

## CHAPTER 4

### EXPERIMENTAL DESIGN

#### 4.1 INTRODUCTION TO CHAPTER 4

This chapter introduces the experimental design of the study. §4.2 presents the motivation for the study; §4.3 discusses the influences affecting the application of HVD. §4.4 presents the dependent variables of the study, the variables being investigated, while §4.5 presents the independent variables of the study, the variables thought to be influencing the dependent variables. Finally, §4.6 gives the methods and procedures utilized to test the effect of the independent variables on the dependent variables.

#### 4.2 MOTIVATION FOR TESTING SR EFFECTS

As noted in the introduction, various studies have upheld the traditional formulation of HVD as a fast-speech rule (Kuriyagawa & Sawashima 1989, Maekawa 1990, Kondo 1994). However, as also noted, there have been indications to the contrary. Vance (1987) anecdotally notes reduced SR effects for some speakers, and Beckman (1994) outright states that ‘devocalization’ occurs even in the most careful read speech. This same reduced effect of SR for some speakers has been noted in experimental situations as well in Kondo (1994), Han (1994), and Varden & Sato (1996). In order to confirm and expand upon these indications, the study being discussed in this dissertation was undertaken to detail the effects of SR on the devoicing of younger Tokyo speakers. The rationale was that whereas most previous studies were based upon the pronunciation of older speakers (often the author or one of the authors of the paper), the three native speakers who participated in the study reported in Varden & Sato (1996) as controls were all younger speakers whose primary dialectal influence was from the Tokyo area. Further, the two younger of these three speakers, ages 21 and 25, devoiced more vowels than the third participant, who was in her mid-30s. This all indicates that the rule of HVD is becoming more widespread in its application.

This follow-up study was therefore designed to test the effects of SR on the devoicing of vowels by younger Tokyo speakers.

#### 4.3 INFLUENCES AFFECTING THE APPLICATION OF HVD

Since the rule appears to be in flux, it was desirable to get as broad a view of the application of HVD as possible. As noted in Han (1994) it is usually desirable to keep the number of variables being investigated to a minimum (usually one or two) in order to avoid one variable compromising the effects of another, known as confounding of effects. For example, using all female participants in a study avoids the necessity of taking into account differences between women and men's speech.

However, in order to get a broad overview of current application of the rule, as many contributing variables as possible were included in the present study. Although this confounds the analysis of the data, methods are available for the analysis of multiple effects; once factors are indicated as significant more detailed research into those particular factors can be undertaken in future work.

Experimental control of the following effects was attempted in the following manners:

- **age and economic bracket effects:** participants were limited to students from our university (ages 19-24; all estimated upper-middle class bracket)
- **dialectal effects:** participants were limited to those who were raised in the Tokyo area
- **2nd language-learning effects:** participants were limited to those who had no significant exposure to 2nd language as judged by interview<sup>1</sup>
- **part-of-speech effects:** all tokens used in the study are nominals
- **syllable structure effects:** all tokens are of the same canonical CVCV form (consonant-short vowel-consonant-short vowel; e.g. *kuki* )

---

<sup>1</sup>Although many Japanese students have highly specific English-to-Japanese translation skills, regrettably lack of listening/speaking ability remains the norm.

- **sentential position effects:** all tokens were placed at the beginning of carrier sentences
- **speaker speech rate variability:** SR was controlled by the presentation program
- **effects of using different carrier sentences:** all carrier sentences contained the same number of mora, the timing units for Japanese<sup>2</sup>
- **practice effects:** a practice session of three repetitions of three randomly-chosen sentences at each of the three speech rates was given
- **effects of changing speech rate:** a 3-2-1 countdown at the new speech rate was given each time the speech rate changed; the same count-down was given in the practice session and the actual data collection.
- **list reading effects:** the order of presentation of the sentences within each repetition set was randomized by the program

The following effects were investigated by manipulating them in the experimental design:

- **gender effects:** both female and male participants were included since there is a very distinct separation of women's and men's speech in Japan
- **effects of moraic position:** tokens were all followed by a syntactic particle beginning with a voiceless stop to allow the 2nd mora vowel to be devoiced as well as the 1st
- **SR effects:** a Hypercard™ stack (program application) was used to display each carrier sentence for three different presentation durations (slow, normal, fast) on a computer monitor

As discussed in Chapter 3, the effect of pitch accent placement was not investigated in this thesis due to the lack of effect on devoicing noted in many studies and the free variation observed in the current data set.

Many of the above effects are discussed in detail below in §4.6 Methods & Procedures.

---

<sup>2</sup>See Kubozono (1995) and Han (1994) for discussion.

#### 4.4 DEPENDENT VARIABLES OF THE STUDY

As discussed in Chapter 1, there are two main points of investigation being pursued in this thesis, namely:

- 1) the relationship between speech rate (as controlled by sentence display times) and the number of devoiced vowels that participants produced at each SR; i.e. did the number of devoiced vowels increase proportionately along with SR, or are devoiced vowels being produced at the same frequency regardless of SR?
- 2) the relationship between token duration (the duration in ms of each target word) and vowel voicing duration (the duration of voicing in ms for each voiced vowel); i.e. did the duration of vowel voicing decrease proportionately with a decrease in token duration?

However, due to a lack of known documentation on using a program such as this for data elicitation, it was first necessary to determine that the program was successful in producing 3 distinct sets of data suitable for statistical analysis. This was done by investigating the effect of the sentence display times (i.e. the controlled SR) on the token duration of tokens produced by the participants.

The following dependent variables are associated with these three points:

- 1) TOKEN DURATION was used to check how successful the program was in eliciting usable data (i.e. were the 'slow', 'normal' and 'fast' token durations separate enough to support statistical analysis) and any other factors affecting these durations;
- 2) V1 VOICED and V2 VOICED, an indicator of whether the 1st or 2nd vowel of a token, respectively, was voiced or not, was used to check the number of vowels devoiced (i.e. not voiced) at each SR; and
- 3) V1 VOICING DURATION and V2 VOICING DURATION, the duration of voicing of the 1st and 2nd vowel, respectively, was used to check whether a phonetic account of devoicing by gestural overlap is upheld.

#### 4.5 INDEPENDENT VARIABLES OF THE STUDY

The following table details the independent variables that are associated with the effects discussed in §4.3 above.

Table 4.1 Independent variables identified in the study.

<i>variable label</i>	<i>description of the variable (possible values are in parentheses)</i>	<i>reason for suspected effect on token duration</i>
BLOCK	repetition block the stimuli were presented in (1, 2)	participants may have performed differently on the second block of repetitions due to fatigue, boredom, etc.
REPETITION	ordinal repetition of each token at each SR in each block (1-3)	participants may not have been ready for changes in display duration; some repetitions may have reflected the previous display duration or an over-anticipation of SR change
PARTICIPANT	participant (10 speakers)	ideolectal differences are always present; they may have affected both SR and frequency of devoicing
GENDER	participant gender (M, F)	females and males may have different average SRs and HVD application characteristics
TOKEN	token stimulus (10 tokens)	the inherent durations of the different segments in the tokens may have affected token durations; different consonantal environments may have affected the application of HVD
CLITIC	the syntactic particle following the token ( <i>ka</i> , <i>to</i> )	two particles were used, <i>ka</i> and <i>to</i> ; the different inherent durations of their segments may have affected token durations

Table 4.1 (continued)

MORA	the ordinal number of a mora (1, 2)	the vowel of the 1st mora of a word may be more readily devoiced than the vowel of the 2nd mora, or vice versa, independent of pitch location
SR	speech rate as determined by the presentation program (slow, normal, fast)	an increase in speech rate necessarily involves shortening of segmental material

In addition, the dependent variable #DEV, the number of devoiced vowels found in a given token (0, 1 or 2), was treated as an independent variable in the calculations even though it is actually a dependent variable itself because its value is dependent on the application of HVD. In statistical terms, #DEV is included in the model as a *covariate regressor*. This point will be discussed further in §5.3.

## 4.6 METHODS & PROCEDURES

### 4.6.1 PARTICIPANTS

Participants were solicited by advertisement from among the student population of the university in Tokyo where the author teaches. Stated requirements of the advertisement included having been raised in the Tokyo area, having parents from the Tokyo area, and having no real conversational ability in another language. Respondents completed a questionnaire roughly detailing their age and linguistic background. Among the 30 or so students who completed the questionnaire, six females and five males were selected for recording after an interview by a native speaker of Japanese who verified that the information on the questionnaires was correct. All participants were in the age range of 18~23 and were paid for their participation. The participants' ages and residences are given in Table A.1 of the Appendix.

A bilingual student assistant from among the participants' peer group was hired to help with the data collection. As much as possible, interaction with the participants was carried out by him to avoid unwanted social interaction between the participants

and the principal investigators.<sup>3</sup> He was explicitly directed and coached to guide the participants toward accomplishing their task in as low-key a manner as possible.

Of the 11 participants selected, one proved unable to consistently complete the data collection phase due to visual difficulties; the data from that participant was not analyzed. The study is therefore based on the productions of 10 speakers, 6 females and 4 males.

## 4.6.2 MATERIALS

### 4.6.2.1 STIMULI

In Varden & Sato (1996), a list of 14 nominal Japanese tokens was chosen as stimuli. However, in analyzing the data of that study it was very difficult to judge the duration of fricative-initial tokens (e.g. *shishi* ‘lion’, *sushi* ‘sushi’) particularly at faster SRs; many times there was a very gradual onset of frication, making it difficult to determine where to start the token duration measurement. For this reason, fricative-initial tokens were not used in the current study; only stop- or affricate-initial tokens were used to provide stop closure releases in the waveforms to use as landmarks for measurement. This resulted in a total of 10 tokens being used to check for vowel devoicing.

As mentioned above in §4.3, all tokens used were nominals to ensure that no part-of-speech effects would occur (e.g. possibly vowels in verbal forms are more resistant to devoicing than vowels in nominal forms). Each token was of the canonical form  $C_1VC_2V$ , where  $C_1$  is a voiceless plosive or affricate,  $C_2$  a voiceless plosive, affricate or fricative, and  $V$  is a high vowel. Further, each token was followed by a syntactic particle (enclitic) that also began with a voiceless plosive, either [t] or [k], resulting in  $C_1VC_2V-C_{t/k}V$  stimuli. This allowed the devoicing of the second vowel of the token as well as the first, since HVD applies across all syntactic and morphological

---

<sup>3</sup>The class division between teachers and students in Japan remains a wide one; many students exhibit a great deal of anxiety when forced to deal directly with their instructors. In addition, the majority of our students have never had direct contact with a foreigner outside of the confines of the classroom, leading to unwanted social anxiety.

boundaries (i.e. it appears to be a P2 rule in the manner of Kaisse 1985, 1990). Tokens were not balanced for pitch accent; tokens may have been accented on either the first or second vowel, or unaccented.

Each token was placed in a coordinate nominal structure of the form “stimulus and something” or “stimulus or something”. These coordinate nominal structures were placed at the beginning of carrier sentences, stimuli tokens first; e.g. *Chichi to haha ga genki da*. ‘Father and mother are healthy.’ In this example the target token and following syntactic particle are *chichi to* ‘father and’. A full list of the tokens and the carrier sentences is given in Appendix A.

A different carrier sentence was used for each token in an attempt to more closely approximate natural speech. Each carrier sentence was balanced for length and was 10 morae long; since Japanese is generally regarded as a mora timed language (see Kubozono 1995, Han 1994 for discussion) each sentence should take about as long to produce as the others at a given SR.

The 10 carrier sentences used to check for vowel devoicing were intermixed with 10 other sentences used to check for consonant elision and V degemination and lengthening. That data was collected for the use of the other investigator in this study. These 10 sentences were also 10 morae long, although they did not utilize the coordinate structure described above. That data is not discussed in this thesis.

#### 4.6.2.2 PRESENTATION PROGRAM

A major concern of the study was how to achieve differing rates of speech without unduly influencing the timing and rhythm of the participants’ productions. Most previous studies have relied on the speakers’ own judgment (e.g. Kuriyagawa & Sawashima 1989; Varden & Sato 1996) although other methods have been tried—e.g. Maekawa (1990) had speakers time their productions between clicks of a metronome, and Kondo (1994, 1997) had speakers time their productions between tones presented over headphones. After literature review and deliberation it was decided to use a Hypercard™ stack<sup>4</sup> that would present the sentences containing the

---

<sup>4</sup>The visual analogy for a Hypercard™ stack is a stack of index cards, with each index card representing each card in the Hypercard™ stack. The program provides

tokens to the participants by leaving the sentences on a computer display for a certain duration of time.

The Hypercard™ stack, written by the current author and entitled *The Stimulizer*, allowed the stimuli sentences to be presented on the computer screen. It displayed the sentences on a 17" computer monitor in large typeface (48-point text at 640 x 480 resolution) at a distance of approximately one meter from the participants. A sample screen is given below, with the Japanese translated into English (the translation was not present on the screen during the data elicitation).

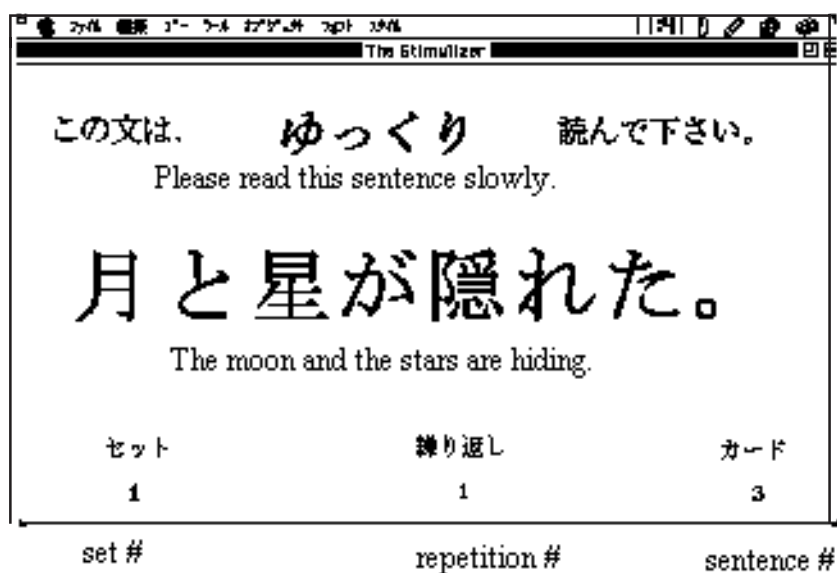


Figure 4.1 Sample screen from the presentation program.

When the stack was opened, it randomized all 10 sentences of a given set of sentences (see §A.2.2 of the Appendix for the list of sentences in each set) and wrote them out to 10 cards, one sentence to each card. These 10 cards were presented to the participant one at a time at each SR. All cards at a given SR remained on the screen for the same period of time; the uniformity of durations for both computers used during recording sessions was verified with a stopwatch before recording began.

---

means of controlling the content of each card, as well as means for navigating between them.

Sentences remained on the screen for three different display durations. These durations were determined during trial runs of the program as being good examples of slow, normal and fast SRs by three native speakers of Japanese, two not involved in the study and the other investigator involved in the project. Durations used for the collection of this data were 4.0 seconds for the slow SR, 2.8 seconds for the normal SR, and 1.7 seconds for the fast SR.<sup>5</sup> Participants were directed to match their production of each sentence to the time it was displayed on the computer screen. Display duration was therefore used as a means of controlling SR.

As noted above, the stack first presented the 10 cards containing the randomized sentences for the slow display duration. The sentences were then randomized among the 10 cards again, and were presented for the normal display duration. Finally, the sentences were randomized a third time before the cards were presented for the fast display duration. This constituted one repetition of each sentence at each SR.

This cycle of one repetition of each sentence at each SR was repeated 3 times using the sentences in the 1st set of sentences, followed by a similar 3 repetition cycles of each sentence in the 2nd set of sentences. This constituted one repetition block. Each set of sentences was then presented for a second set of 3 repetitions in exactly the same manner as just described, constituting the second repetition block (see the A.2.3 of the Appendix for the presentation order of the sentences).

For training, three sentences from Block 1 were displayed once at each display rate; the three sentences were randomly chosen by the program. All participants stated that one training run was adequate to learn the paradigm. After training was completed, the repetition of the sentences began. Except for the one participant whose data was not included in the study,<sup>6</sup> no difficulties completing the elicitation task were observed or reported.

---

<sup>5</sup>These durations are by coincidence quite similar to the 3.3 s, 2.5 s and 1.7 s durations utilized in Kondo's (1994) study. However, since the carrier sentence in that study was 15 morae long, the actual SR imposed on participants in that study would have been faster.

<sup>6</sup>This participant was allowed to participate in the experiment despite having a very weak left eye. However, it proved impossible for him to consistently produce the

#### 4.6.2.3 RECORDING

After arriving at the studio where recording took place, participants were greeted by the investigators in the study. They were then led into the studio where the bilingual assistant was waiting. The participant and assistant were left alone, and the assistant explained the procedure to the participants. Since the carrier sentences were written using Chinese characters as well as Japanese script (the 1st, 3rd and 5th characters of the sample sentence above), as is normal in Japanese writing, a list of the sentences to be displayed was reviewed by the assistant and each participant to ensure that the participant was familiar with each Chinese character.

Training consisted of three sentences from Block 1 displayed once at each display rate, with the three sentences being randomly chosen by the program. When the training run was completed, recording was begun. An analog Aiwa one-point condenser microphone, model CM-S3, and a Sony MD Walkman™ portable mini-CD recorder with a built-in sampling rate of 44.1 kHz, model MZ-R3, were used for recording. The microphone was placed at a distance of approximately 500 centimeters from the participants.

As previously mentioned, there were 20 sentences used in the study, 10 sentences contained tokens for the devoicing research and 10 sentences contained tokens for another study by the other investigator involved in this study. The 20 sentences were intermixed, and presented in two sets of 10 sentences each. All 10 sentences of a set were presented first for the ‘slow’ display duration in random order, then all 10 for the ‘normal’ display duration in random order, and then all 10 for the ‘fast’ display duration in random order. This consisted of one cycle of repetitions. The order of presentation of the sentences is given in the Appendix.

Each cycle of repetitions was repeated three times for each set of 10 sentences. After both sets of 10 sentences had been repeated three times at each of the three display times (9 repetitions of each sentence total), the entire process was repeated again in a second block of repetitions. This resulted in 1800 token repetitions (10 participants

---

full set of sentences at the fast SR, and so his data was excluded. The participant was still paid for his participation.

x 10 tokens x 3 speech rates x 6 repetitions of each token at each speech rate) and 3600 possible devoiced vowels (1800 token repetitions x 2 vowels in each token).

After recording, signals were fed into an Apple™ PowerPC™ 7100 66/AV computer for signal analysis with the program Signalyze™. The analog output of the earphone jack of the MD recorder was used, since this allowed volume control of the signal during sampling by the computer. The reconstructed analog output from this jack was sampled by the computer at 22.05 kHz during input with a standard Apple™ sound card. Due to the analog signal being reconstructed from a 44.1 kHz digital signal, no loss of information was expected during re-sampling at 22.05 kHz.

#### 4.6.3 MEASUREMENTS

Measurements of various parameters of the tokens were taken. These included token durations, vowel voicing durations, formant frequencies, formant frequency durations, and total number of vowels devoiced at each speed. Of these, token durations, vowel voicing durations, and number of devoiced vowels were exhaustive, and will be discussed in the following chapters.

##### 4.6.3.1 *TOKEN DURATION*

As mentioned previously, display time of the carrier sentences was used as a means of controlling SR; token durations were used as a measure of the effect of SR.

Duration of the entire sentence was rejected as a measure of SR because of the SR adjustments that participants made to match their production durations to display durations. Likewise, using mora durations as a measure of SR is unjustified since the word, not the mora, is indicated by previous research as the level at which duration control applies (Homma 1981; Port et al 1987; Han 1994). These studies have found a strong correlation between the number of morae that a Japanese word contains and the overall word duration—words seem to come in whole number duration units in Japanese, corresponding to how many morae there are in the word. Duration of individual morae appear to be adjusted to accommodate inherent segmental durations of other segments in the word in order to achieve the target word duration. (See

Perkell 1997, Farnetani 1997, and Löfqvist 1997 for discussion of articulatory and temporal adjustment mechanisms.)

The durational measurement settled on was one which approximated word duration; i.e. token duration. However, since the initial stop of each token was also the initial segment of the sentence, it was not possible to observe the onset of closure in either the waveform or a spectrogram.

Token durations were therefore measured from the release of the token-initial stop or affricate to the release of the syntactic particle's initial stop. This measurement does not provide a true indication of the duration of the token because it does not include the token-initial stop's closure gesture and does include the particle-initial stop's closure gesture. However, this technique took advantage of the fact that stop and affricate releases are clear landmarks in both waveforms and spectrograms.

A representative token duration measurement is given in the figure below.

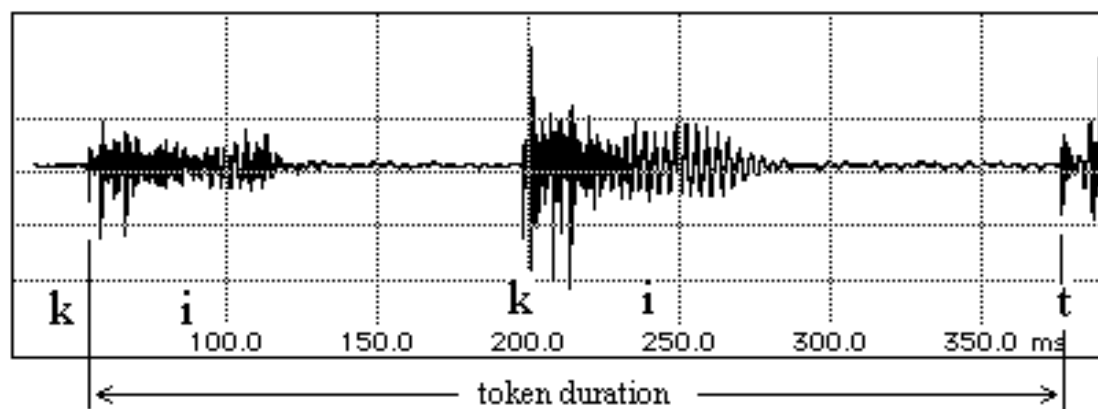


Figure 4.2 Token duration measurement criteria; shown is *kiki to* 'nervousness and'.

When necessary, judgments were aided by examining a wide-band spectrogram of the token/particle pair.

#### 4.6.3.2 VOICING DURATIONS

Like token durations, vowel voicing durations were taken directly from expanded waveforms. However, in order to make the cyclic activity corresponding to voicing

more visible in the waveform, waveforms were filtered with a 6th-order low-pass Butterworth filter set at 1 kHz before vowel voicing duration measurements were taken.

Filtering out the upper frequencies of the signal before taking voicing duration measures was beneficial for two reasons. First, the high frequency sibilance accompanying the onset and offset of vowels often obscured the beginning and/or ending of the cyclic voicing activity—the ‘fuzziness’ of the waveform representing the high frequency components of the signal made it difficult to tell when the cyclic activity corresponding to voicing actually began or ended. This can be seen in the top panel of Figure 4.3 below, where it is quite difficult to judge the onset and offset of voicing activity.

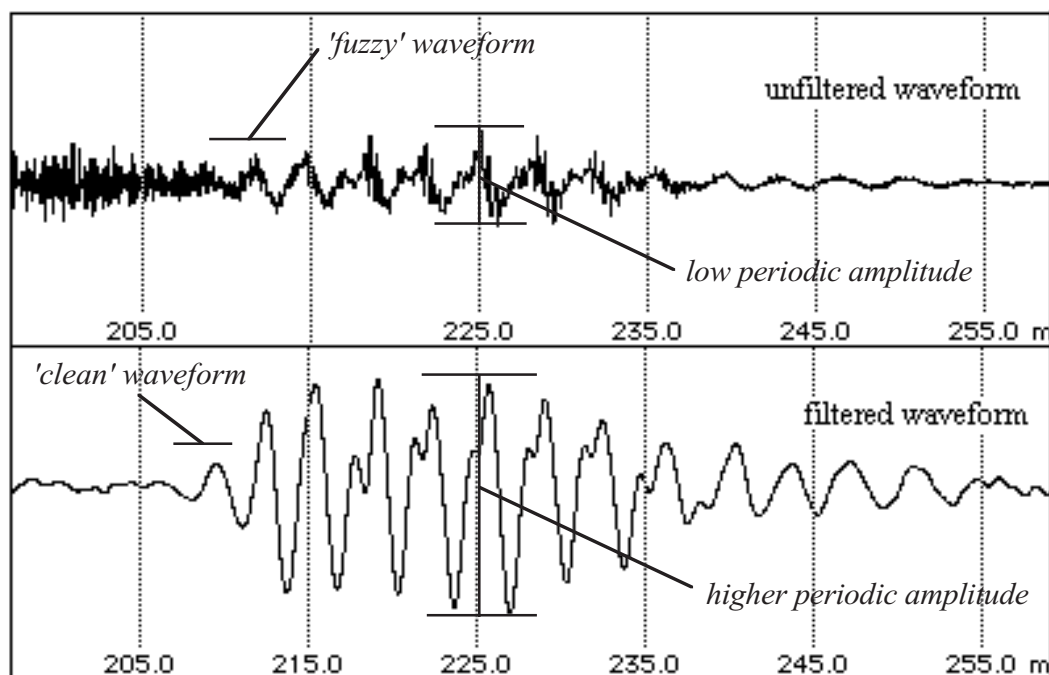


Figure 4.3 Second vowel of token *tsuchi* ‘dirt’ by participant ANa , before and after low-pass filtering at 1 kHz.

In contrast, the onset and offset of voicing is quite clear in the second panel.

Second, low-pass filtering the signals resulted in a magnification of the waveform display of the lower frequencies. As can be seen in the second panel of the figure above, this vertical magnification of the filtered waveform results in a much clearer representation of the lower frequencies, allowing more confident judgments.

After low-pass filtering at 1 kHz, voicing durations were made from the expanded waveforms. The beginning and the end of the voicing were judged on a combination of factors: the ‘cleanness’ of the cyclicity, the sinusoidal envelope of the cyclic activity, and the offset of the cyclic activity.

Judgment calls were still required in many cases. In particular, what might be called "trailing voicing" was not included in measurements for voicing durations. This can be seen somewhat in Figure 4.4 below, where the main voicing period (from about 350 ms to about 445 ms) was followed by several periods of very low amplitude activity. These cycles are judged to be trailing; that is, voicing due to the vocal cords being left in proximity, allowing them the 'flap in the breeze' as it were, rather than being involved in active voicing.

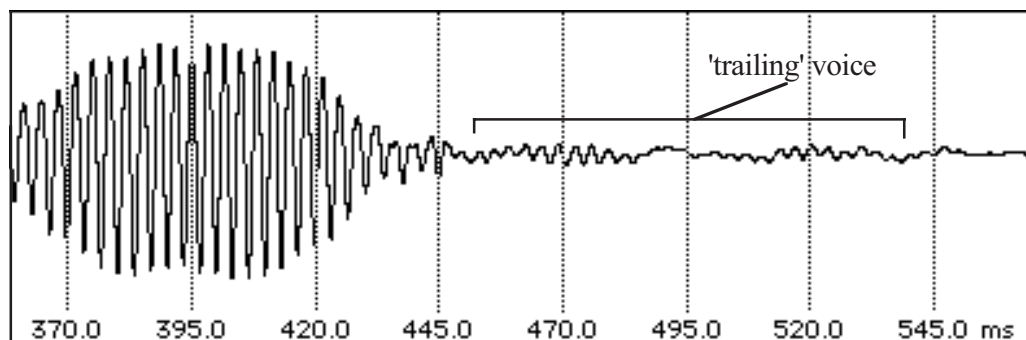


Figure 4.4 Filtered waveform of 2nd vowel of token *kuki* ‘stem’ by participant YU showing ‘trailing voice’.

It appears that the articulatory setting for voicing readiness by the vocal cords, for this participant in particular, is in quite close proximity to the point necessary to initiate voicing, and occasionally the two settings overlap.

In addition, several tokens by participant TO contained what might be called ‘incidental voice’. These were 2-3 cycles of periodic energy at the appropriate frequency to be considered voicing, but they occurred far removed from what would

be considered the location of the vowel in these tokens—much closer to the beginning of the mora than the end. These several bursts of periodic energy were also judged not to be active vowel voicing, but were assumed to be due to an increase in transglottal air flow associated with the release of the stop while the vocal cords were in close proximity in a state of readiness for voicing. It seems reasonable to attribute this participant's unusually located voicing activity to incidental voicing activity due to her heavy use of 'creaky' voice, as described in Ladefoged (1993), during many productions.

#### 4.6.3.3 FORMANT FREQUENCY MEASUREMENTS

As discussed in the previous chapter, one characteristic of the fricative vowels generally observed in this study is a different set of formant frequencies than voiced vowels due to the altered shape of the oral cavity in the formation of the fricative closure. In order to constrain the scope of this thesis, however, exhaustive checks and analysis of these altered formant frequencies is left for further research. It will only be noted here that the measurements that were made were also made from filtered waveforms, as were the fundamental frequencies. In cases where formant frequencies were close to fundamental frequencies for a particular participant's vowel and fricative vowel productions, a combination of zero crossing measures (see §2.3) and auditory impression were used to judge whether the observed cyclic activity was due to voicing or vowel coloring of a fricative vowel.

#### 4.6.4 STATISTICAL ANALYSIS

In order to track the data collected during measurement, a database was constructed using the program FileMaker Pro™. The following information was catalogued for each token (V1 = the first vowel of the token; V2 = the second vowel of the token):

- a token label consisting of the participant's initials, block number, repetition number, the SR, and the token;
- the token duration (ms);
- V1 and V2 status (voiced, devoiced);
- if V1 was voiced, the V1 voicing duration;

- if V2 was voiced, the V2 voicing duration; and
- notes about any unusual activity.

After labeling, values for each of the variables mentioned in this chapter were exported to the statistical programs StatView™ and SuperAnova™ for statistical analysis. Results of the analyses are discussed in the next 3 chapters.