ACOUSTIC PROPERTIES OF PHONEMES IN CONTINUOUS SPEECH FOR DIFFERENT SPEAKING RATE

Hisao Kuwabara

Department of Electronics and Information Science
The Nishi-Tokyo University
Uenohara, Kitatsuru-gun, Yamanashi 409-01, Japan

ABSTRACT

An investigation has been made for individual phonemes focusing mainly on their duration in continuous speech spoken in different rates: fast, normal, and slow. Fifteen short sentences uttered by four male speakers have been used as the speech material which comprises a total of 291 morae. Normal speaking rate (n-speech) is, on average, 150 milliseconds/mora (or 400 morae/minute) and the four speakers have been asked to read the sentences twice as fast as (f-speech) and 1/2 times as slow as (s-speech) the normal speed in reference to the n-speech.

Among consonants, the greatest influence has been found to occur on the syllabic nasal /N/ and the least on the voiceless stop /t/ in f-speech. For the s-speech, /N/ has also been found to be the greatest but the least is voiced stop /d/. The ratio of duration between consonant and vowel of a CV-syllable in the f-speech is kept almost the same as that in the n-speech while vowel lengthening becomes significantly large in the s-speech.

1. INTRODUCTION

Recent progress in speech technology has made it possible to build elaborate systems that can process speech signals more precisely than ever before in many technological areas. Not to mention speech recognition and speech synthesis, there are a variety of applications in speech technology area.

Keeping this technological background in mind, this study has been conducted as a basic research in order to provide an acoustic data for these speech technologies.

Japanese language basically consists of a series of consonant -vowel syllables (CV-syllables). Unlike English or other languages, each syllable corresponds exactly to one Japanese alphabet which is called "仮名."

As it is well known, each syllable in a continuous speech does not carry enough phonetic information to be correctly identified by itself, but rather spread over adjacent phonemes due mainly to coarticulation effects. [1, 2] There are some attempt to recover these reduced ambiguous phonemes. [3, 4] These perceptual evidences must be attributed to such acoustic properties of each phoneme as shortening its duration, reduction of pitch and formant frequencies.

These acoustic properties should, of course, vary from speaker to speaker, from one speaking rate to another. Some studies have so far been made to establish prosodic rules for different speaking rates. [5]

This paper deals with the acoustic properties, chiefly in terms of duration, of individual phonemes in a continuous speech for different speaking rate.

2. SPEECH MATERIAL

Fifteen short sentences have been chosen as the speech material. Table 1 shows a list of these sentences with Romanized transcription. In this table, "N" represents a syllabic nasal which exactly corresponds to a Japanese letter "ん"and "ー" depicts a long vowel like [oo] as in English word "root." "Q" represents the so-called "促音" (soku oN) in Japanese. In the pronunciation "kaQ ta", for example, the vowel /a/ in /ka/ becomes shortened down to abrupt end and followed by a relatively long silent interval before the /ta/ utterance.

1. *ko no ka i do- su ji wa mi ru be ki sa N gyo- mo na ku te tsu do- no be N mo wa ru i*
2. *de N shi ke N bi kyo- de no ka N sa tsu ni hi tsu yo- na shi ryo- zu ku ri o shi te i ru*
3. *su be te no mi chi wa ro- ma e tsu- ji ru*
4. *ki ta to ki to wa ta i sho- no ho- ko- e ha shi ri ha ji me ta*
5. *de- ta o nyu- ryo ku su ru ma e ni jo- ke N seQ te i shi ma su*
6. *ta bi no ra ku da ga tsu ki no sa ba ku o yu ki ma shi ta*
7. *ka re wa na ni ka i i ta so- ni shi te i ta*
8. *ka no jo ga e ki no ho- mu no mu ko- de te o fuQ ta*
9. *to ri a e zu juQ ko no sa N pu ru o ro ku o N si ta*
10. *za N ne N da ke do shiQ pa i si ta*
11. *wa re wa re wa na N ni mo shi na kaQ ta*
12. *o wa ri na go ya wa shi ro de mo tsu*
13. *so re wa ri zu mu a N do bu ru- su to so- ru o be- su ni shi ta*
14. *ma sa ni ka N jo- i nyu- sa re ta mo no daQ ta*
15. *mo de ru no da to- se i o jiQ ke N ni yoQ te sho- me- shi ta*

Table 1: Fifteen Japanese short sentences used in the experiment.

Four male adult speakers who participated in this experiment were asked to read the sentences three times with different speaking rate: normal speech which is referred to as "n-speech" in this paper, fast rate (also referred to as "f-speech") and slow rate ("s-speech").

There is a rhythm when it comes to speak a Japanese sentence. The rhythm, which is sometimes called syllable-timed, is based on the mora which roughly corresponds to a Japanese letter or CV-syllable.

The number of morae per minute defines the speaking rate. Generally, normal speaking rate (n-speech) falls into a speed from 300 to 400 morae per minute but it considerably differ from speaker to speaker, especially between the young and the old.

No special guidance and equipment have been used to control the speed in pronouncing the n-speech, f-speech and s-speech. Therefore, the number of morae per minute sometimes largely differ among the speakers even for the speech in the same rate.

For the f-speech, individual speakers were asked to pronounce the sentences twice as fast as the n-speech that they usually utter in daily conversation. For the s-speech, they were also asked to pronounce half as slow as the n-speech. For each speed, speech data were actually measured later on for speakers individually.

There are 291 morae in the fifteen sentences. Thus, a total of 3,492 (=291morae × 3rates × 4speakers) morae have been gathered to be analyzed.

3. MEASUREMENT OF DURATION

Measurements of duration for individual vowels and consonants have been made to investigate how the duration is affected by the speaking rate. Generally, there are no clear-cut standard positions to define the beginning and the end of each phoneme in a continuous speech except, for example, for plosive consonants where a silent interval usually precedes.

Position of each CV-syllable has been identified first and then consonant- and vowel-parts have been separated for the measurement of length.

Three criteria have been set in order to define the length of phonemes.

- The beginning of such consonants as fricatives, plosives, and affricates is easily determined by inspecting the waveform.
- When it is inseparable between consonant (or vowel) and vowel, transitional part is defined first, and the distinction should be made at the center of the transition.
- Distinction is made between syllabic nasal /N/ and nasal consonant /n/ or /m/ based on hearing.

UNIX workstations have been used in defining the positions of individual phonemes by inspecting both speech waveforms and spectrograms sometimes with a help of hearing. A total of approximately 7,000 distinctions have been made so far.

4. THE RESULTS

Duration data have been pooled for speakers individually and a statistic analysis has been made first, and then these statistic data have been analyzed over all speakers.

Silent intervals, though undoubtedly contribute to the speaking rate, are discarded from the analysis this time except those for /Q/.

4.1. Speaking Rate

Table 2 represents average duration per mora, or approximately one CV-syllable duration, for the three speaking rates. Slow speech tends to be uttered slower than expected, that is, a mora is greater than twice as long as the normal speech (220.7% in reference to the n-speech being n-speech length a 100%).

	fast	normal	slow
Duration (ms)	93.9	156.2	344.8
Rate (%)	60.1	100.0	220.7

Table 2: Average syllable duration and its ratio to the normal speech.

4.2. Consonant vs Vowel

Table 3 depicts the ratios at which an average CV-syllable is divided into consonant- and vowel-parts. Graphical illustration for this is shown in **Figure 1.**

	fast	normal	slow
Consonant (%)	35.7	34.6	24.3
Vowel (%)	64.3	65.4	75.7

Table 3: Average duration ratio between consonant- and vowel-parts in a CV-syllable for different speaking rate.

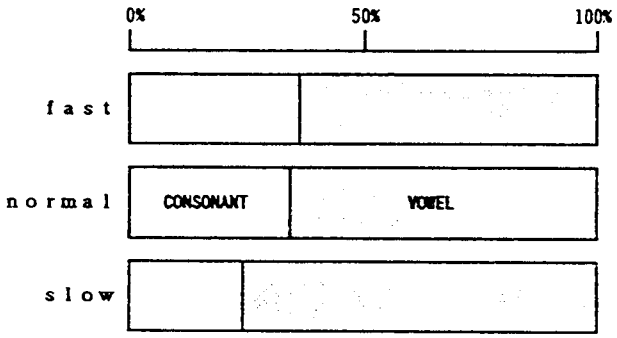

Figure 1: Ratio of duration between consonant- and vowel-parts in a CV-syllable.

For n-speech, the duration of consonant is approximately 1/3 of the total syllable length.

An interesting fact is that almost the same ratio is kept for the f-speech as the n-speech while, in s-speech, vowel part

becomes significantly large. This means that, when we utter f-speech, consonants and vowels tend to be shortened with approximately the same rate as that for the n-speech, but this does not retain any more in the s-speech.

Table 4 represents the average duration for consonant and also average duration for vowel together with the ratio in reference to the n-speech being the duration for n-speech a 100%.

	fast	normal	slow
Consonant (ms)	33.2	50.0	77.2
Rate (%)	66.4	100.0	154.4
Vowel (ms)	62.7	99.0	251.8
Rate (%)	63.3	100.0	254.3

Table 4: Average duration of consonant and vowel.

This Table clearly shows again that, for f-speech, consonant and vowel are approximately the same ratio (66.4% and 63.3%) while, for s-speech, vowel part is significantly lengthened (154.4% Vs 254.3%).

4.3. Vowels and Long -Vowels

Table 5 and **Table 6** represent average five vowel duration and long-vowels duration, respectively. There are no long-vowel samples for /a/ in the speech material.

	fast	normal	slow
/i/ (ms)	51.6	85.3	230.2
/e/ (ms)	72.6	117.8	283.3
/a/ (ms)	78.6	116.5	309.1
/o/ (ms)	55.0	88.1	218.7
/u/ (ms)	53.6	87.7	218.8

Table 5: Average duration of five vowels.

	fast	normal	slow
/i-/ (ms)	113.5	233.3	599.3
/e-/ (ms)	79.6	168.8	416.6
/o-/ (ms)	77.6	161.4	330.7
/u-/ (ms)	59.7	145.7	437.3

Table 6: Average duration of long-vowels.

In **Table 5**, relatively large duration can be observed for /a/ vowel in all speaking rates. In fact, it is the longest in both f- and s-speeches, while vowel /i/ is the shortest of all for f- and n-speech. In **Table 6**, however, vowel /i–/ shows significantly larger duration than any other long vowels in every speaking rate.

4.4. Individual Consonants

Table 7 exhibits duration for individual consonants that have appeared in the speech samples used in this

(1) Voiced plosives (ms)

	fast	normal	slow
/b/	12.1	14.3	29.3
/d/	11.1	13.9	16.3
/g/	15.8	21.0	25.7

(2) Unvoiced plosives (ms)

	fast	normal	slow
/p/	11.9	17.4	21.3
/t/	15.9	17.9	22.9
/k/	25.0	36.9	45.4

(3) Voiced fricatives (ms)

	fast	normal	slow
/z/	36.8	62.9	89.2

(4) Unvoiced fricatives (ms)

	fast	normal	slow
/s/	60.3	101.9	154.6
/h/	51.2	83.6	129.4

(5) Nasals (ms)

	fast	normal	slow
/m/	49.5	69.3	114.8
/n/	36.9	55.0	75.1
/N/	68.1	135.6	348.1

(6) Affricates (ms)

	fast	normal	slow
/ch/	41.8	77.3	102.0
/ts/	57.1	73.4	100.1

(7) Semi-vowels (ms)

	fast	normal	slow
/y/	38.1	52.4	82.7
/w/	39.5	56.9	82.7

(8) Liquid (ms)

	fast	normal	slow
/r/	14.2	19.7	35.1

(9) Silent interval for "so ku o N" (ms)

	fast	normal	slow
/Q/	110.0	226.2	537.3

Table 7: Duration of individual consonants.

experiment. If we look closely at individual consonants' duration, extremely large ones in every speaking rate can be found for the syllabic nasal /N/ for which the duration reaches as high as 348 milliseconds in the s-speech.

Figure 2 again stands for the percentage of individual consonants' duration as compared with the n-speech. Syllabic nasal /N/ again shows the greatest ratios both for f- and s-speeches.

greatest influence has been found to occur on the syllabic nasal /N/ and the least on the voiceless stop /t/.

The ratio of duration between consonant and vowel of a CV-syllable in the fast speech has been found to be almost the same as that for the normal speech. However, this ratio changes a great deal in the slow speech in which duration of vowel-part becomes extremely large.

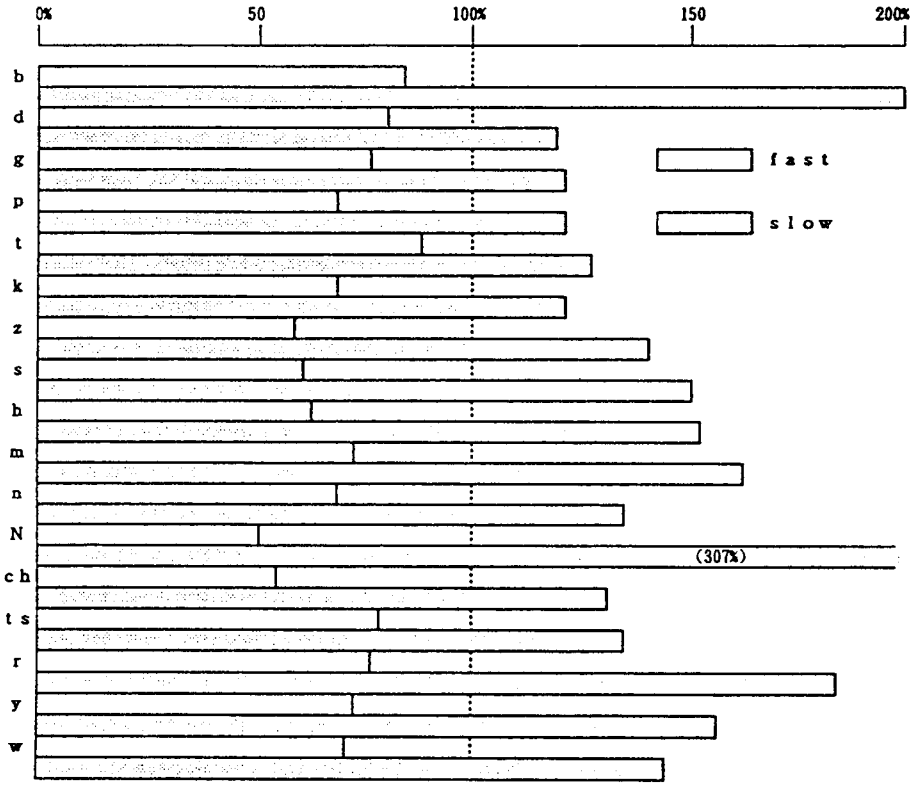

Figure 2: Duration change of individual consonants relative to that of the normal speech. A 100% line indicates duration for the normal speech.

An interesting fact is that the voiced plosive /b/ shows a significantly large duration change for the s-speech (about 200%) while other plosives remain relatively small change.

Liquid consonant /r/, quite different in pronunciation from English retroflex consonant /r/, has also been found to have a greater duration change than other consonants.

5. CONCLUSIONS

Duration of CV-syllables, basic constituents of Japanese language, in continuous speech has been measured for different speaking rate. A statistical analysis of duration has also been made for consonant- and vowel-parts. Four adult male speakers read 15 short Japanese sentences with three different speaking rates; normal speed (about 156 ms/mora), fast speed (about 94 ms/mora) and slow speed (about 345 ms/mora).

As it is expected, vowels receive greater influence than consonants by speaking rate. Among consonants, the

6. REFERENCES

1. Fujimura, O., and Ochiai, K "Vowel identification and phonetic contexts," J. Acoust. Soc. Am., Vo.35, 1889(A), 1963
2. Kuwabara, H. "Perception of CV-syllables isolated from Japanese connected speech," LANGUAGE AND SPEECH, Vol.25, 175-183, 1982
3. Lindblom, B.E.F., and Studdert-Kennedy, M. "On the role of formant transitions in vowel recognition," J. Acoust. Soc. Am., Vol.42, 830-843 1967
4. Kuwabara, H. "An approach to normalization of coarticulation effects for vowels in connected speech," J. Acoust. Soc. Am., Vol.77, 686-694, 1985
5. Miyatake, M., and Sagisaka, Y. "Prosodic characteristics and their control in Japanese speech with various speaking styles," IEICE Trans., Vol.J73-D-II, 1929-1935, 1990