Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 72 (2015) 109–125

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Speech intelligibility tests and analysis of confusions and perceptual representations of Thai initial consonants

Charturong Tantibundhit [a,c,*,1], Chutamanee Onsuwan [b,c]

[a] Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University, Rangsit Campus, Pathumthani 12120, Thailand

[b] Department of Linguistics, Faculty of Liberal Arts, Thammasat University, Rangsit Campus, Pathumthani 12120, Thailand

[c] Center of Excellence in Intelligent Informatics, Speech and Language Technology and Service Innovation (CILS), Thammasat University, Rangsit Campus, Pathumthani 12120, Thailand

Received 4 November 2014; received in revised form 13 April 2015; accepted 12 May 2015
Available online 19 May 2015

Abstract

We methodically designed and developed a subjective intelligibility testing of Thai speech for initial consonants based on the diagnostic rhyme test (DRT). The Thai DRT for initials (TDRT-I) was designed to equally compare 21 phonemes pairwise, which results in 210 stimulus pairs. The test is well-constructed using real monosyllabic words. TDRT-I have main advantages in that percent intelligibility scores in each stimulus pair as well as confusion patterns across all phonemes can be evaluated and compared. To confirm its validity, we carried out a series of experiments. The subjective intelligibility tests were conducted on 28 Thai normal hearing listeners in four SNR levels (−6, −12, −18, and −24 dB) and subsequently on eight sensorineural hearing loss patients (with and without hearing aids) using clean stimuli. Average intelligibility scores, percent correct responses, and confusion matrices were obtained. Comparisons of confusion patterns in both subject groups showed that /r/ is the most confusable phoneme, while /w/, /j/, and /p/ are among the least. Perceptual representation spaces, derived from confusion matrices, yielded five non-overlapping groupings: glide, glottal constriction, nasality, aspirated obstruent, and a combination of liquid and unaspirated obstruent. The results suggested that patients' perceptual difficulty could be attributed to the nasality grouping, normally well separated for normal hearing listeners, shifting close to the glottals and aspirated obstruents. Hearing aids (ITC, BTE, and BW types) seemed to improve perception of all phonemes by 10%, with /tɕ^h/, /k^h/, /s/, and /h/ (all unvoiced) showing significant improvement rate. Lastly, the signal detection theory (SDT) bias values of *c* among all possible 108 pairs of unvoiced vs. voiced phonemes revealed that normal hearing subjects are in favor of unvoiced phonemes. The hearing loss patients (with and without hearing aids) showed the same bias pattern. Interestingly, the hearing aids seem to substantially increase more biases for the unvoiced category.

Keywords: Thai diagnostic rhyme test; Subjective intelligibility; Confusion matrix; Similarity score; Perceptual representation space; *c*-value

1. Introduction

This paper describes and discusses a series of experiments starting with the development of the Thai diagnostic rhyme test for initials (TDRT-I) (Tantibundhit et al., 2011c). Two experiments, using TDRT-I were conducted on normal-hearing listeners (Tantibundhit et al., 2011b) and hearing-loss patients (Tantibundhit et al., 2011a). Experimental results (including confusion matrices) were

* Corresponding author at: Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University, Rangsit Campus, Pathumthani 12120, Thailand.
E-mail address: tchartur@engr.tu.ac.th (C. Tantibundhit).
[1] This author was supported by The Thailand Research Fund (TRF) under the grant number MRG5480272.

http://dx.doi.org/10.1016/j.specom.2015.05.005
0167-6393/

partially given in Tantibundhit et al. (2011a,b,c) but are further analyzed and systematically highlighted here. Moreover, derived perceptual representations are compared and discussed in detail. Importantly, in this work the analysis of signal detection theory (SDT) values of *c* (criterion) is introduced for both sets of data to arrive at the conclusions and implications. In this section, previous and relevant work, related to subjective intelligibility testing, analyses of perceptual confusions, is reviewed.

1.1. Subjective intelligibility testing

Speech intelligibility and speech quality are two distinct properties. Speech quality is subjective in nature and difficult to reliably evaluate. Specifically, it reflects how an utterance is produced and also includes speech attributes (Loizou, 2013). Speech intelligibility, on the other hand, refers to what is being said, i.e., the meaning or the content of the spoken words (Loizou, 2013). Therefore, speech intelligibility is one of the essential attributes of the speech signal and needs to be preserved by speech enhancement algorithms (Tantibundhit et al., 2007, 2010).

Several algorithms have been developed specifically to enhance speech intelligibility in background noise (Tantibundhit et al., 2007, 2010). Evaluating intelligibility of enhanced speech compared with the original is often conducted using a subjective intelligibility testing (Loizou, 2013). Several intelligibility tests have been proposed for English by using rhyming words presented in six-response (House et al., 1965) or in pair-response (Voiers, 1983).

House et al. (1965) developed a test by restricting response choices to a finite set of six rhyming words called the modified rhyme test (MRT). The test was composed of 50 sets, each of which was composed of six monosyllabic consonant–vowel–consonant (CVC) words. Twenty-five sets differed in their initial consonants, while the rest differed in their final consonants (House et al., 1965).

Voiers (1983) refined the MRT and created a diagnostic rhyme test (DRT), which is widely used for a subjective testing for measuring the intelligibility of speech coders (Loizou, 2013). The DRT was an A/B forced comparison test based on word pairs differing in their initial consonants by one of six distinctive features (Voiers, 1983). The DRT test material was composed of a word list of 96 rhyming pairs, e.g., *veal–feel*. As the DRT was developed specifically for English, it has some limitations when evaluating intelligibility of a tonal language such as Chinese (McLoughlin, 2008).

McLoughlin (2008) devloped a New Chinese diagnostic rhyme test (NCDRT). The NCDRT was composed of a test set of phonemes in Chinese, which were classified under six distinctive features similar to the DRT (McLoughlin, 2008). Although subjective intelligibility testing for a tonal language such as Chinese is well underway (McLoughlin, 2008), a testing designed for Thai, a tonal language with acoustic and phonemic differences from that of Chinese (Comrie, 1990), has yet to be developed. With that in mind, in our previous work, we designed and developed an intelligibility testing of Thai speech specifically for its initial (TDRT-I) and final consonants (Tantibundhit et al., 2011c). The test was designed to facilitate an evaluation of percent intelligibility responses in each stimulus pair and to systematically compare confusion responses across all initial and final phonemes (Tantibundhit et al., 2011c). Specifically, several useful frameworks, namely DRT (Voiers, 1983), NCDRT (McLoughlin, 2008), MRT (House et al., 1965), and the analysis method of balanced confusion matrix (Miller and Nicely, 1955) were integrated. Moreover, an A/B forced choice and monosyllabic (CV(V)(C)) rhyming pairs, which differ only in one sound either in an initial or final position (the tone was kept identical) were used. The words were well-selected from real and commonly used words in the language (Tantibundhit et al., 2011c).

1.2. Analyses of perceptual confusions

Analyses of perceptual confusions among phonemes (speech sounds) provide valuable information in determining and understanding speech perception in general and cross-linguistically (Johnson, 2003). By and large, there are two main motivations behind these types of analysis. First of all, confusion patterns provide essential clues for the understanding of how speech signals are auditorily processed and transformed as some parts of the signals will become more distinct while others suppressed (Stevens, 1981). This insight is crucial for a number of areas in speech research, including speech recognition (Mermelstein, 1976). Secondly, a number of cross-linguistic perception experiments have shown that perception of speech sounds is not only limited to the input from the auditory system, but also the result of perceptual representations, which are largely shaped by listener's language experience (Strange, 1995). Importantly, perceptual confusion patterns, which generally reflect phonological predisposition of speech sounds, will provide a more reasonable explanation for a connection between language, i.e., its sound inventory and (human) auditory constraints (Stevens, 1981).

A number of studies have focused on confusion analyses of English consonants, e.g., Miller and Nicely (1955). Among them, a classic report from Miller and Nicely (1955), where perception of English word-initial consonants (16 phonemes) in an open-response task was conducted under different bandwidths of nonsense syllables (in between 200 and 6500 Hz) and different signal to noise ratios (SNRs) (−18, −12, −6, 0, 6, and 12 dB).

Shepard (1972) proposed a method to assess a psychological representation of speech sounds by computing similarity and distance scores from confusion matrices. He applied his formula and method to the English perceptual data from Miller and Nicely (1955). The analysis showed that the perceptual representation of English consonants