Ann.Bull.RILP(1985)No.19, 133-138

# AN OBSERVATION ON SYNTHETIC JAPANESE SPEECH SOUND IN UNITS OF FIXED DURATION

Yukie Masuko and Shigeru Kiritani

## Introduction

Naturalness in speech sounds is thought to depend importantly on the duration of each segment which composes the speech sound chain. Conceptually, the durations of moras in Japanese are thought to be equal. But, practically, it is known that the results of measurement show that the duration of moras differs from each other under various conditions.

Some researches have shown how the duration of each segment or each mora is changed by the condition of the preceding or following segments, or by the kinds of segments, or by the existence of pauses and breaths. Some trials have been also made to find the rules controlling duration.

This paper investigates, by the method of LPC analysis, how and to what extent the naturalness of speech sounds is influenced by changes in the duration of segments or moras according to various conditions. This is to say, speech sounds were synthesized that had all units equal in duration, for example the duration of all the moras was the same, and to what extent the naturalness is lowered was examined.

## Method

The speech sample used in this experiment was the first three sentences of "Momo-Taroo", which is a Japanese folk tale.

1. mukashimukashi arutokoro-ni ojiisan-to obaasan-to-ga arimashita
2. ojiisan-wa yama-e takigi-o tori-ni obaasan-wa kawa-e sentaku-ni ikimashita
3. obaasan-ga sentaku-o shiteimasu-to kawakami-no hoo-kara momo-ga tsunburatsunbura-to nagarete-kimashita

These three sentences were pronounced by an adult man (a professional announcer). These original sounds were converted to A/D under the condition of 10kHz, and a 12bit LPC analysis was performed. The parameters for the LPC synthesis were stored in a computer. The order of the analysis was 12, and the interval of each frame was 10msec.

On the basis of these data the durations were controlled according to the methods shown below, and the synthesized speech sounds were listened to and examined.

Method 1. The duration of Consonant + Vowel is fixed.
The duration from the beginning of a consonant to the end of the following vowel is treated as one unit. In the case where a vowel is followed by a mora without a consonant, Consonant + $Vowel_1$ + $Vowel_2$ is treated as two units. $Vowel_1$ and $Vowel_2$ are lengthened/contracted equally. /N/, a mora phoneme, is treated in the same way, so Consonant + Vowel + /N/ is two units.

Method 2. The duration of $\frac{1}{2}Vowel_1$ + Consonant + $\frac{1}{2}Vowel_2$ is fixed.
The duration from the middle of a vowel to the middle of the following vowel is treated as one unit. In the case where a vowel is followed by a mora without a consonant, a temporary boundary is made between the two vowels. $\frac{1}{2}Vowel_1$ + Consonant + $Vowel_2$ is treated as one and a half units, and also the following $Vowel_1$ + Consonant + $\frac{1}{2}Vowel_2$. As for /N/, a mora phoneme, $\frac{1}{2}Vowel_1$ + /N/ + Consonant + $\frac{1}{2}Vowel_2$ is treated as two units.

Method 3. The duration of $\frac{1}{2}Consonant_1$ + Vowel + $\frac{1}{2}Consonant_2$ is fixed.
The duration from the middle of a consonant to the middle of the following consonant is treated as one unit. In the case where a vowel is followed by a mora without a consonant, $\frac{1}{2}Consonant_1$ + $Vowel_1$ + $Vowel_2$ + $\frac{1}{2}Consonant_2$ is treated as two units. /N/, a mora phoneme, is treated in the same way.

In order to control durations according to the methods mentioned above, the duration of each segment, such as vowels and consonants, was measured on a sonograph. Based on the data, the duration of each unit was calculated and the average of the units in a sentence was calculated. The duration of every unit in a sentence was controlled as the average; in other words, every unit was lengthened/contracted to be equal to the average. It was the vowels that were lengthened/contracted, and frames of 10msec were inserted or deleted in the middle of the vowels.

In Method 2 the amount to be lengthened/contracted in one unit was shared by the first vowel and the second vowel equally.

The amounts lengthened/contracted are shown in Table 1,2 and 3.

## Results and Conclusions

An informant listened to the synthesized sounds and checked how the naturalness was damaged. The results are shown in Table

4,5 and 6. When a segment was felt unnatural because it was too long, a "↑" mark was attached. When a segment was felt unnatural because it was too short, a "↓" mark was attached. In the other cases in which a segment was felt unnatural, a "*" mark was attached.

Our conclusions after examination of Table 1,2,3,4,5 and 6 are as follows.

The percentage of the damage to diphthongs, long vowels and vowels before /N/ was higher than that of damage to the others. This should be because the unit that contains a diphthong, a long vowel or an /N/, is treated as two.

A comparison was made between the segments after a pause (that is, the beginning of a sentence or the beginning of a phrase just after a breath) and the segments before a pause (that is, the end of a sentence or the end of a phrase followed by a breath). In the original sound, the units before a pause were almost always long. Though units longer than the average were contracted, not a vowel was heard to be unnatural at the end of a sentence or a phrase. However, some were heard to be unnatural after pauses. This may show that position in a phrase or a sentence has an influence on the limits within which the change in duration of a segment is heard to be natural.

In the cases other than those mentioned above, it cannot positively be said that the unnaturalness depends on the amount of the change in the duration. Further investigation will be made on what kinds of factors have a relation to naturalness.

Moreover, there were some cases where, not the lengthened/contracted vowel itself but the preceding or the following segments were heard to be short/long or deformed. The cause of this fact will also be investigated in the future.

Now, the number of vowels lengthened/contracted 30msec or more than 30msec was 83 out of 9 sentences. In spite of these changes, the number of those felt to be unnatural was smaller than we expected. This may be partly because the synthesized sounds used in this experiment differs from the sounds commonly synthesized by rules on these two points; one is that the former had the same pitch contour as the original sounds, the other is that the average speed of a sentence was the same as the original sentence.

In addition, there were some cases where, not the lengthened/contracted vowel itself, but the following segments or the preceding segments or the whole word or the word with a particle that contained the lengthened/contracted vowel were heard to be unnatural. Further research will be made on the extent of influence that changes in duration can have.

Table 1. The amount lengthened/contracted in the first sentence

(msec)

| | mu | ka | shi | mu | ka | shi | a | ru | to | ko | ro | ni | o | jii | san | too | baa | san | to | gaa | ri | ma | shi | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method 1 | +70 | -30 | -40 | +30 | -30 | -110 | +60 | +60 | 0 | -20 | +10 | -80 | +40 | +50 | 0 | +20 | +30 | +20 | -60 | +10 | +30 | -20 | -30 | 0 |
| Method 2 | 0 | -30 | -20 | +10 | -50 | -50 | +20 | +20 | 0 | +10 | 0 | -30 | +10 | 0 | +10 | +50 | +20 | -10 | +40 | +70 | +20 | -20 | -10 | 0 |
| Method 3 | +50 | -60 | -10 | +20 | -60 | 0 | 0 | +30 | 0 | +20 | -10 | 0 | 0 | +20 | 0 | +40 | 0 | -10 | -20 | +20 | +20 | -30 | -10 | 0 |

Table 4. The segment felt unnatural in the first sentence

felt long: ↑ felt short: ↓ felt deformed: *

| | mu | ka | shi | mu | ka | shi | a | ru | to | ko | ro | ni | o | jii | san | too | baa | san | to | gaa | ri | ma | shi | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method 1 | ↑ | * | | | * | | ↑ | ↑ | | ↓ | | | ↑ | * | | | | | | | | | | |
| Method 2 | | * | | | * | | ↑ | | | | | | | | | | | | ↑ | | | | | |
| Method 3 | ↑ | * | | | * | | | | | | | | | | | | | | | | | | | |



Table 2. The amount lengthened/contracted in the second sentence

(msec)

| | o | jii | san | wa | ya | mae | ta | ki | gio | to | ri | ni | o | baa | san | wa | ka | wae | sen | ta | ku | nii | ki | ma | shi | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method 1 | +60 | +50 | 0 | -80 | +10 | -20 | -40 | -50 | +20 | -30 | +20 | -60 | +80 | +60 | -100 | +40 | +10 | +20 | +30 | -40 | 0 | +60 | -20 | -30 | 0 | +40 |
| Method 2 | +10 | 0 | -10 | 0 | 0 | -20 | -50 | -10 | 0 | 0 | 0 | -20 | +20 | +10 | -20 | 0 | 0 | 0 | +20 | +10 | +30 | +20 | 0 | 0 | +10 | +10 |
| Method 3 | 0 | +50 | 0 | 0 | +20 | -40 | -50 | 0 | 0 | 0 | 0 | 0 | 0 | +40 | -70 | 0 | 0 | 0 | +40 | -40 | +30 | +20 | +10 | -20 | +10 | 0 |

Table 5. The segment felt unnatural in the second sentence

felt long: ↑ felt short: ↓ felt deformed: *

| | o | jii | san | wa | ya | mae | ta | ki | gio | to | ri | ni | o | baa | san | wa | ka | wae | sen | ta | ku | nii | ki | ma | shi | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method 1 | ↑ | * | | | | | | | | | | | ↑ | | ↓ | | | | | * | * | | | | | |
| Method 2 | | | ↓ | | | | | * | | | | | | | ↓ | | | | | | | | | | | |
| Method 3 | | | | | | | | * | | | | | | | ↓ | | | | | * | * | | | | | |

Table 3. The amount lengthened/contracted in the third sentence

(msec)

| | o | baa | san | ga | sen | ta | kuo | shi | tei | ma | su | to | ka | wa | ka | mi | no | hoo | ka | ra | mo | mo | ga | tsun |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method 1 | +80 | +40 | +30 | -20 | +60 | -30 | +20 | +10 | +80 | -20 | -40 | -40 | +50 | +50 | -30 | +20 | 0 | +30 | -40 | -20 | -10 | -30 | -100 | -30 |
| Method 2 | +20 | +10 | +20 | +20 | +10 | +20 | +20 | +40 | +40 | -10 | -30 | -10 | 0 | 0 | +10 | +20 | 0 | -10 | 0 | +10 | -10 | -50 | -40 | 0 |
| Method 3 | 0 | +20 | +40 | -30 | +40 | -40 | +20 | +20 | +60 | -30 | -30 | 0 | +30 | +30 | -10 | +20 | -30 | +10 | 0 | 0 | -10 | -40 | 0 | -10 |

Table 6. The segment felt unnatural in the third sentence
felt long: ↑ felt short: ↓ felt deformed: *

| | o | baa | san | ga | sen | ta | kuo | shi | tei | ma | su | to | ka | wa | ka | mi | no | hoo | ka | ra | mo | mo | ga | tsun |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method 1 | ↑ | ↑ | | | ↑ | ↓ | * | | | | | | ↑ | ↑ | | | | | | | | | | ↑ |
| Method 2 | | | | | | | ↓ | | | | | | | | | | | | | | | | | |
| Method 3 | ↓ | ↑ | ↓ | | ↑ | ↓ | * | | | | | | ↑ | | | | | | | | | | | ↑ |



Table 3. (continued)

(msec)

| | bu | ra | tsun | bu | ra | to | na | ga | re | te | ki | ma | shi | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method 1 | +30 | 0 | -50 | +20 | +20 | -90 | +40 | -10 | +30 | -20 | -10 | -30 | -10 | +30 |
| Method 2 | +10 | 0 | -20 | 0 | -30 | -40 | 0 | +10 | 0 | -30 | -10 | 0 | 0 | +10 |
| Method 3 | +30 | -70 | -10 | +20 | -20 | 0 | +20 | 0 | -10 | -30 | +10 | -20 | 0 | 0 |

Table 6. (continued)

| | bu | ra | tsun | bu | ra | to | na | ga | re | te | ki | ma | shi | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method 1 | ↑ | | ↓ | | ↑ | | ↑ | | | | | | | |
| Method 2 | | | | | | | | | | | | | | |
| Method 3 | ↑ | | | | | | | | | | | | | |

References

1. Sagisaka,Y. and Tohkura,Y. (1980); Characteristics of Segmental Durations in Connected Speech, Transactions of Committee on Speech Research, The Acoustical Society of Japan, S80-34, 267-273.
2. Higuchi,N. and Fujisaki,H. (1980); Durational Control of Segmental Features in Connected Speech, Transactions of the Committee on Speech Research, The Acoustical Society of Japan, S80-40, 315-321.
3. Higuchi,N. and Fujisaki,H. (1981); Influence of Neighboring Phonemes upon Duration of Vowels in Connected Speech, Transactions of the Committee on Speech Research, The Acoustical Society of Japan, S80-96, 741-748.
4. Sagisaka,Y. and Tohkura,Y. (1981); Rule of Segmental Durations Using Statistical Features of Segment, Transactions of Committee on Speech Research, The Acoustical Society of Japan, S80-72, 561-568.
5. Sagisaka,Y. (1981); The Control of Segmental Duration and its Perception, Transactions of the Committee on Speech Research, The Acoustical Society of Japan, S81-22, 169-175.
6. Hoshino,M. and Fujusaki,H. (1983); A Study of Perception of Changes in Segmental Durations, Transactions of the Committee on Speech Research, The Acoustical Society of Japan, S82-75, 593-599.