STUDIA PHONOLOGICA XXVI (1992)

Distinction between Vowels and Unvoiced Stops using Features Observed in Speech Waveform

Daiichiro UEMURA AND Shigeyoshi KITAZAWA*

ABSTRACT

In the preceding recognition by spectrum parameters, there were considerable errors that unvoiced stops tended to be added before the initial vowels. We propose new speech waveform parameters that characterize stop/vowel contrasts, discriminations of the consonants (vowel, /p/, /t/, /k/), and complements to the earlier parameters (the power spectrum). The extracted parameters were the DC bias, the zero crossing, the decreasing pitch, and the wave-form envelope. The results of experiment with combination parameters of the spectrum and the waveform were improved for three kinds of phoneme environments; syllables, CV words, and VCV words. In the waveform features, the DC bias and the decreasing pitch outperformed the spectrum.

1. INTRODUCTION

In this study we propose new acoustic parameters for machinery recognition. In the preceding recognition by spectrum parameters of French and Japanese consonants, there were considerable errors that unvoiced stops tended to be added before the initial vowels (Table 1).

Table 1. Japanese consonants recognition by spectrum [1].

		Number of samples classified						
Consonant	Rate (%)	ʔ	p	t	k	b	d	g
ʔ	84.4	346	38	9	14	0	1	2
p	82.0	30	336	29	10	3	0	2
t	82.4	4	30	338	37	0	1	0
k	92.4	7	4	19	379	0	0	1
b	83.9	11	32	0	0	344	15	8
d	81.7	0	5	31	0	21	335	18
g	79.5	3	5	7	44	13	12	326

*Daiichiro UEMURA (上村大一郎): Graduate student, Department of Computer Science, Faculty of Engineering, Shizuoka University.

Shigeyoshi KITAZAWA (北澤茂良): Associate professor, Department of Computer Science, Faculty of Engineering, Shizuoka University.

The authors have done the research under the direction of Dr. Shuji Doshita, Professor of Information Science, Kyoto University.

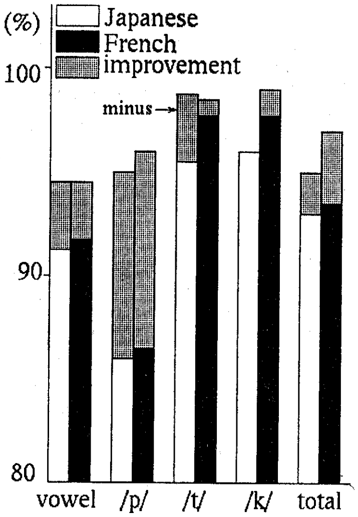


Fig. 1. Distinction rate of the French and Japanese syllable initial phoneme by combination of spectrum and waveform.

Considering that the feature of the speech waveform contains independent one from that of the spectrum, we experimented using these combined features. The performance for syllables was improved except for /t/ in Japanese [2] (Fig. 1). Furthermore those waveform features, the decreasing pitch after the burst and the onset DC bias, outperformed those of the spectrum. The results were common in two languages, Japanese and French. Then we tried to extend the context to words (initial consonants and intermediate consonants) from the initial consonant of an isolated CV syllable. In this new situation we contrasted the initial vowels and the intermediate unvoiced stops, since this is the preliminary step toward continuous speech where the distinction between the vowels after a voiceless segment and the unvoiced stops would be necessary.

2. EXPERIMENTAL METHOD

The data used this time is disylabic and trisyllabic words (including loan words) in which unvoiced stops except double consonants occur in the initial or the intermediate (from the second to the third) syllable. The initial vowels were distinguished from the initial unvoiced stops, and the initial vowels were also distinguished from the intermediate unvoiced stops. Since our interest is focused on the onset of a phoneme segment preceded with a silent interval, we excluded such segments like intermediate vowels that are not preceded by a voiceless section. Intermediate unvoiced stops need segmentation. We do segmentation accurately with inspection as shown in Fig. 2. Concerning intermediate stops, we regarded the onset or the burst point after a short silence (30~100 ms) as start point of analysis. The number of data is 1640 for syllables, 600 foı initial stops and vowels (CV word) and 698 for

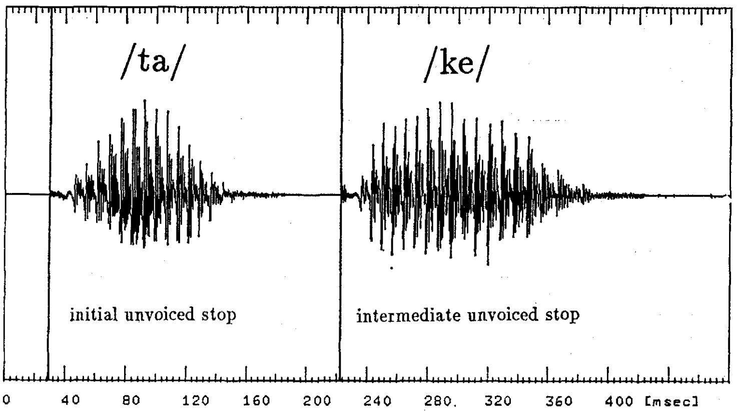


Fig. 2. Phoneme segmentation by inspection.

intermediate stops (VCV words). We got the spectrum by LPC analysis, and the waveform features by enumeration: DC bias (SP, SM), zero crossings (ZC), pitch's change (PT), LPC cepstrum (C[0]~C[5]). In all speech waveform features except PT, we searched for the most effective analysis sections by maximizing distinction rate with the linear discriminant analysis. Pitch periods were measured by inspection. We evaluated the effectiveness of these features from the result of the distinction rate. Then we selected the parameters according to the evaluation, and compared the results of words with that of syllables. The way to make the parameters of the spectrum and the speech waveform for words is the same as for syllables [2].

3. Speech Waveform Features

The features extracted from the speech waveform are the following:

(1) DC bias near the onset or the burst point (SP, SM),

(2) Zero crossing near the onset or the burst point (ZC),

(3) Pitch's change after the onset or the burst point (PT),

(4) Amplitude's envelope development at the transition part (LPC cepstrum C[0]~C[5]). Now we will explain these parameters of the speech waveform.

(1) $SP=\sum_{i=1}^{N}(|S_i|+S_i)/2\,,\quad SM=\sum_{i=1}^{N}(|S_i|-S_i)/2\,,$

$S_i(i=1, 2, \cdots, N)$: Samplings of the speech signal within the analysis window 1 to N.

(2) ZC: Zero crossing count within the analysis window.

(3) $PT=PT_2-PT_1$ (ms),

PT_1 (ms): the first pitch period after the burst point,

PT_2 (ms): the second pitch period after the burst point.

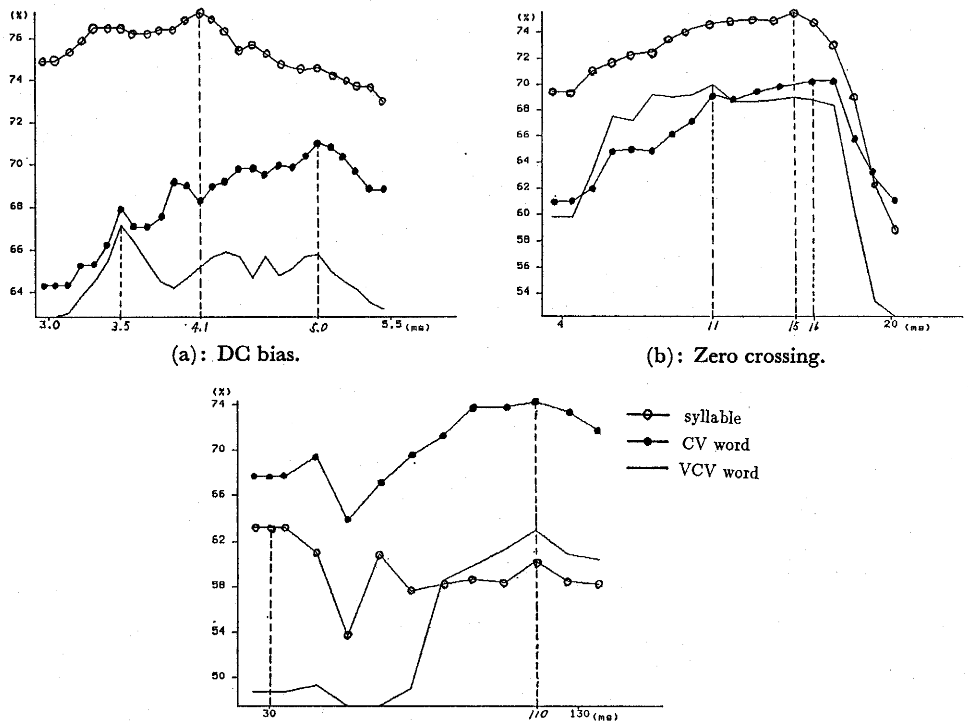
(4) C [0]~C [5]: The LPC cepstrum coefficients derived from the 5-th ordered LPC coefficients of the waveform envelope resampled at every five ms.

Table 2. The means and the analysis intervals of parameters.

context	phoneme, analysis interval	SP-SM	ZC	PT (ms)	C[0]
syllable	vowel	-57.7	23.5	-3.8	16.2
	unvoiced stop	445.5	50.8	12.6	16.3
	analysis interval (ms)	4.1	15	-	110
word initial	initial vowel*	-99.0	20.8	-5.7	16.8
	initial unvoiced stop	534.3	41.8	18.8	18.2
	analysis interval (ms)	5.0	16	-	110
word internal	initial vowel**	-25.6	15.2	-5.1	16.9
	intermediate unvoiced stop	244.5	30.6	17.6	17.4
	analysis interval (ms)	3.5	11	-	110

***The number of this data is 249. **The number of this data is 194, the subset of *. Speech waveform is digitized in 16 bits at 16 kHz sampling rate.**

For the PT, the analysis section is not fixed, but variable by inspection. The C[0] represented the feature of amplitude's building up toward the following vowel. The SP minus SM of vowels biased to the minus, while those of unvoiced stops biased to the plus. The ZC of unvoiced stop is about twice of that of the vowel. The PT of unvoiced stop is positive signed and greater than that of the vowel (usually



(a): DC bias.

(b): Zero crossing.

(c): Envelope.

Fig. 3. Relations between analysis interval and distinction rate.

negative signed). The C[0] of unvoiced stop is greater than that of the vowel. The features of DC bias and zero crossing in words were weaker than in syllables because the differences of them became smaller. The analysis section of SP, SM and ZC was shorter for word internal position than for word initial position. The features of the pitch and the envelope were stronger for words than for syllables because the differences of the means of words were much larger than those of syllables. Fig. 3 shows the distinction rate in relation to the length of the analysis section. In the syllables, the distinction rate by LPC cepstrum is maximum at 30 ms analysis section, however, we considered 110 ms, where another distinction peak is observed, is appropriate judging from the results of words.

4. Experimental Results

Here shows the result of distinction rate and parameter selection experimented by the linear discriminant analysis. Then the consonant discrimination is shown using the most appropriate features in combination.

4.1 Distinction Between Vowels and Unvoiced Stops

The following three combinations of parameters were examined for Japanese syllables and words (Table 3).

(1) All of the waveform parameters, SP, SM, ZC, PT and C[0]~C[5] (10 dimensions),

(2) Spectrum parameters (176 dimensions).

(3) Combination of spectrum and waveform (176+10=186 dimensions).

The features of decreasing pitch and envelope are much stronger for words than for syllables, so the distinction rate of intermediate phoneme using waveform features is better than syllables or the initial phoneme of words. The distinction rates of intermediate unvoiced stop were less than that of initial one. Among the distinction

Table 3. Results concerning distinction between vowels and unvoiced stops.

context	feature	vow. (%)	uvs. (%)	total (%)
syllable	speech waveform	83.2	79.2	80.2
	spectrum	91.2	93.6	93.0
	combination	94.4	95.2	95.0
CV word	speech waveform	87.1	92.0	90.0
(initial vowel vs.	spectrum	89.6	94.0	92.2
initial unvoiced stop)	combination	93.6	99.1	96.8
VCV word	speech waveform	76.3	83.3	81.4
(initial vowel vs.	spectrum	93.8	94.6	94.4
intermediate unvoiced stop)	combination	95.4	97.8	97.1

vow.: vowel; uvs.: unvoiced stops.

rates with spectrum feature, the intermediate one came first to the syllable initial or word initial one. The distinction rate using both spectrum and speech waveform features was much better than the initial unvoiced stop.

4.2 Parameter Selection

Table 4 shows the results of the parameter selection. Both PT and SP (or SM) were selected as more effective parameters than the spectrum here concerning the words as they were to the syllables. ZC was the most effective parameter for the examination with respect to the waveform parameters. On the other hand, for the combination examination of spectrum and waveform, ZC was less effective parameter. The result, we considered, can be explained that ZC and spectrum were not independent but correlated.

Table 4. Selection speech waveform parameters listed along the order of significance.

context	feature	selected parameters(within 10-th)
syllable	speech waveform	ZC,PT,SP,C[0],SM,C[2]
	combination	PT(2),SP(3),ZC(8),SM,C[0～5]
CV word	speech waveform	PT,C[0],ZC,SM,SP,C[2],C[5],C[3]
	combination	PT(2),SM(3),C[0](6),C[2],SP
VCV word	speech waveform	PT,C[5],ZC,SP,C[0]
	combination	PT(3),SP(5),ZC(6),C[5],C[0]

4.3 Consonants Discrimination

We experimented for consonants (vowel, /p/, /t/, /k/) discrimination on the same condition of the parameters and analysis intervals as for distinction between vowels and unvoiced stops (see Fig. 4). The recognition rate was improved with

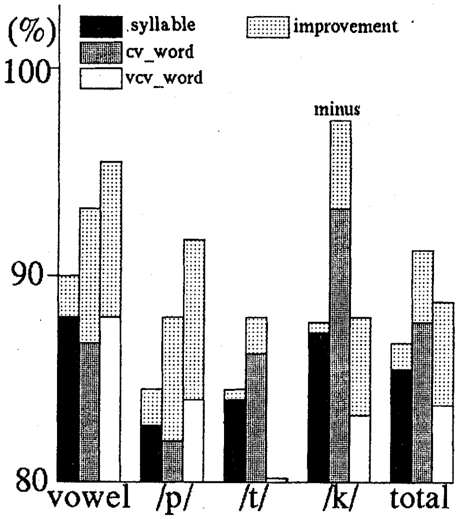


Fig. 4. Japanese consonants discrimination by combination of spectrum and waveform.

combination of spectrum and waveform for all consonants except initial /k/ and intermediate /t/. Best of all, vowel and /p/ were improved. The total recognition rate was 86.8% (+1.3% gain to spectrum features) for syllables, 91.2% (+3.4%) for initial consonants and 88.2% (+4.8%) for intermediate consonants.

5. Conclusion

In this study we extracted the features of vowels and unvoiced stops from the speech waveform that is different from the spectrum. We achieved improvement in the distinction for words (both initial and intermediate) also for syllables. Best of all, the distinction rate of vowel and /p/ was improved. We found these features in the speech waveform depend a little on phoneme environment. This time we did phoneme segmentation and extraction of pitch period by inspection. Considering the practical use, we want to do them automatically in the future. The results concerning voiced stops will be soon obtained.

Acknowledgement

We would like to thank Assistant Tatsuya Kawahara of faculty of engineering of Kyoto University for providing us a linear discriminant analysis program. We would also like to thank to Professor Shuji Doshita who has supported our research in many ways.

References

[1] Norito Nakamura: "Extraction of features for consonants and evaluation by perception experiments", graduation thesis Faculty of Engineering Shizuoka University, 1992 (in Japanese).

[2] Uemura, Kitazawa: "Distinction between vowels and unvoiced stops by using features observed in speech waveform", Technical Report of IEICE. SP91-121, 1992 (in Japanese).