Applied Acoustics 114 (2016) 79–91

Contents lists available at ScienceDirect

Applied Acoustics

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

Speech intelligibility of English, Polish, Arabic and Mandarin under different room acoustic conditions

Laurent Galbrun *, Kivanc Kitapci

School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom

ARTICLE INFO

Article history:
Received 30 May 2016
Accepted 6 July 2016

Keywords:
Speech intelligibility
Room acoustics
Speech transmission index
Diagnostic rhyme test

ABSTRACT

This paper examines the impact of room acoustic conditions on the speech intelligibility of four languages (English, Polish, Arabic and Mandarin). Listening test scores (diagnostic rhyme tests, phonemically balanced word tests and phonemically balanced sentence tests) of the four languages were compared under four room acoustic conditions defined by their speech transmission index (STI = 0.2, 0.4, 0.6 and 0.8). The results obtained indicated that there was a statistically significant difference between the word intelligibility scores of languages under all room acoustic conditions, apart from the STI = 0.8 condition. English was the most intelligible language under all conditions, and differences with other languages were larger when conditions were poor (maximum difference of 29% at STI = 0.2, 33% at STI = 0.4 and 14% at STI = 0.6). Results also showed that Arabic and Polish were particularly sensitive to background noise, and that Mandarin was significantly more intelligible than those languages at STI = 0.4. Consonant-to-vowel ratios and languages' distinctive features and acoustical properties explained some of the scores obtained. Sentence intelligibility scores confirmed variations between languages, but these variations were statistically significant only at the STI = 0.4 condition (sentence tests being less sensitive to very good and very poor room acoustic conditions). Overall, the results indicate that large variations between the speech intelligibility of different languages can occur, especially for spaces that are expected to be challenging in terms of room acoustic conditions. Recommendations solely based on room acoustic parameters (e.g. STI) might then prove to be insufficient for designing a multilingual environment.

1. Introduction

In a modern and globalised world, the interaction between multilingual and multicultural people in public, commercial and social spaces is gaining importance, and oral communication is at the centre of this interaction. In the literature, only few studies have been comparing differences between physical measures and subjective measures of speech intelligibility for native speakers of varying languages [1–5], and most of these focused on comparisons between English and Chinese (i.e. Mandarin) [2–5]. Additionally, design guidelines used for speech intelligibility always focus on physical parameters only (e.g. speech transmission index (STI), reverberation time, signal-to-noise ratio (S/N)), disregarding the possibility of having interactions between room acoustics parameters and languages. Investigating the relations between commonly used objective speech intelligibility measures and subjective intelligibility scores of different languages may clarify how each language performs in a given acoustics condition, and help designing the acoustic environment appropriately for a specific language, or a combination of languages.

Houtgast and Steeneken [1] investigated the speech intelligibility of various languages by examining differences between rank orders obtained across the languages, for different room acoustic conditions. The research examined 11 western languages (English, Finnish, French, German, Hungarian, Italian, Dutch, Maori, Polish, Swedish and Slovak) under 16 acoustic conditions which were varied in terms of reverberation time and signal-to-noise ratio. The main purpose of this study was to validate the rapid speech transmission index (RASTI), which is a simplified version of the STI, by comparing this physical measure of speech intelligibility with the articulation index (AI) obtained from listening tests. Differences between the test materials used for each language did not make it possible to compare word intelligibility percentages obtained from the different languages. However, correlations between rank orders were carried out, and these highlighted differences in speech intelligibility between the languages. It was suggested that these may be caused by several effects, including talker specific effects, phoneme or language specific effects, as well as absence

* Corresponding author.
E-mail address: L.G.U.Galbrun@hw.ac.uk (L. Galbrun).

http://dx.doi.org/10.1016/j.apacoust.2016.07.003
0003-682X/

of (or subtle differences among) the carrier phrases, and level mismatch between the tests [1]. The research presented here focuses on language specific effects.

Another highly relevant study was conducted by Kang [2], who compared the intelligibility of English and Mandarin in two spaces (a seminar room and a corridor), under different room acoustic conditions. It was found that for a relatively high STI, the word intelligibility of Mandarin was better than English (around +5% at STI = 0.6), and for a low STI, the intelligibility of English was better (around +10% at STI = 0.2). It is interesting to note that these significant differences were observed in the corridor, but not in the seminar room (almost no differences for STIs below 0.5 and only around +2% for Mandarin at STI = 0.6 and above). Converted sentence intelligibility showed even more marked differences both in the corridor and in the seminar room, especially at low STI values. This led the author to state that Mandarin is slightly better than English under reverberant conditions, and English is considerably better than Mandarin under noisy conditions. Kang suggested that the greater dynamic range of English might explain its better scores at low STI values, while the tonality of Mandarin might have been helpful at high STI values. Peng [3] also compared the word intelligibility of Mandarin and English as a function of the STI, and found English to be more intelligible than Mandarin across most STI conditions (+2–4%), with the exception of STIs of approximately 0.3 and below, where Mandarin was marginally more intelligible. More recently, Zhu et al. [4] found that the word intelligibility of English is slightly better than that of Mandarin up to an STI of 0.7 (typically around +2–3%, with a maximum difference of +4.5% at STI = 0.4), after which the scores are very similar. Overall, the studies [2–4] indicate that English tends to be slightly more intelligible than Mandarin under most room acoustic conditions, although some contradictions are observed between the findings of these studies, especially for either very poor or very good room acoustic conditions. These contradictions have been mainly attributed to the use of different test materials [4].

Ji et al. [5] investigated the correlation between objective measures of speech intelligibility and subjective intelligibility scores of Chinese, Japanese and English. The research found that the objective measures providing the best correlations varied depending on the language considered, suggesting that a single objective measure cannot accurately predict the intelligibility of different languages. Unlike the work presented here, the research focused on correlations and did not examine variations between the subjective scores of the three languages examined.

A number of other researchers also examined native and non-native speech intelligibility [6–9], main findings being that non-native speakers tend to perform lower under any type of masking condition [6,8] and that the linguistic content of background noise can also affect speech intelligibility [7,9].

Overall, the review of previous work shows that the number of studies that investigated the relationship between languages and speech intelligibility is quite limited, most comparisons having been made between English and Mandarin. Although it is known that there can be speech intelligibility variations between languages, little is known about the extent of these variations and their statistical significance. The present study aims to develop this knowledge by comparing the speech intelligibility of four languages representative of a wide range of linguistic properties (English, Mandarin, Polish, and Arabic) under various room acoustic conditions. The comparisons have been based on a physical measure of intelligibility (STI) and word/sentence intelligibility scores. More specifically, these four languages have been tested under four room acoustic conditions (varying in terms of reverberation time and signal-to-noise ratio), and diagnostic rhyme tests (DRT), phonemically balanced word tests (PB word), and phonemically balanced sentence tests (PB sentence) have been used to determine speech intelligibility scores. It is important to point out that both word and sentence tests have some limitations with regard to comparisons between languages. For example, Kang [2] pointed out that English PB words, especially monosyllabic ones, represent the English words with relatively few phonemes and letters, unlike Mandarin PB words that represent all type of words in Mandarin. In that sense, the use of sentences provides a more direct way to compare the speech intelligibility of different languages, but sentence scores tend to be high under good acoustic conditions and not very sensitive to small changes in listening conditions [10], i.e. less sensitive to identifying variations across languages. For these reasons, both word and sentence tests have been used in the research; their respective limitations should however be kept in mind when analysing results.

The paper first presents the methodology used in the study, followed by the illustration and analysis of results, a discussion, and conclusions.

2. Methodology

This section describes the selection of languages, the word and sentence lists, the recording procedure, the post-processing, and the listening tests used in the research. All the intelligibility tests were carried out under four different room acoustic conditions that were defined in terms of different speech transmission index values (STI = 0.2, 0.4, 0.6 and 0.8). According to the STI qualification ratings of ISO 9921 [11], these corresponded respectively to "bad", "poor", "good" and "excellent" speech intelligibility conditions (Bad: STI 0–0.3; Poor: STI 0.3–0.45; Fair: STI 0.45–0.6; Good: STI 0.6–0.75; Excellent: STI 0.75–1.0).

2.1. Selecting the languages

Languages representative of a wide range of linguistic properties were selected from different language families such as the Indo-European (e.g. English, German, Polish, Spanish, and Farsi), Uralic (e.g. Turkish), Afro-Asiatic (e.g. Arabic), and Sino-Tibetan (e.g. Mandarin) language families. Five criteria were applied for identifying the languages to be tested:

(1) The selected languages had be representative of real multilingual environments, such as those often found in large western cities.
(2) A significant variability between the consonant-to-vowel ratios of the languages was aimed for, as the speech intelligibility is affected by the loss of consonants [12], and as such variability would allow examining whether languages with a high consonant-to-vowel ratio are more sensitive to poor room acoustic conditions. Consonant-to-vowel ratios of languages are calculated from consonant and vowel inventories which are elements of phonology of a language [13]. Inventories are not limited to the letters specified as consonants and vowels in an alphabet, as a combination of several letters might produce a single consonantal or vowel speech sound, such as 'th' or 'ch' in English. The total numbers of such sounds create the consonant and vowel inventories. Depending on the language, the number of consonants in a consonant inventory varies between 6 and 122, and the number of vowels in a vowel inventory varies between 2 and 14 [13]. Consonant-to-vowel ratios are calculated by dividing the number of consonants by the number of vowels in an inventory, resulting in a number between 1 and 29. The results are divided into 5 categories, which have been used when selecting the languages of the research presented: low (smaller than or equal to 2), moderately low (between