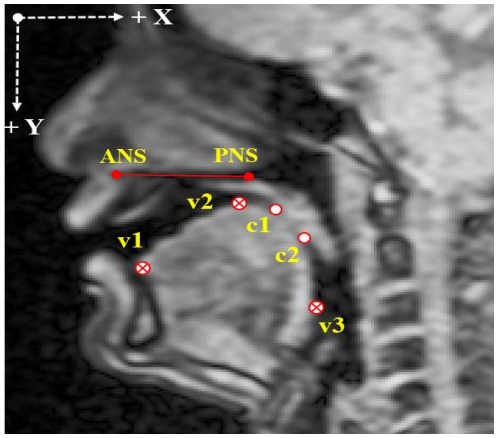

Fig. 1: Measurement points

Because the size of the vocal tract differs considerably depending on the speakers, the measured values were normalized for each speaker with reference to the locations of the anterior nasal spine (ANS) and posterior nasal spine (PNS) [11]. ANS and PNS are shown using filled circles connected by a real line in Fig.1.

In the original MRI image file, the origin of the coordinate was in the upper-left corner as shown by the dotted arrows in Fig. 1. The new origin was set at the location of ANS, and all measurement points were rotated so that the line connecting ANS and PNS became parallel to the original x-axis. The distance from the new origin was remeasured using the ANS-PNS distance as the unit length on both axes.

Although automatic annotation of the tongue and other articulators are underway [12], measurement for the current study was conducted manually using two tools. The rtMRI movie annotation tool developed by Asai, Kikuchi and Maekawa [13] and the ImageJ [14].

Measurement was conducted for a frame of the MRI movie that had the typical characteristics of /N/, that is, distinctly lowered velum and clear contact between the upper and lower articulators. Because the temporal reconstruction rate of the current movie is relatively low (a frame corresponds to time interval of about 70ms), it was not difficult to choose a frame that maximally met these criteria. For the measurement of the preceding vowel, a frame that was best approximated by the typical articulatory position of the vowel was selected for measurement. In most cases, the frame that was two frames before the one chosen for the /N/ was selected. Among the total of 186 instances of /N/, 25 were realized as nasalized vowels. These instances were excluded from the current analysis.

3. ANALYSIS

3.1. Medial /N/

First, the samples of word-medial /N/ were analyzed. Fig. 2 shows the distribution of normalized c1 point for the medial /N/. The data are pooled over three speakers. Here, samples are classified according to the places of articulation of the immediately following consonants. The abbreviations in the legend "Alv", "Lab", "Pal", and "Vel" stand respectively for "alveolar" ([t],[ts],[tʃ],[s],[n],[z],[ɾ]), "labial" ([b], [Φ]), "palatal" ([j], [ç]) , and "velar" ([k]). Fig.2 shows that there is nearly perfect separation of samples due to the following consonants; this finding is congruent with the description of ordinary (i.e., word-medial) /N/ in the literatures.

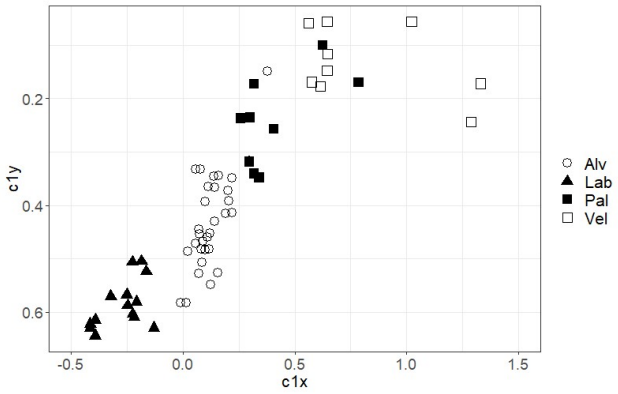

Fig. 2: C1 distribution of medial /N/ as a function of the place of articulation of the following consonants

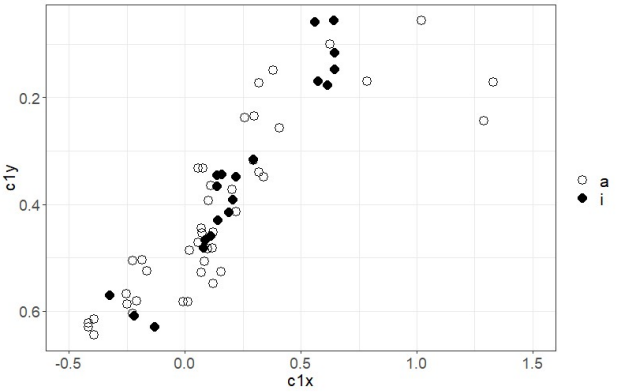

Fig. 3: C1 distribution of medial /N/ as a function of the preceding vowels

Fig.3 is the same as Fig. 2 except that the samples are classified with respect to the preceding vowel of the /N/. It is clear from the figure that the influence of the preceding vowels on /N/ is negligible.

3.2. Final /N/

Next, the samples of utterance-final /N/ were analyzed. Fig. 4 shows the distribution of normalized c1 of the final /N/ samples; the data are pooled over three speakers, and the samples are classified according to the immediately preceding vowels. Unlike Fig. 3, here, most samples are distributed in the right half of the figure corresponding to the articulatory regions of the hard palate, soft palate, and uvula. Obviously, the place of articulation for the final /N/ cannot be fixed to a single location. At the same time, however, there seems to be moderate correspondence between the location of the final /N/ and that of the preceding vowels. The capital letters in Fig. 4 show the mean v2 locations of the preceding vowels for reference.

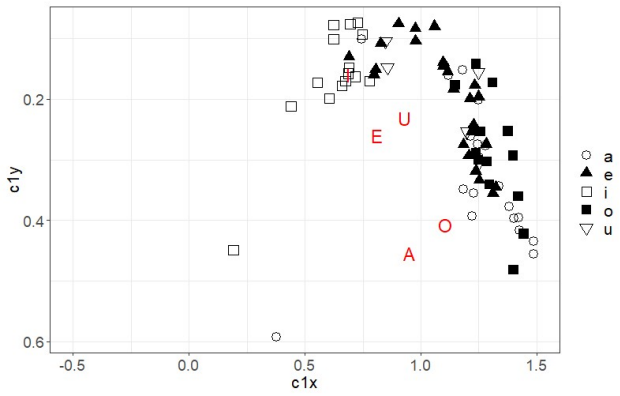

Fig. 4: C1 distribution of final /N/ as a function of the preceding vowels

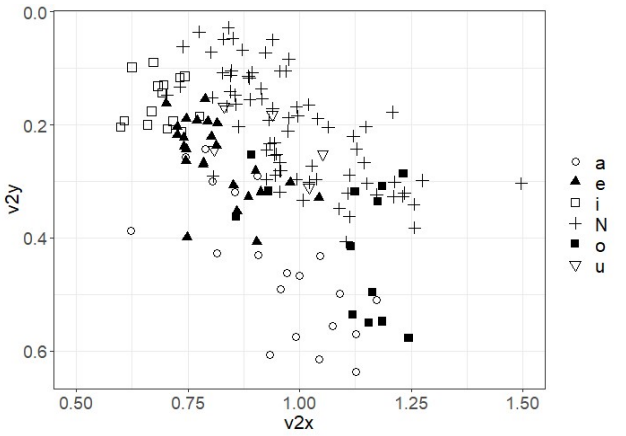

Fig. 5: V2 distribution in the preceding vowels and the final /N/

Fig. 5 shows the distribution of the v2 values of the preceding vowels; the distribution of the v2 values in the final /N/ is also shown for reference (by the symbol "+"). It is clear from Figs. 4 and 5 that, while articulating the final /N/, the tongue has a higher position than in the preceding vowels. Fig. 6 examines this issue more closely. Arrows in this figure are the vectors connecting the v2 of the preceding vowel and that of the final /N/. In the cases of /u/, the length of the vectors is generally short. This means that the tongue remains almost the same between the preceding vowels and final /N/. In the case of /i/, the vector length is short, but the directions of the vectors are consistently backward and upward. The same tendency is observed in /e/ as well. The vectors of /o/ are characterized by relatively large perpendicular upward movement. Last, the vectors of /a/ have the largest length. With respect to the direction, however, there seem to be two subgroups. One is characterized by /o/-like perpendicular upward movement, and the other is characterized by backward and upward movement such as in the cases of /i/ and /e/. These observations will be discussed later in section 4.

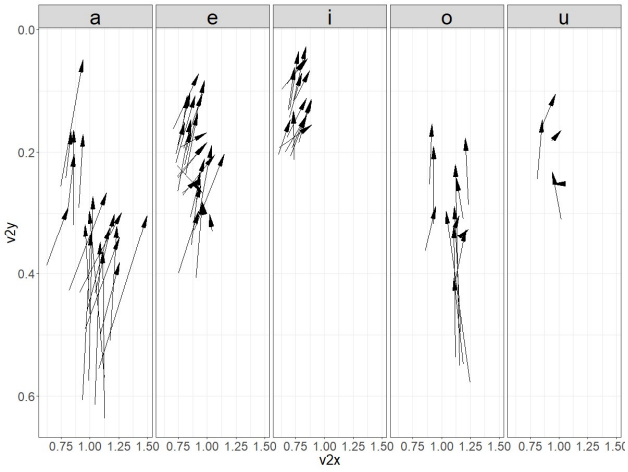

Fig. 6: Distributions of the vectors connecting the v2 locations of the preceding vowel and final /N/

3.3. Statistical modeling

The analyses presented in the previous sections strongly suggest that the location of the final /N/ is predictable from the membership of the preceding vowel. To examine this hypothesis, a generalized-linear-mixed-effect-model-based regression model was constructed to predict the c1x value of the /N/. For comparison, both medial and final /N/ were analyzed. The membership of the consonants that immediately followed the medial /N/, and, the vowels that immediately precede the final /N/ were used as independent variables, and speakers and words were used as the random effect variables. The model was constructed by the lme4 library (ver. 1.1-19) of the R language (ver. 3.5.1) [15, 16].

Fig. 7 shows the correlation between the observed and predicted c1x values. The results of medial and final /N/ are shown together. The Pearson correlation coefficients are .986 and .887 respectively for the medial and final /N/. Further, the mean and SD of the prediction error are 0.094 and 0.100, respectively, for the final /N/, and 0.047 and 0.039, respectively, for the medial /N/.

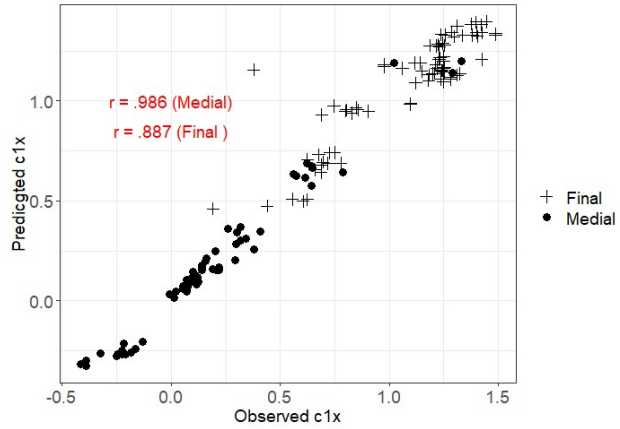

Fig. 7: Correlation between the observed and predicted c1x values

3.4. Cross validation of the model

Although the result shown in Fig.7 seems to be quite successful, it is based upon so-called closed data. To evaluate the generalisability of the statistical models to new data, leave-one-out cross validation was conducted. The results are summarized in Table 1. Prediction performance as shown by the overall mean and SD is slightly lower, but it is on the same level as that of the closed data model noted at the end of the previous section. Additionally, the correlation coefficients between the observed and predicted values of the test set samples are 0.808 and 0.963, respectively, for the final and medial /N/. These results show that models can treat test data reasonably well; hence, the model can be generalizable to new data.

Table 1: Results of leave-one-out cross validations

Type of /N/	Prec. vowel / Fol. cons.	Mean error	SD of error	N
Final	/a/	.177	.217	20
	/e/	.127	.076	25
	/i/	.074	.083	15
	/o/	.060	.039	13
	/u/	.119	.040	5
	Overall	.118	.130	78
Medial	Alv	.048	.040	32
	Lab	.085	.047	15
	Pal	.105	.079	9
	Vel	.122	.097	9
	Overall	.075	.064	65

4. DISCUSSION AND CONCLUSION

The analyses conducted above revealed that the place of articulation of utterance-final /N/ in Japanese is largely predictable from the membership of the immediately preceding vowels. As shown by Fig. 6, the final /N/s are phonetically realized by upward (and occasionally rightward) movement of the tongue from the location of the immediately preceding vowels. When coupled with the descending movement of the velum, the vocal tract closure for the final /N/ can be created in various regions encompassing the hard palate, soft palate, and uvula.

The linguistic consequence is that any trial to specify a fixed place of articulation for the utterance-final /N/ is of no use. In this respect, the description of final /N/ by Saito [2], who stated that "the place of articulation is velar after front vowels and uvular after back vowels" (translation by Hashi [4]) seems to be the most accurate among the existing literature, but even this description is problematic in that it cannot properly handle the wide dispersion of /N/ after the /a/ vowel, as seen in Figs. 4 and 5.

As pointed out at the end of section 3.2, the movement of the tongue from /a/ to /N/ is not uniform. This is due probably to the phonological property of the /a/ vowel in Japanese that does not have the phonological opposition of [FRONT] and [BACK]. In Japanese, /a/ is realized as either front [a] or back [ɑ] depending on the phonological context and paralinguistic setting [17]. As the result, the /N/ after the vowel can be realized as a palatal, velar, or uvular consonant. The fact that the mean prediction error of the final /N/ is the largest in the samples preceded by /a/ (see Table 1) is the direct consequence of this variability.

Although this study needs to be strengthened by an analysis of samples where the /N/ is realized as nasalized vowels, it would be safe to conclude from the current analytical results that it is inappropriate to treat the articulatory variation of utterance-final /N/ as a conditioned allophony. Rather, it is to be understood as the variation caused at the level of coarticulation. Put differently, it belongs to phonetics rather than phonology.

Acknowledgement: Work supported by KAKENHI grant (17H02339), and the collaborative research project of the Center for Corpus Development, NINJAL. The author thanks the staffs of the ATR-BAIC, especially, Yasuhiro Shimada, Yukiko Nota, and Shinobu Masaki, for their technical supports and advises. He also thanks Kiyoshi Honda for the advice on the measurement of ANS and PNS.

5. REFERENCES

[1] Shibatani, Masayoshi. *The languages of Japan*. Cambridge Univ. Press, 1990.

[2] Saito, Yoshio. *Nihongo Onseigaku Nyuumon Kaiteiban* (Introduction to Japanese Phonetics Revised edition). Tokyo: Sanseido, 2011.

[3] Vance, Timothy J. *The Sounds of Japanese*. Cambridge Univ. Press, 2008.

[4] Hashi, Michiko. "Articulatory variability in word-final Japanese moraic-nasals: An x-ray microbeam study." *Onseikenkyu* (*The Journal of the Phonetic Society of Japan*), 20 (1), 77-87, 2016.

[5] Kiritani, Shigeru., Kenji Ito and Osamu Fujimura. "Tongue-pellet tracking by a computer-controlled x-ray microbeam system." *The Journal of the Acoustical Society of America*, 57(6 Pt 2), 1516–1520, 1975.

[6] Perkell, Joseph S. et al. "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements." *Journal of the Acoustical Society of America*, 92 (6), 3078-3096, 1992.

[7] Northern Digital Inc. *Electromagnetic Articulography: measuring movements of the tongue, jaws and lips during speech* (https://www.ndigital.com/msci/wp-content/uploads/sites/17/2016/10/8300352-Rev001-Electromagnetic-Articulography.pdf), 2016.

[8] Narayanan, Shrikanth, et al. "An approach to real-time magnetic resonance imaging for speech production". *Journal of Acoustical Society of America*, 115 (4), 1771-1776, 2004.

[9] Niebergall, Aaron, et al. "Real-time MRI of speaking at a resolution of 33 ms: Undersampled radial FLASH with nonlinear inverse reconstruction". *Magnetic Resonance in Medicine*, 69, 477–485, 2013.

[10] Maekawa, Kikuo et al. "Nihongo-hatsuon no chooon-onseigaku-teki-kijutsu no seichika" (Refinement of the articulatory phonetic description of Japanese moraic nasal). *2018 Spring meeting Acoustical Society of Japan*, 1247-1248, 2018.

[11] Honda, Kiyoshi. *Zikken-onsei-kagaku* (Experimental speech science). Tokyo: Korona-sha, 2018.

[12] Goto, Tsuabsa et al. "Kikai-gakushuu ni yoru rtMRI-dooga ni okeru hatsuwa-kikan no rinkaku-chuushutsu-hoohoo no kentoo" (Examination of edge detection of speech organs from real-time MRI movie by a machine learning method). *2018 Autumn Meeting Acoustical Society of Japan,* 813-814, 2018.

[13] Asai, Takuya, Hideaki Kikuchi, and Kikuo Maekawa. "Chooon-undoo anoteeshon sisutemu no kaihatsu" (Development of an annotation system for speech articulation movies). *2018 Autumn Meeting Acoustical Society of Japan,* 1235-1238, 2018.

[14] Ferreira, Tiago and Wayne Rasband. *ImageJ User Guide (IJ 1.46r)*. (https://imagej.nih.gov/ij/docs/guide/user-guide.pdf), 2016.

[15] R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/, 2018.

[16] Bates, D. et al. *Package 'lme4'* (https://cran.r-project.org/web/packages/lme4/lme4.pdf), 2019.

[17] Maekawa, Kikuo and Takayuki Kagomiya. "Influence of paralinguistic information on segmental articulation". *Proceedings of ICSLP 2000*, Beijing, 349-352, 2000.