Forensic Science International 249 (2015) 116–122

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Cross-language differences of articulation rate and its transfer into Japanese as a second language

Kanae Amino *, Takashi Osanai

National Research Institute of Police Science, Fourth Section of Forensic Science, 6-3-1 Kashiwanoha, Kashiwa, Chiba 2770882, Japan

ARTICLE INFO

Article history:
Received 24 August 2014
Received in revised form 20 January 2015
Accepted 24 January 2015
Available online 3 February 2015

Keywords:
Forensic phonetics
Foreign accents
Speech tempo
Accent identification

ABSTRACT

Recently, the articulation rate has been attracting attention in forensic speech investigation as an acoustic feature that varies across speakers, dialects, and languages. The present study investigates how cross-language differences in the articulation rate are transferred into Japanese as a second language. Participants were speakers of Japanese, Chinese, Korean, and Thai. They were recorded while they read a passage in their native language and in Japanese. Local and global articulation rates were calculated based on the number of syllables as well as the number of morae for Japanese speech. When we compared the articulation rate of the native languages, Japanese was the fastest, then Korean, Chinese, and Thai in that order. Also, a significant positive correlation was observed between the articulation rate of the native language and that of the second language. A gender difference was found in the articulation rate of some languages, with males speaking faster than females. The effect of age was limited to Thai speakers only. Accent discrimination and identification experiments were conducted and the results revealed that native and non-native accents could be correctly discriminated just by the articulation rate.

1. Introduction

Speech accents are defined as deviations in pronunciation and can be classified into social and regional accents. Regional accents are further classified into native (intra- and inter-national) and non-native (foreign) accents. Foreign accents occur due to the interference of the speaker's native language (L1), or incomplete- or false-interpretation of the phonological system of the target language (L2, when it is the second language) [1]. Investigations on foreign accents are beneficial for research areas such as teaching L2. For L2 learners, who wish to attain a "native-like" accent, they must overcome their foreign accent. Also, improvement of L2 pronunciation is favourable to listeners, since foreign-accented speech is hard to process for those unfamiliar with that accent [2].

In practical forensic speech investigations, however, foreign accents provide valuable information about the speaker. A speech accent, including a foreign accent, is one of the most significant features that forensic investigators use to identify a speaker [3,4]. Accents can separate innocent speakers from suspects, since a consistent accent is a prerequisite for showing that two speech samples were uttered by the same speaker [5].

In order to exploit the characteristics of foreign accents in forensic science, we have to extract them as acoustic features. In previous studies, voice onset time (VOT), consonant clusters, vowel qualities, and deviant tones were reported to be the acoustic correlates of foreign accents (see [6]). Acoustic features used in forensic speech investigation need to satisfy several conditions: (1) have small within-speaker variation, (2) be easy to measure, (3) be robust against noise and transmission differences, and (4) be robust against disguise. Among these conditions, (3) and (4) become especially problematic in practical situations. Forensic speech samples are often recorded through a telephone line and contain extraneous noise [5,7]. It is not uncommon for speech samples to have negative signal-to-noise ratios. Furthermore, a foreign accent is often used in disguised speech. Neuhauser [8] investigated speech samples uttered by native speakers of German in an attempt to imitate French learners of German, and found that they mostly changed articulatory features, and some attempted to change pitch as well. Neuhauser also reported that speakers were unable to perform consistently. Rogers [9] introduced a forensic case that investigated an English speaker with a Cantonese accent. He compared two speech samples, one from a criminal and the other from a suspect to conduct linguistic-phonetic analyses. He found that the criminal's speech contained some effort to disguise the voice; thus, he decided not to use phonation properties including fundamental frequency, breathiness, and jitter when yielding the forensic conclusion.

* Corresponding author. Tel.: +81 0 4 7135 8001.
E-mail address: amino@nrips.go.jp (K. Amino).

http://dx.doi.org/10.1016/j.forsciint.2015.01.029
0379-0738/

Considering the above mentioned problems, prosodic features other than phonatory features are suitable for forensic speech investigation. Therefore, the articulation rate (hereafter AR) is one of the prospective features in forensic science. AR can be calculated by the number of linguistic units divided by the time unit (e.g., syllables per second). AR differs from speech rate in that any pauses and other fluency interruptions are excluded in its calculation. Recent studies on speech tempo have shown that AR has large between-speaker and between-dialect variations, as well as small within-speaker variations, and may be useful for forensic speech investigation [6,10–13]. In addition, when a speech sample contains some kind of voice disguise, the speech rate can enable us to recognise the speaker's identity [14].

Jessen [11] provided the forensic reference data on German AR. He analysed read and spontaneous speech uttered by 100 native speakers of German. Cao and Wang [12] also recorded 100 Mandarin Chinese speakers talking over the telephone and analysed within- and between-speaker variations of AR for forensic reference purposes. Both of these studies [11,12] reported that AR is normally distributed; however, their mean values differ across languages. The mean AR (syllables per second) for German spontaneous telephone speech is 5.19 [11], whereas that for Mandarin Chinese is 6.58 [12]. Jacewics et al. [13] analysed read and free-conversational speech uttered by 94 speakers of American English from Wisconsin and North Carolina. They found that the main effect of regional dialect was significant. Gut [6] analysed the AR of native and non-native German and English speech and found a significant difference between native and non-native speech. The mean ARs (syllables per second, read speech) values of native German and English speakers were both 4.10, and those of non-native German and English speakers were 3.30 and 3.25, respectively. Furthermore, Yuan et al. [15] analysed the speech rate (called the "speaking rate" in their study) of L2 English speech uttered by speakers with a different L1, and found the possibility that the speech rate in L2 depends on the speaker's L1.

This study first investigates the cross-language differences of AR in L1. We also examine how these cross-language differences are reflected in L2 Japanese. AR is analysed for four languages including Japanese, and also for three foreign accents of L2 Japanese. Finally, we conduct small experiments on discriminating native and non-native accents and identifying foreign accents by using AR.

2. Methodology

2.1. Speech materials

As shown in Table 1, 32 native and 42 non-native speakers of Japanese participated in the recordings. All speakers read a Japanese passage twice that consisted of six sentences (shown in the Appendix) in an anechoic room at the National Research Institute of Police Science in Chiba, Japan. Non-native speakers read another passage ("The Northwind and the Sun") [16] in each speaker's L1. When reading the passages, speakers were instructed to read as naturally as possible. They were allowed to practice before the recording started. The non-native speakers' level of Japanese was between higher-intermediate and advanced. Almost all of them had passed N2 of the Japanese-Language Proficiency Test [17].

Their speech data were recorded through a microphone (Sony ECM-23F5) and a telephone (Nitsuko ST-D2002K) using a PCM recorder (Marantz PMD 670). All speech data were sampled at 44.1 kHz with 16 bit resolution and resampled at 16 kHz before the analysis.

Table 1
Characteristics of the speakers.

Native language	Gender	Population	Mean age (range) [yr]
Japanese	Male	15	23.3 (18–33)
	Female	17	24.6 (19–48)
Chinese	Male	8	24.5 (21–27)
	Female	12	26.8 (20–46)
Korean	Male	6	24.3 (20–37)
	Female	8	22.9 (19–29)
Thai	Male	2	24.5 (24–25)
	Female	6	23.2 (25–32)

2.2. Analysis of articulation rate

Since telephone speech is more common in forensics, speech materials recorded through the telephone line were used in the analysis. In order to calculate AR, we must decide the time and linguistic units to use. For the time unit, most previous studies in forensic science have chosen seconds instead of minutes, because forensic speech materials usually last less than one minute [10]. A syllable, considered to be the most basic linguistic unit, has been commonly used in previous studies concerning AR [6,10–13]. Speech in Japanese can be segmented into morae as well as syllables [18]. In this study, we calculated syllables per second for all speech materials and also morae per second for Japanese speech. We counted the syllables and morae on the basis of canonical syllables and morae instead of the realised ones in the speech.

The acoustic features shown in Table 2 were analysed manually using Praat [19]. An example of an annotation is shown in Fig. 1. The local AR (LAR) was calculated for each sentence, and the global AR (GAR) for the entire passage.

2.3. Accent identification experiment

Using the acoustic features shown in Table 2, we conducted two experiments: native vs. non-native accent discrimination and foreign accent identification. Two additional features, the average of *LARs* within the passage and that of *LARm* were also used as the global features in the experiment.

In both experiments, logistic regression analysis with ten-fold cross-validation was conducted using the machine learning toolkit WEKA [20]. The results were evaluated by the percentages of correct discrimination and identification.

3. Results and discussion

3.1. Articulation rate of L1 speech

Distributions of *LARs* and *GARs* for each speaker group reading a passage in L1 are shown in Figs. 2 and 3, respectively. Both features

Table 2
AR analysed in this study.

Scope	Feature name	Definition
Local	*LARs*	The number of syllables per second
	LARm	The number of morae per second (only for Japanese speech)
Global	*GARs*	The number of syllables per second
	GARm	The number of morae per second (only for Japanese speech)