A Forensic Phonetic Investigation into the Duration and Speech Rate

KyungHwa, Kim

Supreme Public Prosecutor's Office, Korea

E-mail: savoix@sppo.go.kr

Abstract

In this paper, the usefulness of the duration and rate of speech is examined as speaker identification parameter in forensic phonetics. Speech samples from 2 actual criminal cases were chosen. The duration and the duration ratio of each syllable of "yeo-bo-se-yo(hello)" and other repeated phrases in Korean delivered by the criminals and the suspects were measured. The results showed that duration ratio from the speech samples of one person were almost identical and were different from the other speaker.

Subjects spoke at different speeds but, for more convincing results, it is necessary to consider the speaker’s speaking habits of pause and hesitation, as well as different speech occasions and styles, which are also potential for speaker identification.

1. Introduction

To identify people by their voices, we use several acoustic features which we believe are useful, but it has not yet been clearly proven which feature is the most effective and decisive one among them.

The theoretically ideal parameter in forensic speaker identification should: show high between-speaker variability and low within-speaker variability; be resistant to attempted disguise or mimicry; have a high frequency of occurrence in relevant materials; be robust in transmission; and be relatively easy to extract and measure. [1]

In speaker identification of forensic phonetics, formant frequency and fundamental frequency (F0) are generally considered as useful acoustic parameters. In real crimes cases, however, audio samples collected for analysis often have poor sound quality due to various noises, such as other voices around the speaker, TV sounds, street noises, and the sound of clothes rubbing. These noises together contribute to the poor sound quality of speech samples, which are usually taped by hidden recorders. In these cases, the information of formant and F0 is lost and hard to analyze. F0 is also affected by the speaker's emotion and the occasion.

When two audio materials, analyzed in a real criminal case, were judged to be of the same person or of different people, there also were auditory similarities or dissimilarities in the duration of utterance and speech rate, for the same word or phrases. Speech duration and rate are considered to be one of the potential factors in speaker identification, but there has not been enough research on how a speaker can be discriminated from others on the basis of the duration and rate of utterance.

The temporal vector, as well as formant and fundamental frequency, can be a useful vector in speaker identification and to measure this vector, (1) Total
Speech Time, (2) Speaking Time Ratio, (3) Silent Interval, (4) Speech Rate, (5) Consonant/Vowel Duration Ratios should also be analyzed, and the results can be used to identify the speaker. [2]

The purpose of this pilot study, based on real forensic cases, was to measure the duration and the rate of speech of speakers who were judged as the same person or as different individuals, and to see whether the results support the original decisions and if these parameters are useful for speaker identification. In addition, we considered the speaker’s speaking habits such as specific intonation patterns and articulations which are closely connected to the duration of utterance.

2. Method

2.1. Subjects and Material

An experiment was carried out involving 5 male voices from 2 real criminal cases.

In Case 1(assumed as the same person), repeatedly spoken materials, such as a telephone recording of criminal A and the suspect B’s reading of the same text were extracted: the word “yeo-bo-se-yo(hello)” spoken six times and three same texts were each delivered once – “I trusted you. / I really trusted you. / Oh, this is the last time.”1

In Case 2(assumed to be either one of two people), a telephone recording of criminal C, its text read by the first suspect D, and the casual speech of the second suspect E, were chosen. The word “yeo-bo-se-yo” spoken seven times by C and D, and “-gu(with)” spoken by C and E, were extracted from the material and compared.

A and B were presumed to be the same person, and C and D were thought to be different speakers. There was possibility that C and E were the same person.

To measure the duration of syllable, “ yeo-bo-se-yo” and 3 sentences spoken by A and B were used in Case 1. “yeo-bo-se-yo” by C and D, and 2 word-final endings “-gu” which were lengthened from D and E, were chosen in Case 2.

All subjects, A, B, C, D, E, spoke Standard Korean.

2.2. Procedure

From each individual material, the duration of the whole utterance and each syllable within it were measured. The reason for the measurement of the duration ratios of each syllable in the utterance was to see between- and within-speaker difference.

The duration was measured using PitchWorks and syllable segmentation was made by viewing a waveform, spectrogram and listening.

To measure the speech rate, 5 sentences spoken by A and B from Case 1 and by C, D and E from Case 2 were selected. The total duration of each sentence was measured, and the number of syllables per second were counted.

3. Results

3.1. The duration of utterance

Case 1. (A, B were presumed to be the same speaker)

Table 1. Duration(ms) and duration ratio of “yeo-bo-se-yo(hello)” of A and B. Parenthesis indicates standard deviation.
Table 3. Duration(ms) and duration ratio of “hae-ga-ku” and “wa-ga-ku” of C, E

<table>
<thead>
<tr>
<th></th>
<th>hae</th>
<th>ga</th>
<th>ku</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>87.5</td>
<td>122.2</td>
<td>463.8</td>
<td>673.5</td>
</tr>
<tr>
<td></td>
<td>13.0%</td>
<td>18.1%</td>
<td>68.9%</td>
<td>100%</td>
</tr>
<tr>
<td>E</td>
<td>95.9</td>
<td>148.9</td>
<td>548.2</td>
<td>793.0</td>
</tr>
<tr>
<td></td>
<td>12.1%</td>
<td>18.8%</td>
<td>69.1%</td>
<td>100%</td>
</tr>
</tbody>
</table>

The results matched some factors that are known to govern duration. They are phrase-final lengthening, vocal effort and emotion. [3]

In Case 1, the duration of each syllable "yeo-bo-se-yo (hello)" and duration ratios in the utterance were very similar, and the last syllable 'de' of a connecting or a word-final ending '-neun-de (and then)' was lengthened almost 2 times with high intensity from both A and B. A and B spoke in emotionally annoyed mood and showed similar rising-fall intonation pattern.

In Case 2, C and D have different syllable duration ratios in "yeo-bo-se-yo" and, C and E have lengthening of the word-final ending '-gu(and)' of the auxiliary verb with almost 70% duration ratios of the utterance with rising-fall intonation.

3.2. Speech rate

Number of syllables per second of each speaker is shown in Table 4.

Table 4. Speech rate of 5 speakers

<table>
<thead>
<tr>
<th></th>
<th>Speaker</th>
<th>syl/sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>A</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>7.7</td>
</tr>
<tr>
<td>Case 2</td>
<td>C</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>7.0</td>
</tr>
</tbody>
</table>
B and D read speech transcription while A, C, and E spoke in casual speech. It was observed that there was a difference of speaking rate between casual speech and reading speech. In this experiment, reading speech was faster than casual speech. B’s and D’s recordings were made in front of police officers and the atmosphere might have affected their speech.

A’s was casual speech, but he was threatening the victim through the phone and spoke clearly with lengthening of some speech sounds and avoiding reduced forms of consonant and vowel segments.

Here, we have to consider that people speak at different speeds according to their speaking styles and on different occasions.

A previous study found that different speakers have different articulation rates, and claimed that articulation rate is a possible forensic phonetic parameter. [4]

Also, people characteristically speak at different speeds and speakers can differ not only with respect to how fast they speak, but also in fluency, which is the number and types of pauses a speaker makes. There are filled and silent pauses. Filled pauses are habitual pauses and prolongation of the last segment of a word. [5]

In Korean, there are filled pauses such as [ʌː], [tumː], [tumː], and [gutligoːː](and). [ntumdeːː](and then). The second examples usually appear with the speaker specific intonation pattern. This individuality of filled pauses and hesitation could be emphasized in forensic phonetics.

4. Conclusion and future works

In this study, we have investigated whether the speech duration and rate can be used as speaker identification parameter in forensic phonetics. As a pilot study, we performed the experiment with 2 speech samples from real criminal cases, and compared the results to see if they supported the decisions made earlier.

The results showed that speakers differed in total duration of utterance and in syllable-syllable duration ratios within a word or a phrase, and that between-speaker difference was larger than within-speaker difference.

Each subject showed difference in speech rate, but, because the rate even in the same speaker can vary according to materials, occasions, and other reasons, further study with more subjects and materials is needed to support the usefulness of rate of speech in discrimination speakers.

The next stage of this study will focus on the relationship between speech rate and the segment duration when the speech speed is changed, and on speaker specific syllable-syllable and consonant-vowel duration ratios. And tonal intonation and pause pattern of individuals will also be considered.

References