PERCEPTUAL EFFECT OF VOWEL DEVOICING AND ITS WORKING RANGE

Makiko Aoyagi

Dokkyo University
aoyagi@dokkyo.ac.jp

ABSTRACT

This study reports a perceptual effect of a devoiced vowel in Japanese as an assimilative outcome of connected speech. A focus is placed on how such an effect changes its magnitude as the source and the recipient of the effect are separated.

A devoiced vowel *ki-* has a strong effect of causing voiceless judgment in the following *da-ta* VOT variations with natural closure durations. However, such an effect is weakened as the closure interval is expanded beyond the original one for a voiceless stop. In this case, the original voicing judgment of the *da-ta* variations in isolation gradually returns. Also, the more voiced the *da-ta* variation itself, the earlier and farther it deviates from the influence of the devoiced vowel. Phonetic variation resulting from connected speech serves to aid segmental perception, but in turn the effect obtains in a 'connected' speech event.

Keywords: vowel devoicing, Japanese, voicing judgment, perception, closure duration.

1. INTRODUCTION

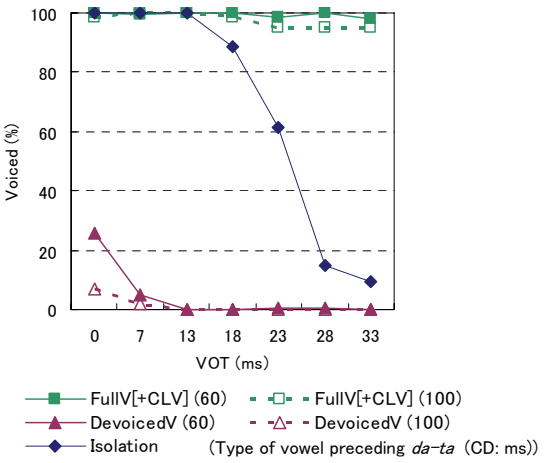
Segmental sequences cause phonetic variation according to the contextual coordination of speech. For example, the first vowel in /apa/ is manifested quite differently from that in /aba/. Some differences are a shorter duration and an acute termination of the vowel for /apa/ and a longer duration and the presence of continuing voicing during closure for /aba/. When the vowel is /i/, as in /ipa/, it is even shorter and more weakened. Furthermore, the vowel /i/ can be devoiced, in Japanese for example, when preceded by another voiceless consonant, as in /kipa/ [kipa]. These regular phonetic variations in turn may naturally help the listener restore the segmental sequence.

1.1. Japanese vowel devoicing

In Japanese, unaccented high vowels between voiceless obstruents are devoiced. This typical devoicing is almost mandatory except in a few environments where devoicing is avoided [7, 8].

As the environment for devoicing is thus predictable, Aoyagi [1] and Aoyagi and Komatsu [3] hypothesized that a listener anticipates a voiceless consonant to follow upon hearing a devoiced vowel. They demonstrated a strong effect of a devoiced vowel inducing voiceless judgment in the following stop. Figure 1, from [1], shows voicing judgments of *da-ta* VOT variations in isolation and those when preceded by [ki̥] with the voice devoiced, and by [ki] with closure voicing present, both spliced with two closure durations. The graph shows that the *da-ta* variations preceded by a devoiced vowel are mostly judged voiceless across the VOT range, showing a strong effect of the preceding devoiced vowel, compared with the line for the *da-ta* variations in isolation.

Figure 1: Voiced judgment rates of *da-ta* VOT variations in isolation and when preceded by [ki̥] (DevoicedV) and by [ki] (FullV[+CLV]), spliced with the closure durations (CD) 60 and 100 ms – from Aoyagi [1]



Another finding in the studies was the effect of closure duration. As has been pointed out [6, 10], a longer closure duration caused more voiceless judgments, and a shorter one more voiced judgments, reflecting the patterns in production.

1.2. Objective of study

Given that a devoiced vowel induces voiceless judgment in the following stop and a longer

closure adds to the effect, it is of interest to investigate "how long" a closure duration helps the effect. If devoicing is a result of concatenation of adjacent segments, the said anticipatory effect for voicelessness should dissipate as the source and the recipient of the effect become farther apart. In order to find out the relations between the devoiced vowel's effect and the closure interval, a perceptual experiment was conducted.

2. EXPERIMENT

A devoiced syllable [k̥i] was spliced with *da-ta* VOT variations with increasing closure intervals. The listener made voicing judgments between *kita* and *kida*. The result was compared with the voicing judgments of *da-ta* variations in isolation.

2.1. Stimuli

The stimuli were the same as those in Aoyagi [1]. The words *ta*, *da*, *kita* and *kida* (in the LH pitch pattern) were recorded many times by a native Japanese speaker and digitized (sampling at 22 kHz, 16-bit quantization) for later editing. The variations below were derived from *da* and *kita* that had a similar duration and F0 in their [a].

2.1.1. Devoiced ki-

A devoiced *ki-* syllable was extracted from *kita* [k̥ita] naturally uttered in the LH pitch pattern.

2.1.2. Da-ta continuum

The *da-ta* continuum was created using the method by Repp and Lin [9]: by removing the voicing lead of *da*, and replacing each of the initial glottal cycles of *da* with an aperiodic segment of *ta* of equal duration. Six stimuli were made with the VOT values of 0, 7, 13, 18, 23 and 28 ms.

2.1.3. Splicing

The devoiced *ki-* and the six *da-ta* syllables were spliced with nine silence intervals, or closure duration (CD), of 60, 100, 120, 140, 180, 220, 260, 300, 340 ms to make 54 stimuli (1×6×9=54).

2.1.4. Experimental stimuli

The *k̥i-da/ta* stimuli were mixed with 20 dummy tokens which had a variety of voicing durations, amplitudes and termination manners in *ki-*. Each of the 74 stimuli (54+20=74) was randomly arranged five times in the computer-assisted instruction program *FCAI* (74×5=370).

2.2. Subjects and procedure

Nineteen native speakers of Japanese from Tokyo and its neighboring areas (14 males and 5 females, aged from 19 to 40 with a mean age of 21.8) participated in the experiment.

After a practice session, each subject heard each stimulus a single time through a headset and judged he or she heard *kita* or *kida*. No feed-back was provided. After the 370 stimuli, the subject heard *da-ta* variations in isolation, randomly arranged five times each on the program (6×5=30), to make voicing judgments. This session was only after the *kita/kida* session to lessen the possibility of exclusive focus on the *da-ta* part in the *kita/kida*.

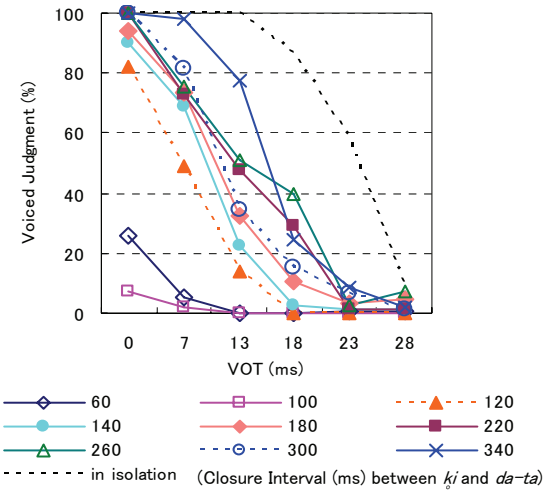
Each of the 74 *k̥i - da/ta* stimuli and the 6 *da-ta* VOT variations was judged 95 times for voicing judgment rates (5 times × 19 subjects = 95).

3. RESULTS

3.1. Voiced judgment rates

Figure 2 shows the mean voiced judgment rates of *da-ta* variations preceded by devoiced *ki-* with various closure intervals. First, the far right broken line is for the *da-ta* variations in isolation, showing a pattern similar to those for categorical perception. This serves as the baseline for comparison.

Figure 2: VOICED judgment rates of [devoiced *ki-* + *da-ta* variations] with increasing closure intervals (60 - 340 ms)



The bottom two lines for 100 and 60 ms closure intervals (CD) show a clear contrast to the baseline. For example, the *da-ta* at 7 and 13 ms VOT were judged 100% voiced when presented in isolation, but the same stimuli were judged voiceless at fairly high rates when preceded by devoiced *ki-* with those intervals. As the 100 ms was the original CD

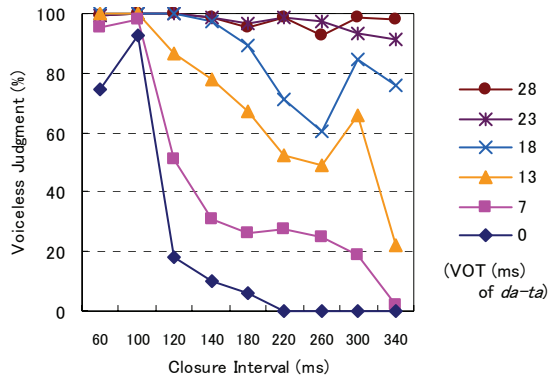
for *kita* and 60 ms was for *kida*, the case with the shorter shows a higher voiced judgment rate.

Except the above cases, i.e. the 100 ms interval shows lower voiced judgment rates than the 60 ms one, the voiced rates increase as the interval becomes longer, especially at shorter to midrange VOT. A general trend is that the rates draw back towards the baseline with increasing closure intervals. Apparently, the preceding vowel's effect is weaker on *da-ta* stimuli with shorter VOT and on those with longer intervals. Still there remains the effect even with the 340 ms interval, showing a downward shift for voicelessness from the baseline.

3.2. Voiceless judgment rates

Figure 3, transformed from Figure 2, better illustrates the decreasing effect of the preceding devoiced vowel. It shows the 'voiceless' judgment rates of *da-ta* VOT variations preceded by devoiced *ki-* in function with the closure interval.

Figure 3: VOICELESS judgment rates of [devoiced *ki-* + *da-ta* variations (0-28 ms VOT)] with increasing intervals.



The top two lines are cases with 23 and 28 ms VOT. They were judged voiceless at 41.7% and 91.7% respectively in isolation, but they were judged mostly voiceless, placed after devoiced *ki-*. They stay near the 100% line, indicating that they are still judged as voiceless even when placed 340 ms away from the preceding devoiced vowel.

The other lines show general declines of voiceless judgment rates, though with peaks at 300 ms. After showing the highest voiceless judgment rate at 100 ms, which is the original CD for *kita*, the rates decrease with longer intervals. The original voicing judgments of *da-ta* variations return as the distance between the devoiced *ki-* and the *da-ta* syllable becomes larger.

Also, the shorter the VOT of the *da-ta*, the earlier and farther it deviates from 'voiceless' to more 'voiced.' The difference in voiceless rates due to VOT of the *da-ta* syllables is the smallest at 100 ms and expands steadily with longer intervals, with the steepest decline for the *da-ta* of 0 ms VOT.

3.3. Statistical analyses

A two-way repeated-measure ANOVA revealed the main effects of CD [$F(8, 64) = 85.145$, $p < .001$] and of VOT of the *da-ta* syllable [$F(5, 40) = 210.355$, $p < .001$] with interaction between CD and VOT [$F(40, 320) = 18.758$, $p < .01$].

Next a one-way repeated-measure ANOVA for each VOT of the *da-ta* syllable, taking CD as the factor, revealed that the effect of CD was significant for *da-ta* with 0, 7, 13 and 18 ms VOT, but not for 23 and 28 ms VOT. Then another one-way repeated-measure ANOVA for each CD type, taking the VOT of the *da-ta* syllable as the factor, revealed that the VOT was a significant factor except at 100 ms CD. These support the foregoing graph interpretations of decreasing voiceless judgment rates over intervals and the difference in the decrease degree by the *da-ta* variation.

A further analysis shows that the interaction between VOT and CD is in a significant linear relation [$F(8, 8) = 183.594$, $p < .001$], supporting the observation of steeper declines for *da-ta* syllables with shorter VOT.

4. DISCUSSION

This study reconfirms Aoyagi's [1] finding that the preceding devoiced *ki-* had a strong effect of causing voiceless judgment in the following *da-ta* syllable. The magnitude of the effect was as strong as completely reversing the voicing judgment. The effect was by far greater than that reported in Repp and Lin [9]. In their study, a precursor ending with a voiceless consonant or a whispered vowel shifts the *p-b* categorical boundary for slightly shorter VOT for more voiceless judgment.

The effect of the preceding devoiced *ki-* reaches its peak at 100 ms CD, which is considered to be optimal for a voiceless stop. Therefore, the voicing judgment of all VOT conditions showed nearly 100% voiceless judgments at that interval. Thereafter, however, the effect gradually decreases with increasing intervals. The longer intervals gradually weakened the vowel's effect and allowed the original voicing judgment of *da-ta* to return. The vowel's effect decays as the source and the recipient of the effect are placed farther away from each other. This is analogous to Summerfield [10], which demonstrated that the closure interval had an effect on the voicing judgment of a following

stop as long as it did not exceed the range of the natural interval period. Kashino [6] also claimed that the closure duration effect obtained when the pre- and post- closure signal in (C)V-CV, *ape* for example, was interpreted as a single speech event. The characteristics of connected speech seem to operate on perception when the signal is perceived as a connected speech unit.

As shown earlier, VOT and the interval interact with each other; i.e. the change in the preceding vowel's effect varies with the VOT of following *da-ta* variations. Variations with long VOT are more in line with the voicelessness and are subject to the devoiced vowel's effect over extending intervals whereas variations with shorter VOT, more voiced in their original nature, start to resist the effect much sooner with shorter intervals.

It is speculated that the rises in the voiceless rate at 300 ms in Figure 3 may be due to the phonological restriction on voiced geminates in Japanese. The listeners may have had the linguistic bias and judged the stimuli with the closure, long enough to be a geminate, as voiceless. As they are not as conspicuous in their own voicing nature as those with 0 and 28 ms for example, their behavior seems to be influenced more by the context.

Of several factors involved in the effect of the preceding devoiced vowel: the closure interval, the VOT value of respective *da-ta* variations and possibly the restriction on voiced geminates, the first two were statistically significant, but it is not known with our data if the last condition is valid.

Aoyagi [1] and Aoyagi and Komatsu [3] demonstrated other phonetic cues related to the voicing judgment of *da-ta* variations. As the preceding /ki/ is gradually reduced from fully voiced with closure voicing to minimally voiced and to devoiced, the voicing judgment of the following syllable becomes more voiceless, also gradually. Thus, phonetic variations as contextual cues lie in various temporal ranges – from one as local as VOT, to a wider one of closure duration and then of transitions from the preceding segment. Then these cues in temporal changes of speech are referred to within some natural range of connected speech and used by listeners to restore segments.

Aoyagi [2] also reports that an identical phonetic signal [-kita-] is perceived differently by phonological specifications. This suggests that the listener is sensitive to contextual variations in connected speech, but reference to phonological restrictions may change the way the phonetic information is used.

5. CONCLUSIONS

The present study first reconfirmed a strong effect of a devoiced vowel on the voicing judgment of the following stop. A devoiced *ki-* caused mostly voiceless judgments in the following *da-ta* variations across the VOT range. As devoicing occurs between voiceless consonants, such anticipation is a reflection of a production pattern.

However, the preceding devoiced vowel's effect obtained as long as the closure duration, or interval, was within a natural range for that of a stop. The effect decayed as the distance between the vowel and the stop, the source and the recipient of the effect, were placed farther away from each other. With increasing intervals, the original voicing judgment of the *da-ta* VOT variations in isolation gradually returned. Furthermore, the shorter the VOT of *da-ta*, the earlier and farther it deviated from the preceding devoiced vowel's effect towards its voiced nature.

Phonetic variations are a result of segmental concatenation in connected speech. The listener is sensitive to such temporal changes of speech for segmental identification, but such phonetic information appears to operate on perception when the signal is perceived as a 'connected speech' unit.

* This is a revised version of part of Aoyagi [2].

6. REFERENCES

[1] Aoyagi, M. 2002. Contextual effect on voicing judgment of following stop consonant in Japanese. *Proc. 7th ICSLP* Boulder, 1129-1132.

[2] Aoyagi, M. 2006. *Perceptual Effects of Vowel Devoicing and Reduction in Japanese.* Ph.D. Dissertation. Sophia University.

[3] Aoyagi, M., Komatsu, M. 2003. Vowel reduction and voicing judgment of following stops: comparison between Jpn and Eng speakers. *Proc. 15th ICPhS* Barcelona, 203-206.

[4] Han, M. 1962. The feature of duration in Japanese. *Study of Sounds* 10, 65-80.

[5] Homma, Y. 1981. Durational relationship between Japanese stops and vowels. *J. Phonetics* 9, 273-281.

[6] Kashino, M. 1991. Effects of auditory streaming on voicing judgment of Japanese intervocalic stop consonants. *J. Acoust. Soc. Japan* 90 (4), 2361.

[7] Kondo, M. 1997. *The Mechanism of Vowel Devoicing in Japanese.* Ph.D. Dissertation. Edinburgh University.

[8] Maekawa, K. 1989. Boin no museika [Vowel devoicing]. In: Sugito, M. (ed.) *Koza Nihongo to Nihongo Kyoiku*, vol. 2, *Nihongo no Onsei,On'in*. Tokyo: Meiji Shoin, 135-153.

[9] Repp, B., Lin, H.B. 1991. Effects of preceding context on the voice-onset-time category boundary. *J. Exp. Psychology: Human Perc. and Perf.* 17, 289-302.

[10] Summerfiled, Q. 1981. Articulatory rate and perceptual constancy in phonetic perception. *J. Exp. Psychology: Human Perc. and Perf.* 7, 1074-1095.