# A4

# The phonetic reduction of nasals and voiced stops in Japanese across speech styles

Yoichi Mukai, Benjamin V. Tucker (University of Alberta)
mukai@ualberta.ca, bvtucker@ualberta.ca

## 1 Introduction

In daily conversations, speakers often produce speech in a casual manner. Casual speech, also referred to as spontaneous or conversational speech, contains a high degree of variation as compared to more careful speech styles (Ernestus and Warner, 2011). One important aspect of casual speech leading to this high variability is phonetic reduction, resulting in words being pronounced with fewer segments, shorter durations, and assimilation. For example, *yesterday* pronounced carefully could be something like /jɛstə˞deɪ/ but in casual speech it could be pronounced [jɛʃeɪ] (Tucker, 2007). Reduced pronunciation variants has been studied cross-linguistically with evidence being reported in American English (e.g., Johnson, 2004; Warner and Tucker, 2011), Dutch (e.g., Ernestus et al., 2002), French (e.g., Brand and Ernestus, 2015), Finnish (e.g., Lennes et al., 2001), German (e.g., Kohler, 1990), and Japanese (e.g., Arai et al., 2007; Maekawa, 2005).

In the present study, we use a large-scale speech corpus, the Corpus of Spontaneous Japanese (Maekawa, 2003), to examine the phonetic variability found in nasals and voiced stops and to describe how that variation and reduction occurs across speech styles in Japanese. Using the Corpus of Spontaneous Japanese, we analyzed the duration and intensity difference of target segments across four styles of speech: academic presentations, simulated public speech, dialogues, and read speech. The intensity difference was defined as the difference between the minimum intensity of the target segment to the averaged maximum intensity of surrounding segments (Tucker, 2011; Warner and Tucker, 2011). We hypothesized that we would observe stronger reduction (more approximant-like productions), as indicated by shorter duration and smaller intensity difference, as speech style becomes more casual. In other words, the shortest duration and the smallest intensity difference would be found for nasals and voiced stops in dialogues (most casual) and the longest duration and the largest intensity difference in read speech (least casual).

## 2 Methods

### 2.0.1 Data

We used the Corpus of Spontaneous Japanese, which contains approximately 44 hours of speech (about half million words) from four different speech styles: academic presentations, simulated public speech, dialogues, and read speech (Maekawa, 2003). All acoustic analysis was performed using Praat (Boersma and Weenink, 2016) and the predefined segmental boundaries provided in the corpus.

### 2.0.2 Analysis

We used linear mixed-effects models with lme4 and lmerTest packages (Bates et al., 2017; Kuznetsova et al., 2015) in R (R Core Team, 2017) to measure whether duration and intensity difference of nasals and voiced stops differs across speech styles, as well as to predict the relative duration and intensity difference of these segments across speech styles. Furthermore, we also ran Bonferroni adjusted post hoc comparisons between SpeechStyle and Phoneme using the multcomp and lsmeans packages in R (Lenth, 2017; Torsten Hothorn, 2016). The variables of interest were as follows:

- Dependent variables: LogDuration; LogIntensityDifference (A log-transformation was applied to attenuate skewness)
- Main predictors: Phoneme (nasals: /m/, /n/, /ɴ/; voiced stops /b/, /d/, /g/); SpeechStyle (AcademicPresentation, SimulatedPublicSpeech, Dialogue, and ReadSpeech)
- Control variables: SpeakerAge, WordDuration, and PhonemeEnvironment (Word-initial, -medial, -final)

All the control variables were included in the models as long as the variables significantly contributed to the fit of the model. We also included Speaker as a random intercept and SpeechStyle by Speaker as a random slope.

## 3 Results & Discussion

### 3.1 Duration

Statistical analysis of nasal duration across speech styles, visualized in Figure 1, revealed that there is a main effect of SpeechStyle [F(3,8)=13.3, $p<0.001$] and Phoneme [F(2,138229)=2829.1, $p<0.001$] as well as an interaction between SpeechStyle and Phoneme [F(6,137122)=44.0, $p<0.001$]. Individual comparisons of /m/ across speech styles revealed that the duration of /m/ is longer for the read speech in comparison to simulated public speech ($t$=-4.7, $p<0.001$) and academic presentations ($t$=-5.97, $p<0.001$). The /m/ nasal durations in simulated public speech are also significantly longer than in academic presentations ($t$=3.77, $p<0.01$). The comparisons between dialogues and academic presentations, as well as dialogues and simulated public speech, are not significant. We also identified that /ɴ/ follows a similar pattern to /m/ where the

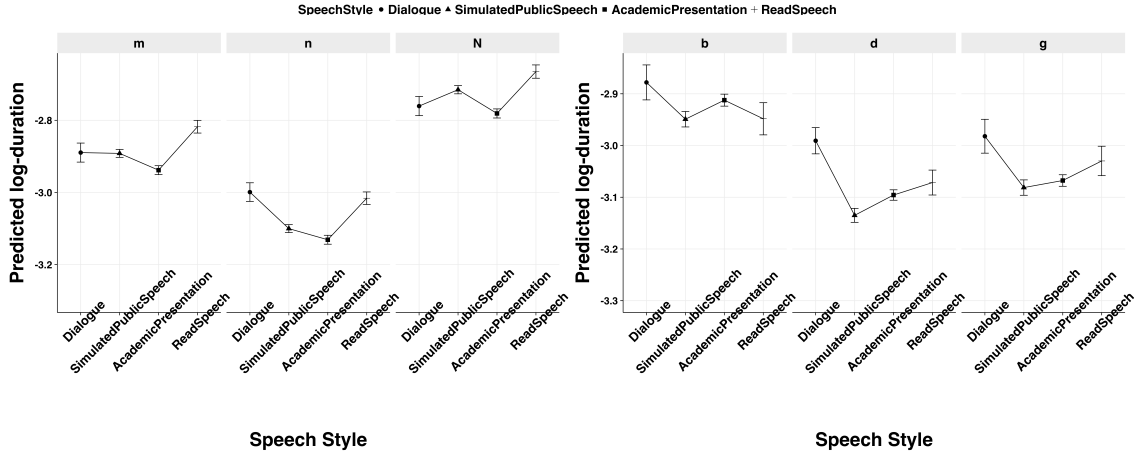


Figure 1: Interaction effect of Phoneme and SpeechStyle on log-duration

duration of /ɴ/ for read speech is the longest among the other speech styles [Dialogue and ReadSpeech ($t$=-3.61, $p<0.01$); AcademicPresentation and ReadSpeech ($t$=-5.56, $p<0.001$); SimulatedPublicSpeech and ReadSpeech ($t$=-3.05, $p$=0.05)]. We also found that the /ɴ/ durations of simulated public speech are significantly longer than academic presentations ($t$=5.24, $p<0.001$). The comparisons between dialogues and academic presentations, as well as dialogues and simulated public speech, are not significant. Analyses of /n/ revealed a different pattern from both /m/ and /ɴ/ in which the /n/ duration for both dialogues and read speech are longer (no significant difference between Dialogue and ReadSpeech) than simulated public speech and academic presentations [Dialogue and SimulatedPublicSpeech ($t$=4.97, $p<0.001$); Dialogue and AcademicPresentation ($t$=4.55, $p<0.001$); SimulatedPublicSpeech and ReadSpeech ($t$=-5.52, $p<0.001$); AcademicPresentation and ReadSpeech ($t$=-5.77, $p<0.001$)]. The comparisons between dialogues and read speech, as well as academic presentations and simulated public speech, are not significant. Figure 1 indicates in the left panel that dialogues show the shortest durations and read speech displays the longest durations for both /m/ and /ɴ/, as predicted. However, simulated public speech and academic presentations indicate no difference or an opposite relationship and does not fit with our prediction, in which simulated public speech should have longer duration than academic presentations. A possible explanation for this is that academic presentations are often prepared speech that are highly contentful, allowing speakers to deliver a very dense talk faster than unprepared speech. This might be the source of the shorter durations in academic presentations. Research has also found that the faster speech rate leads to more instances of reduction (Brand and Ernestus, 2015). Additionally, as seen on the left side of Figure 1 in the middle panel, /n/ shows an unexpected pattern in which there is no difference between dialogues and read speech, and both simulated public speech and academic presentations display shorter duration than dialogues and read

speech. It is possible that there is a particular word that is common in dialogues but that does not occur in the other speech styles causing this effect. However, further analysis is needed to explore this pattern and confirm our hypotheses about this effect.

Analyses of the duration of voiced stops across speech styles also revealed a main effect of SpeechStyle [F(3,10)=7.53, $p<0.01$] and Phoneme [F(2,37910)=125.99, $p<0.001$] as well as an interaction between SpeechStyle and Phoneme [F(6,36380)=3.40, $p<0.01$]. Voiced stops display different patterns, both from what we predicted and the pattern of nasals. Individual comparisons of the /b/ duration across speech styles revealed that all the comparisons are not significant. We also found that the duration of /d/ is longer for dialogues than for both academic presentations ($t$=3.57, $p<0.05$) and simulated public speech ($t$=6.05, $p<0.001$), but other comparisons are not significant. Although /g/ shows a similar pattern to /d/, none of the comparisons among the speech styles in /g/ are significant in this data set. As shown on the right side of Figure 1, the results do not follow our predictions for speech style. In addition, fewer comparisons in voiced stops reached significance as compared to in nasals. In order to further examine these tendencies, we conducted additional analyses by segmenting the stops into two parts: closure duration and release duration. The closure duration was defined as the duration from the offset of the preceding segment to the onset of a burst release, and the release duration was defined as the duration from the onset of a burst release to the onset of the following segment. However, due to the approximated articulation of stops, especially for dialogues, a number of the stops did not have closure and release durations as a separate unit, meaning that the boundary between the offset of closure duration and the onset of release duration was unclear. Table 1 illustrates the numbers of the stops with clear boundary between closure and release durations and with no boundary between the two durations across speech styles. As expected, the number of the stops with no boundary is the highest in dialogues and the lowest in read speech. For the following analysis, the stops with clear boundary (83,957 stops) were utilized.

Table 1: Number of voiced stops with ClearBoundary and NoBoundary and their ratio

| SpeechStyle | ClearBoundary | NoBoundary | Total | BoundaryRatio | NoBoundaryRatio |
|---|---|---|---|---|---|
| Dialogue | 4164 | 236 | 4400 | 94.64 | 5.36 |
| SimulatedPublicSpeech | 39839 | 1889 | 41728 | 95.47 | 4.53 |
| AcademicPresentation | 36556 | 1408 | 37964 | 96.29 | 3.71 |
| ReadSpeech | 3398 | 40 | 3438 | 98.84 | 1.16 |

### 3.1.1 Closure duration

Figure 2 illustrates an interaction between phonemes and speech style on release and closure durations. Statistical analysis of closure duration across speech styles demonstrated a main effect of SpeechStyle [F(3,15)=3.85, $p<0.05$] and Phoneme [F(2,36253)=184.78, $p<0.001$] as well as an interaction between SpeechStyle and Phoneme [F(6,33554)=5.936, $p<0.001$]. Individual comparisons of /b/ across speech styles revealed that none of the differences in closure duration across speech styles reached significance. For /d/, we found that relationships among the speech styles differ from that of /b/ where the closure duration of /d/ for dialogues is longer than for simulated public speech ($t$=7.07, $p<0.001$), and we also identified that the /d/ closure duration for simulated public speech is shorter than for both academic presentations ($t$=-4.0500, $p<0.01$) and read speech ($t$=-3.82, $p<0.01$). The other comparisons are not significant. The /g/ closure duration displays similar relationships among the speech styles to /b/ but the degree of durational differences across speech style are small; therefore, none of the differences reached significance. Figure 2 in the left panel indicates that closure duration shows a similar pattern to the entire stop duration where the durations are the longest in dialogues and the shortest in simulated public speech, but the degree of durational differences across speech styles for closure durations are smaller than for the entire stop durations.

### 3.1.2 Release duration

Likewise, analyses of release duration across speech styles revealed that there is an main effect of SpeechStyle [F(3,12)=4.12, $p<0.05$] and Phoneme [F(2,38921)=578.30, $p<0.001$] as well as an interaction effect between SpeechStyle and Phoneme [F(6,38632)=15.16, $p<0.001$]. We identified that the release duration of /b/ in dialogues is the longest among the other speech styles [Dialogue and SimulatedPublicSpeech ($t$=4.12, $p<0.01$); Dialogue and AcademicPresentation ($t$=4.83, $p<0.001$); Dialogue and Read-

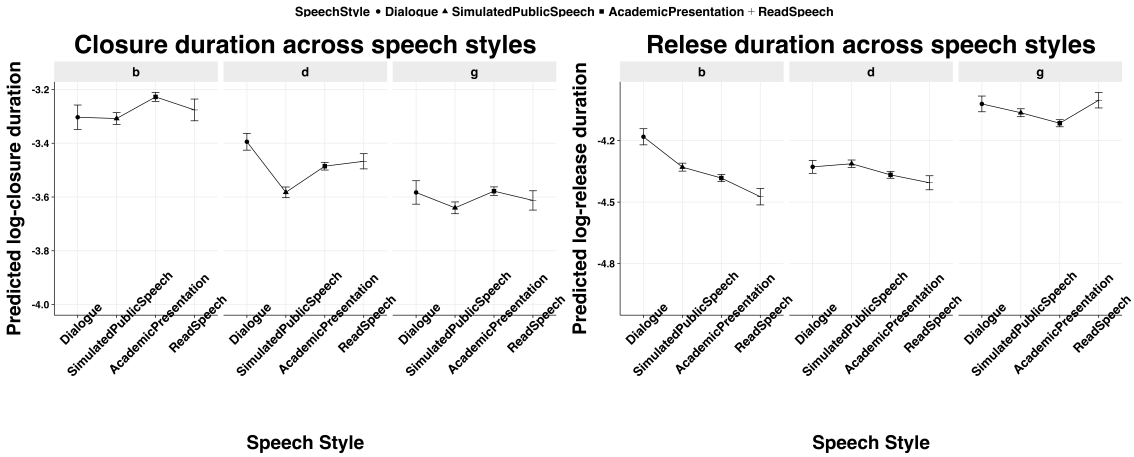


Figure 2: Interaction effect of Phoneme (voiced stops) and SpeechStyle on log-release and closure duration

Speech ($t$=5.17, $p$<0.001)]. We also found that the /b/ release duration for simulated public speech is longer than for read speech ($t$=4.06, $p$<0.01). The comparisons between academic presentations and read speech, as well as academic presentations and simulated public speech, are not significant. Individual comparisons of /d/ across speech styles revealed that release duration of /d/ displays similar relationships among the speech styles to /b/ but the degree of the differences are smaller. The /d/ release duration for simulated public speech is longer than for read speech ($t$=3.30, $p$<0.05), but all other comparisons did not reach significance. As shown by Figure 2 in the right panel, the /g/ release duration also indicates similar relationships among the speech styles to the other two stops except that read speech is relatively longer. Similarly, the degree of durational differences for /g/ across speech styles are small; therefore, none of the differences reached significance. As in the results of the entire duration of voiced stops, we identified an overall tendency towards the release duration being longer in casual speech (i.e., dialogues and simulated public speech) as compared to careful speech (i.e., academic presentation and read speech) except /g/. Interestingly, although the overall tendency was opposite to what we expected, release duration displayed consistent durational differences across speech styles. Additionally, the number of significant differences were greater than that of the entire duration of voiced stops. As a result, our findings here suggest that speech style difference is better reflected in release duration than the entire duration of the voiced stop. Importantly, as shown by the both panels of Figure 2, the relationships among the speech styles in closure and release durations vary, suggesting that the way in which speech styles impact release and closure durations differ.

### 3.2 Intensity difference

Figure 3 illustrates an interaction between phoneme and speech style on intensity difference. Analyses of intensity difference in nasals across speech styles demonstrated that there is a main effect of SpeechStyle [F(3,12)=16.8, $p$<0.001] and Phoneme [F(2,138612)=5168.7, $p$<0.001] as well as an interaction between SpeechStyle and Phoneme [F(6,137983)=51.7, $p$<0.001]. Individual comparisons of /m/ across speech styles revealed that the intensity difference of /m/ in academic presentations are greater than in read speech ($t$=3.73, $p$<0.01). Other comparisons did not reach significance. Analysis of /n/ revealed that relationships among the speech styles in /n/ is similar to that of /m/ in which the /n/ intensity difference for dialogues is smaller than for academic presentations ($t$=-3.24, $p$<0.05). We also found that the /n/ intensity difference for simulated public speech is greater than both for read speech ($t$=3.43, $p$<0.05) and academic presentations ($t$=7.42, $p$<0.001). The /ɴ/ intensity difference shows different relationships among the speech styles from that of both /m/ and /n/ but none of the differences reached significance. As shown by Figure 3 in the left panel, /m/ and /n/ display a similar pattern in which dialogues contain a smaller intensity difference as compared to both simulated public speech and academic presentations, as well as simulated public speech possesses a smaller intensity difference as compared to academic presentations. However, unlike our prediction, the intensity difference of read speech is smaller than that of both simulated public speech and

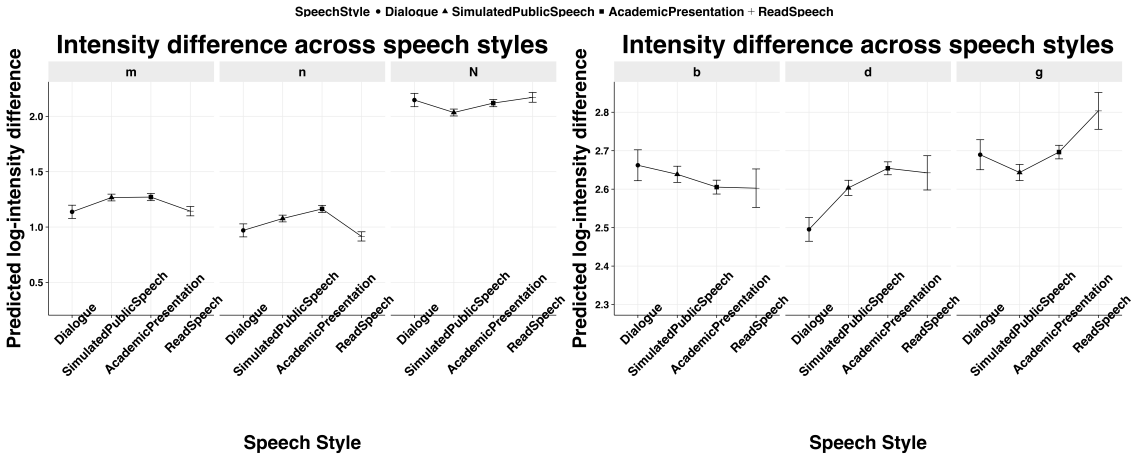


Figure 3: Interaction effect of Phoneme and SpeechStyle on log-intensity difference

academic presentations in both /m/ and /n/. Additionally, the intensity difference of /N/ are similar across speech styles except that simulated public speech shows a smaller intensity difference than the others. A possible explanation for this could be that due to the sonority hierarchies, the intensity difference between stops (oral stops) and neighbouring sounds (most likely vowels) is expected to be smaller than the intensity difference between nasals (nasal stop) and neighbouring sounds (most likely vowels). This means that inherently voiced stops have more space to indicate reduction by a smaller intensity difference than nasals. It is possible that the intensity difference measurement is more appropriate to measure reduction for voiced stops (oral stops) than nasals (nasals). This account is reflected in the results of voiced stops.

Analyses of intensity difference in voiced stops revealed that there is no main effect of SpeechStyle but there is a main effect of Phoneme [$F(2,38816)=61.61$, $p<0.001$] and an interaction between SpeechStyle and Phoneme [$F(6,38831)=25.55$, $p<0.001$]. Individual comparisons of /b/ across speech styles indicate that none of the differences among the speech styles reached significance. Analyses of the /d/ intensity difference show that the relationships among the speech styles in /d/ differ from that of /b/ where the /d/ intensity difference for dialogues is smaller than for both academic presentations ($t$=-5.51,$p<0.001$) and read speech ($t$=-3.25, $p<0.05$). We also found that the /g/ intensity difference in simulated public speech is smaller than in read speech ($t$=-3.22,$p<0.05$). The other comparisons are not significant. As shown by the right panel in Figure 3, the relationships among the speech styles in /d/ and /g/ display, to a great extent, what we expected and the intensity differences across speech styles are relatively consistent. The intensity difference is smaller in more casual speech styles (i.e., dialogues and simulated public speech) as compared to careful speech (i.e., academic presentation and read speech). Our findings here support what we discussed above that the intensity difference measure is more useful for voiced stops than for nasals.

## 4 Conclusion

In the present study, we used a large-scale spontaneous speech corpus to examine the phonetic variability of nasals and voiced stops in Japanese. We attempted to account for how the variation and reduction of nasals and voiced stops occur across speech styles by measuring the duration and the intensity difference of target segments. We hypothesized that the shortest duration and the smallest intensity difference would be observed in dialogues and the longest duration and the largest intensity difference in read speech. Unlike what we predicted, the phonetic variability exhibited complex patterns across both phonemes and speech styles. As a result, our findings revealed a few important aspects of the phonetic variability and the effect of speech styles. First, the way in which speakers reduce segments is not consistent across both speech style and phoneme. That is, the way speakers implement reduction is variable depending not only on speech style but phoneme. Second, the relationships among the speech styles in closure and release durations vary. In other words, the way in which speech styles influence closure and release durations differ. Third, the intensity difference measure is more useful for voiced stops than for nasals. Further research is needed

to investigate the instances where the segment is deleted and the segments that are realized as different phonemes (e.g., /d/ → [ɾ]).

## References


Arai, T., Warner, N., and Greenberg, S. (2007). "Analysis of spontaneous Japanese in a multi-language telephone-speech corpus". *Acoustical Science and Technology*, 28:1, 46–48.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2017). "Package lme4". *R package version*, 1.1-13.

Boersma, P. and Weenink, D. (2016). "Praat: doing phonetics by computer [Computer program]".

Brand, S. and Ernestus, M. (2015). "Reduction of obstruent-liquid-schwa clusters in casual french". In *the 18th International Congress of Phonetic Sciences (ICPhS 2015)*.

Ernestus, M., Baayen, H., and Schreuder, R. (2002). "The recognition of reduced word forms". *Brain and Language*, 81:1–3, 162 – 173.

Ernestus, M. and Warner, N. (2011). "An introduction to reduced pronunciation variants". *Journal of Phonetics*, 39:3, 253–260.

Johnson, K. (2004). "Massive reduction in conversational American English". In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.

Kohler, K. J. (1990). "Segmental reduction in connected speech in german: Phonological facts and phonetic explanations". *Speech production and speech modelling*, 55, 69–92.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). "Package lmerTest". *R package version*, 2.

Lennes, M., Alarotu, N., and Vainio, M. (2001). "Is the phonetic quality of unaccented words unpredictable? an example from spontaneous finnish". *Journal of the International Phonetic Association*, 31:1, 127–138.

Lenth, R. (2017). "Package lsmeans". *R package version*, 2.26-3.

Maekawa, K. (2003). "Corpus of Spontaneous Japanese: Its design and evaluation". In *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*.

Maekawa, K. (2005). "Toward a pronunciation dictionary of Japanese: Analysis of CSJ". In *Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2005)*, (pp. 43–48).

R Core Team (2017). "R: A language and environment for statistical computing". *R Foundation for Statistical Computing. Vienna, Austria*.

Torsten Hothorn, Frank Bretz, P. W. R. M. H. A. S. S. S. (2016). "Package multcomp". *R package version*, 1.4-6.

Tucker, B. V. (2007). *Spoken word recognition of the reduced American English Flap*. phdthesis, The University of Arizona.

Tucker, B. V. (2011). "The effect of reduction on the processing of flaps and /g/ in isolated words". *Journal of Phonetics*, 39:3, 312–318.

Warner, N. and Tucker, B. V. (2011). "Phonetic variability of stops and flaps in spontaneous and careful speech". *The Journal of the Acoustical Society of America*, 130:3, 1606–1617.