SPEECH INTELLIGIBILITY IN MULTILINGUAL SPACES

Kıvanç Kitapcı

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Energy, Geoscience, Infrastructure and Society

August 2016

ABSTRACT

This thesis examines speech intelligibility and multi-lingual communication, in terms of acoustics and perceptual factors. More specifically, the work focused on the impact of room acoustic conditions on the speech intelligibility of four languages representative of a wide range of linguistic properties (English, Polish, Arabic and Mandarin). Firstly, diagnostic rhyme tests (DRT), phonemically balanced (PB) word lists and phonemically balanced sentence lists have been 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). Additionally, perceived speech intelligibility and soundscape perception associated to these languages was also analysed in three multi-lingual environments: an airport check-in area, a hospital reception area, and a café. Semantic differential analysis showed that perceived speech intelligibility of each language varies with the type of environment, as well as the type of background noise, reverberation time, and signal-to-noise ratio. Variations between the perceived speech intelligibility of the four languages were only marginally significant (*p* = 0.051), unlike objective intelligibility results. Perceived speech intelligibility of English appeared to be mostly affected negatively by the information content and distracting sounds present in the background noise. Lastly, the study investigated several standards and design guidelines and showed how adjustments could be made to recommended STI values in order to achieve consistent speech intelligibility ratings across languages.

ACKNOWLEDGEMENTS

First of all, I would like to express the deepest appreciation to my supervisor Dr. Laurent Galbrun for his invaluable supervision and unending support throughout the preparation of the study. It has been a pleasure to know him both as a talented researcher and as a close friend. Without his guidance and persistent help in every instant of my life, this thesis would not have been possible.

Secondly, I would like to thank Prof. Graham Turner and Dr. Bernadette O'Rourke for guiding and supporting me in the area of socio-linguistics.

A special thank goes to the technician Alastair MacFarlane for his assistance during the laboratory tests. Many thanks also go to all my colleagues and staff of the School of Energy, Geoscience, Infrastructure and Society for their assistance and support.

I would like to thank all my family for their continuous support. I owe special thanks to my parents Tulin and Yavuz Kitapci, my sister Basak Kaman, my lovely niece Ekin Kaman, and her father Cem Kaman for their invaluable motivation and trust. I am grateful to my dearest friends Hakan Guney, Abdullah Alkan, Alphan Bayindir, and Erginbay Kirbas for their invaluable help, continuous patience, and moral support.

ACADEMIC REGISTRY

Research Thesis Submission

Name:	KIVANÇ KİTAPCI		
School/PGI:	EGIS		
Version: *(i.e. First, Resubmission, Final)*	FINAL	Degree Sought (Award **and** Subject area)	PhD in Construction

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	11-08-2016

Submission

Submitted By *(name in capitals)*:	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals)*:			
Chapter 1Method of Submission *(Handed in to SSC; posted through internal/external mail)*:			
Chapter 2*E-thesis Submitted* ***(mandatory for final theses)***			
Signature:		Date:	

Please note this form should bound into the submitted thesis.

Updated February 2008, November 2008, February 2009, January 2011

TABLE OF CONTENTS

CHAPTER 1 Introduction

CHAPTER 2 Literature Review

CHAPTER 3 Methodology

CHAPTER 4 Room acoustics and speech intelligibility of English, Polish, Arabic, and Mandarin

CHAPTER 5 Soundscape and speech intelligibility of English, Polish, Arabic, and Mandarin

CHAPTER 6 Standards and design guidelines of spaces used for speech and their relation with multilingual intelligibility

CHAPTER 7 Conclusions

LIST OF FIGURES

FIGURES CREDITS

Figure 2.4 Correlation between the decay of sound and the sound pressure level with time (Galbrun, 2011). 19

Galbrun, L. (2011). Architectural acoustics lecture notes, Heriot-Watt University.

Figure 2.5 The correlation between common speech intelligibility indices (speech transmission index (STI), and articulation index (AI)) and the signal to noise ratio (S/N) (Houtgast *et al.*, 1980) 21

Houtgast, T., and Steeneken, H. J. M. (1980). *Predicting speech intelligibility in rooms from the modulation transfer function. 1. General room acoustics. Acustica,* **46,** 60-72, 1980.

Figure 2.6 The relation between the sound pressure level and the distance between the source and the receiver. The reverberant field becomes dominant after the critical distance *r** (Galbrun, 2011). 23

Galbrun, L. (2011). Architectural acoustics lecture notes, Heriot-Watt University.

Figure 2.7 An example model of a concert hole to be used for ray tracing technique (Galbrun, 2011). 24

Galbrun, L. (2011). Architectural acoustics lecture notes, Heriot-Watt University.

Figure 2.8 Results of typical intelligibility tests (Miller *et al.*, 1951). 25

Miller, G. (1951). Language and communication, McGraw-Hill, New York.

Figure 2.9 An example calculation and the weighting factors of the articulation index (Kryter, 1970). 27

Kryter, K. D. (1970). *The effects of noise on man*. New York: Academic Press.

LIST OF TABLES

GLOSSARY OF SYMBOLS

A	Absorption surface (m^2)
C_{80}	Early/late-arriving sound energy ratio
D_{50}	Early to total energy ratio
F	F-ratio
H	Kruskal-Wallis statistic
I	Sound intensity (W/m^2)
L_{Aeq}	A-weighted equivalent continuous noise level (dBA)
L_I	Sound intensity level (dB re 10^{-12} W/m^2)
L_p	Sound pressure level (dB 2×10^{-5} Pa)
L_{SN}	Signal-to-noise ratio (dB)
L_{SNapp}	Apparent signal-to-noise ratio (dB)
L_w	Sound power level (dB re 10^{-12} Watts)
Q	Directivity
R	Number of correct responses
S	Surface area (m^2)
	Adjusted percent-correct responses
S_T	Total surface area (m^2)
T	Averaging time period (s)
	Reverberation time (s)
	Total number of responses
V	Room volume (m^3)
W	Sound power (Watts)
	Number of incorrect responses
W_0	Reference sound power (10^{-12} Watts)
d	Source-receiver distance (m)
f	Frequency (Hz)
f_m	Modulation frequency (Hz)
m	Modulation reduction factor
p_0	Reference pressure (2×10^{-5} Pa)
r	Source-receiver distance (m)
	Amount of reflected energy (EREF/EINC)
r^*	Critical distance (m)

t	Time (s)
	t-statistic
w_i	Weighting for octave bands
α	Absorption coefficient
$\bar{\alpha}$	Average absorption coefficient
ρ	Spearman’s correlation
AI	Articulation index
CVC	Consonant-vowel-consonant
DRT	Diagnostic Rhyme Test
ESII	Extended Speech Intelligibility Index
HP	High predictability
KMO	Kaiser-Meyer-Olkin statistic
LP	Low predictability
LVI	Language Vitality Parameter
MRT	Modified Rhyme Test
MTF	Modulation Transfer Function
PA	Public Address
PB	Phonemically balanced
PCA	Principal Component Analysis
RASTI	Rapid Speech Transmission Index
RP	Received Pronunciation
SD	Standard deviation
S/N	Signal-to-noise ratio (dB)
SII	Speech Intelligibility Index
SPIN	Speech Perception in Noise
SPL	Sound pressure level (dB)
STI	Speech transmission index
WPM	Words per minute

LIST OF CONFERENCES AND PUBLICATIONS

Kitapci, K., Galbrun, L., O'Rourke, B., Graham, H. T., Acoustics and speech intelligibility in multilingual spaces, *BAAL 2013*, Edinburgh, UK, 5-7 September 2013.

Kitapci, K., Galbrun, L., O'Rourke, B., Graham, H. T., Speech intelligibility in multilingual spaces, *Proceeding of InterNoise 2013*, Innsbruck, Austria, 15-18 September 2013.

Galbrun, L., and Kitapci, K, Accuracy of speech transmission index predictions based on the reverberation time and signal-to-noise ratio. *Applied Acoustics*, **81**, 1-14, 2014.

Kitapci, K., Galbrun, L., Comparison of speech intelligibility between English, Polish, Arabic and Mandarin, *Proceeding of Forum Acusticum 2014*, Krakow, Poland, 7-12 September 2014.

Kitapci, K., Galbrun, L., Subjective speech intelligibility and soundscape perception of English, Polish, Arabic and Mandarin, *Proceeding of InterNoise 2015*, 9-12 August 2015, San Francisco, California, USA.

Galbrun, L., and Kitapci, K, Speech intelligibility of English, Polish, Arabic and Mandarin under different room acoustic conditions. *Applied Acoustics*, 114, 79-91, 2016.

CHAPTER 1

Introduction

1.1 General introduction

In a modern and globalized world, the interaction between multilingual and multicultural people in public, commercial and social spaces is gaining importance, and communication is that the centre of this interaction. The aim of this thesis is to find out possible relations between speech intelligibility and multi-lingual communication, in terms of acoustics, linguistics, and perceptual factors. In order to investigate the multi-dimensional structure of the intelligibility of speech in multi-lingual spaces, the project was divided into two main phases.

In the first phase of the project, the interaction of room acoustics with the speech intelligibility of different languages was investigated. The second phase of the study investigated how soundscape affects perceived speech intelligibility of different languages. In multilingual spaces, socio-lingual factors can affect speech intelligibility and communication between people, and the perception of the sound environment can in fact become as important as the quality of sound itself. The combination of physical and perceptual factors can be taken into account by the soundscape approach developed by Schafer (1977), which considers all the sound present within a space and the perception of that sound environment. The soundscape methodology is therefore a valuable approach which has been used in the present study to evaluate the multiple factors affecting multi-lingual communication. The combination of the results obtained from both phases could ultimately lead to design guidelines and spatial design solutions for the use of service and product providers in order to minimise communication problems between end users.

It has been found out that the number of studies that investigated the relationship between languages and speech intelligibility is limited. Except the studies that compared some languages in relation to room acoustic conditions (Houtgast and Steeneken, 1984; Kang, 1998; Peng, 2011; Ji *et al.,* 2014; Zhu *et al*., 2014), a direct comparison between a wide range of languages is lacking and soundscape has not been examined in relation to speech intelligibility. The present study develops research in those areas.

1.2 Justification of the research

The research aims to contribute to the literature by combining different perspectives of multilingual speech intelligibility, such as room acoustics and soundscape. The literature review showed that there is a missing link between room acoustic and the perceptual aspects of speech intelligibility. This project aims to combine these aspects and widen the range of languages to be compared in terms of speech intelligibility.

Previous studies showed that the intelligibility of speech differs depending on the language considered. In the previous literature, multiple western languages and Chinese (i.e. Mandarin) were tested and compared under several room acoustic conditions, in order to examine potential differences in speech intelligibility (Houtgast and Steeneken, 1984; Kang, 1998). According to the results of these studies, there are differences between the speech intelligibility of varying languages, and the differences might be caused by the fact that each language has different phonemic and linguistic properties. For instance, Chinese is a tonal language and English has a wider sound pressure level range compared to Chinese. Therefore, Chinese and English are not equally intelligible, Kang (1998) having found that Chinese is slightly better than English under reverberant conditions, whilst English is considerably better than Chinese under noisy conditions.

Peng (2011) also compared the word intelligibility of Chinese and English as a function of the speech transmission index (STI), and found English to be more intelligible than Chinese across most STI conditions, with the exception of STIs of approximately 0.3 and below, where Chinese was marginally more intelligible. More recently, Zhu *et al.* (2014) found that the word intelligibility of English is slightly better than that of Chinese up to an STI of 0.7, after which the scores are very similar. Overall, the studies (Kang, 1998; Peng, 2011; Zhu *et al.*, 2014) indicate that English tends to be more intelligible than Chinese 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 (Zhu *et al.*, 2014).

Ji *et al.* (2014) also 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 (Garcia Lecumberri *et al.*, 2006; Van Engen and Bradlow, 2007; Garcia Lecumberri *et al.*, 2010; Van Engen, 2010), main findings being that non-native speakers tend to perform lower under any type of masking condition (Garcia Lecumberri *et al.*, 2006; Garcia Lecumberri *et al.*, 2010) and that the linguistic content of background noise can also affect speech intelligibility (Van Engen and Bradlow, 2007; Van Engen, 2010).

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 Chinese. The first phase of the study aims to bridge that gap 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 being based on a physical measure (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 quantify the speech intelligibility.

Within the context of multilingualism, there are also many recent studies on socio-linguistics and multilingual communication, but the relevance of these tends to be limited, as they do not analyse speech intelligibility in any detail. Most socio-linguistic studies evolve around topics such as health issues, communication disorders, visual communication, information technologies, and linguistic landscapes. For example, a recent study examined urban multilingualism in Europe (Extra and Yağmur, 2011). The research analysed the cultural and linguistic diversity of Europe, and carried out an extensive investigation on multicultural European cities (Goteborg, Hamburg, The Hague, Brussels, Lyon and Madrid). Another example is given by the work of Wodak *et al.* (2012) who conducted a study on language choice and code-switching in institutions of the European Union (the European Parliament and the European Commission). These examples highlight the gap between socio-linguistic work and speech intelligibility work.

The second phase of the study partially bridges that gap by investigating the role of soundscape and perceived speech intelligibility of different languages. The combination of physical and perceptual factors can be taken into account by the soundscape approach developed by Schafer (1977), which considers all the sound present within a space and the perception of that sound environment. The soundscape methodology is therefore a valuable approach which is used in the present study to evaluate the multiple factors affecting multi-lingual communication. The second phase of the study was carried out using listening tests involving sixty native speakers of English, Polish, Arabic, and Mandarin (fifteen participants per language). In the tests, listeners were asked to subjectively evaluate three acoustic environments (an airport, a hospital, and a café) using eleven semantic attributes (intelligibility, speech level, speech pleasantness, noisiness, annoyance, relaxation, comfort, environmental pleasantness, eventfulness, excitement, and familiarity). The tests were undertaken for three room acoustic conditions defined by a different speech transmission index (STI).

To sum up, the study aimed to identify relations between speech intelligibility and multilingual communication, in terms of room acoustic properties and speech intelligibility of the four languages examined. The combination of the results obtained from both phases develops the knowledge and understanding of multilingual communication, also in relation to existing standards and design guidelines.

1.3 Aims and objectives

The main aim of the thesis is to find out possible relations between speech intelligibility and multi-lingual communication, in terms of acoustics, linguistics and perceptual factors. More specifically, the work focuses on the impact of room acoustic conditions on the speech intelligibility of four languages representative of a wide range of linguistic properties (English, Polish, Arabic and Mandarin). Additionally, perceived speech intelligibility and soundscape associated to these languages are also analysed. Lastly, the study investigates several standards and design guidelines of spaces used for speech and their relation with multilingual intelligibility from the perspective of the outcomes of the study.

The objectives of the research are:

1. To understand relations between language specific effects and speech intelligibility, as well as relations between room acoustic properties and speech intelligibility of the four languages (Chapter 4).
    a. To analyse differences in the word intelligibility of English, Polish, Arabic and Mandarin, for a range of room acoustic conditions.
    b. To compare the intelligibility of distinctive features within and across languages tested with Diagnostic Rhyme Tests (i.e. English, Arabic and Mandarin), in order to obtain an insight into word intelligibility variations.
    c. To analyse differences in the sentence intelligibility of English, Polish, Arabic and Mandarin, for a range of room acoustic conditions, and compare those with the differences observed for word intelligibility.
2. To analyse the effects of soundscapes on the perceived intelligibility of the four languages (Chapter 5).
    a. To identify the variability of perceived speech intelligibility across different environments (an airport check-in area, a hospital reception and a café), for a range of room acoustic conditions.
    b. To examine differences between perceived intelligibility and actual intelligibility.
    c. To identify correlations between perceived speech intelligibility and semantic attributes characterizing communication as well as the acoustic environment, in order to ascertain the importance of multiple factors with regard to multilingual communication.
3. To review relevant standards and design guidelines of spaces used for speech and to critically analyse their relation with multilingual intelligibility from the perspective of the outcomes of the study (Chapter 6).

The findings obtained could ultimately be used to inform design guidelines and spatial design solutions of multilingual spaces, in order to minimise communication problems between end users.

1.4 Methodology

In order to achieve the main objectives of this research project, three different methodological approaches have been used:

1. Word (DRT and PB-word) and sentence (PB-sentence) intelligibility test (to address objectives 1, 2, and 3).
2. Semantic differential tests (to address objectives 4, 5, and 6).
3. Calculating STIs based on the room acoustic parameters (reverberation time and signal-to-noise ratio) presented in the standards and design guidelines reviewed, and comparing suggestions with the outcomes of the study (to address objective 7).

1.4.1 Word and sentence intelligibility tests

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). Five criteria were applied for identifying the languages to be tested: real environment depiction, consonant-to-vowel ratio, tonality, native speakers' population, and availability of subjects. Four languages were selected following these criteria. These were English, Mandarin, Arabic, and Polish.

To assess the speech intelligibility of each language, diagnostic rhyme tests (DRT), phonemically balanced (PB) word lists and phonemically balanced sentence lists were used. DRT and PB word tests were employed to examine word intelligibility, whilst PB sentence tests were used for the analysis of sentence intelligibility. PB word tests were used for only Polish, because of the lack of DRT material in Polish.

The listening tests were conducted in one of the chambers of the acoustic laboratory of Heriot-Watt University. 3 male and 3 female listeners were selected from native speakers of each language, in order to achieve equal gender representation. The listeners of each language were selected from the same regions/countries of the speakers. Prior to the listening tests, hearing tests were carried out in the anechoic chamber of Heriot-Watt University to ensure that participants had normal hearing abilities. For DRT tests, listeners had to identify the spoken words within the pairs of words provided on a list (by ticking), whilst for PB words and PB sentences, these had to be written down. Each listening test was repeated for four different acoustic conditions (STI = 0.2, 0.4, 0.6 and 0.8), by changing the reverberation time and signal-to-noise ratio.

A total of six statistical analysis methods were applied to the data sets in order to test several hypothesis of the current study; these methods were Intra-Class Correlation analysis, one-way Analysis of Variance (one-way ANOVA), factorial Analysis of Variance (factorial ANOVA), Spearman's RHO correlation analysis, and Principal Component Analysis (PCA). Consistency within the test participants taking part in the intelligibility tests was analysed by using the Intra-Class Correlation analysis. The difference between languages was statistically analysed by using the One-way ANOVA test. Factorial ANOVA was used in order to analyse the combined effects of languages and room acoustic conditions. Spearman's RHO correlation analysis was performed to investigate correlations between the consonant-to-vowel ratio of languages and word/sentence intelligibility scores. Finally, the interaction between the distinctive feature intelligibility scores (language specific word intelligibility scores) were investigated by using the PCA.

The results obtained were used to investigate the impact of room acoustic conditions on the speech intelligibility, as well as the relationship between the room acoustic parameters and distinctive features of four languages (English, Polish, Arabic and Mandarin). Additionally, the results allowed identifying the correlation between consonant-to-vowel ratios and speech intelligibility, and the relationship between word and sentence intelligibility.

1.4.2 Semantic differential tests

The aim of the second phase of the study was to investigate relations between speech intelligibility and soundscape of the native speakers of English, Polish, Arabic, and Mandarin. 15 participants per language (i.e. a total of 60) were asked to subjectively evaluate acoustic environments by answering nine questions on a five-point semantic scale, under three room acoustic conditions, in three digitally simulated multi-lingual environments. The three multi-lingual environments were an airport check-in area, a hospital reception area, and a café, i.e. three spaces where communication is crucial and which were representative of a variable context. The speech samples were uniquely designed for each environment in order to achieve an appropriate context. Six sentences were created for each environment, and the samples were recorded by four native speakers (two males and two females) of each language in the anechoic chamber of Heriot-Watt University. The three room acoustic conditions were created digitally by adding contextually appropriate background noise and reverberation to the speech recordings. The finalised speech recordings were then presented to the participants in combination with the visuals of environments in the anechoic chamber of Heriot-Watt University, where they were asked to subjectively evaluate the audio-visual material using eleven semantic attributes (intelligibility, speech level, speech pleasantness, noisiness, annoyance, relaxation, comfort, environmental pleasantness, eventfulness, excitement, and familiarity). The results of the experiment were statistically analysed in order to identify statistically significant differences and correlations between the attributes tested.

The results obtained were used to identify the semantic attributes affecting perceived speech intelligibility of the 4 languages, to examine the differences between perceived intelligibility and actual intelligibility for various environments and room acoustic conditions, the type of environment, and the type of background noise, and to identify the semantic components affecting perceived speech intelligibility of the 4 languages.

1.5 Outline of the thesis

Chapter 2 initially explains language and sociolinguistics related definitions, as well as the concept of multilingualism. This is followed by room acoustics factors that affect the speech intelligibility, such as reflection, scattering, absorption, reverberation time and signal-to-noise ratio. Then, the objective and subjective evaluations, as well as the factors affecting speech intelligibility, are explained. This is followed by a critical review of the previous research on room acoustics, speech intelligibility and the factors affecting it, a review of the soundscape approach relevant to the study, and a description of the sociological factors affecting the research. Lastly, a critical discussion of the information provided is given.

Chapter 3 presents the methodology of both phases of the study. Initially, the selection process of the languages is described. For the first phase, a description is given on the word and sentence lists that were used, the recording and post processing of these word and sentence lists, together with details on the laboratory space used and the equipment used, as well as the listening test procedure. For the second phase, the followings are presented: selection process of the cases, preparation of the sentence lists, recording and post processing of the sentence lists and the background noise samples, preparation of the visual materials, details on the laboratory space and the equipment used, and information on semantic differential analysis. For both phases of the study, the statistical analysis methods used to analyse results are also described.

Chapter 4 discusses comparisons of the subjective listening test scores obtained for four languages (English, Polish, Mandarin, and Arabic), under different room acoustic conditions defined by their speech transmission index (STI=0.2, STI=0.4, STI=0.6, STI=0.8). Overall intelligibility scores, language specific intelligibility scores of distinctive features, and sentence intelligibility scores are presented and analysed in order to understand relations between language specific effects and speech intelligibility, as well as relations between room acoustic properties and speech intelligibility of the different languages.

Chapter 5 presents and discusses how soundscape might affect the perceived speech intelligibility of English, Polish, Arabic, and Mandarin, by comparing the subjective assessment of three multi-lingual spaces (an airport, a hospital, and a café) tested under three room acoustic conditions (STI=0.4, STI=0.5, and STI=0.6). Results of the semantic differential analysis and principal component analysis are also given in this chapter.

Chapter 6 presents an overview of 5 standards and 2 design guidelines that can be consulted in the process of designing various multilingual spaces, from the perspective of the outcomes of the present study, more specifically, the results of Chapter 4. Each standard and design guideline are presented and discussed in terms of importance given to speech intelligibility, specifically to room acoustic parameters (i.e. reverberation time (T), signal-to-noise ratio (L_{SN}), and ultimately speech transmission index (STI)), and multilingual communication. The signal-to-noise ratio and reverberation time information presented in such documents are converted to STI values by using the modulation transfer function (MTF), and a comparison of the STIs calculated are presented in Section 6.3 in relation to the results of Chapter 4. The chapter investigates the effectiveness of the current standards and design guidelines in terms of speech intelligibility in multi-lingual environments.

Chapter 9 provides a summary of the conclusions, and describes the impact of the research and suggestions for future work, as well as limitations of the current work.

Appendix A presents the diagnostic rhyme test word lists for English, Arabic, and Mandarin and Appendix B presents the phonemically balanced word lists for Polish. The phonemically balanced sentence lists are given in Appendix C, Appendix D shows the sentence lists used for the second phase of the research and Appendix E presents the questionnaires used for the second phase of the study, including listening test instructions.

CHAPTER 2

Literature Review

2.1. Introduction

This chapter describes, discusses, and critically analyses the previous studies and background information required for the research on acoustics and speech intelligibility in multilingual spaces. The chapter first explains language and sociolinguistics related definitions and the concept of multilingualism. Furthermore, it illustrates room acoustics factors that affect speech intelligibility, such as reflection, scattering, absorption, reverberation time and signal-to-noise ratio. Then, the definition, objective and subjective evaluation and additional factors affecting speech intelligibility are explained. This is followed by a critical review of the previous research on room acoustics, speech intelligibility and the factors affecting it, a review of the soundscape approach relevant to the study, and a description of the sociological factors affecting the research. Lastly, a critical discussion of the information provided is presented.

2.2. Language and Sociolinguistics

The communication system between two or more people employs a code, which is known as a language. Each speaker knows the system (i.e. the grammar) which linguists try to define; however, it is not an easy process to define the grammar. The process also involves psychological, social and genetic factors (Wardhaugh, 2006).

Chomsky (1965) claimed that “Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance”. However, the knowledge of language that speakers have is more than the knowledge of language, it is more abstract. Proper use of the language is situational and communication between talkers and listeners depends on the common cultural and sociological background. Certain messages are hard to understand for one group compared to the other because of the cultural differences rather than language differences (Wardhaugh, 2006).

In the following section, basic definitions of language, linguistics and socio-linguistics will be given, as well as the basic linguistic items such as phonemes, syllables, and words.

2.2.1. Definitions

Linguists define a language as it is composed of a set of items that are called linguistic items, such as sounds, words, grammatical structures and so on. Sapir (1921) gave the basic definition of language as "Language is a purely human and non-instinctive method of communication ideas, emotions, and desires by means of voluntarily produced symbols. These symbols are, in the first instance, auditory and they are produced by the so-called organs of speech". Furthermore, it is not possible to define a language by using only linguistic items; it needs to be combined with the culture. Mesthrie *et al.* (2009) claims that languages vary according to social class, status, region of origin, and gender. Social theorists, especially sociologists use concepts as identity, power, class, status, solidarity, accommodation, face, gender, and politeness. Combining these two sets of definitions and studying how they relate to each other is not an easy process (Wardhaugh, 2006).

It is important to note that language and dialect are ambiguous terms and easy to confuse. The general definition of dialect is that it is a regional variety of a language that has a literary tradition (Wardhaugh, 2006). It is also argued that dialect is more than a variety of language; it is also excluded from polite society. Because the terms language, style, or dialect arouse emotions in many ways, the more neutral term "code" is used frequently. The code term was taken from information theory, and used for defining any system used for communication of two or more people (Wardhaugh, 2006). The most common language-contact phenomena, code-switching, is also based on codes, and will be explained in the next section (Martin-Jones *et al.*, 2012).

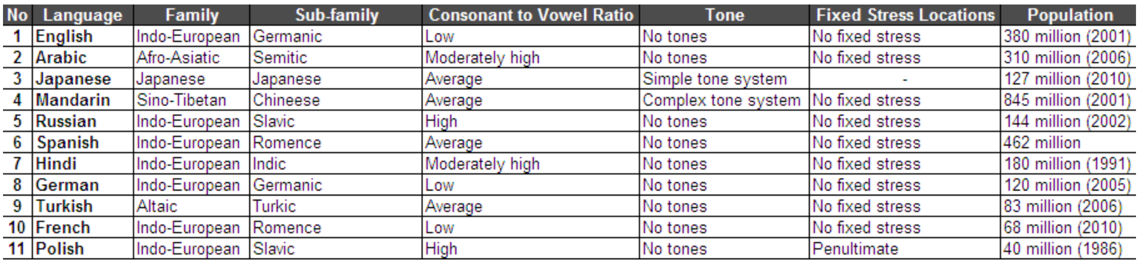
Sociolinguists are trying to build connections between language and culture, seeking the relationship between linguistic items and speakers' understanding of their environment. Before going into detailed definitions about language and sociolinguistics, it is important to give a basic definition of culture. A very well-known definition of culture was given by Goodenough (1957) that “a society’s culture consists of whatever it is one has to know or believe in order to operate in a manner acceptable to its members, and to do so in any role that they accept for any one of themselves”. Therefore, culture is not a genetic endowment, it is learned, in order to get through the task of daily living (Wardhaugh, 2006).

Sociolinguistics cannot be defined solely as a combination of linguistics and sociology. Holmes (1992) claimed that “the sociolinguistic’s aim is to move towards a theory which provides a motivated account of the way language is used in a community and of the choices people make when they use language”. Sociolinguistics is the study of the variety of language among speakers, and analyses the relationship between language variety and the speakers’ knowledge of the language (Wardhaugh, 2006).

It is also stated that sociolinguistics is the study of language use within or among groups of speakers, which sociolinguists define as speech communities. Individuals are members of various speech communities, either discrete, or overlapping. The intersection of speech communities results in linguistic variation that reflects a need which individuals must be seen as the same as other individuals for some occasions and as different from other individuals on other occasions (Wardhaugh, 2006).

Throughout the thesis, the linguistic terms such as vowels, consonants and phonemes will be mentioned frequently. Therefore, it is important to define these terms in order to move on. The most basic unit of speech is a phoneme. Phonemes are perceptually distinct units of speech that distinguish one word from another. There are three types of phonemes. The first type is the vowels. The vowels are phonemes where air flows through the mouth unobscured (i.e. a, e, i, o, and u for the English language). The second type of phonemes is the consonants. The consonants are phonemes marked by closure in the breath channel (i.e. letters other than a, e, i, o, and u in the English language) (Phonics, 2012). Last type of phonemes is the diphthongs. The diphthongs are combination of vowels (i.e. [aI] phoneme contained in the word “ride”, which is a combination of the vowels [a] and [I]) (Mesthrie et al., 2009).

The linguistic items and the linguistic properties of a language vary depending on the language. In the present study, several languages will be investigated in terms of the



No	Language	Family	Sub-family	Consonant to Vowel Ratio	Tone	Fixed Stress Locations	Population
1	**English**	Indo-European	Germanic	Low	No tones	No fixed stress	380 million (2001)
2	**Arabic**	Afro-Asiatic	Semitic	Moderately high	No tones	No fixed stress	310 million (2006)
3	**Japanese**	Japanese	Japanese	Average	Simple tone system	-	127 million (2010)
4	**Mandarin**	Sino-Tibetan	Chineese	Average	Complex tone system	No fixed stress	845 million (2001)
5	**Russian**	Indo-European	Slavic	High	No tones	No fixed stress	144 million (2002)
6	**Spanish**	Indo-European	Romence	Average	No tones	No fixed stress	462 million
7	**Hindi**	Indo-European	Indic	Moderately high	No tones	No fixed stress	180 million (1991)
8	**German**	Indo-European	Germanic	Low	No tones	No fixed stress	120 million (2005)
9	**Turkish**	Altaic	Turkic	Average	No tones	No fixed stress	83 million (2006)
10	**French**	Indo-European	Romence	Low	No tones	No fixed stress	68 million (2010)
11	**Polish**	Indo-European	Slavic	High	No tones	Penultimate	40 million (1986)

Figure 2.1 Comparison table for the most common languages.

intelligibility of speech. In order to see the possible differences between several languages, a comparison table was prepared, comparing the number of native speakers, consonant-to-vowel ratios, tonal properties, and stress properties of the languages. The table represents a brief overview of the most common languages (Figure 2.1).

Ball composed an extensive collection of the sociolinguistic works on the world languages (Ball, 2010). The study includes languages form the Americas, Asia, Australasia, Africa, the Middle East, and Europe. The research methods and other contents differ in-between languages, because research interests of each region varies.

After giving the basic definition of language, linguistics, and sociolinguistics, in the next section, the basic information on multilingual communication is given.

2.2.2. Multilingual communication

Many western countries gave importance to define a territory by its own language, since the period of intense nationalism (Mesthrie et al., 2009). The effects of globalization, population flows between nations, technology, and the new political and economic landscape of different parts of the world caused significant linguistic, cultural and demographic changes. This phenomenon grew the international interest in multilingualism (Martin-Jones *et al.*, 2012).

The meanings of the terms multilingual and bilingual depend on the context. There is no clear borderline that indicates multi- prefix is used for more than two and bi- prefix is used for two, or the opposite. However, in this thesis, multi- prefix is used for more than two and bi-prefix is used for two, as in the policy of International Journal of Multilingualism (Mesthrie *et al.*, 2009).

Communication between two or more people happens in a time-frame within a space, and the "socio" in sociolinguistics addresses these spatial aspects and dimensions of language and communication. Space, in this particular case, is not a passive phenomenon; it is a part of the context, and context defines the communication between people. Entering a space sets the norms and rules of the communication *(Blommaert et al.*, 2005). Therefore, space itself has an active role in multilingual communication. It is stated that communication problems in a multilingual environment are the result of injecting ones communicative skills in a space with specific linguistic norms and rules (Blommaert *et al.*, 2005).

The space and the context affects the way a person speak. Talkers tend to switch codes depending on the context and the environment. Code-switching (CS) is explained by Gardner-Chloros as "the use of several languages or sociolects in the same conversation or sentence" (Gardner-Chloros, 2009). Heller defined CS in a more social context and defined it as "moving away from a focus on the whole bounded units of code and community, and towards a more processual and materialist approach which privileges language as a social practise, speakers as social actors and boundaries as products of social action" (Heller, 2007). Code-switching is one of the core aspects of the multilingual phenomenon and its interaction with the space and the context; however, it will not be taken into account in the present research.

Communicating in a multilingual environment is affected by the physical environment, as well as the language and the social context. In the following section, room acoustic factors will be explained in order to understand the effects of the physical environment on multilingual communication.

2.3 Room Acoustics

Room acoustics is the science of measuring, calculating and/or predicting the movement of sound waves in enclosures by inspecting its interaction with surfaces for several reasons, mainly to maintain good sound. In an enclosure, any speech generated by a talker is transmitted to the listener through a sound path. Generally, this sound path is open to distortions caused by the acoustic parameters of the enclosure. Therefore, the speech received by the listener is not an identical copy of the speech generated by the talker (Houtgast and Steeneken, 1985).

In this section, the acoustic properties that affect the intelligibility of speech in an enclosure, which are reflection, scattering, absorption, reverberation time (*T*) and signal to noise ratio (*S/N*) are discussed.

2.3.1 Sound Absorption

When the sound wave interacts with a surface, some of the sound is absorbed by the surface by being converted into heat or mechanical vibrations, some is transmitted to the back side of the surface, and the remaining part is reflected back to the enclosure (Figure 2.2). In room acoustics, the sound energy that is reflected back to the enclosure is crucial to control in order to maintain high acoustic comfort and to improve speech intelligibility. It is also hypothesised that different amounts of absorption might have varying effects on intelligibility of different languages, as investigated in Chapter 4 of the thesis.

The absorption coefficient, α, is used to describe absorptive properties of a material. This quantifies the amount of sound absorbed by that material. The absorption coefficient, α, can be determined from;

$$\alpha = 1 - r \tag{2.1}$$

where r is the amount of reflected energy, and is equal to E_{REF} / E_{INC}.

The absorption coefficient α is a function of frequency, and can take a value from 0 to 1. Materials that are reflective, for example stone, ceramics or marble have low absorption coefficients, while absorbent materials such as wool or carpet have high absorption coefficients. A material of $\alpha = 0.2$ means that it absorbs 20% of the incident sound energy and reflects 80% of it.

In order to calculate the absorption provided by a surface, the absorption coefficient α and the area of that surface S have to be known. By multiplying the absorption coefficient with the area of the surface, absorption is found.

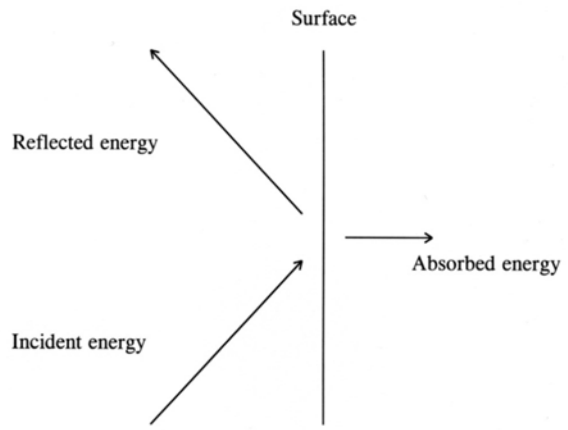


Figure 2.2 Surface reflection and absorption diagram.

The combination of materials in a room leads to a composite absorptive/reflective environment and the total absorption in a room, A, can be found from

$$A = \sum_i S_i \alpha_i \tag{2.2}$$

where S_i is the surface area and α_i is the absorption coefficient. In order to increase the absorption of a room, various sound absorber materials can be used. Sound absorbers are defined under the three main categories of porous (dissipative), membrane and cavity (Helmholtz) absorbers. Each of the absorber types has different absorptive properties for different frequency ranges (Figure 2.3). Porous absorbers are effective at higher frequency ranges and their effectiveness increases with the thickness of the material. The friction between the air particles and material causes the sound energy to dissipate into heat, hence absorbing the sound. There are some wide varieties of porous materials such as rock wool, glass fibre and common building materials such as carpets and curtains.

The membrane absorbers are effective at lower frequencies, especially below 500 Hz. They are mostly used by combination with an air space behind them, to absorb sound waves by the resonance in the cavity that causes vibration of the material, hence converting the sound energy into heat. Effectiveness and frequency range can be changed by modifying the panel mass and/or air space behind the absorber.

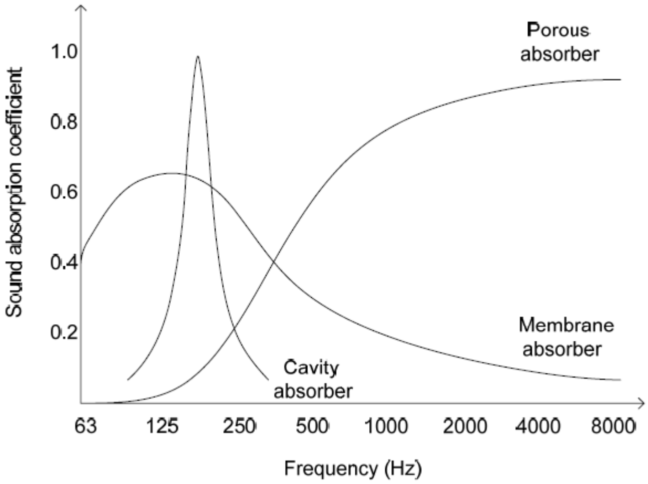


Figure 2.3 Absorption coefficients of sound absorbers.

The Helmholtz absorbers are used for absorbing sound energy at more specific frequencies. Their frequency range is much narrower than other types of absorbers; however, the effectiveness at that specific frequency is significant. They are composed of a cavity with a narrow opening, that takes sound inside and absorb the energy by multiple reflections within the cavity.

To achieve a better absorption, usually different absorbers are used in combination. The aim is to take advantage of the effective frequency ranges of every absorber, such as the effectiveness of panel absorbers at low frequencies and the effectiveness of porous absorbers at high frequencies. For instance, suspended ceilings combine the effectiveness of porous ceiling tiles for high frequencies and the air gap between the tiles and ceilings serve as a low frequency absorber.

2.3.2 Reverberation (T)

One of two disturbances to the speech transmission path between a talker and a listener is the reverberation time (*T*). As soon as a sound wave is generated by a talker, it travels and spreads in different directions, and due to the surface reflections within an enclosure, listeners receive not only the direct sound, but also reflections of the same sound. This combination of direct and indirect sounds eventually reduces the intelligibility of speech (Houtgast and Steeneken, 1980). Additionally, as further investigated in Chapter 4 of the thesis, it is possible that reverberation time might have varying effects on different languages.

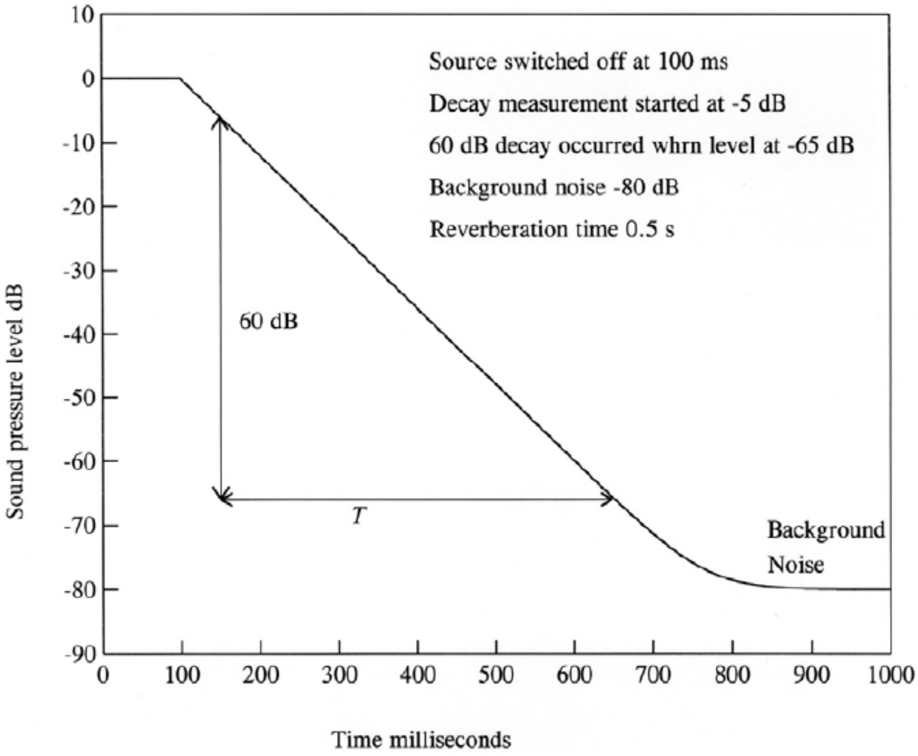


Figure 2.4 Correlation between the decay of sound and the sound pressure level with time (Galbrun, 2011).

The most common measure of the acoustic wave strength is the sound pressure level (SPL) (Long, 2006). In a room, the decline of SPL of a sound source after turning it off is not an immediate action, it takes time to decrease. At every reflection a certain amount of sound energy is absorbed as a function of the absorption coefficient α of reflecting surfaces. Figure 2.4 shows the decay of sound energy with time. This effect took the attention of Wallace Clement Sabine, and his studies showed that the persistence of sound energy in an enclosure is related with the size of the room, finishing materials of the surfaces and the occupants. He first named this effect as "duration of audibility of residual sound" (Egan, 1988).

The reverberation time of an enclosure is defined as the period of time which is needed for reducing the sound pressure level of a continuous noise by 60 dB, after the source has been switched off. In most of the situations, the sound pressure level difference between the background noise and the main sound source is below 60 dB. In order to measure reverberation time correctly in those situations, the decay time for a 30 dB decrease is measured and then multiplied by two. There is no standard for a suitable reverberation time; it depends on the purpose of the enclosure. However, a too long or a too short reverberation time is usually uncomfortable for the audience. For instance, enclosures that are built specifically for music performance require higher reverberation times (1.5 sec – 2.5 sec), and in contrast, enclosures that are built for speech require lower reverberation times (0.5 sec – 1.0 sec).

Reverberation time (*T,* seconds) is a function of the room's volume (*V*, m^3) and surface absorption of the room (A, m^2). It can be computed by Sabine's formula (2.3);

$$T = \frac{0.161V}{A} = \frac{0.161V}{\sum S\alpha} \tag{2.3}$$

where S are the areas (m^2) of surfaces making up the room and α represent the corresponding absorption coefficients for each surface. Furthermore, there is another computing method called Eyring's formula. The difference between equations is that Sabine's formula assumes a diffuse field whilst Eyring's formula is based on intermittent decays. The Eyring's formula is computed as;

$$T = \frac{0.161V}{-S_T \ln(1-\overline{\alpha})} \tag{2.4}$$

where S_T is the total surface area (m^2) and $\bar{\alpha}$ is the average absorption coefficient. It is claimed that it is better for more absorptive rooms, because with Sabine formula, to achieve zero reverberation time in an enclosure, all of the surface materials need to have an infinite absorption coefficient. However, with Eyring formula, surface materials with an absorption coefficient of 1.0 leads to zero reverberation time (Beranek, 2006).

2.3.3 Signal-to-Noise Ratio (S/N)

The intelligibility of speech depends on the correct perception of speech sounds that are varying in terms of frequency and intensity. However, the speech signal is not the only sound that is perceived by the ear. Every unwanted sound, or in other words, noise, has a masking effect on the speech signal, and decrease the sensitivity of the ear to the speech sounds (French and Steinberg, 1947). In Chapter 4 and Chapter 5 of the thesis, objective and subjective effects of varying levels of *S/N* and varying types of background noise on different languages are investigated.

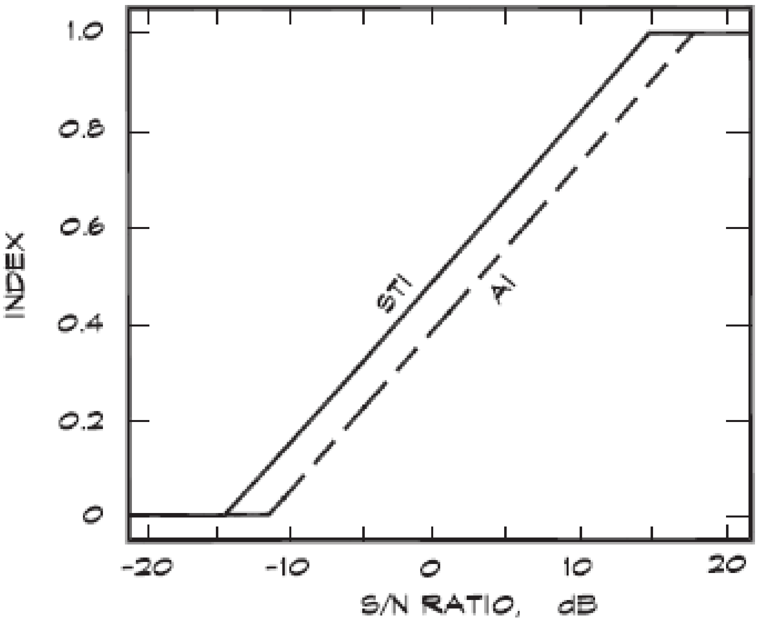


Figure 2.5 The correlation between common speech intelligibility indices (speech transmission index (STI), and articulation index (AI)) and the signal to noise ratio (S/N) (Houtgast *et al.*, 1980)

A sound transmission path between a talker and a listener is sensitive to the interfering ambient noise, either generated inside of an enclosure or penetrating from outside. The sound pressure level difference between a talker generated sound and an ambient noise, in other words the signal-to-noise ratio (S/N), has a crucial effect on the intelligibility of speech (Figure 2.5) (Houtgast and Steeneken, 1980).

There are other studies in the literature that have sought a connection between other room acoustics parameters such as clarity, which is a measure of the early/late-arriving sound energy ratio (C_{80}) and definition, which is a measure of the early to total energy ratio (D_{50}). However, there is no significant evidence that C_{80} or D_{50} affect the intelligibility of speech (Bradley, 1986). Therefore, these parameters are not included in the present research.

2.3.4 Sound power level and sound pressure level (SPL)

A common indicator of the acoustic wave strength is the sound pressure level (SPL) (Long, 2006). Along with the other parameters such as reverberation time and signal-to-noise ratio, these parameters should be calculated at the design stage of the enclosures, including multilingual spaces. The SPL of a sound source depends on the absorption of the walls and the floors of an enclosure. It can be found from

$$L_P = L_W - 10\log\frac{V}{T} + 14 = L_W + 10\log\frac{4}{A} \tag{2.5}$$

where, L_P is the sound pressure level (dB re 2×10^{-5} Pa), L_W is the sound power level (dB re 10^{-12} W), V is the room volume (m^3), T is the reverberation time (s), and A is the total absorption (m^2). This equation can either be used for calculating the sound pressure level of a sound source at a known distance, or to calculate the power of a source in a room of known reverberation time and volume (Galbrun, 2011).

The sound that reaches the receiver directly from the sound source is called the direct sound. After being reflected by the room surfaces for one or more times, it is known as the reverberant sound. If a point source is placed in the middle of a room, then the sound waves will spread from the source spherically, and the SPL can be calculated from

$$L_P = L_W - 20\log r - 11 \tag{2.6}$$

where r is the source-receiver distance (m). Corrections have to be made according to the source position and the sound radiation. For instance, if the source is in a ventilation duct, the sound radiation is hemi-spherical and the SPL can be calculated from

$$L_P = L_W - 20\log r - 8 \tag{2.7}$$

The SPL of the reverberant field can be calculated from

$$L_P = L_W + 10\log\frac{4(1-\bar{\alpha})}{A} \tag{2.8}$$

where $\bar{\alpha}$ is the average absorption coefficient of the room, and A is the total absorption of the room (m^2). In order to find the SPL at a distance from the source in a room, the two equations can be combined, so that

$$L_P = L_W + 10\log\left(\frac{1}{4\pi r^2} + \frac{4(1-\bar{\alpha})}{A}\right) \tag{2.9}$$

where r is the distance between the source and the receiver. In this equation it is assumed that the reverberant sound field is diffuse, the sound radiation is spherical and that the sound source directivity is equal to 1. For other cases, the first term in the bracket should be corrected accordingly (Long, 2006).

The relation between the sound pressure level and the source-to-receiver distance is given in Figure 2.6. The direct field is dominant when the receiver is close to the source, and the reverberant field is dominant when the receiver is far from the source.

The critical distance (r^*) is where the direct field is equal to the reverberant field. It can be calculated by

$$\frac{1}{4\pi r^{*2}} = \frac{4(1-\bar{\alpha})}{A} \tag{2.10}$$

from which

$$r^* = \sqrt{\frac{A}{16\pi(1-\bar{\alpha})}} \tag{2.11}$$

The SPL can also be calculated using specific software. The ray tracing method and auralisation technique are explained in the following chapter.

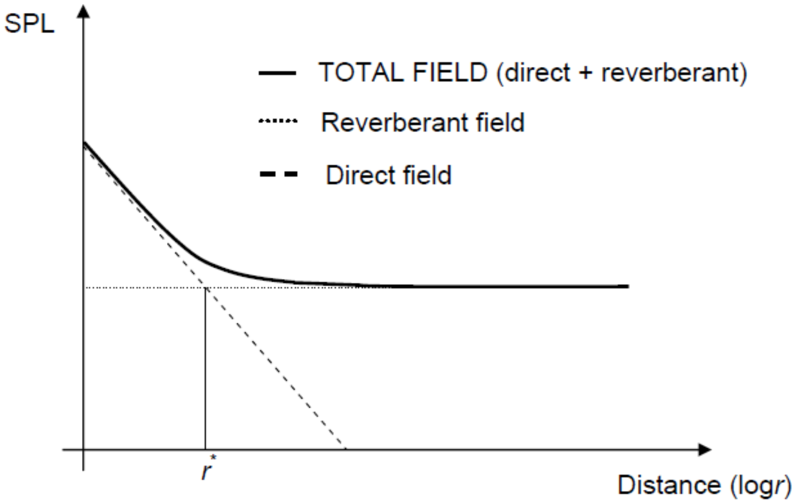


Figure 2.6 The relation between the sound pressure level and the distance between the source and the receiver. The reverberant field becomes dominant after the critical distance r* (Galbrun, 2011).

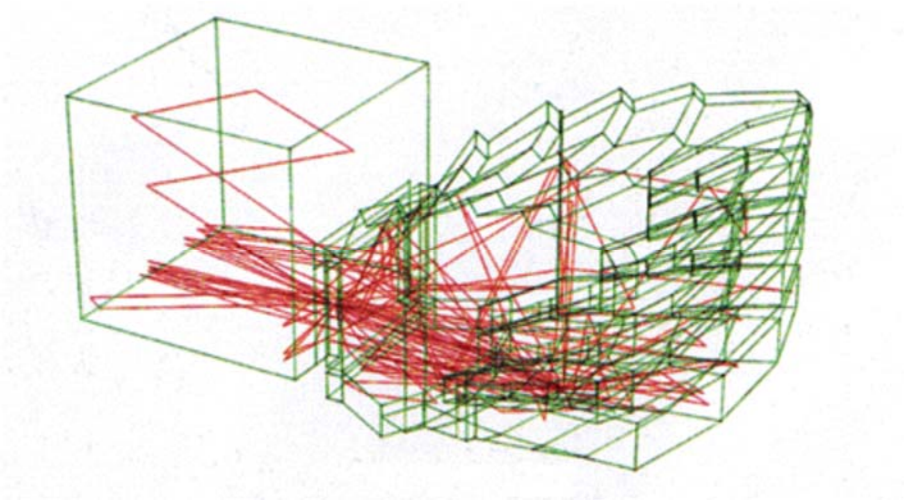


Figure 2.7 An example model of a concert hole to be used for ray tracing technique (Galbrun, 2011).

2.3.5 Ray tracing models and auralisation

The ray tracing technique uses specular reflections of sound rays at each surface of a room to calculate the impulse response of the enclosure. Once surface absorptions are known, a computer model can be used in order to simulate both the reflections and the direct sound emitted from the sound source. The accuracy of the model is directly related with the number of sound rays used in the model. An example computer model of a concert hole is given in Figure 2.7.

Auralisation is the acoustic equivalent of visualization. The impulse response that is derived from the ray tracing model is used in this technique. By combining the room impulse response and anechoic chamber recordings of a source signal, it is possible to evaluate the acoustic design of a room by listening.

2.4 Speech Intelligibility

To compare the differences among languages under varying acoustic conditions in terms of speech intelligibility, it is necessary to define what speech intelligibility is. Regardless of the language considered, a conversation in an enclosure requires an adequate amount of sound energy transferred from a talker to a listener without changing. Speech intelligibility is a measure of successful transmission of specific information carriers, for example words or sentences, between a talker and a listener by oral communication (Long, 2006). The most common method of measuring speech intelligibility is using carrier sentences that contain a keyword. Three kinds of test materials can be used: sentences, words, and syllables. These carrier sentences are either recorded to be played to a listener group or read to a listener group face-to-face, and listeners are required to identify the keywords (Long, 2006).

The intelligibility test can be presented under various room acoustic conditions such as different reverberation times (*T*) and signal-to-noise ratio (S/N). Different intelligibility test materials have sensitivities to room acoustic conditions. If the test material is shorter, it is more difficult to understand in higher S/N. For instance, sentences are the most intelligible and syllables are the least intelligible. The comparison of Figure 2.8 demonstrates the difference between intelligibility test materials (Miller *et al.*, 1951).

The reason behind the differences between test materials is the predictability of the keyword. Previous studies have shown that, if the sentence is presented in a specific context that is established before, then the intelligibility of the keyword is higher than when it is used in a more neutral context (Kalikow and Stevens, 1977).

Besides acoustics properties of an enclosure, talkers' or listeners' native language, social background and current psychological situation may also affect the intelligibility of speech, but these factors are rarely quantified (Houtgast and Steeneken, 1984; Davies *et al.*, 2009b). The assessment of intelligibility is explained and discussed in the following sections.

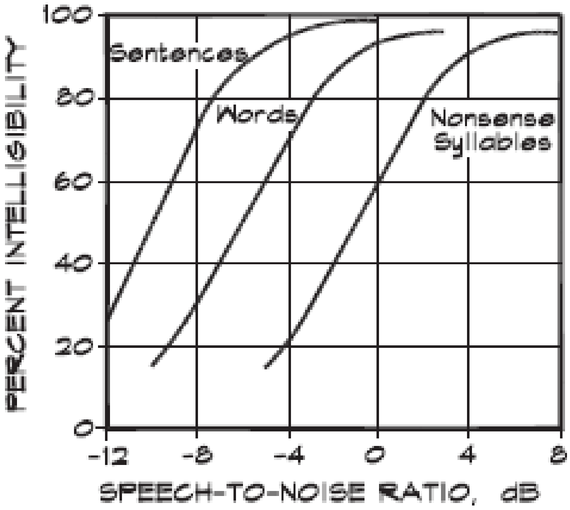


Figure 2.8 Results of typical intelligibility tests (Miller *et al.*, 1951).

2.4.1 Assessment of Speech Intelligibility

A large variety of speech intelligibility tests and experiments have been developed in order to assess speech transmission systems, including face to face speech communication in an enclosure. Some of these speech intelligibility tests such as the Articulation Index (AI) rely on subjective word scores, and others, such as the Speech Transmission Index (STI) are measured and/or calculated physically (Steeneken and Houtgast, 2002).

In the following sections, objective and subjective methods of measuring and predicting speech intelligibility are discussed.

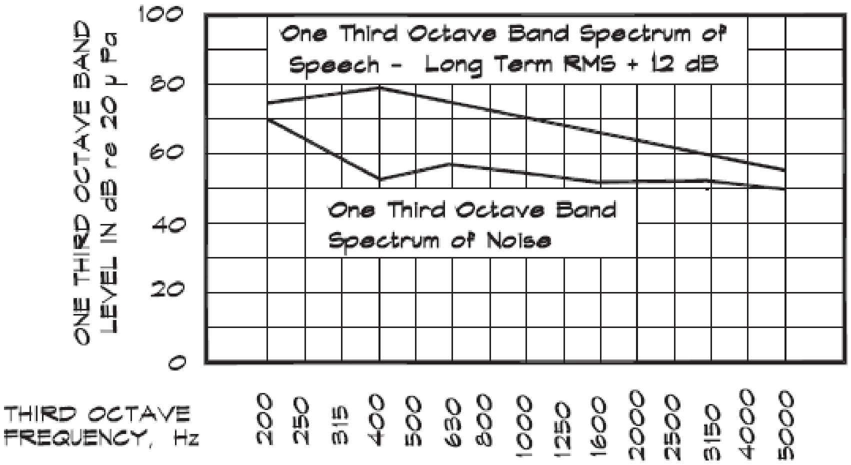
2.4.1.1 Subjective Assessment

Subjective assessment of speech intelligibility is based on reading spoken word or sentence lists either face-to-face or by playing the lists from a recording to the listeners. The listeners are required to write what they heard, and the number of correct answers leads to a rating of speech intelligibility (Peng, 2005).

The articulation index (AI) was developed by French and Steinberg (1947), and is a measure of speech intelligibility. It is generally used for speech privacy purposes. AI is tested by reading meaningless consonant-vowel-consonant (CVC) words or phonemes in carrier sentences to human subjects. Subjects are required to write the keywords they hear, and the fraction of correct answers gives the articulation index (AI). It can also be calculated by using signal-to-noise ratio (S/N) data. The calculation method is illustrated in Figure 2.9.

In both methods discussed, the minimum value of AI is 0, which is no intelligibility of speech, and the maximum value is 1, which is excellent (Long, 2006).

The intelligibility test materials are made of specially selected words or sentence lists. A comparison between the different types of test materials which can be used is shown in Figure 2.10.



BAND	SPEECH PEAKS MINUS NOISE (dB)	WEIGHT	COLUMN 2 x 3
200	5	0.0004	0.0020
250	12	0.0010	0.0120
315	18	0.0010	0.0180
400	26	0.0014	0.0364
500	23	0.0014	0.0322
630	18	0.0020	0.0360
800	17	0.0020	0.0340
1000	16	0.0024	0.0384
1250	15	0.0030	0.0450
1600	14	0.0037	0.0518
2000	12	0.0037	0.0444
2500	10	0.0034	0.0340
3150	8	0.0034	0.0272
4000	6	0.0024	0.0144
5000	5	0.0020	0.0100
			AI = 0.4358

Figure 2.9 An example calculation and the weighting factors of the articulation index (Kryter, 1970).

The smallest pronounceable unit of speech is a syllable, which consists of a minimum of one vowel, or a combination of a vowel and one or more consonants. Monosyllabic words consist of a single syllable and polysyllabic words consist of two or more syllables. Monosyllabic meaningless words are also called logatoms. The articulation test materials can be composed of a variety of speech elements including monosyllabic meaningful words, monosyllabic meaningless words (logatoms), polysyllabic words, or sentences (ISO TR 4870, 1991).

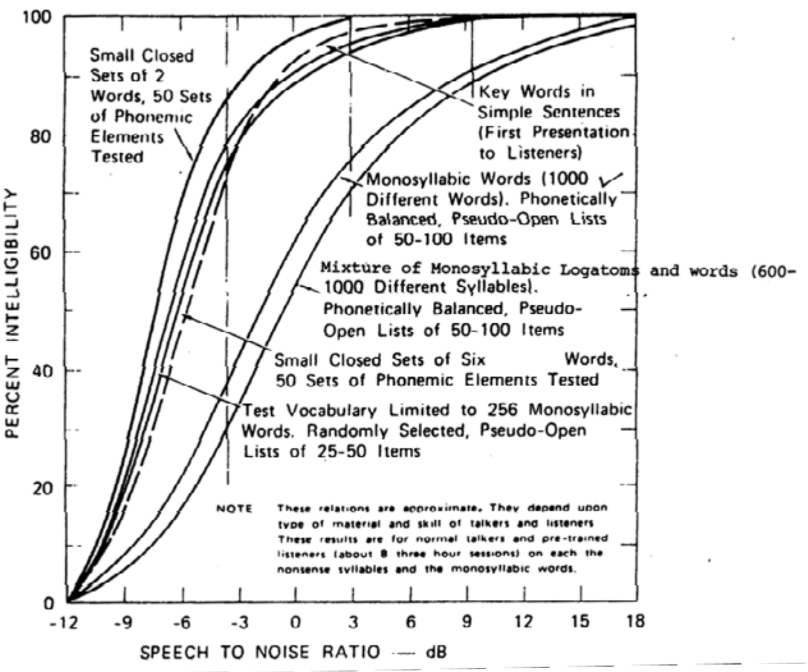
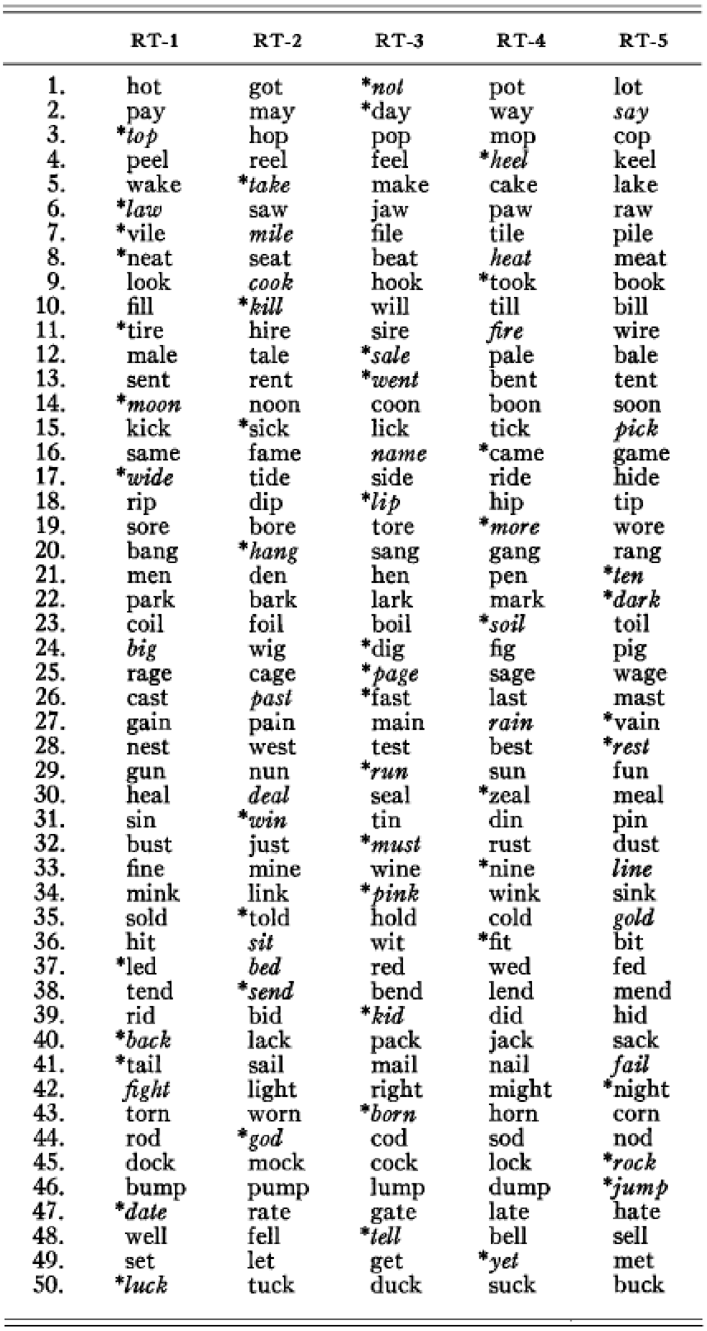


Figure 2.10 Comparison between various speech intelligibility test materials (ISO TR 4870, 1991).

The first method discussed in this review is the Rhyme Test, which was developed by Fairbanks (1958). This articulation test includes 50 sets of monosyllabic words, where each set includes 5 rhyming monosyllabic words. Within each 5 rhyming words, the initial consonant differs. An example set of rhyme test can be seen in Figure 2.11. All 50 sets of words were recorded on a tape by the author, by reading rhyming words consecutively without stopping. The final recording was then divided into 5 sub-recordings (10 sets of words per sub-recording), which could be presented to 5 different listener groups in various acoustic conditions. Each sub-recording consisted of 50 words. Subjects were given a response sheet of 50 words with their initial letter left blank, and asked to complete the missing letter while listening to the recording (Fairbanks, 1958).



	RT-1	RT-2	RT-3	RT-4	RT-5
1.	hot	got	*_not_	pot	lot
2.	pay	may	*day	way	_say_
3.	*_top_	hop	pop	mop	cop
4.	peel	reel	feel	*_heel_	keel
5.	wake	*_take_	make	cake	lake
6.	*_law_	saw	jaw	paw	raw
7.	*vile	_mile_	file	tile	pile
8.	*neat	seat	beat	_heat_	meat
9.	look	_cook_	hook	*took	book
10.	fill	*_kill_	will	till	bill
11.	*tire	hire	sire	_fire_	wire
12.	male	tale	*_sale_	pale	bale
13.	sent	rent	*_went_	bent	tent
14.	*_moon_	noon	coon	boon	soon
15.	kick	*sick	lick	tick	_pick_
16.	same	fame	_name_	*came	game
17.	*_wide_	tide	side	ride	hide
18.	rip	dip	*_lip_	hip	tip
19.	sore	bore	tore	*_more_	wore
20.	bang	*_hang_	sang	gang	rang
21.	men	den	hen	pen	*_ten_
22.	park	bark	lark	mark	*_dark_
23.	coil	foil	boil	*_soil_	toil
24.	_big_	wig	*dig	fig	pig
25.	rage	cage	*_page_	sage	wage
26.	cast	_past_	*fast	last	mast
27.	gain	pain	main	_rain_	*vain
28.	nest	west	test	best	*_rest_
29.	gun	nun	*_run_	sun	fun
30.	heal	_deal_	seal	*zeal	meal
31.	sin	*_win_	tin	din	pin
32.	bust	just	*_must_	rust	dust
33.	fine	mine	wine	*nine	_line_
34.	mink	link	*_pink_	wink	sink
35.	sold	*told	hold	cold	_gold_
36.	hit	_sit_	wit	*fit	bit
37.	*led	_bed_	red	wed	fed
38.	tend	*_send_	bend	lend	mend
39.	rid	bid	*_kid_	did	hid
40.	*_back_	lack	pack	jack	sack
41.	*tail	sail	mail	nail	_fail_
42.	_fight_	light	right	might	*night
43.	torn	worn	*_born_	horn	corn
44.	rod	*_god_	cod	sod	nod
45.	dock	mock	cock	lock	*_rock_
46.	bump	pump	lump	dump	*_jump_
47.	*_date_	rate	gate	late	hate
48.	well	fell	*_tell_	bell	sell
49.	set	let	get	*_yet_	met
50.	*_luck_	tuck	duck	suck	buck

Figure 2.11 An example set of rhyme test. Italics: biased with high familiarity, asterisks: matched the consonant distribution of language (Fairbanks, 1958).

The research presented above was built only for the US English language, but it can be modified for other languages. Another limitation of this study is the lack of evidence of representing the language, in this case US English. To represent a specific language, a word list should include approximately the same phonetic properties and sound types as they appear in the language. Word lists that have these properties are called "phonetically balanced" (PB). It should however be pointed out that monosyllabic consonant-vowel-consonant (CVC) words are too short to represent a language, even if they are phonetically balanced. Therefore, it is recommended to use a combination of monosyllabic and polysyllabic PB word lists if multiple languages are to be tested (ISO TR 4870, 1991).

In the literature Fairbank's rhyme test was modified for several languages including Chinese, Czech, and Spanish (ISO/TR 4870, 1991). A recent study by Alias and Trivino (2007) modified the rhyme test for Catalan language by balancing the CVC words phonetically. Because of the lack of meaningful monosyllable words in the Catalan language, obtaining a full list was a challenging process. Keeping this in mind, attention should be given to monosyllabic words while modifying rhyme tests for other languages as stated in ISO/TR 4870 (1991).

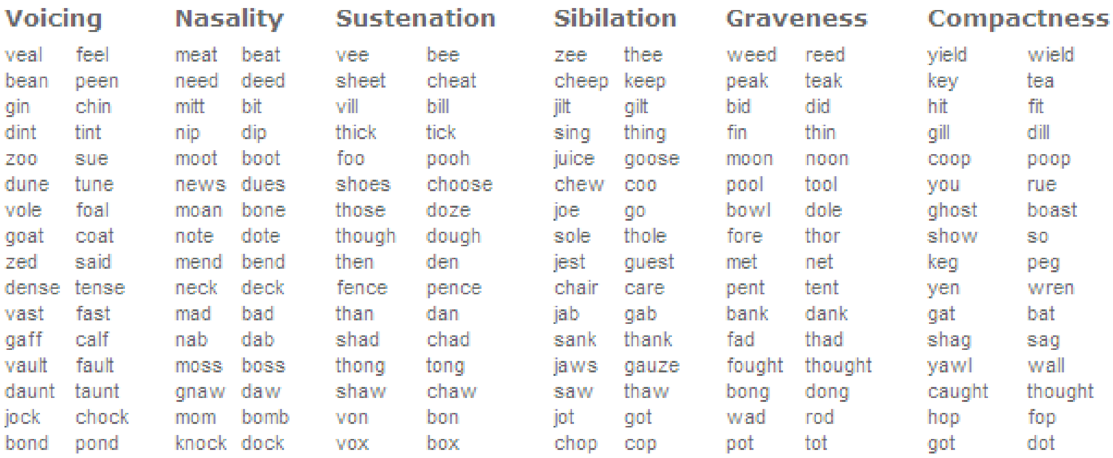
Two types of test methods are suggested for assessing word intelligibility in ISO/TR 4870 (1991); firstly large set tests and small closed set tests. Large set test materials use either 1000 meaningful words or at least 650 logatoms for testing intelligibility under varying conditions. These 1000 words should be divided into 50 word lists. During the test, subjects are expected to write down what they hear. While preparing small closed sets, test materials are organized into smaller sub-sets of 2 to 10 words, depending on the test material. In each sub-set, one word differs from the others by one letter of the same position in the word (ex. Rhyme test). Subjects receive a response form, and are expected to select the word that they heard (ISO TR 4870, 1991).

The third speech intelligibility test material type is the sentence lists. Sentence lists as a speech intelligibility testing material are mostly used for assessing the stress pattern, inflection, and maintenance of loudness rather than the intelligibility (Beranek, 1949). While preparing sentence lists as a test material, special attention should be given to balance the sentences in a way that they represent an average everyday conversation in a language. It is also important to control the effect of predictability caused by the content, context and the prosodic structure of sentences. When a sentence is presented to the listener in a specific context, previous words might carry an utterance about following words. Also, prosodic structure of a sentence carries meanings that grammatical and phonetic structures do not. In order to overcome the mentioned effects of predictability Kalikow and Stevens (1977) suggested to use a balanced combination of low-predictability sentences (ex. "John was discussing the ...") and high-predictability sentences (ex. "The boat sailed across the ..."). Subjects are given a response sheet, in which the key words are left blank. Assessment and rating of the sentence intelligibility test is based on the keyword (usually the last word of each sentence) (Kalikow and Stevens, 1977).

Based on the work of Fairbanks (1958) on rhyme tests, Voiers (1965) developed the diagnostic rhyme test (DRT). It is claimed that the diagnostic rhyme test is minimally affected by listeners' familiarity with the voice of the speaker. The DRT is a closed-set test, which is composed of 96 pairs of rhyming monosyllabic consonant-vowel-consonant (CVC) words (Figure 2.12). The word pairs differ only by their initial consonants, and no carrier sentence is used. In the literature, it is found that not all consonants are equally understandable in various positions in the word; however, the feature of the consonant does not change (Voiers, 1977). The word pair is presented visually to the listeners, and they are asked to identify the word they heard (ANSI/ASA S3.2, 2009). The chosen stimulus word could be any of the two words in the pair. The choice does not affect the identity of the feature, it only affects the state (positive/negative, i.e., voiced/unvoiced) (Voiers, 1977). It was advised to use more than one speaker for the recordings, and to use between eight to ten listeners. The test material and the response options on the listeners' response forms could be randomised, without changing the main structure of the test. The rate of word presentation was suggested as one word per 1.4 seconds. The scores are usually calculated for six major diagnostic features, and a total score that represents the average of the diagnostic scores. All of the DRT scores are calculated by computing the formula below, which is adjusted for familiar correction of guessing:

$$S = \frac{100(R - W)}{T} \tag{2.12}$$

where S is the adjusted percent-correct responses, R is the number of correct responses, W is the number of incorrect responses, and T is the total number of responses (Voiers, 1977). The DRT was also validated to test the speech intelligibility under noisy conditions. Voiers (1977) found several differences between the consonant features of English in terms of the speech intelligibility. The results showed that voicing and nasality are the least vulnerable features to the noise masking, and graveness is extremely vulnerable. The DRT was developed in several languages besides English



Voicing		Nasality		Sustenation		Sibilation		Graveness		Compactness	
veal	feel	meat	beat	vee	bee	zee	thee	weed	reed	yield	wield
bean	peen	need	deed	sheet	cheat	cheep	keep	peak	teak	key	tea
gin	chin	mitt	bit	vill	bill	jilt	gilt	bid	did	hit	fit
dint	tint	nip	dip	thick	tick	sing	thing	fin	thin	gill	dill
zoo	sue	moot	boot	foo	pooh	juice	goose	moon	noon	coop	poop
dune	tune	news	dues	shoes	choose	chew	coo	pool	tool	you	rue
vole	foal	moan	bone	those	doze	joe	go	bowl	dole	ghost	boast
goat	coat	note	dote	though	dough	sole	thole	fore	thor	show	so
zed	said	mend	bend	then	den	jest	guest	met	net	keg	peg
dense	tense	neck	deck	fence	pence	chair	care	pent	tent	yen	wren
vast	fast	mad	bad	than	dan	jab	gab	bank	dank	gat	bat
gaff	calf	nab	dab	shad	chad	sank	thank	fad	thad	shag	sag
vault	fault	moss	boss	thong	tong	jaws	gauze	fought	thought	yawl	wall
daunt	taunt	gnaw	daw	shaw	chaw	saw	thaw	bong	dong	caught	thought
jock	chock	mom	bomb	von	bon	jot	got	wad	rod	hop	fop
bond	pond	knock	dock	vox	box	chop	cop	pot	tot	got	dot

Figure 2.12 The diagnostic rhyme test (DRT) word list (Voiers, 1977).

such as, Arabic (Boudraa et al., 2008), Chinese (Li et al., 2000) and Japanese (Kondo, 2012).

Each of the mentioned tests has its own limitations as discussed. Rhyme tests are not phonetically balanced, so these are not representing a language (Fairbanks, 1958). Modified rhyme tests (MRT) are phonetically balanced, however, using only monosyllables is not recommended for testing multiple languages. Sentence intelligibility tests are not sensitive enough for discriminating the differences between languages because of the effect of predictability (Kalikow and Stevens, 1977). Voiers (1977) explained the limitations of the phonetically balanced (PB) word lists, the rhyme tests, and the modified rhyme tests in comparison to diagnostic rhyme test (DRT). First of all, the PB word lists require extensive durations of training, in order to control the effects of listeners' familiarity. The increased training duration leads to a decrease of the difficulty of the listening test. Familiarity with the PB list might also affect qualitative changes in the listeners' performance. Therefore, the results of the listening test do not reflect the actual intelligibility level of speech.

The rhyme test and the MRT use single phoneme as word lists, which provide improved control over inter-phonemic constraints. However, the rhyme test cannot eliminate other contextual factors; the MRT has the extensive control over these factors by restricting the response options. Restricting the listeners' options also has its own limitations, such as, the listeners may be forced to give misleading responses, or the listener may be cannot find the appropriate response among the options. The above limitations could be controlled by using two-choice test items, as it is in the DRT. Moreover, the extensive amount of data derived from the results of the DRT yields more than 24 independents score (the diagnostic features of consonants and the total score) (Voiers, 1999). These results could be compared among various languages, to see the effects of the background noise and the reverberation time on the linguistic features.

2.4.1.2 Objective Assessment

Traditionally, the intelligibility of speech has been tested by using some test material such as word or sentence lists on listeners in enclosures (Long, 2006). In further studies, direct measuring and predicting methods to assess the intelligibility of speech were developed to overcome human errors caused by the use of human listeners as subjects (Bradley, 1986).

One of the most common methods used to predict the intelligibility of speech is the modulation transfer function (MTF). The MTF method was introduced as a measure to examine the effects of the enclosures acoustic properties on the intelligibility of speech. The MTF was originally used for the assessment of optical system performance by calculating the MTF using spatially sine-wave modulated light patterns (Houtgast and Steeneken, 1973). This technique was then adapted to room acoustics, especially using sine-wave modulation on a speech signal. The main reason for the success of calculating speech intelligibility by using the MTF is the influence of reverberation time, excessive echoes and interfering noise on the function (Houtgast and Steeneken, 1973).

Preservation of speech is implied by the preservation of intensity modulations and the method was built on this principle. The modulation transfer function (MTF) points out the amount of preserved intensity modulations. The main idea behind its use is that modulated bands of sound create speech. Vibrating vocal cords create sounds and the mouth modulates the sound in different frequencies to form words. The MTF simulates the idea behind the creation of words, and uses an octave-band wide source of noise to modulate it with a low-frequency tone. In other words, the creation process of words is converted into a mathematical formula (Houtgast and Steeneken, 1980).

The speech signal is affected by two factors until reaching the listener, and these factors cause a reduction of the intelligibility of speech. Firstly, the background noise or in other words signal-to-noise ratio (S/N), which is independent from the modulation

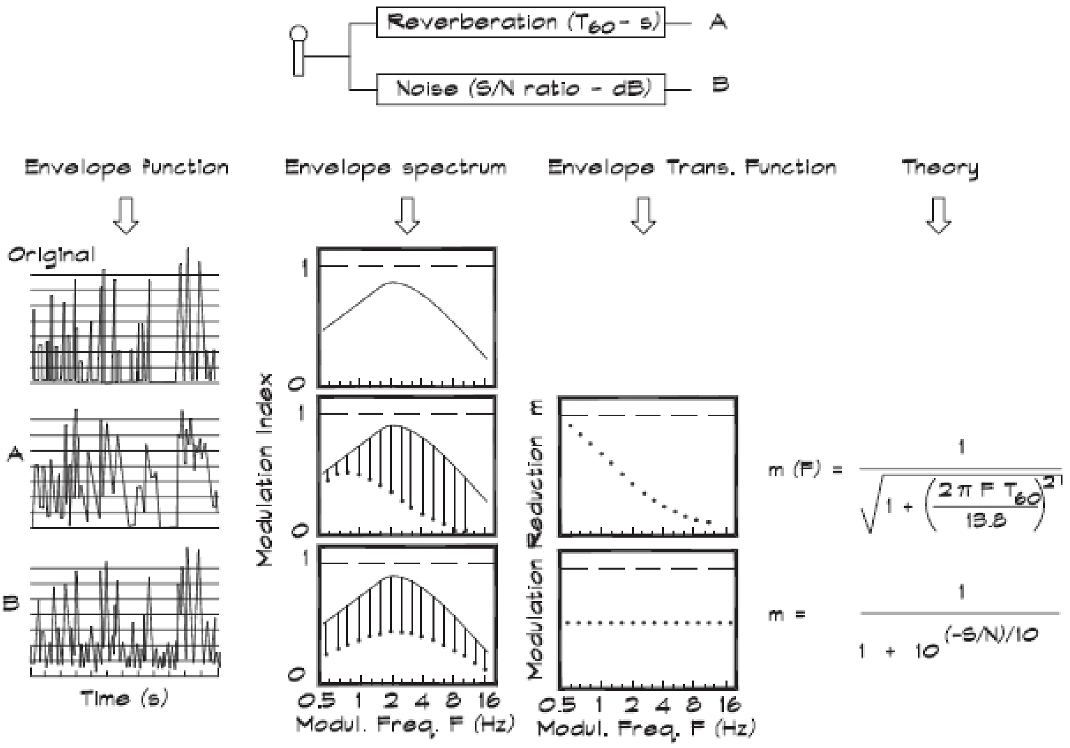


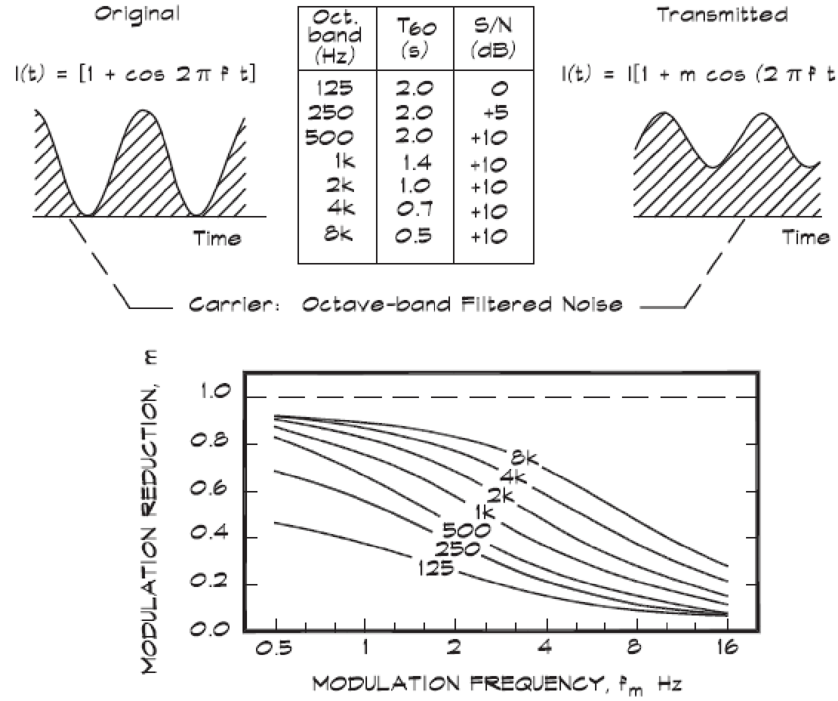
Figure 2.13 Modulation Transfer Function Theory (Houtgast and Steeneken, 1985).

frequency; secondly the reverberation time (*T*), which has a different effect on every frequency. Figure 2.13 illustrates the MTF theory, and shows the modulation formulae used for S/N and T. Figure 2.14 shows the modulation reduction between the original and transmitted signal.

The modulation reduction factor m($\mathrm{f_m}$) of Figure 2.13 defines the decrease in the modulation caused by acoustical conditions and is a function of the modulation frequency $\mathrm{f_m}$ (Long, 2006). First of all m($\mathrm{f_m}$) should be calculated by using the formula below (Houtgast and Steeneken, 1980),

$$\mathrm{m}(\mathrm{f_m}) = \frac{1}{\sqrt{1+\left[2\pi f_m \frac{T_{60}}{13.8}\right]}} \cdot \frac{1}{1+10^{(-0.1L_{SN})}} \tag{2.13}$$

where, m($\mathrm{f_m}$) is the modulation reduction factor, L_{SN} is the signal-to-noise ratio (dB), f_m is the modulation frequency (Hz) and T_{60} is the reverberation time of the enclosure. m($\mathrm{f_m}$) should be calculated for each of the 14 modulation frequencies between 0.63 Hz and 12.5 Hz for 7 octave bands, a total of 98 m values (Long, 2006).



Oct. band (Hz)	T60 (s)	S/N (dB)
125	2.0	0
250	2.0	+5
500	2.0	+10
1k	1.4	+10
2k	1.0	+10
4k	0.7	+10
8k	0.5	+10

Figure 2.14 Calculation of modulation reduction using MTF according to the reverberation time and signal-to-noise data given in the table (Long, 2006).

2.4.1.2.1 Speech Transmission Index (STI)

After having explained the MTF, the connection between the MTF and speech intelligibility is described in this section. The missing link between the MTF and speech intelligibility is the speech transmission index (STI), which is a direct measure of speech intelligibility comparable to the articulation index (AI). The relation between the STI and AI was shown in Figure 2.5.

The first step for calculating the speech transmission index (STI) is using an algorithm that transforms a set of *m* values into an apparent signal-to-noise ratio.

$$\mathrm{L}_{\mathrm{SNapp}} = 10 \log \frac{m}{1-m} \tag{2.14}$$

where $\mathrm{L}_{\mathrm{SNapp}}$ is the apparent signal-to-noise ratio and *m* is the modulation reduction factor. Then STI is derived by calculating the weighted average of 98 apparent signal-to-noise ratios (14 modulation frequencies in 7 octave bands) after normalizing the $\mathrm{L}_{\mathrm{SNapp}}$ values such that;

$$\mathrm{STI} = 1.0 \text{ when } \mathrm{L_{SNapp}} \geq 15 \text{ dB} \tag{2.15}$$

$$\mathrm{STI} = 0.0 \text{ when } \mathrm{L_{SNapp}} \leq -15 \text{ dB} \tag{2.16}$$

and

$$\overline{\mathrm{L_{SNapp}}} = \sum_{i=1}^{7} \mathrm{w_i} \left(\mathrm{L_{SNapp}}\right)_i \tag{2.17}$$

where $\overline{\mathrm{L_{SNapp}}}$ is the average apparent signal-to-noise ratio (dB) and w_i is the weighting for octave bands from 125 Hz to 8000 Hz, which are 0.13, 0.14, 0.11, 0.12, 0.19, 0.17, and 0.14 respectively. The w_i values are derived from the articulation index method, and modified in order to weight phonetically balanced (PB) word scores and speech transmission index (STI) (Steeneken and Houtgast, 1980). Finally, to calculate the value of the speech transmission index (STI) the formula below is used,

$$\mathrm{STI} = \left[\overline{\mathrm{L_{SNapp}}} + 15\right]/30 \tag{2.18}$$

The STI varies between 0 (no intelligibility) and 1 (complete intelligibility). The relationship between the signal-to-noise ratio, reverberation time, and speech transmission index is shown in Figure 2.15.

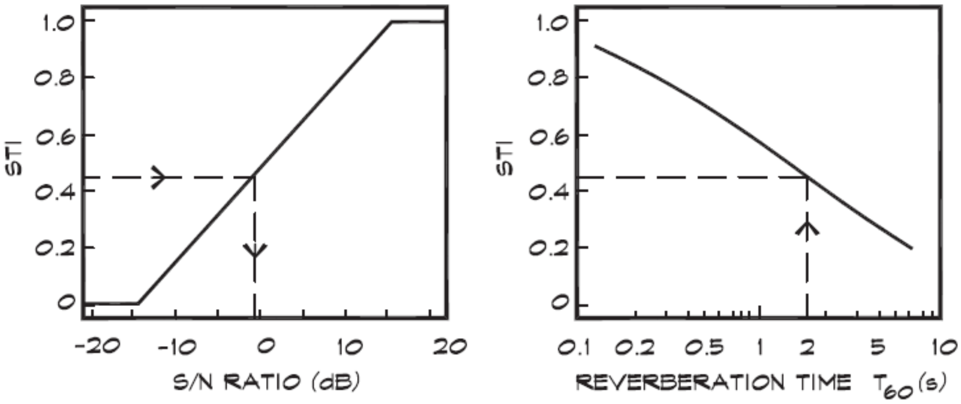
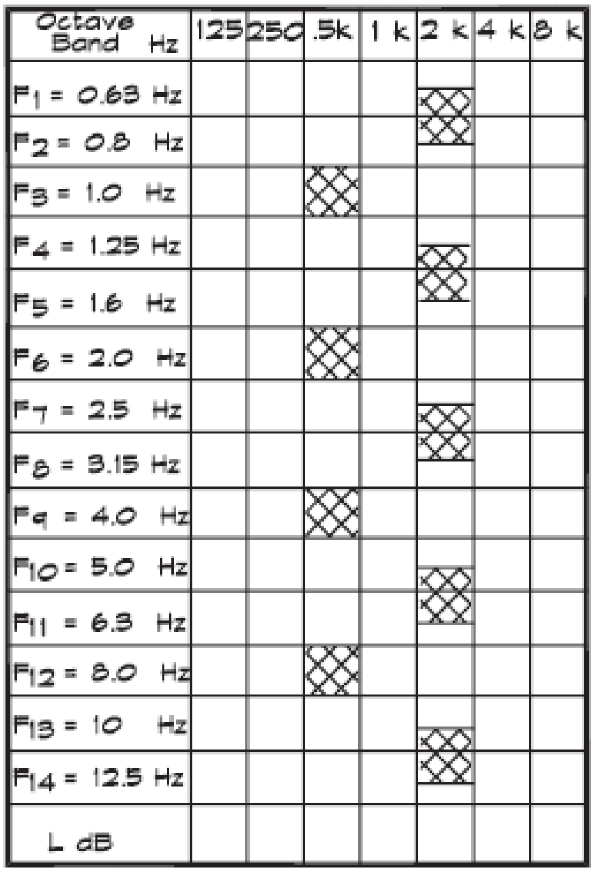


Figure 2.15 Relationship between signal-to-noise ratio (S/N), reverberation time (T_{60}) and speech transmission index (STI) (Houtgast and Steeneken, 1985).

The MTF method used to calculate STI makes it possible to establish intelligibility tests by using a test signal rather than human subjects, therefore preventing any human errors (Long, 2006).

2.4.1.2.2 Rapid Speech Transmission Index (RASTI)

The rapid speech transmission index (RASTI) is a simplified method of the speech transmission index (STI), which is used to measure the intelligibility of speech. Rather than calculating the modulation reduction factor (*m*) for 98 variations of modulation frequencies, values of *m* are calculated for nine modulation frequencies, which are 1, 2, 4, and 8 Hz. at 500 Hz, and 0.7, 1.4, 2.8, 5.6, and 11.2 Hz. at 2000 Hz (Figure 2.16) (Long, 2006). RASTI is computed quicker than STI, because of the fewer calculations of *m*.



Octave Band Hz	125	250	.5k	1 k	2 k	4 k	8 k
F_1 = 0.63 Hz					▨		
F_2 = 0.8 Hz					▨		
F_3 = 1.0 Hz			▨				
F_4 = 1.25 Hz					▨		
F_5 = 1.6 Hz					▨		
F_6 = 2.0 Hz			▨				
F_7 = 2.5 Hz					▨		
F_8 = 3.15 Hz					▨		
F_9 = 4.0 Hz			▨				
F_{10} = 5.0 Hz					▨		
F_{11} = 6.3 Hz					▨		
F_{12} = 8.0 Hz			▨				
F_{13} = 10 Hz					▨		
F_{14} = 12.5 Hz					▨		
L dB							

Figure 2.16 9 out of 98 modulation frequencies that are used for calculating the rapid speech transmission index (Houtgast and Steeneken, 1985).

Similarly, to the calculation of the speech transmission index (STI), the m values are converted into apparent signal-to-noise ratios (L_{SNapp}) to be averaged in order to find the RASTI by using the following equation,

$$\mathrm{RASTI} = \left[\overline{\mathrm{L_{SNapp}}} + 15\right]/30 \tag{2.19}$$

where, RASTI is the rapid speech transmission index and $\overline{\mathrm{L_{SNapp}}}$ is the average apparent signal to noise ratio (dB). RASTI values can be calculated to evaluate both unamplified talker and amplified sound system situations (Long, 2006).

2.5 The Soundscape Approach

Each listener-talker pair has its own communication channel which is affected by various individual factors, for example an enclosure, a public address system or even a phone line (van Wijngaarden *et al.*, 2004). These individual factors might be background noise, reverberation time, or individuals' disabilities such as hearing loss as well as distortions in communication devices (Christiansen *et al.*, 2010). Furthermore, another factor affecting speech intelligibility is the differences between languages. Kang (1998) claimed that the articulation index (AI) and speech transmission index (STI) are based on Western languages, therefore it cannot be said that if the acoustical condition of an enclosed space is ideal for speech, for example, in English, it is also ideal for other languages. According to a previous research held by Houtgast and Steeneken (1984), one of the reasons behind the disparity between various tests could be language specific effects.

It is also possible that the speech communication channel between a talker and a listener is affected by social, physical and perceptual factors. These factors are related to listening, capturing, feeling and being in a sound environment, or in other words, in a soundscape. To evaluate any social effects on speech perception, a qualitative approach such as soundscape theory should be followed (Davies *et al.*, 2009b).

Perceptual factors and room acoustic conditions are the two main aspects considered by the present research. While dealing with multi-lingual factors and communication between people, the perception of the sound becomes as important as the quality of the sound itself. Schafer (1977) built the framework of the idea of combining perceptual factors and the acoustic conditions of a space, which is called the soundscape theory. Soundscape is a terminology used to define a collection of sounds within a space and the perception of that sound environment. Current soundscape research therefore considers and encompasses both the physical and perceptual properties of the aural environment. In this section, the background information, and methodologies of the soundscape theory will be discussed.

2.5.1 Background

Soundscape research is widely interdisciplinary and can involve areas as varied as engineering, social sciences, environmental psychology and arts. Acousticians and psycho-acousticians focus on the movement of sound and the perception of it by the human brain, sociologists and psychologists try to understand the relationship between human behaviour with sound and how human behaviour is influenced by the sound, and engineers as well as artists (mainly musicians) use their knowledge and abilities to design the ideal soundscape. Attention should also be given to the human behaviour patterns in different sonic environments to investigate cross-cultural interaction in a soundscape (Schafer, 1977).

For assessment of sound environments and soundscape quality in both indoor and outdoor spaces, a Working Group of ISO/TC 43/SC 1 was established in 2008. The major aim of the group has been developing a standardised method for assessment of soundscape environments and to create enhanced guidelines for both professionals and researchers who are working on architecture, acoustics, social sciences and any other research areas that are seeking an understanding of a connection between human perception and sonic environments. The standard BS ISO 12913-1 (2014) resulted from this work and was published in 2014. This standard defines the soundscape as the "acoustic environment as perceived or experienced and/or understood by a person or people, in context" (BS ISO 12913-1, 2014).

Kang (2010) published an extensive review of the literature on soundscape theory, from the basic concept of the theory to the design of a soundscape. The review covers the framework and the recent activities such as research projects, standardization, practice and publications on definition, evaluation, description, modelling, data collection, standardizing, and design of a soundscape.

Brown *et al.* (2011) discussed basic definitions of the soundscape terminology, possible key features and methodologies for standardising the assessment of the sonic environment. At the time, it was argued that a single strict definition of soundscape cannot be established, however, some observations were discussed in order to build a framework. It was suggested that the word 'soundscape' includes both the perception of the acoustic environment and the total collection of sounds within a space (Figure 2.17). Brown *et al.* (2011) classified all sound sources in an acoustic environment to create a framework for a general identification of sources. A possible taxonomy of the acoustic environment has been designed. The main categories of the taxonomy were indoor acoustic environment and outdoor acoustic environment. Then, the outdoor sonic environment category divided into four sub-categories: urban, rural, wilderness and underwater. The taxonomy lists all possible sound sources under these sub-categories.

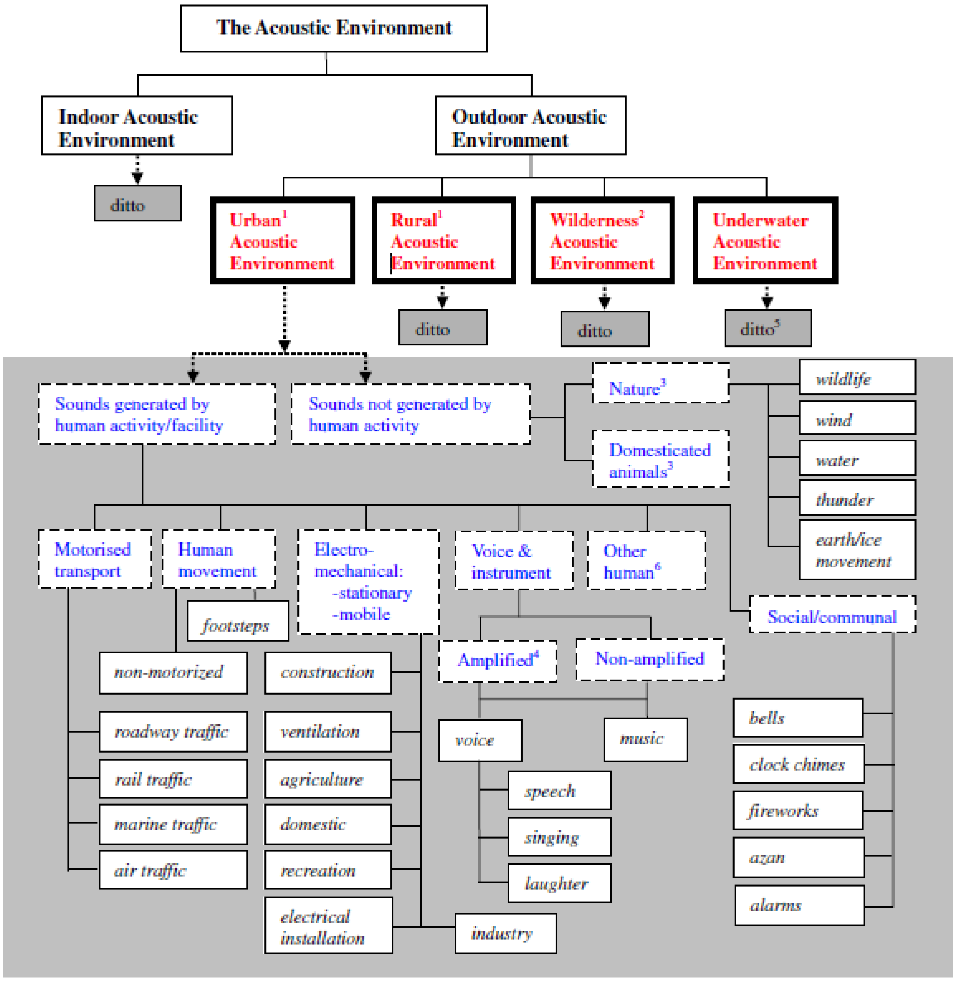
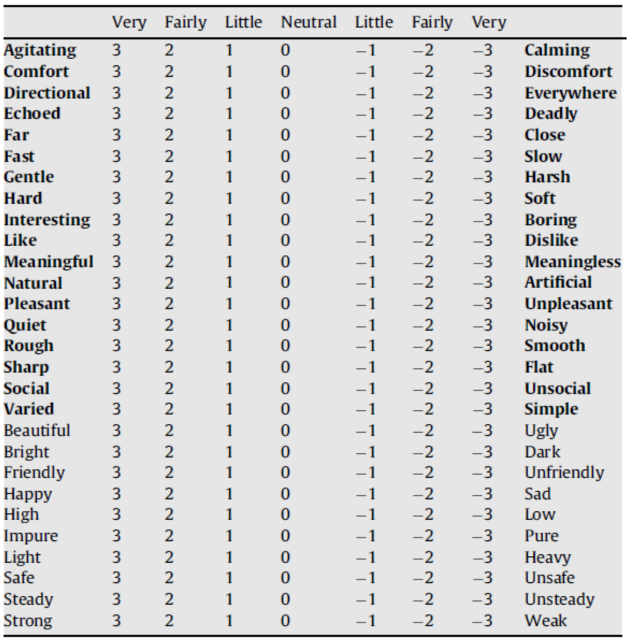


Figure 2.17 A possible taxonomy of the acoustic environment. Bold boxes: Categories of places, dashed boxes: categories of sound sources, italics: sound sources. (Brown *et al.*, 2011).

2.5.2 Methodologies and Soundscape

The complexity of a soundscape is the result of its interdisciplinary nature, therefore it is crucial to identify the key points which explain the soundscape. The soundscape theory concentrates on the perception of the sonic environment in terms of physical, psychological and social factors. Perception of a sonic environment is achieved by two types of processing in terms of listening. The first type is holistic listening, which does not differentiate between sounds and perceives the sonic environment as a whole. The second type is the descriptive listening, which aims to identify every single sound source contained in the sonic environment (Kang and Zhang, 2010). Raimbault (2006) pointed out the differences between holistic and descriptive listening according to the results of the soundscape study that compares two French cities (Lyon and Nantes). The aim of the present study is to evaluate sounds in a sonic environment in terms of intelligibility and other qualitative descriptors. This can be established by using the semantic differential technique, which depends on descriptive listening and identifies the emotional meaning of words. An example response form is given in Figure 2.18. A typical application of the methodology is represented by a soundscape walk, where participants listen carefully, and assess sounds they are hearing. The auralization technique can also be used to evaluate soundscapes (Peng, 2005; Peng *et al.*, 2011). These assessments are based on subjective measures (Kang and Zhang, 2010).

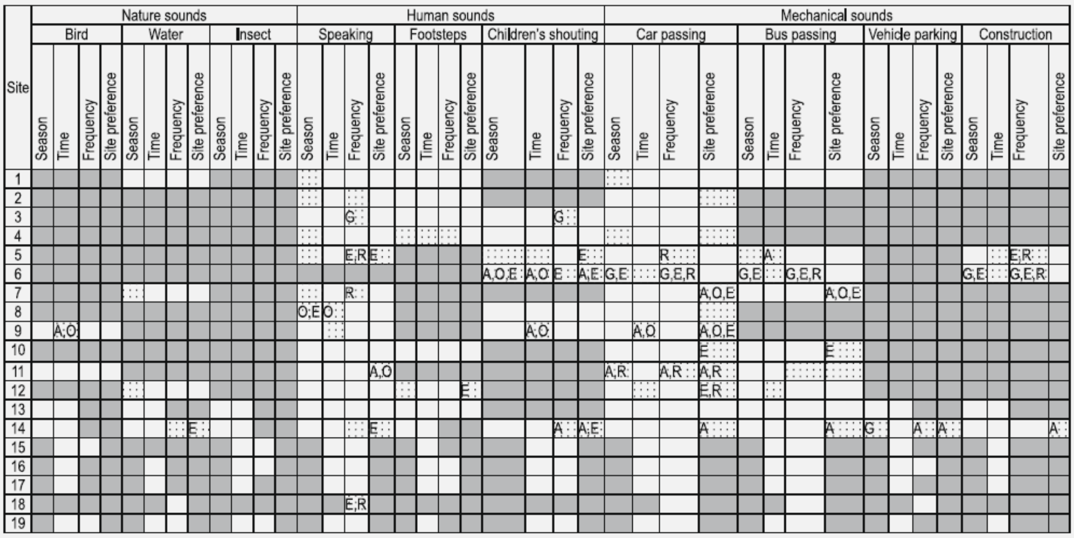


	Very	Fairly	Little	Neutral	Little	Fairly	Very	
Agitating	3	2	1	0	−1	−2	−3	**Calming**
Comfort	3	2	1	0	−1	−2	−3	**Discomfort**
Directional	3	2	1	0	−1	−2	−3	**Everywhere**
Echoed	3	2	1	0	−1	−2	−3	**Deadly**
Far	3	2	1	0	−1	−2	−3	**Close**
Fast	3	2	1	0	−1	−2	−3	**Slow**
Gentle	3	2	1	0	−1	−2	−3	**Harsh**
Hard	3	2	1	0	−1	−2	−3	**Soft**
Interesting	3	2	1	0	−1	−2	−3	**Boring**
Like	3	2	1	0	−1	−2	−3	**Dislike**
Meaningful	3	2	1	0	−1	−2	−3	**Meaningless**
Natural	3	2	1	0	−1	−2	−3	**Artificial**
Pleasant	3	2	1	0	−1	−2	−3	**Unpleasant**
Quiet	3	2	1	0	−1	−2	−3	**Noisy**
Rough	3	2	1	0	−1	−2	−3	**Smooth**
Sharp	3	2	1	0	−1	−2	−3	**Flat**
Social	3	2	1	0	−1	−2	−3	**Unsocial**
Varied	3	2	1	0	−1	−2	−3	**Simple**
Beautiful	3	2	1	0	−1	−2	−3	Ugly
Bright	3	2	1	0	−1	−2	−3	Dark
Friendly	3	2	1	0	−1	−2	−3	Unfriendly
Happy	3	2	1	0	−1	−2	−3	Sad
High	3	2	1	0	−1	−2	−3	Low
Impure	3	2	1	0	−1	−2	−3	Pure
Light	3	2	1	0	−1	−2	−3	Heavy
Safe	3	2	1	0	−1	−2	−3	Unsafe
Steady	3	2	1	0	−1	−2	−3	Unsteady
Strong	3	2	1	0	−1	−2	−3	Weak

Figure 2.18 An example response form of semantic differential analysis (Kang and Zhang, 2010).

The semantic differential analysis of Figure 2.18 was designed by Kang and Zhang (2010) to assess outdoor sonic environments. Further studies are needed to adopt the semantic differential analysis method for indoor usage, and for speech intelligibility research.

Kang and Yu (2010) analyzed gathered data of 14 European and five Chinese urban spaces, to evaluate cultural and demographical effects on sound preferences. The database consisted of age, gender, education, residential status; sounds often heard at home, frequency of coming to the site, and reason for coming to the site. Participants evaluated a large variety of sounds under three categories. The first category was natural sounds, which include bird, water, and insect sounds. The second category was human sounds, which are speaking, footsteps and children's shouting. The last category was mechanical sounds, which are car passing, bus passing, vehicle parking, and construction sounds. According to the data analysis it was found that social/demographical and physical/behavioural/psychological factors have some correlations (Figure 2.19).



Site	Nature sounds												Human sounds												Mechanical sounds															
	Bird				Water				Insect				Speaking				Footsteps				Children's shouting				Car passing				Bus passing				Vehicle parking				Construction			
	Season	Time	Frequency	Site preference	Season	Time	Frequency	Site preference	Season	Time	Frequency	Site preference	Season	Time	Frequency	Site preference	Season	Time	Frequency	Site preference	Season	Time	Frequency	Site preference	Season	Time	Frequency	Site preference	Season	Time	Frequency	Site preference	Season	Time	Frequency	Site preference	Season	Time	Frequency	Site preference
1																																								
2																																								
3															G								G																	
4																																								
5															E,R	E								E			R			A									E,R	
6																					A,O,E	A,O	E	A,E	G,E		G,E,R		G,E		G,E,R						G,E		G,E,R	
7															R													A,O,E				A,O,E								
8													O,E	O																										
9		A,O																				A,O				A,O		A,O,E												
10																												E				E								
11																A,O									A,R		A,R	A,R												
12																				E								E,R												
13																																								
14								E								E							A	A,E				A				A	G		A	A				A
15																																								
16																																								
17																																								
18															E,R																									
19																																								

Figure 2.19 The relationship table of physical/behavioural/physical factors, and social/demographical factors. The white areas indicate the investigated sites, and the dotted white areas indicate that the relationship is statistically significant. (A. age; G, gender; E, education level; R, residence) (Yu and Kang, 2010).

In particular, age and education levels were strongly correlated with sound preference; however, the correlation amount depended on the types of urban open spaces and sounds (Yu and Kang, 2010). It is important to note that, in this study, sound/noise sources were evaluated individually, throughout descriptive listening. In the present study, one of the aims is to find out perceptual effects on the speech intelligibility, with combinations of sounds contributing to the soundscape.

Previous research by Cain *et al.* (2011) analysed soundscape in a combination of two dimensions. The first dimension contained the descriptors of sounds (psychoacoustics), such as loudness, clarity and spatiality. The second dimension used the emotional descriptors that focus on the feelings of listeners. The main difference between the two descriptors was that the descriptors of sound evaluate the hearing of a listener; whilst the emotional descriptors of sound evaluate the feelings of a listener. This is interesting in relation to a multi-lingual environment, as it highlights the importance of addressing the effects of feelings in spaces. In the literature, there are few other studies that focused on emotional descriptors, such as the work of Davies *et al.* (2009a) who created a framework to discriminate psycho-acoustical and emotional dimensions of a soundscape. In this study, two parts of the word soundscape, sound and scape, were considered as two main dimensions. The sound word was related with the description of physical properties of the sound itself, and the scope word was used for explaining the emotional interaction between the listeners and the environment (Figure 2.20).

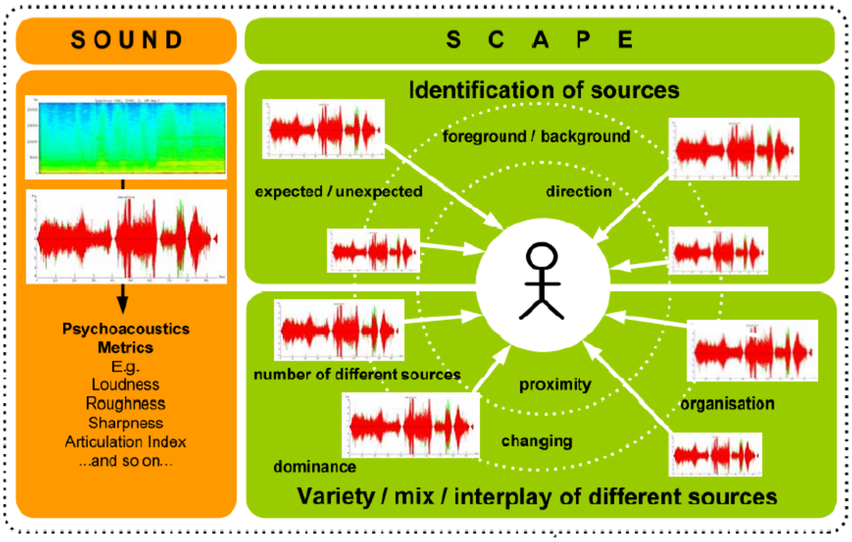


Figure 2.20 The graphical explanation of the sound-scape framework (Davies *et al.*, 2009a)

To expand the previous literature on soundscape theory and speech intelligibility, enclosures should be considered by future research. Most common multi-lingual speech environments persist in enclosures, such as airports, courts, and social gathering spaces. Dokmeci and Kang (2010) built a theoretical and conceptual framework to define soundscape theory for enclosures. The main point of the study was to understand the main aspects of indoor soundscapes. The research argued that the assessment of sound environments consists of three main aspects, which are objective analysis, subjective analysis and the assessment of the built entity. Combination of these three aspects leads us to a psychological point of view rather than physical noise control technique to understand users' sensory experience in the specific sound environment or in other words, the soundscape (Dokmeci and Kang, 2010).

The assessment of an indoor soundscape is very similar to room acoustic analysis, but it also differs in many areas. There are four factors that should be considered while assessing a sound environment. These factors are objective factors, subjective factors, social factors and sonic factors (Figure 2.21) (Dokmeci and Kang, 2010). The objective factors were previously discussed in section 2.1, and they do not differ from room acoustic measurements. However, the subjective, spatial and sonic factors are different in terms of assessment of the sound environment.

The sonic factor includes all of the sound sources in a sonic environment, including the ambient noise (i.e. background noise). The nature and balance of each sound source, as well as the history of the present sonic environment are considered as a sonic factor of the soundscape study (Dokmeci and Kang, 2010).

While considering multi-lingual speech intelligibility as a focal point of the present study, the spatial assessment of a soundscape differs according to the type of noise interference, the function of the enclosure and the user type. The questionnaires or surveys should be prepared accordingly (Dokmeci and Kang, 2010).

The indoor soundscaping is a crucial part of the present study, in order to include the effects of varying languages and socio-lingual backgrounds to the speech intelligibility research. By considering the four factors of indoor soundscaping, particularly user profile and the function of the enclosure, to prepare the surveys and questionnaires, the effects may be clearly observed.

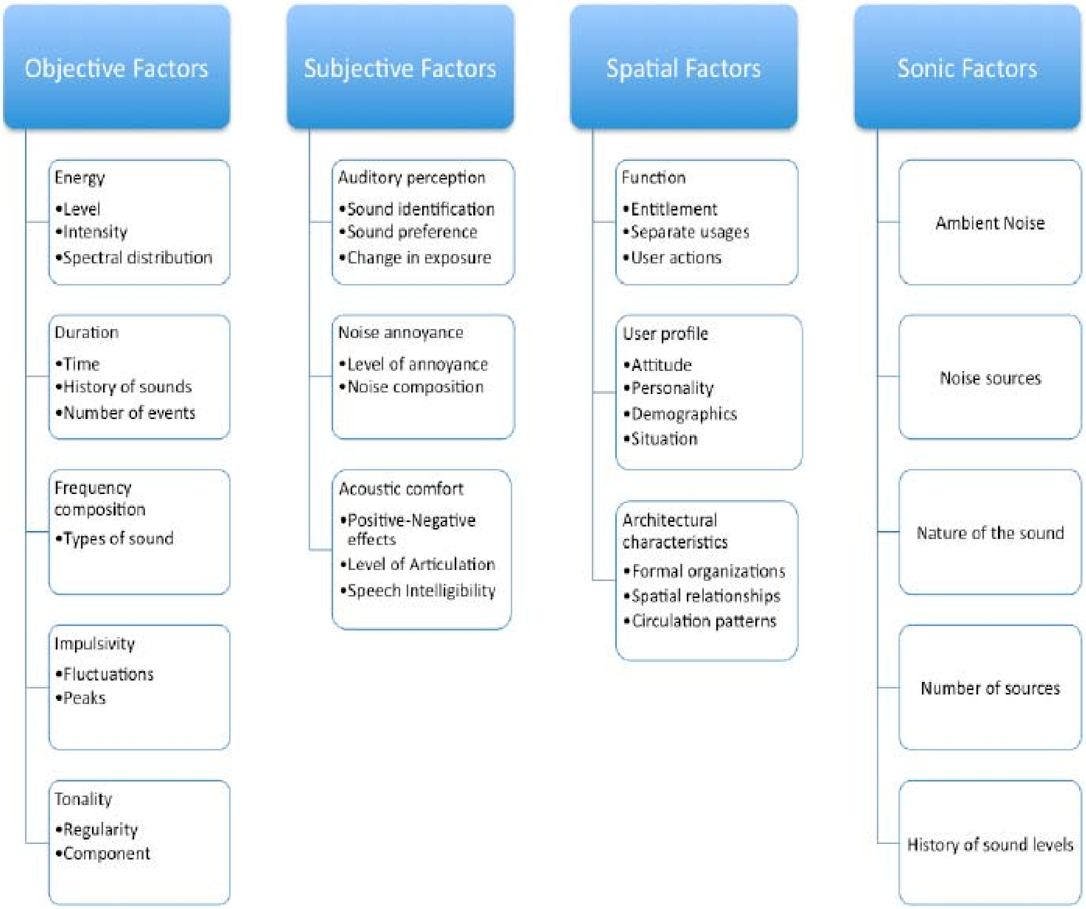


Figure 2.21 The four factors of the indoor soundscaping (Dokmeci and Kang, 2010).

2.6 Socio-linguistic Approach

Even if a talker and a listener are from the same community and have the same native language, their speech recognition abilities could differ because of socio-lingual effects. Two talkers cannot speak the same language, because their experiences with that language cannot be the same. Experience with a specific culture and a specific language modifies each individual's speaking habits and language (Hudson, 1996).

Each individual is unconsciously generating a mind map of their own community, and this mind map leads to numerous differences and similarities in-between them, in terms of various aspects of linguistics. Therefore, people who have a different socio-linguistic background will communicate in that language accordingly. This phenomenon also applies for listeners and for the activity of listening. Listeners' past experience and mind maps work as a filter for incoming speech and the intelligibility of that speech (Hudson, 1996).

To inspect possible socio-lingual effects on speech intelligibility, it is possible to include soundscape analysis to the study. By using the soundscape analysis, subjects' intelligibility response to a speech could be analysed under various sound environments that trigger cognitive filtering process.

2.7 Key Studies

This section examines some of the important studies carried out on room acoustics, soundscape, and socio-linguistics, that are related to the present research. In the literature of both research areas, multi-lingual speech intelligibility is analysed from different perspectives. For instance, recent studies have focused on the creation of articulation test word lists and rhyme tests using several languages (Alias and Trivino, 2007; Li *et al.*, 2000; Abushariah *et al.*, 2011) or comparing the results of speech intelligibility tests between two languages (Kang, 1998). Davies *et al.* (2009b) also examined speech intelligibility in a soundscape, and investigated perceptual factors affecting the intelligibility of speech. A discussion on how the soundscape approach and speech intelligibility are related was given in Section 2.5 (pages 38-39), where the complexity of soundscape work in relation to speech intelligibility was highlighted. Furthermore, it should be noted that the comparison of multiple languages increases this complexity by adding linguistic and socio-cultural factors to the wide range of factors affecting perception. In the present research, the three areas of speech intelligibility research (room acoustics, socio-linguistics and soundscape theory) will be combined in order to understand the acoustical and perceptual effects on speech intelligibility.

2.7.1 Room Acoustic Studies

One of the most relevant works on comparing the speech intelligibility between various languages focused on the differences between Chinese (i.e. Mandarin) and English (Kang, 1998). The main reason behind choosing Mandarin to be compared with English was that Chinese is a tonal language and English is a non-tonal language. It was hypothesised that the Articulation Index (AI) and the Speech Transmission Index (STI) would show significant differences between a tonal language and a non-tonal language. The research was conducted by undertaking experiments using two test material sources, which were, phonetically balanced (PB) word lists in Chinese and English, and public address (PA) messages from the Honk Kong Mass Transit Railway (MTR). It was emphasized that two different languages' PB word scores could not be compared, because of the several differences between word structures. Therefore, a new parameter called converted speech intelligibility was developed, which was the conversion of PB word intelligibility scores into sentence intelligibility scores (Kang, 1998). Word lists and PA messages were recorded on a tape with the help of 4 native speakers of each language. By using both speech materials and subjective ratings of speech intelligibility, four results were obtained: word intelligibility, converted sentence intelligibility, PA sentence intelligibility and subjective rating. For subjective rating of the acoustic conditions, listeners were required to assess the articulation test on a five-point scale from one to five (1-bad, 2-poor, 3-fair, 4-good, and 5-excellent). Experiments were conducted in two types of spaces, in order to include results from both diffuse and non-diffuse fields. The first space was a corridor simulating a train or metro station (non-diffuse field). The second space was a seminar room and had a simpler rectangular shape (diffuse field). In the corridor, three cases were tested. Case C1, one speaker and without noise; case C2, one speaker and with noise and case C3, two speakers and with noise. In the seminar room, two cases were tested. Case S1, noise source off and case S2, noise source on. It was found that the word intelligibility of Chinese is better than the word intelligibility of English in the high STI condition that corresponds to a high S/N ratio (Figure 2.22). The reason of this could be that some English consonants are not intelligible under reverberant conditions; however, tones in the Mandarin language are helpful for increasing the word intelligibility. Furthermore, the word intelligibility of English language was better than that of Mandarin in lower STI values caused by a decreased S/N ratio. It was hypothesized that because of the wider SPL range of the English language, some English words could be more intelligible by picking up only the high peaks (Kang, 1998).

It was also seen that the converted speech intelligibility of English is better than that of Mandarin at most STI conditions in the seminar room (Figure 2.23). It was hypothesised that the main reason behind this phenomenon could be the fact that, all kinds of Chinese words can be represented by Mandarin PB words; however, only relatively short English words can be represented by English PB words.

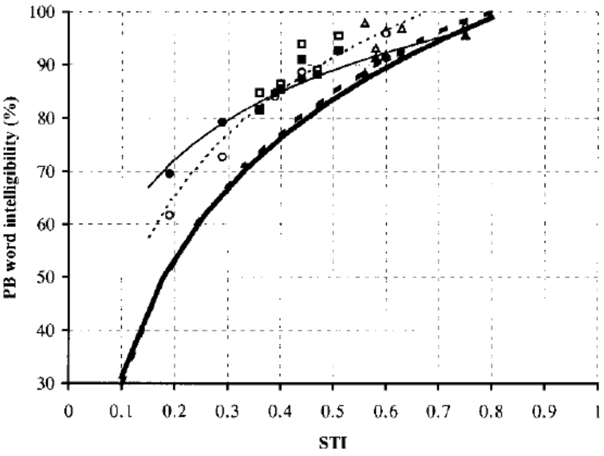


Figure 2.22 Comparison of the word intelligibility between Mandarin and English. ▲ and Δ, Case C1; ● and ○, Case C2; ■ and □, Case C3. English in the corridor: —; English in the seminar room: ▬; Mandarin in the corridor: ....; Mandarin in the seminar room: -------- (Kang, 1998).

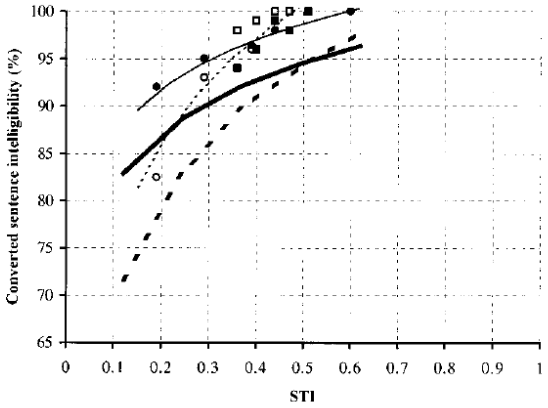


Figure 2.23 Comparison of the converted sentence intelligibility between Mandarin and English. ▲ and Δ, Case C1; ● and ○, Case C2; ■ and □, Case C3. English in the corridor: —; English in the seminar room: ▬; Mandarin in the corridor: ....; Mandarin in the seminar room: -------- (Kang, 1998).

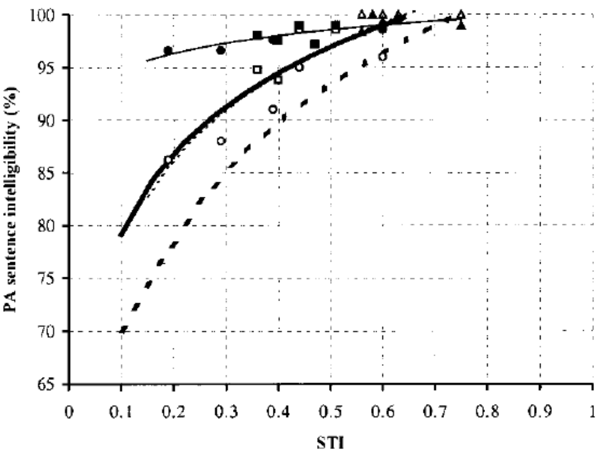


Figure 2.24 Comparison of the PA sentence intelligibility between Mandarin and English. ▲ and Δ, Case C1; ● and ○, Case C2; ■ and □, Case C3. English in the corridor: —; English in the seminar room: ▬; Mandarin in the corridor: ....; Mandarin in the seminar room: -------- (Kang, 1998).

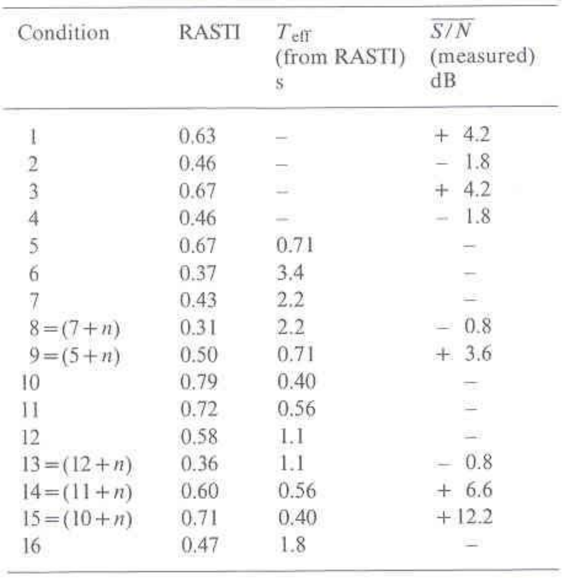
It was found that in the corridor (non-diffuse field) there are significant differences between English and Mandarin in terms of PA sentence intelligibility (Figure 2.24), especially in cases C2 and C3, in which the noise source is on. The main reason behind this difference could be that there are significantly less sounds in Mandarin than that in English. Both corridor and seminar room results pointed at better intelligibility of English across most room acoustic conditions.

In order to compare several languages' PB intelligibility scores, it is helpful to convert the PB word scores into sentence intelligibility scores by finding relationships between tested languages in terms of sentence and word intelligibility. It was also suggested to develop PB sentence lists as test materials (Kang, 1998). Converting PB word scores to compare two languages could be achieved by using the word and sentence structures; however, the conversion process for more than two languages is a challenging process. Furthermore, this study was focused on public announcement systems, and the methodology of the research should be modified in order to be used for face-to-face communication.

In order to discuss the speech intelligibility in every aspect, it is necessary to understand the communication channel between a talker and a listener. Additionally, it is also important to understand the talker and the listener. The spoken language used for communication has a crucial role on the intelligibility of speech in any acoustical condition. It is hypothesised that each language has its own properties, which can lead to variable rates of speech intelligibility in a fixed acoustic condition (Houtgast and Steeneken, 1984, Kang, 1998).

Only a few studies investigated the relation between various languages and the speech intelligibility. Houtgast and Steeneken (1984) carried out a study using 11 western languages (English, Finnish, French, German, Hungarian, Italian, Dutch, Maori, Polish, Swedish and Slovak) in 16 acoustic conditions. They collected recorded speech material from the selected laboratories of the 11 participating countries. These recordings consisted of 16 blocks with 10 minutes of speech each. One speech block was used per acoustic condition. Acoustic conditions were varied in terms of reverberation time and signal to noise ratio. The full details of acoustic conditions are presented in Figure 2.25. The main purpose of this study was to validate the rapid speech transmission index (RASTI), which is a simplified version of the speech transmission index (STI), by comparing the results obtained with the articulation index (AI). Differences between the test materials 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 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 of (or subtle differences among) the carrier phrases, and level mismatch between the tests. The research presented here focuses on language specific effects.

Peng (2011) investigated Mandarin speech intelligibility under different listening conditions (diotic and dichotic) by using the PB word test in computer simulated virtual rooms. The simulated rooms were 4 classrooms, 3 report halls, and a church. The reverberation time ranged from 0.36s to 5.16s in between these rooms. The results suggested that the STI method can evaluate Mandarin speech intelligibility without any algorithm modifications. Furthermore, the word intelligibility of Mandarin and English were also compared 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. In other studies (Peng, 2005; 2007; 2010) Peng also demonstrated the usefulness of auralization for conducting speech intelligibility tests.



Condition	RASTI	T_{eff} (from RASTI) s	$\overline{S/N}$ (measured) dB
1	0.63	–	+ 4.2
2	0.46	–	– 1.8
3	0.67	–	+ 4.2
4	0.46	–	– 1.8
5	0.67	0.71	–
6	0.37	3.4	–
7	0.43	2.2	–
$8=(7+n)$	0.31	2.2	– 0.8
$9=(5+n)$	0.50	0.71	+ 3.6
10	0.79	0.40	–
11	0.72	0.56	–
12	0.58	1.1	–
$13=(12+n)$	0.36	1.1	– 0.8
$14=(11+n)$	0.60	0.56	+ 6.6
$15=(10+n)$	0.71	0.40	+ 12.2
16	0.47	1.8	–

Figure 2.25 The detailed acoustic conditions information of Houtgast and Steeneken’s study (1984).

More recently, Zhu *et al.* (2014) conducted a similar research on Mandarin speech intelligibility. A total of 64 subjects were tested by using the Mandarin PB word lists in 4 different general rooms. The results suggested that when a spectrum of Chinese is adopted and the revised STI method is used, with similar STI values, the PB word test scores can vary. It was also 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.

Ji *et al.* (2014) 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 (Garcia Lecumberri *et al.*, 2006; Van Engen and Bradlow, 2007; Garcia Lecumberri *et al.*, 2010; Van Engen, 2010), main findings being that non-native speakers tend to perform lower under any type of masking condition (Garcia Lecumberri *et al.*, 2006; Garcia Lecumberri *et al.*, 2010) and that the linguistic content of background noise can also affect speech intelligibility (Van Engen and Bradlow, 2007; Van Engen, 2010).

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, and the present aims to develop this knowledge. The studies of Kang (1998), Peng (2011) and Zhu *et al.* (2014) 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 (Zhu *et al.*, 2014). It is important to state that the results of the listening tests were compared with the scores of different tests from previous studies; therefore, the comparisons should be considered with caution.

2.7.2 Soundscape Studies

Previous research pointed out that speech communication and intelligibility is strongly related to the soundscape theory (Davies *et al.*, 2009b). The area of interest on acoustics is shifting its perspective into human related observation, listening, feelings and other emotional and cultural criteria, while conserving and improving the transmission of sound energy within public spaces. While listening in a specific environment, the cognitive process affects the interpretation of the sound energy by the brain and this extra cognitive load may result in either reinforcing or reducing the intelligibility of speech. In that case, human speech communication contributes to the structure of the soundscape, which becomes a background setting for our conversations. It is hypothesised that this soundscape setting should be relaxing, comforting and positive in order to achieve intelligible speech (Davies et al., 2009b).

Davies *et al.* (2009b) compared the extended speech intelligibility index (ESII) with subjective ratings of speech intelligibility by using the soundscape methodology. Those subjective ratings were derived from two types of tests. The first test was a direct intelligibility test that used carrier sentences of meaningful words, and the listeners were assigned to write down the key word of each sentence. The second test was a questionnaire where listeners were required to assess the presented sound environments in terms of quality and clarity on a five-point scale. Speech samples were gathered from Speech Perception in Noise (SPIN) sample sets, which is a standardized word list used for speech intelligibility tests. Noise recordings were taken from St. Ann's Square in Manchester, and included traffic, water, walking, and tyre noises. Two soundscape recordings were taken. The first recording (soundscape 1) did not include traffic and other kinds of impact sounds, however, the second recording (soundscape 2) included those sounds. The final recording was a mix of speech samples and either soundscape 1 or soundscape 2 recording that was chosen randomly. It was found that the predicted speech intelligibility corresponded to measured speech intelligibility for soundscape 1, however the impact sounds included in soundscape 2 had a distracting extra cognitive load, and the prediction performance decreased.

As a second experiment, clarity and quality ratings were compared with the speech intelligibility index (SII) ratings. It was found that the SII cannot predict either clarity or quality, because the physical characteristics of the sound are not the only effect on intelligibility of speech. The cognitive process and cognitive features of recorded sounds also had an extensive effect (Davies *et al.*, 2009b). Speech intelligibility tests, either word or sentence intelligibility scores or objective measurement methods, were affected by the quality and characteristics of background noise. Therefore, while conducting research on speech intelligibility using the soundscape technique, it is important to control the types of sounds involved in the recording to prevent extra cognitive load, and misleading results. The background noise could be assessed by using the semantic differential analysis method; therefore further information on both separate noise sources' effects and social/psychological/linguistic effects on the intelligibility of speech could be observed. The methodology used in this research could be modified in order to be used for indoor soundscaping.

Raimbault and Dubois (2005) compared viewpoints of urban planners and city users in their study, to find out the effects of social and psychological phenomena on the soundscape theory. It was claimed that, decreasing or removing noise levels is not enough to achieve a good soundscape quality, an approach that urban planners and noise pollution protocols are nowadays concentrating on. Some low-level background noises might be higher than more pleasant sounds, for instance in a case that foot-step noises are dominant over the sound of birds on a tree. However, in that case, foot-step sounds may trigger a feeling of pleasantness of a pedestrian area. Therefore, focusing solely on sound pressure level is not sufficient to assess a soundscape. A questionnaire was prepared to evaluate the opinions of ten city planners. The results of linguistic analysis showed that, urban planners are focused on sound pressure levels, rather than social and emotional aspects. It was suggested to modify and improve the design process and noise pollution protocols by including a humane perspective of perception of the sound (Raimbault and Dubois, 2005).

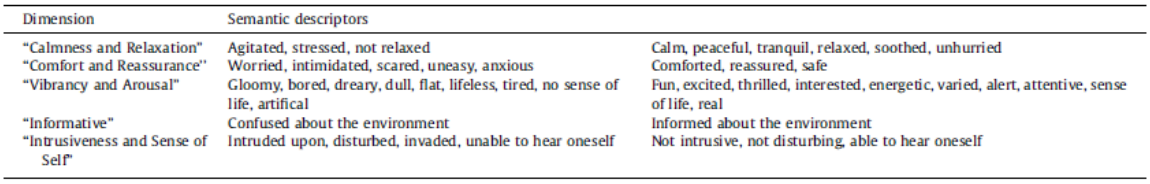
Raimbault (2006) conducted research on comparing eight locations' soundscapes in two French cities (Lyon and Nantes), by using both verbal comments on open-ended questions and the semantic differential analysis. These eight locations varied in terms of activities, characteristics and functions. Open-ended questions were focused on spatial attributes such as place and location. The semantic differential analysis focused on psycho-acoustical descriptors rather than emotional descriptors. The descriptors used were loudness (quite/loud), spatial attributes (near/far, organised/disorganised, and very present/not present), temporal attributes (steady/unsteady, and established/changing), intelligibility (unclear/distinct), activity features (monotonous/varied), and appraisal (pleasant/unpleasant). A total of 296 subjects evaluated the soundscapes. A psycholinguistic analysis was applied to the verbal comments, and the results identified three types of judgements: acceptable, usual, and resigned. The verbal comments' results showed that the difference among the soundscape quality of eight areas was mainly dominated by noise sources, especially traffic noise. The semantic differential analysis results did not show consistency among subjects, mainly because of the interpretation differences of selected descriptors between the subjects. It was argued that variance between holistic listening and descriptive listening might have caused the inconsistency, which is related to the listeners' experience, attitudes and expectations (Raimbault, 2006). These effects should be included in further studies, in order to identify the correlation between personal attributes and the soundscape phenomenon. In more controlled conditions, listeners could be leaded to holistic or descriptive listening, according to the need of further analysis. In the present research, the personal experience and cultural variety is expected to be closely related with the speech intelligibility of multi-lingual spaces. Therefore, it is also an important task for this research to choose the appropriate descriptors and listening modes.

Nilsson and Berglund (2006) developed another tool to assess soundscape quality in urban residential areas, including indoor soundscape quality. Four regions in Sweden were analysed, where three of them had high traffic noise. Soundscape walks were held for six conditions in each area. The conditions were indoors in a room with a closed window facing to the high traffic noise façade, the same room with an open window, indoors in a room with closed window facing to the less noisy yard, the same room with an open window, outdoor that is close to a less noisy yard, and outdoor that is close to a high traffic noise street. A total of 106 participants evaluated these areas by using twelve attributes on a 0-100 scale. The attributes were soothing, pleasant, light, dull, eventful, exciting, stressful, hard, intrusive, annoying, noisy, and loud. All of the attributes except loud and noisy were emotional descriptors. Every participant was exposed to each of the sound environments for 30 seconds, and a sound recording were taken at the same time. The results were analysed by using a three component Pearson's Correlation Analysis (PCA). These showed that there is a high correlation between noise levels and evaluation attributes, such as loud and dull. It was claimed that an indoor environment with a closed window, which faces a high traffic area, is less pleasing than an outdoor low-noise environment. The reasons behind this difference should be analysed in future studies, in order to further understand the importance of emotional or behavioural effects. This study is one of the few studies that focused on a comparison between outdoor and indoor soundscaping.

Cain *et al.* (2011) carried out research that analysed the emotional dimensions of a soundscape. In the literature, soundscape was defined by three or four dimensions, such as loud, harshness, clarity and spatiality. It was hypothesized that limiting the dimensions to two is more practical to describe and analyse a soundscape. Furthermore, describing a soundscape in a two dimensional (2-D) space is more useful for designers who could read the 2-D data analysis of a soundscape study, and improve their design accordingly. To select the starting set of descriptors, Cain *et al.* (2011) used previous data which were responses from questionnaires, and the answers were coded and clustered into several categories. Descriptors that defined the feelings of an individual about the space were included in the shortlist. After preparing the descriptors, two experiments were conducted. In the first experiment, urban soundscape recordings such as streets with busy traffic, urban parks with wildlife and café atmospheres were used.

In the second experiment, soundscape recordings from the University of Warwick campus were used. The contents of campus soundscape recordings were similar to the first experiment. Participants evaluated the recordings on an eight point scale, using the descriptors shown in Figure 2.26. By applying a Principal Component Analysis to the results of both experiments, two major dimensions were selected as emotional descriptors, which were calmness and vibrancy.

In the second phase of the research, a 2-D perceptual space of calmness and vibrancy were used to describe thirteen different soundscapes. Participants used the same semantic descriptors of the first phase of the study to evaluate the soundscapes, and then the results were interpreted by the 2-D perceptual space of calmness and vibrancy. It was pointed out that the results could be used to show how various soundscapes can be positioned in a 2-D perceptual space, to explore the correlation between design decisions and perception of a soundscape and to set targets for soundscape design. Figure 2.27 shows how it can be used for target-setting in soundscape design. The



Dimension	Semantic descriptors	
"Calmness and Relaxation"	Agitated, stressed, not relaxed	Calm, peaceful, tranquil, relaxed, soothed, unhurried
"Comfort and Reassurance"	Worried, intimidated, scared, uneasy, anxious	Comforted, reassured, safe
"Vibrancy and Arousal"	Gloomy, bored, dreary, dull, flat, lifeless, tired, no sense of life, artifical	Fun, excited, thrilled, interested, energetic, varied, alert, attentive, sense of life, real
"Informative"	Confused about the environment	Informed about the environment
"Intrusiveness and Sense of Self"	Intruded upon, disturbed, invaded, unable to hear oneself	Not intrusive, not disturbing, able to hear oneself

Figure 2.26 Starting set of dimensions (Cain *et al.*, 2011).

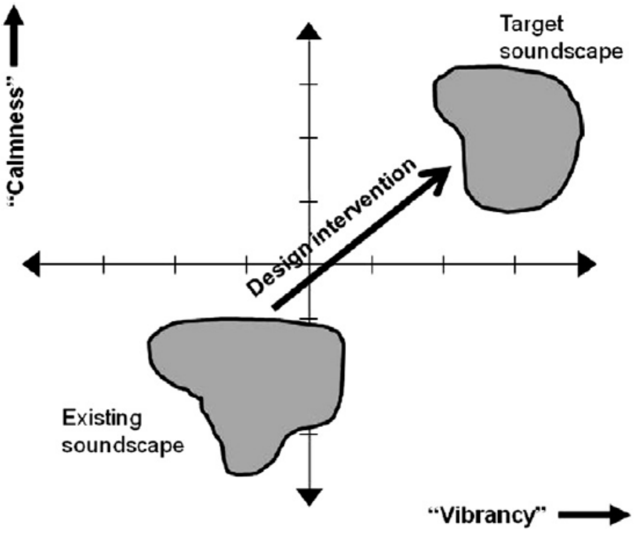


Figure 2.27 An example 2-D perceptual space of Calmness and Vibrancy used for target setting in soundscape design (Cain *et al.*, 2011).

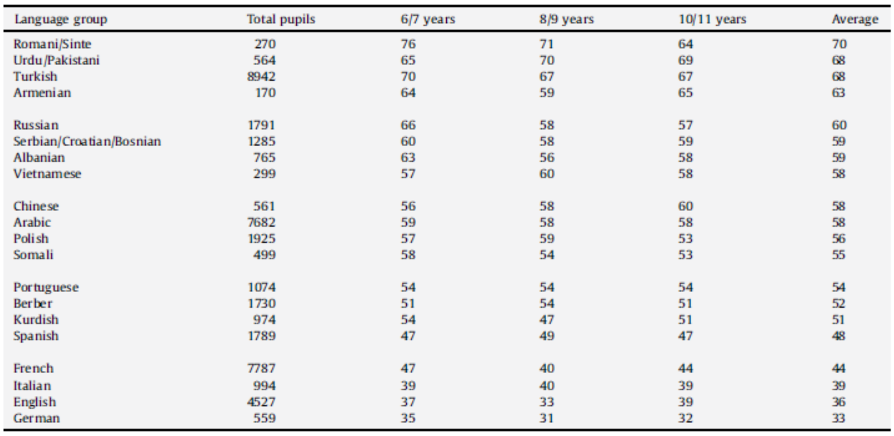
findings of the research show that a soundscape can be evaluated by using two emotional dimensions and the results could be used for designing better soundscapes. It is also important that the connection between multilingual communications in an environment is related with demographic issues and emotions, and revealing this connection is suggested for further research in this area.

2.7.3 Socio-linguistic Studies

Although there are many recent studies on socio-linguistics and multilingual communication, the number of studies relevant to the present research is limited. Most of these studies evolve around topics such as health issues, communication disorders, visual communication, information technologies, and linguistic landscapes. Oral multilingual communication studies are mostly theory based, and combined with globalisation, politics and economy.

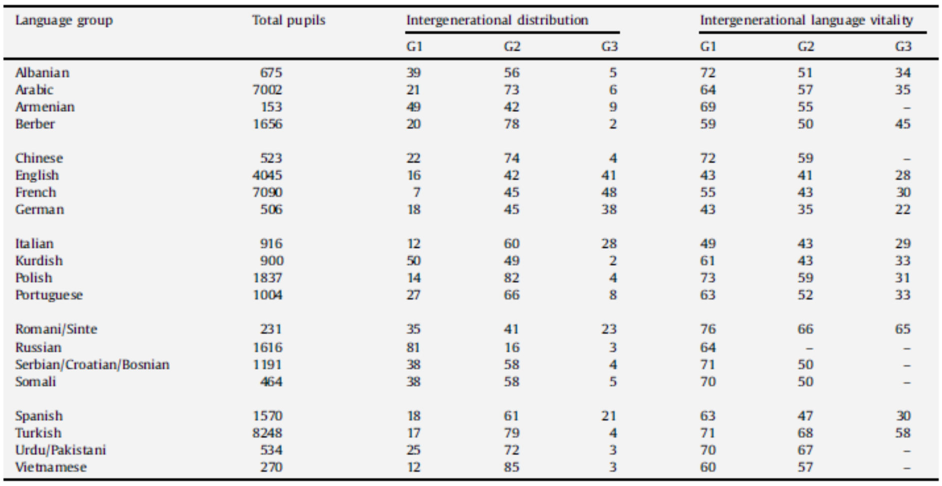
One study examined urban multilingualism in Europe (Extra and Yağmur, 2011). The article expresses the cultural and linguistic diversity of Europe, and establishes an extensive research by investigating six multicultural European nation-states, such as Goteborg, Hamburg, The Hague, Brussels, Lyon and Madrid. The study was named the Multilingual Cities Project, and the participant cities were selected according to the variety of immigrant minority groups and their multicultural environment. The focal point of the project was the minority language / home language distribution of the selected European cities. The collected data focused on four perspectives that were the demographic perspective, the sociolinguistic perspective, the educational perspective, and the economic perspective. The data collection process was held by using questionnaires and the participants were selected among multicultural school populations. The number of participants was 160.000 and the language vitality parameter (LVI) was calculated for each participant. The LVI is a combination of four scores. The first score is the language proficiency, which is the extent to which the minority language under consideration is understood. The second score is the language choice, which is the preferred language to communicate with the mother at home. The third score is the language dominance, which is the minority language that is spoken best. The last score is the language preference, which is the minority language that is preferred for oral communication. The LVI is the mean value of these four scores. The results were presented in two tables. The first table shows the language vitality per language and age group (Figure 2.28). The second table shows the intergenerational distribution and intergenerational LVI (Figure 2.29) (Extra and Yağmur, 2011).

The results of the Multilingual Cities Project revealed the distribution and language vitality of immigrant minority languages at home across European cities. It found that an increasing number of children are using more than one language. Between one third and more than half of the participant children responded that they are using other languages than the mainstream language at home (Extra and Yağmur, 2011).



Language group	Total pupils	6/7 years	8/9 years	10/11 years	Average
Romani/Sinte	270	76	71	64	70
Urdu/Pakistani	564	65	70	69	68
Turkish	8942	70	67	67	68
Armenian	170	64	59	65	63
Russian	1791	66	58	57	60
Serbian/Croatian/Bosnian	1285	60	58	59	59
Albanian	765	63	56	58	59
Vietnamese	299	57	60	58	58
Chinese	561	56	58	60	58
Arabic	7682	59	58	58	58
Polish	1925	57	59	53	56
Somali	499	58	54	53	55
Portuguese	1074	54	54	54	54
Berber	1730	51	54	51	52
Kurdish	974	54	47	51	51
Spanish	1789	47	49	47	48
French	7787	47	40	44	44
Italian	994	39	40	39	39
English	4527	37	33	39	36
German	559	35	31	32	33

Figure 2.28 Language vitality (LVI) (%) per language group and age group (Extra and Yağmur, 2011)



Language group	Total pupils	Intergenerational distribution			Intergenerational language vitality		
		G1	G2	G3	G1	G2	G3
Albanian	675	39	56	5	72	51	34
Arabic	7002	21	73	6	64	57	35
Armenian	153	49	42	9	69	55	–
Berber	1656	20	78	2	59	50	45
Chinese	523	22	74	4	72	59	–
English	4045	16	42	41	43	41	28
French	7090	7	45	48	55	43	30
German	506	18	45	38	43	35	22
Italian	916	12	60	28	49	43	29
Kurdish	900	50	49	2	61	43	33
Polish	1837	14	82	4	73	59	31
Portuguese	1004	27	66	8	63	52	33
Romani/Sinte	231	35	41	23	76	66	65
Russian	1616	81	16	3	64	–	–
Serbian/Croatian/Bosnian	1191	38	58	4	71	50	–
Somali	464	38	58	5	70	50	–
Spanish	1570	18	61	21	63	47	30
Turkish	8248	17	79	4	71	68	58
Urdu/Pakistani	534	25	72	3	70	67	–
Vietnamese	270	12	85	3	60	57	–

Figure 2.29 Intergenerational distribution (%) and intergenerational language vitality (%) per language group. G1: pupil + father + mother born abroad; G2: pupil born in country of residence, father and/or mother born abroad; G3: pupil + father + mother born in country of residence (Extra and Yağmur, 2011).

Wodak *et al.* (2012) conducted a study on language choice and code-switching in European Union (EU) institutions. The institutions were the European Parliament (EP) and the European Commission (CEC). It was claimed that these two institutions are representing the European population, therefore reflecting the same multilingual characteristics. The study hypothesised that various contextual settings and different language ideologies affects multilingual communication. In order to collect the experimental data, different levels of multilingual communication, such as semi-official communication and internal everyday communication were observed and recorded. As a result, it was understood that various multilingualisms are being simultaneously performed in the investigated EU contexts. The topic of the meetings, interactional dynamics and language regulations of the particular organisations play an important role on power. Also, it was claimed that code-switching depends on technical jargon, the language of the preceding speaker, politeness phenomena, language-ideology, power-related factors, and personality/relationship related factors (Wodak *et al.*, 2012).

The two studies reviewed highlight the connections between multilingual communication and social/cultural/political settings, and exemplify two different kinds of methodologies of sociolinguistic studies. In particular, the study of Wodak *et al.* (2012) demonstrates the multilingual and multicultural setting in an enclosure, and is therefore a relevant study in relation to the present work.

2.8 Discussion

This chapter provided the background information required to carry out the work presented in this thesis. Definitions of language and sociolinguistics were initially given, followed by room acoustic concepts, speech intelligibility, assessment of speech intelligibility, and the factors affecting speech intelligibility such as socio-lingual factors and soundscape theory. Finally, the previous literature was critically analysed within the areas of room acoustics, soundscape research and socio-linguistics.

Both objective and subjective assessment methods of speech intelligibility have been reviewed. For objective measurements, details about the speech transmission index (STI) were given, as this is the parameter used in the present research to define different room acoustic conditions (and is the parameter commonly used in previous work (e.g. Houtgast and Steeneken (1984), Kang (1998), Zhu *et al.* (2014))). Subjective tests showed how different test materials can easily be responsible for significant variations within a language (e.g. nonsensical vs. meaningful words), highlighting the importance of using comparable test materials when comparing intelligibility across languages. In that respect, the Diagnostic Rhyme Test (DRT) has been identified as being suitable for the current work. The decision of using the DRT followed guidance given in the standard ANSI/ASA S3.2 (2009): the DRT specifically allows examining distinctive features of speech through the discrimination of phonemes, and comparing those features across languages (unlike other tests). Furthermore, this is one of the few tests for which materials have been developed for several languages using consistent procedures (pairs of words based on a Consonant-Vowel-Consonant (CVC) sequence, with words varying in the first consonant only) (ANSI/ASA S3.2, 2009) and the test is known to give stable intelligibility scores (Voiers, 1977). The DRT method is universal and can be applied to any language, and the concept of distinctive features was indeed developed through the analysis of multiple languages (Jakobson, 1952). Unfortunately, DRT lists were not available for one of the languages examined (Polish), and PB word lists were used instead for that case. Furthermore, sentence intelligibility comparisons have also been made in the current work. Further details about intelligibility tests used in the current work are described in Chapter 3.

Overall, the review of previous work showed that the number of room acoustic studies that investigated the relationship between languages and speech intelligibility is quite limited, most comparisons having been made between English and Mandarin. In particular, the studies of Kang (1998), Peng (2011) and Zhu *et al.* (2014) indicated that English tends to be slightly more intelligible than Mandarin under most room acoustic conditions, although some contradictions have been observed between the findings of these studies, especially either very poor or very good room acoustic conditions. These contradictions have been mainly attributed to the use of different materials (Zhu *et al.*, 2014). In general, it can be said that 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.

Soundscape studies were also reviewed, as the soundscape approach was identified as an appropriate method for perceptual evaluation. The latter might in fact be of particular importance when different language groups are considered, as differences can be expected between objective evaluations of speech intelligibility and subjective evaluations of different language groups. Semantic attributes are normally used in soundscape work, and can allow identifying qualitative aspects important to communication. For example, Davies *et al.* (2009b) showed that certain sounds can be more distractive (e.g. impact sounds) and this can in turn affect speech intelligibility due to additional cognitive load of those distractive sounds. The fact that decreasing noise levels does not necessarily guarantee achieving a good soundscape quality was pointed out by Raimbault and Dubois (2005), and the importance of using emotional descriptors was also highlighted in a number of studies (Raimbault, 2006; Nilsson and Berglund, 2006; Cain *et al.*, 2011). Overall, the review of such studies indicated that soundscape analysis can provide a further insight into the speech intelligibility of multilingual spaces.

Finally, a review of socio-linguistic studies on the topic of multilingualism highlighted the very different perspective used within such studies, compared to room acoustic studies dealing with speech intelligibility. Socio-linguistic research tends to be more concerned with topics such as globalisation, politics and economy and does not quantify speech intelligibility. Within that context, the soundscape approach can be seen as a technique bridging socio-linguistic research and traditional room acoustic research.

2.9 Conclusions

The present study aims to contribute to the literature by combining several perspectives of speech intelligibility that were discussed in this chapter. Overall, the review of previous work showed 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. Furthermore, the literature review showed that there is limited knowledge about the perception of speech intelligibility and the factors affecting it. The research presented aims to combine both objective and perceptual aspects and widen the range of languages to be compared by investigating the effects of room acoustic properties and soundscapes on the intelligibility of four different languages (English, Polish, Arabic, and Mandarin). The effects of room acoustic parameters are investigated by conducting objective listening tests (DRT and PB word tests, and PB sentence tests) (Chapter 4) and soundscape analysis is conducted using subjective listening tests and performing semantic differential analysis (Chapter 5). Ultimately, the results of the both experimental phases are compared with the standards and design guidelines available on the intelligibility of speech (Chapter 6).

CHAPTER 3

Methodology

3.1 Introduction

The study consists of two experimental phases. The first phase, ‘room acoustics and speech intelligibility of multiple languages’, focuses on the effects of the room acoustic properties (reverberation time and signal-to-noise ratio) on the intelligibility of English, Polish, Arabic, and Mandarin. The second phase, 'assessment of soundscape perception on speech intelligibility of multiple languages', investigates the relationship between soundscape perception and intelligibility of the four languages, by using the soundscape methodology. This chapter presents the methodology of both phases of the study. Initially, the selection process of the languages is described. For the first phase, a description is given on the word and sentence lists that were used, the recording and post processing of these word and sentence lists, together with details on the laboratory and the equipment used, as well as the listening test procedure. For the second phase, the following are presented: selection process of the cases, preparation of the sentence lists, recording and post processing of the sentence lists and the background noise samples, preparation of the visual materials, details on the laboratory and the equipment, and information on semantic differential analysis is presented. For both phases of the study, the statistical analysis methods are also described.

The results of the first phase of the study are presented and discussed in Chapter 4, and the results of the second phase of the study are presented and discussed in Chapter 5.

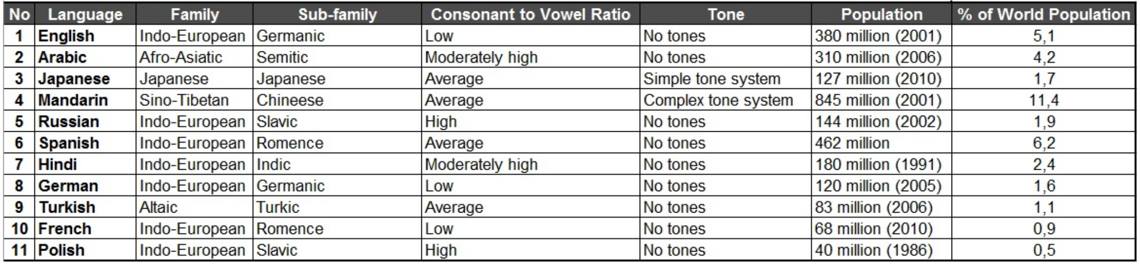
3.2 Selecting the languages (English, Polish, Arabic, and Mandarin)

This section explains the selection process of the languages that were used in the study. The study was carried out using four sample groups, in which the native language of each sample group was the variable. Languages representative of a wide range of linguistic properties were planned to be 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).

The selection process of languages had various criteria. First of all, it was decided that the selected languages should be representative of a multilingual environment in a western city and should cover a large percentage of the world population, as the current research is applied rather than theoretical. Table 3.1 shows that the total percentage of the world population covered is 21.2% (English=5.1%, Polish=0.5%, Arabic=4.2%, and Mandarin=11.4%). Secondly, 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 (Peutz, 1971), 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 (Maddieson, 2008). 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 (Maddieson, 2008). 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 2 and 2.75), average (between 2.75 and 4.5), moderately high (between 4.5 and 6.5), and high (larger than or equal to 6.5) consonant-to-vowel ratio (Maddieson, 2008).

Tonality was identified as a linguistic factor that can clearly differentiate languages (Maddieson, 2013), which is why at least one tonal language had to be selected. Tone is the change of the meaning of a word by the change of pitch, and in that respect languages can be subdivided into three categories: no tones, simple tonal system, and complex tonal system (Maddieson, 2013). Languages with a simple tonal system utilise only two-way contrast in terms of tones (i.e. high pitch - low pitch), but languages with a complex tonal system, such as Mandarin, can also use an ascending or descending pitch. 307 out of 527 languages utilise no tones, whilst 132 have a simple tonal system and 88 have a complex tonal system (Maddieson, 2013). To examine the effects of the tonal system of a language on the speech intelligibility, at least one tonal language had to be selected.

Table 3.1 The comparison table of the common languages of the world.



No	Language	Family	Sub-family	Consonant to Vowel Ratio	Tone	Population	% of World Population
1	**English**	Indo-European	Germanic	Low	No tones	380 million (2001)	5,1
2	**Arabic**	Afro-Asiatic	Semitic	Moderately high	No tones	310 million (2006)	4,2
3	**Japanese**	Japanese	Japanese	Average	Simple tone system	127 million (2010)	1,7
4	**Mandarin**	Sino-Tibetan	Chineese	Average	Complex tone system	845 million (2001)	11,4
5	**Russian**	Indo-European	Slavic	High	No tones	144 million (2002)	1,9
6	**Spanish**	Indo-European	Romence	Average	No tones	462 million	6,2
7	**Hindi**	Indo-European	Indic	Moderately high	No tones	180 million (1991)	2,4
8	**German**	Indo-European	Germanic	Low	No tones	120 million (2005)	1,6
9	**Turkish**	Altaic	Turkic	Average	No tones	83 million (2006)	1,1
10	**French**	Indo-European	Romence	Low	No tones	68 million (2010)	0,9
11	**Polish**	Indo-European	Slavic	High	No tones	40 million (1986)	0,5

The native speakers' population of each language also had to be taken into account. The research should in fact be representative of a wide range of people; therefore, the languages with higher native speaker populations were selected. The availability of native speakers for the selected languages was also considered, and the languages selected had to comply with high number of participants that could be found at Heriot-Watt University. A comparison table was prepared to identify the differences between some of the major languages used in the world (Table 3.1). The linguistic properties of languages, such as consonant-to-vowel ratio, tone and fixed stress locations, and the population of native speakers of the languages are presented in the comparison table.

Based on the above mentioned criteria of consonant-to-vowel ratio, tonal properties, native speaker population, and availability of subjects, four languages were selected. These were English (low consonant-to-vowel ratio, wide-spread usage around the world), Mandarin (complex toned system, high native speaker population), Arabic (moderately high consonant-to-vowel ratio, high native speaker population), and Polish (high consonant-to-vowel ratio, and availability of speakers).

3.3 First phase – Room acoustics and speech intelligibility of multiple languages

The aim of the first phase of the study was to find out possible relations between speech intelligibility and multi-lingual communication, in terms of acoustics and linguistic properties of the four languages selected (English, Polish, Arabic, and Mandarin).

These four languages have been tested under four room acoustic conditions (STI = 0.2, 0.4, 0.6 and 0.8), varying in terms of reverberation time and signal-to-noise ratio (S/N). To measure speech intelligibility, diagnostic rhyme tests (DRT), phonemically balanced word tests (PB word), and phonemically balanced sentence tests (PB sentence) have been used. The standard ANSI/ASA S3.2 (2009) suggests that the minimum number of listeners should be 5. In the current study, in order to achieve equal gender representation, 3 male and 3 female listeners were selected from native speakers of each language. Attention was given to select the listeners from the same regions/countries of the speakers. The listening tests were conducted in one of the acoustic chambers of the Heriot-Watt University (described in Section 3.3.3). The results of the speech transmission index (STI, a physical measure of speech intelligibility (described in Section 2.4.1.2)) measurements and word/sentence intelligibility test scores were then compared in order to find out the relations between the room acoustic properties and the linguistic properties of each language.

The methodology used in the first phase of the study is described in detail in the following sections. The results of the listening tests, statistical analysis of the results and discussions are presented in Chapter 4.

3.3.1 Preparing the word and sentence lists

In this section, the word and sentence tests that were used to assess speech intelligibility in the first phase of the study are described. Three types of tests were used. The first word test was the Diagnostic Rhyme Test (DRT) and the second word test was the Phonemically Balanced word test (PB word). Additionally, Phonemically Balanced sentence lists (PB sentence) were used in order to compare the word and sentence test scores, and to identify any possible effects of context on speech intelligibility.

3.3.1.1 Diagnostic Rhyme Test (DRT)

The DRT is a listening test consisting of 192 words arranged in 96 pairs (ANSI/ASA.S3.2, 2009). The words are common, monosyllabic words, and most of them have three sounds ordered in a consonant-vowel-consonant (CVC) sequence. The word pairs differ only in their initial consonants, therefore discrimination of each distinctive feature of a given language can be analysed in relation to the room acoustic properties. The DRT is a closed-set type of test, which requires participants to select the

Table 3.2 Example lists of DRT.

(a) An example list of the English DRT (Voiers, 1977).

Voicing		**Nasality**		**Sustention**	
Veal	Feel	Meat	Beat	Fence	Pence
Bean	Peen	Need	Deed	Sheet	Cheat
Gin	Chin	Mitt	Bit	Foo	Pooh
Sibilation		**Graveness**		**Compactness**	
Zee	Thee	Weed	Reed	Gill	Dill
Cheep	Keep	Peak	Teak	Key	Tea
Jilt	Gilt	Bid	Did	Hit	Fit

(b) An example list of the Arabic DRT (Boudraa *et al*, 2008).

Tenseness		**Nasality**		**Flatness**	
Sir	Zir	Lad	Nad	Taf	T’af
Kil	Qil	Lab	Nab	Tab	T’ab
Ar	Bar	Lud	Nud	Dur	D’ur
Mellowness		**Graveness**		**Compactness**	
Dam	Zam	Fil	Sil	Bal	Qal
Hil	Fil	Mil	Nil	Bud	Qud
Hud	Xud	Bar	Dar	Ful	Xul

(c) An example list of the Mandarin DRT (Fu *et al*., 2011).

Airflow		**Nasality**		**Sustention**	
Cang1	Zang1	Man3	Ban3	Fan2	Pan2
Chen4	Zhen4	Man4	Ban4	Fang2	Pang2
Cheng1	Zheng1	Mang3	Bang3	Len2	Pen2
Sibilation		**Graveness**		**Compactness**	
Can2	Chan2	Ban1	Dan1	Gang3	Dang3
Can3	Chan3	Bang1	Dang1	Gong3	Dong3
Cong2	Chong2	Bang4	Dang4	Guan3	Duan3

correct word out of the provided pair of words. No carrier sentence is needed to conduct the DRT test (ANSI/ASA.S3.2, 2009).

Jakobson et al. (1952) suggested that the sounds of languages can be identified by using a set of distinctive features, which does not exceed twelve distinctive features. The test focuses on consonants in order to compare distinctive features between two sounds. The English DRT test focuses on voicing, nasality, sustention, sibilation, graveness, and compactness (ANSI/ASA.S3.2, 2009); the Arabic DRT test focuses on tenseness, nasality, mellowness, flatness, graveness, and compactness (Boudraa *et al*., 2008), and the Mandarin DRT test focuses on airflow, nasality, sustention, sibilation, graveness, and compactness (Fu *et al*., 2011). Full lists and detailed information on distinctive features are given in Chapter 4, Section 4.2.2. An example set of DRT material for each language is presented in Table 3.2 (a), Table 3.2 (b), and Table 3.2 (c). All the DRT material is available in Appendix A.

The lack of DRT material in Polish is a limitation of the current study, although comparisons between DRT and PB English words data (the former being taken from the current study and the latter from (Anderson and Kalb, 1987)) indicate that the variability between DRT and PB scores tends to be fairly small (Figure 3.1), suggesting that comparisons between DRT and PB results are acceptable (as also pointed out in ANSI/ASA S3.2-2009 (ANSI/ASA.S3.2, 2009)). Figure 3.1 shows that DRT and PB scores of English have an average difference (calculated from absolute values) of 2.4% across the four STI conditions considered, with a maximum difference of 5.5% observed at STI = 0.4. This is well below the large differences observed between languages that are presented in section 3 (which are as high as 33% at STI = 0.4), indicating that these inaccuracies are not expected to have affected the main findings obtained when comparing Polish PB word scores to DRT scores of the other languages. It is however accepted that some inaccuracies should be expected and are unfortunately not quantifiable for Polish, and that the variations between DRT and PB word scores of Polish could be higher than those presented for English in Figure 3.1. The data taken from (Anderson and Kalb, 1987) was based on the standard Harvard PB word test, which is commonly used in the United States. It should also be noted that comparability of DRT and PB scores can be achieved only by removing the effect of guesswork in the calculation of DRT scores (see equation (4.1)), as rhyme tests are closed tests that are otherwise expected to provide higher scores (Peng, 2011).

Table 3.3 An example list of the Polish PB words (Ozimek *et al.*, 2007).

Dres	Jas	Jacht	Mech	Biust	Tak
Glos	Wyz	Pyl	Plaszcz	Won	Krzak
Zal	Kant	Krem	Wodz	Klops	Bron

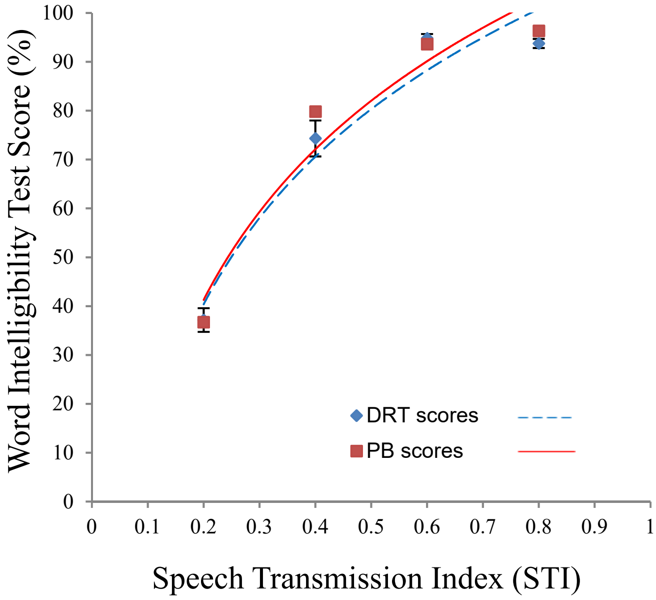


Figure 3.1 DRT (present study) and PB (Anderson and Kalb, 1987) word intelligibility scores of English obtained under different STI conditions (data markers, standard errors of the means (for DRT data only) and logarithmic regression lines).

3.3.1.2 Phonemically balanced word test (PB word)

A listener of a spoken language has the knowledge of linguistic information, such as, phonological, lexical, syntactic, and semantic properties of that language (Kalikow et al., 1977). Improved intelligibility of such linguistic properties provide a context, eventually decreasing the dependency on the perfectly transmitted acoustical signal. Therefore, a speech intelligibility test that demonstrates an everyday speech must assess both the acoustic-phonetic and the linguistic-structural components of the process (Kalikow et al., 1977). In order to achieve this tests, similar to the phonemically balanced word tests (PB word) were created in order to represent linguistic properties of languages.

PB word lists represent a language by having approximately the same phonetic properties and sound types of that language. It should be noted that in order to represent a specific language, all of the word lists must be phonemically balanced. Therefore, the DRT is a phonemically balanced test as well. The difference between the DRT and PB word tests is that the former focuses on the differentiation of initial consonants, and the latter focuses on the intelligibility of the whole word (ANSI/ASA.S3.2, 2009). The Polish PB word lists used in the current study consists of 4 sets of 48 words, with one set used for each acoustic condition in the current study. An example set of words is presented in Table 3.3. All the Polish PB words are available in Appendix B.

3.3.1.3 Phonemically balanced sentence test (PB sentence)

In the current study, the effects of context on speech intelligibility were also investigated. Typically, intelligibility of a given word increases when it is presented in a sentence, compared to an isolated word, when context of a sentence is adequately transferred to the listener (Miller et al., 1951). The context of a sentence is dependent on the number of possible alternative words that are available for a given sentence in that particular word location. More alternatives lead to a less predictable sentence, and less alternatives lead to a higher degree of predictability (Kalikow *et al.*, 1977). It should be noted that levels of predictability were not taken into account for the statistical analysis of the research presented, due to the fact that the PB sentence lists were not designed accordingly (except the English PB sentence test) and the scores were calculated by counting the number of correct words in the whole sentence, rather than evaluating only the last word. The PB sentence lists that were used in the current study were composed of everyday sentences, therefore representing a wide range of predictability. The English PB sentence pool consisted of 6 high predictability (HP) and 4 low predictability (LP) sentences. Polish, Arabic, and Mandarin sentences pools consisted of sentences that represent an everyday conversation. The number of words in each sentence varied with the language. In the English PB sentence list, there were a minimum of 6 and a maximum of 7 words in each sentence. In the Arabic PB sentences the minimum number of words was 3 and the maximum was 6. In the Mandarin list, all of the sentences had 7 words. Lastly, in the Polish PB list the number of words was 5. The PB sentence lists consisted of a total of 10 sentences. 2 sentences were used for STI = 0.8 and STI = 0.6 conditions, and 3 sentences for STI = 0.4 and STI = 0.2 were selected randomly from the sentences pool (Kalikow *et al.*, 1977) (Ozimek, 2009) (Boudraa *et al.*, 2000) (Fu *et al.*, 2011). An example set of sentences for each language is presented in Table 3.4 (a) – (d). All the sentences used are available in Appendix C.

Table 3.4 Example lists of PB sentences.

(a) An example list of the English PB sentences (Kalikow *et al.*, 1977).

The watchdog gave a warning growl	HP
She made the bed with clean sheets	HP
The old man discussed the dive	LP

(b) An example list of the Polish PB sentences (Ozimek, 2009).

Znowu ta winda nie dziala
Najpierw zwabilo go swiatlo
Wracam pozno do hotelu

(c) An example list of the Arabic PB sentences (Boudraa *et al.*, 2000).

أَحفَظُ مِنَ الأَرضِ
أَينَ المُسَافِرُونَ ?
لَا لَمْ يَستَمتِع بِثَمَرِهَا

(d) An example list of the Mandarin PB sentences (Fu *et al.*, 2011).

今天的阳光真好
节假日不用门票
晚上一块去跳舞

The recording procedure of the word and sentence lists, specifications of the speakers and the equipment that was used are presented in the next section.

3.3.2 Recording process

The word lists were recorded in the anechoic chamber of Heriot-Watt University, using native speakers for each language (3 males and 3 females). The interior dimensions of the chamber were 4.3 m (length) × 4.3 m (width) × 4.6 m (height). In the standard ANSI/ASA S3.2 (2009), the minimum number of speakers is stated as 5; however, in order to achieve equal gender representation, 6 speakers were used in the current study. Because of the significant variety of accents within languages, attention was given to the origin of the speakers. The English speakers had to speak English with Received Pronunciation (RP) (Jones, 1917), which is normally associated with formal speech and tends to be spoken in the south of England. The Arabic speakers were selected from Syria, although the origin of Arabic speakers was not crucial, as the Arabic material was written and recorded in modern standard Arabic (al-fuṣḥá) (Bourdaa *et. al.,* 2000)

(Bourdaa *et. al.,* 2008), for which the pronunciation is independent from accents and dialects. Care was also taken in the selection of Polish and Mandarin speakers, so that they could produce formal speech material. Before the actual recordings, a practice list was read by each speaker, to make them familiar with the process, and to train them in producing normal vocal effort and average rate of speaking, which is 160 to 190 words per minute (*WPM*) for British English (Pimsleur *et al.*, 1977). It should be noted that average speaking rates may vary due to the syllable structure of a language (Kowal *et al.*, 1983). All the lists were read by the speakers prior to the actual recordings. The speaking rates and average sentences' durations were comparable across languages. These were, respectively, 187.5 *WPM* and 2.22 s for English, 187.5 *WPM* and 1.62 s for Polish, 153.8 *WPM* and 1.77 s for Arabic, and 240 *WPM* and 1.81 s for Mandarin (Table 3.5).

The word and sentence recordings were then calibrated in terms of sound pressure level, by using a custom made head and torso model with microphones (Brüel & Kjaer 4189 (Naerum, Denmark)) placed inside its ears and connected to a sound level meter (Brüel & Kjaer 2250). The material to be calibrated was played through Beyer Dynamics DT150 headphones placed over the head of the model. Audio files were then prepared for the listening tests, including randomisation in the sequencing of words and editing of gaps between words. For the DRT tests (English, Arabic, and Mandarin), the word selected between a pair was simply ticked on a list provided, and the word frequency was set to one word per 1.4 seconds, following guidance by Cohen (1965). For the Polish PB word tests, the gap between words was set to 5 seconds, to give a convenient amount of time for writing down the whole word. Although there is no standard for the frequency of words in PB word tests, Diaz *et al.* (1995) suggested the frequency of one word per 4 seconds for Spanish PB word tests. This was adapted to 5 seconds for Polish, based on trial and error. For sentence tests, each new sentence was played after the listener had finished writing down the sentence just heard (no predefined frequency/duration).

Table 3.5 Speaking rates used in the tests.

	English	Polish	Arabic	Mandarin
Words per minute (WPM)	187.5	187.5	153.8	240
Sentences durations (sec)	2.22	1.62	1.77	1.81

3.3.3 Listening tests

The listening tests were conducted in one of the chambers of the acoustic laboratory of Heriot-Watt University. The dimensions of the chamber were 6.8 m (length) × 4.0 m (width) × 3.0 m (height) (Figure 3.2). All the surfaces were made of reflective materials (brick walls, concrete floor and ceiling), and the room had no windows. The minimum number of listeners stated in the standard ANSI/ASA S3.2 (2009) is 5, but 3 male and 3 female listeners were selected from native speakers of each language, in order to achieve equal gender representation. As the listeners of each language were selected from the same regions/countries of the speakers, it is important to note that the results of the study are not representative of all the dialects of the languages selected. Age distribution of the listeners was as follows: English participants ranged from 23 to 42 yr (mean 32.3 yr and standard deviation 6.7 yr), Polish from 24 to 33 yr (mean 29.3 yr and standard deviation 3.1 yr), Arabic from 30 to 33 yr (mean 31.7 yr and standard deviation 1.4 yr) and Chinese from 21 to 32 yr (mean 26.2 yr and standard deviation 5.2 yr). It should be noted that the variability between different age groups is outside the of scope of the present study. Furthermore, the listeners taking part in the tests do not represent a wide range of social backgrounds, as these were university students and researchers (i.e. the findings cannot be generalised to overall populations). The hearing threshold level of the participants was tested using the simple *AudioCheck* online hearing test (Audiocheck, 2015), results showing that all the participants had normal hearing. Hearing tests were carried out in the anechoic chamber of Heriot-Watt University using Beyerdynamic DT 150 closed headphones.

The recorded material was presented through a loudspeaker (KEF Coda III (Maidstone, UK)) placed at 1 m from one of the 4 m wide walls and was positioned over a small table with a propagating height of 1.2 m (mid-way between the woofer and tweeter). Listeners were seated at a distance of 2 m from the loudspeaker, and the speech level was adjusted to 65 dB(A), 1 m on axis from the loudspeaker and 1.2 m above floor level. The level was calibrated using uninterrupted speech material (gaps removed between words) and the sound level meter Brüel & Kjaer 2250.

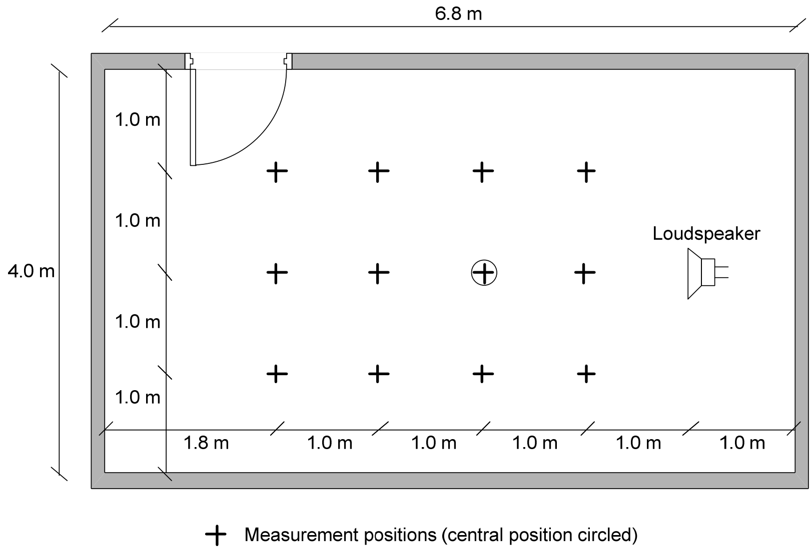


Figure 3.2 Floor plan of the acoustic chamber.

For DRT tests, listeners had to identify the spoken words within the pairs of words provided on a list (by ticking), whilst for PB words and PB sentences, these had to be written down. Each listening test was repeated for four different acoustic conditions (STI = 0.2, 0.4, 0.6 and 0.8), by changing the reverberation time and signal-to-noise ratio. The order of the acoustic conditions tested was always highest (STI = 0.8) to lowest (STI = 0.2), in order to minimise auditory fatigue. This fixed order was not expected to be responsible for learning effects for three reasons: 1) Word familiarity can be neglected in DRT tests (Voiers, 1977); 2) The sequence of DRT words and talkers was randomised; 3) The number of DRT words heard in each condition was quite large (96 for English and Mandarin and 72 for Arabic), so that words were unlikely to be easily learnt. The reverberation time was controlled by adding or removing foam and glass-wool panel absorbers on the walls. The use of different absorbers was due to not having enough identical panel absorbers for achieving the STI = 0.8 condition. The panel absorbers were distributed evenly across the room and were used only at the STI = 0.8 and STI = 0.6 conditions.

The signal-to-noise ratio was controlled by adding artificial noise to the speech signal, using the white noise generator Brüel & Kjaer 1405 (S/N = +5 dB for STI = 0.4, and S/N = -5dB for STI = 0.2). No artificial noise was used at the STI = 0.6 and STI = 0.8 conditions. The STI conditions could then be described as follows; STI = 0.8: no artificial noise and low reverberation time; STI = 0.6: no artificial noise and medium reverberation time; STI = 0.4: S/N = +5 dB and high reverberation time; STI = 0.2: S/N

= -5 dB and high reverberation time.

The physical evaluation of speech intelligibility was made using the speech transmission index (STI), which was measured using the commercial Maximum Length Sequence System Analyser (MLSSA) software. The computer used to run MLSSA was connected via its sound card to the loudspeaker KEF Coda III and to a half inch microphone Brüel & Kjaer 4190, which was in turn connected to a microphone power supply Brüel & Kjaer 2804.

3.3.4 Statistical analysis

The statistical analysis methods that were used in the first phase of the study are presented in this chapter. Three sets of data were gathered from the word (DRT and PB word) and sentence (PB sentence) intelligibility tests. The first set of data considered was the overall word intelligibility scores, which included the results of the DRT (for English, Arabic, and Mandarin) and PB word tests (for Polish). The second set of data was the distinctive features scores, which are the consonant specific results of the DRT. Therefore, Polish was not included in the comparisons of the distinctive features scores. The last set of data was the sentence intelligibility scores, which are the results of the PB sentence tests. The calculation method of intelligibility scores can be found in Chapter 4.

A total of six statistical analysis methods was applied to the data sets in order to test several hypothesis of the current study; these methods are Intra-Class Correlation analysis, one-way Analysis of Variance (one-way ANOVA), factorial Analysis of Variance (factorial ANOVA), Kruskal–Wallis H. test, Spearman's RHO correlation analysis, and Principal Component Analysis (PCA). All of the above mentioned statistical analysis methods were computed by using the Statistical Package for Social Sciences (SPSS) v.22 software.

Consistency within the test participants taking part in the intelligibility tests was analysed by using the Intra-Class Correlation analysis (Field, 2009). In order to assess between subjects reliability, the Intra-Class Correlation Coefficient (ICC) was computed for the participants of each language by using the SPSS software. In the current analysis the correlations between the test participants have been investigated, which means that the variable is the participants. ICC measures the correlation between two or more variables, when the variables measure the same phenomenon, which suits the current analysis (Field, 2009). It should be noted that and ICC value greater than 0.720 is usually considered as an acceptable value for social sciences (Shrout *et al.*, 1979).

One-way ANOVA tests the null hypothesis that there is no difference between cases, when only one independent variable is present. However, it tests for an overall experimental effect, it does not provide detailed information on the differences between pairs of cases. Therefore, additional analysis is required in order to understand how an independent variable affects the experiment. When the confidence interval is set to 95%, $p < 0.05$ indicates that the null hypothesis is not valid (the alternative hypothesis is valid) (Field, 2009). In the present study, the independent variable was the language. Prior to the analysis, null and alternative hypothesis must be set (Field, 2009). The null hypothesis was set to 'there is no effect of the languages on word/sentence intelligibility' and the alternative hypothesis was 'the change of language affects word/sentence intelligibility'. The one-way ANOVA was applied to each room acoustic condition separately for both word and sentence intelligibility scores.

Factorial ANOVA is a type of analysis of variance method that tests the validity of an hypothesis when two or more independent variables (with a minimum of two levels) are present. When the confidence interval is set to 95%, $p < 0.05$ indicates that the null hypothesis is not valid (the alternative hypothesis is valid) (Field, 2009). In the first phase of the study the two independent variables were the language (four levels: English, Polish, Arabic, and Mandarin) and the room acoustic conditions (four levels: STI=0.8, STI=0.6, STI=0.4, and STI=0.2). There were three sets of data to be analysed. The first set of data was the word intelligibility test scores, the second set of data was the distinctive features (language specific) intelligibility test scores and the last set of data was the sentence intelligibility test scores. Factorial ANOVA was applied to each set of data separately in order to test three null hypothesis; these null hypothesis were 'the change of language does not affect word/distinctive features/sentence intelligibility', 'the change of room acoustic conditions does not affect word/distinctive features/sentence intelligibility', and 'there is no interaction between the languages, the room acoustic conditions, and word/distinctive features/sentence intelligibility'. The alternative hypothesis were the opposite of the null hypothesis; these are 'the change of language affects word/distinctive features/sentence intelligibility', 'the change of room acoustic conditions affects word/distinctive features/sentence intelligibility', and 'there is an interaction between the languages, the room acoustic conditions, and word/distinctive features/sentence intelligibility'.

Spearman's RHO correlation analysis is a non-parametric statistical method and can be used when the data is ordinal, which means that the data is sorted in categories with a meaningful order (i.e. when data is non-normally distributed) (Field, 2009). In the current study it was performed to investigate a possible correlation between the consonant-to-vowel ratio of languages and word/sentence intelligibility scores. In this case, the consonant-to-vowel ratios of languages are non-normally distributed ordinal data. Four categories were used for the consonant-to-vowel ratio, which were low, average, moderately high, and high. Spearman's RHO correlation analysis was applied to the word and sentence intelligibility scores separately.

Principal Component Analysis (PCA) is a statistical method used for investigating underlying dimensions of a set of data. It is commonly used for extracting any linear components existing within the data set, and for examining how a particular variable interacts with the component (Field, 2009). In the first phase of the study the interaction between the distinctive feature intelligibility scores (language specific word intelligibility scores) were investigated by using the PCA. In order to interpret extracted factors more conveniently, Varimax rotation was applied to the data set. It loads a smaller number of variables onto each factor, therefore factor clusters become more visible (Field, 2009). The sampling adequacy of the data was tested by the Keiser-Meyer-Olkin measure of sampling adequacy (KMO). KMO values vary between 0 and 1, where 0 means that the PCA is not reliable and 1 means it is reliable. According to Hutcheson and Sofroniou (1999) values between 0.5 – 0.7 are considered mediocre, values between 0.7 – 0.8 are considered good, values between 0.8 – 0.9 are considered great, and values above 0.9 are considered superb for the KMO.

The results of the statistical analysis are presented along with the word, distinctive features word, and sentence intelligibility test results in Chapter 4.

3.4 Second phase – Effects of soundscape perception on speech intelligibility

The aim of the second phase of the study was to investigate relations between speech intelligibility and soundscape perception of the native speakers of English, Polish, Arabic, and Mandarin. 15 participants per language (i.e. a total of 60) were asked to subjectively evaluate acoustic environments by answering nine questions on a five-point semantic scale, under three room acoustic conditions, in three digitally simulated multi-lingual environments. The three multi-lingual environments were an airport check-in area, a hospital reception area, and a café. The speech samples were uniquely designed for each environment in order to achieve an appropriate context. Six sentences were created for each environment, and the samples were recorded by four native speakers (two male and two female) of each language in the anechoic chamber of the Heriot-Watt University. The three room acoustic conditions were created digitally by adding contextually appropriate background noise and reverberation to the speech recordings. The finalised speech recordings were then presented to the participants in combination with the visuals of environments in the anechoic chamber of Heriot-Watt University, where they were asked to subjectively evaluate the audio-visual material. The results of the experiment were statistically analysed in order to find any possible effects of socio-cultural backgrounds of the participants on speech intelligibility.

In the next section, the methodology used in the second phase of the study is described. The results of the listening tests, statistical analysis of the results and discussions are presented in Chapter 5.

3.4.1 Selecting the cases

In today's global cities, the majority of the public environments are multi-cultural, therefore multi-lingual speech communication is common. The aim of the second phase of the study was to investigate the relations between speech intelligibility and soundscape perception of native speakers of English, Polish, Arabic, and Mandarin, by subjectively evaluating the acoustic environment of such spaces, in which the context of speech is rather crucial. Therefore, the following criteria were applied for selecting the cases: oral communication must be at the centre of attention, the environments should represent a variety of acoustic conditions; and the test participants from England, Poland, Syria, and China should have an experience of the selected environments.

The first case selected was an airport check-in area. airports are common public environments in the majority of global cities, where oral communication between a passenger and a check-in desk attendant is often crucial. The simulated Airport enclosure is typically large and spacious with a high ceiling to accommodate several activities such as check-in, waiting, and circulation; therefore leading to a reverberant acoustic environment. The background noise is typically fairly steady, and occasionally there are public announcements (PA) and other impact sounds (i.e. footsteps and luggage wheels), and hubbub speech noise, which interfere with the speech content. Regarding the above mentioned criteria, the airport check-in area was chosen as an example of a high reverberation time and high background noise acoustic environment, which is common in a majority of global cities, with a crucial content of multi-lingual speech communication.

The second case selected was a hospital reception area. The speech content of a hospital reception area is usually crucial due to the fact that the context is about health issues. Conversations between a patient and a receptionist can accommodate critical information, which cannot be risked to be unintelligible. Compared to the other public spaces selected for the experiments, the hospital reception area simulated in the present study was a medium sized enclosure leading to a medium to low reverberation time, with a relatively low continuous background noise mostly composed of hubbub speech noise. Other noises such as a telephone ringing were also present.

The last case selected was a café. Although the speech content in a café environment is not as crucial as the other cases, conversation still is at the centre of attention. Additionally, especially in global cities, cafés are one of the most multi-cultural and multi-lingual public spaces. It is also a relaxed environment, as opposed to the stressful environments represented by the airport check-in area and the hospital reception. The simulated café environment considered in this study was a medium to large sized space, with a moderately-high reverberation time and continuous background noise, which was mostly composed of hubbub speech noise.

Detailed information on the room acoustic properties of the simulated environments, the preparation process of the simulated acoustic environments and accompanying visuals presented to the participants in the second phase of the study are presented in the next section. These include the preparation of the sentence lists, recording process of the speech samples, and the post-processing.

3.4.2 Preparing the test material

The first step to prepare the audio-visual materials was deciding on the room acoustic conditions that would be tested. After analysing the data from the first phase of the study (detailed analysis is presented in Chapter 4), the largest variation of word/sentence intelligibility scores was observed at STI = 0.4. In practice, this represents a poor room acoustic condition (Barnett and Knight, 1995) and most spaces should be expected to perform with higher STI values. Therefore, three STI values were aimed, which are STI = 0.4, STI = 0.5, and STI = 0.6. This guaranteed investigating conditions, where variations in intelligibility between languages are highest, but representative of real cases (i.e. lower STI conditions are rare and STI conditions above 0.6 are not expected to show significant differences between languages).

The second step was designing the sentence lists. Different sentence lists were prepared for each of the three cases. Each case was containing six sentences that were uniquely designed to match the context of the environment. For English, each sentence contained approximately 50 syllables (minimum 44 syllables and maximum of 51 syllables). The sentences were then translated to Polish, Arabic, and Mandarin by native speakers of the languages. During the translation, attention was given to the syllable counts of the sentences to be comparable within the languages. It was also important that the listeners felt as part of the conversation; therefore, the sentences were designed to simulate active engagement of the participants, either by directing a question, or by illustrating a task. Examples of the sentence lists used are presented in Table 3.6. All the sentences used are available in Appendix D.

The third step was recording the sentences and the background noise samples that were used in the final audio files. The word lists were recorded in the anechoic chamber of Heriot-Watt University, using four native speakers of each language (2 males and 2 females). For the speakers' accents and the recording process, the same criteria that were implemented for the first phase of the study were applied (see Section 3.3.2).

The airport and hospital background noise samples could not be recorded at the location because of the security restriction; therefore, previously recorded high-quality sound samples were used. After subjectively reviewing the catalogue of the 'audiosparx.com' website in terms of audio quality, sample length, and the availability of sound marks related to the environments, one background noise sample was selected for each of the environments. The airport and hospital background noises were 24 seconds long samples that were selected out of a 2 minutes and 27 seconds long (AJ Pro Audio, 2014) audio recording of an airport and a 1 minute and 36 seconds long (X-Ray Sound Studio, 2014) audio recording of an hospital. Both of the audio files were high-quality wave sound files (44.100 Hz, 16 bit). The café background noise sample was recorded at the canteen, which is located in Heriot-Watt University. It is a medium-large sized enclosure and attracts many people from the university. The café background noise sample was recorded using a digital sound recorder Zoom H4n (Fig. 3.3(a)) during the lunchtime, which is the most crowded time period of the day. After reviewing the recording in terms of homogeneity of sound events, a 3 minutes long section of the recording was selected to be used in the final audio mixes.

Table 3.6 Example list of English sentences.

Airport check-in area	I am afraid the luggage allowed on this flight is two pieces maximum, regardless of the maximum weight permitted. The charge per extra luggage is fifteen Euros, which you can pay at the airline's counter.
Hospital reception area	In order to book an appointment, I first need you to fill in this form and submit it to me when completed. Please write down your name, date of birth, phone number and health insurance number if available.
Café	I am really looking forward to the weekend. Yesterday I spent some time planning a two hour hike in the mountains, as well as a short boat trip on the lake, if the weather is good. Would you be interested in coming with me?

(a) Zoom H4n sound recorder

(b) M-Audio MobilePre USB soundcard

Figure 3.3 Equipment that were used during the second phase of the study.

The next step was mixing the speech and background noise sound samples, and finalising the sound files by adding reverberation to the speech samples in order to achieve the aimed STI values (STI = 0.4, STI = 0.5, and STI = 0.6) for each of the three environments. Digital audio processing was carried out by using Studio One 2 audio production software (PreSonus audio electronics), installed on a personal computer (PC) connected to an external M-Audio USB sound card (Fig. 3.3(b)). Sound pressure level measurements of the speech and the background noise samples were carried out by connecting a sound level meter Brüel and Kjaer Type 2250 to the master sound output of M-Audio USB sound card.

The STI values were computed individually for each of the 288 speech recordings (6 sentences, 3 environments, 4 speakers, and 4 languages) by using the modulation transfer function (MTF) method (Houtgast and Steeneken, 1973). Detailed information on the MTF was given in Chapter 2, Section 2.4.1.2. Due to the fact that the airport, hospital, and café environments vary in terms of overall volume of the spaces, the direct field contribution was included in the calculation of the modulation reduction factor *m*. In order to calculate the direct field contribution, the critical distance and source-to-receiver distance had to be identified for each of the acoustic environments. The critical distance is the point where the direct sound pressure level is equal to the reverberant field sound pressure level (Long, 2006), and can be computed by using the following equation, when assuming spherical propagation of sound:

$$r_c = \sqrt{\frac{A}{16\pi(1-\overline{\alpha})}} \tag{3.1}$$

where A is the total absorption in the room (m^2) and α is the average absorption coefficient. The source-to-receiver distance was assumed as 1 m for all of the acoustic environments. After finding the critical distance and knowing the source-to-receiver distance, the modulation reduction factor m was computed by using the following equation (Long, 2006)

$$m(\omega_m) = \frac{(A^2 + B^2)^{1/2}}{C}[1 + 10^{-0.1L_{SN}}]^{-1} \tag{3.2}$$

with

$$A = \frac{Q}{r^2} + \frac{Q}{r_c^2}\left[1 + \left(\frac{\omega_m T_{60}}{13.8}\right)^2\right]^{-1} \tag{3.3}$$

$$B = \frac{\omega_m T_{60}}{13.8}\frac{Q}{r_c^2}\left[1 + \left(\frac{\omega_m T_{60}}{13.8}\right)^2\right]^{-1} \tag{3.4}$$

$$C = \frac{Q}{r^2} + \frac{Q}{r_c^2} \tag{3.5}$$

where Q is the source directivity, r is the source-to-receiver distance (in meters), r_c is the critical distance (in meters), ω_m is the modulation angular frequency (in Hz) ($\omega_m = 2\pi f_w$), L_{SN} is the signal-to-noise ratio (in dB), and T_{60} is the reverberation time (in seconds). The L_{SN} was calculated by using the sound pressure level measurements of the speech and the background noise samples, which were obtained by using a sound level meter Brüel and Kjaer Type 2250 connected to the master output of the M-Audio USB sound card. Computing the STI based on the modulation reduction factors is described in Chapter 2, Section 2.4.1.2.

As the STI values were previously decided based on the first phase results of the study (STI = 0.4, STI = 0.5, and STI = 0.6), the reverberation time on the speech recordings and the signal-to-noise ratios were adjusted to achieve the desired STI values. Based on the comparative volumes of the three environments and in order to achieve a variety of reverberation times in between three cases, the airport check-in area, which is the enclosure with the highest volume was modelled to have a reverberation time of 1.5s, the café, which is the medium-large sized enclosure was modelled to have a reverberation time of 1.2s, and the hospital reception area, which is the medium-sized enclosure was modelled to have a reverberation time of 1.0s, across all frequencies. The signal-to-noise ratios were then set manually by adjusting the sound pressure level of the background noise samples in order to achieve the desired STI values. The audio files were finalised by adjusting the sound pressure level of the speech sample to 65dB (A), and mixing it with the background noise sample in order to achieve the desired signal-to-noise ratios.

After the recording and post-processing procedure, a total of nine sound environments were created. The acoustic properties of the environments are presented in Table 3.7.

In combination with the sound samples, case specific visuals were presented to the test participants. For the airport check-in area (Fig. 3.4(a)) and the hospital reception area (Fig. 3.4(b)), one high-resolution photograph for each of the environments was selected from online searches. For the café, high-resolution photographs were shot of the canteen within Heriot-Watt University (Fig. 3.4(c)) during lunch-time (during the same period of time with when the background noise samples were recorded). Attention was given to take photographs with a general area view and with no distracting focal points.

The audio samples and the visuals of the environments were compiled in the form of a slide-show for the listening tests. The slide shows were prepared by using the software Microsoft PowerPoint 2013. Each slide-show consisted of a total 28 slides, including

Table 3.7 The list of tested acoustical environments.

Environment	Reverberation Time		S/N	STI
Airport	Across all frequencies	1.5s	-2,90 dB	0.4
Airport		1.5s	0,06 dB	0.5
Airport		1.5s	3,12 dB	0.6
Hospital		1.0s	-2,33 dB	0.4
Hospital		1.0s	1,00 dB	0.5
Hospital		1.0s	5,26 dB	0.6
Café		1.2s	-2,72 dB	0.4
Café		1.2s	0,15 dB	0.5
Café		1.2s	3,11 dB	0.6

detailed instructions on the listening test sessions. Both the speech samples used for each acoustic environment and the order of the acoustic environments presented to each participant were randomised to avoid order effects. Detailed information on the listening test procedure is presented in the next section (Section 3.4.3).

The next section describes the listening test procedure, including the selection of participants and facilities/equipment that have been used.

3.4.3 Listening tests

The listening tests were conducted in the anechoic chamber of Heriot-Watt University. 15 participants were selected from native speakers of each language (total of 60 participants). Each participant was paid with Amazon.com vouchers after the tests were completed. The listeners of each language were selected from the same regions/countries of the speakers (see section 3.3.2).

(a) Airport check-in area (Easy-jet check-in area, n.d.)

(b) Hospital reception area (Saint Paul's hospital lobby, n.d.)

(c) Cafe

Figure 3.4 The visuals of the tested environments.

The sound samples were presented through a pair of Beyerdynamic DT 150 headphones, which were connected to the PC through an M-Audio USB sound card. The sound output level was adjusted to 65 dB(A) for the speech signal. The power-point presentations were displayed on a 27 inch Samsung flat-screen computer monitor (Fig. 3.5) placed on a standard office desk, and the participants were seated on an upholstered office chair. Listeners' input the commands for skipping to the next slide through a standard PC keyboard. There were no distracting objects in the anechoic chamber other than the mentioned test equipment and furniture, and the background noise level in the chamber was 21 dB(A) (including noise from the computer).

The evaluation forms were prepared by implementing the semantic differential technique. The semantic differential technique was first developed by Osgood *et al* (1957) to identify emotional meanings of the words. In the previous literature semantic differential analysis has been adopted for soundscape analysis by identifying sounds and their linguistic and psychological meanings (Kang and Zhang, 2010).

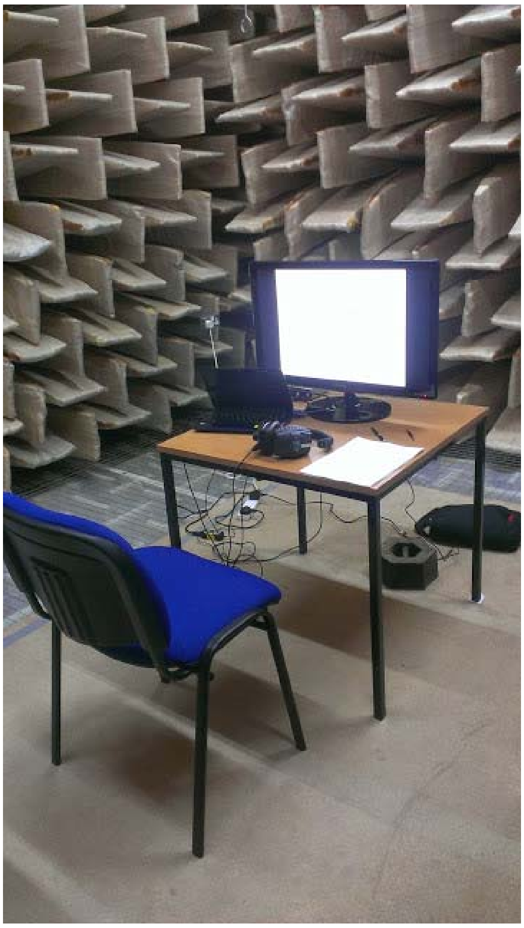


Figure 3.5 The test setup in the anechoic chamber.

In the current study, the participants were asked to fill a subjective evaluation form after listening to each acoustic environment. Detailed instructions were presented both on-screen (Appendix E) and orally prior to the listening tests. They were allowed to listen to each acoustic environment only once. Each evaluation form consisted of 11 5-point scale semantic questions (3 semantic questions for assessing the speech and 8 semantic questions for assessing the overall acoustic environment). The attributes tested were speech intelligibility, speech loudness, pleasantness (of speech), noisiness, annoyance, relaxation, acoustic comfort, pleasantness (of the environment), eventfulness, excitement, and familiarity. After completing the evaluation form, participants were asked to proceed to the next acoustic environment by pressing the assigned key on the keyboard. The process was repeated until all of the acoustic environments (Table 3.6) had been evaluated.

The methods used for analysing the data gathered from the evaluation forms are described in the next section.

3.4.4 Statistical analysis

A total of four statistical analysis methods were applied to the data sets in order to test several hypothesis of the current study; these methods were Intra-Class Correlation analysis (ICC), one-way Analysis of Variance (one-way ANOVA), repeated measures Analysis of Variance (repeated measures ANOVA), and Principal Component Analysis (PCA). All of the above mentioned statistical analysis methods were computed by using the Statistical Package for Social Sciences (SPSS) v.22 software.

First of all, consistency between the test participants was analysed by calculating the Intra-Class Correlation Coefficient (ICC) as explained in Section 3.3.4. Next, the difference between the results of listening tests of the four languages were statistically analysed for each room acoustic condition and for each attribute by using the one-way Analysis of Variance (one-way ANOVA) method with the help of the Statistical Package for Social Sciences (SPSS) software. In the current phase of the study, the independent variable was the language. The null hypothesis was 'there is no effect of the languages on the participants’ subjective ratings of the acoustic environments' and the alternative hypothesis was 'the change of languages affects the participants’ subjective ratings of the acoustic environments'. The confidence interval was set to 95% for the one-way

ANOVA. Detailed information on one-way ANOVA was presented in Section 3.3.4.

After looking at the differences between languages for each case separately, the individual and combined effects of the three independent variables (language, environment, and STI conditions) on the results were analysed by the repeated measures Analysis of Variance (ANOVA) method; therefore, an explanation for variance of the scores of each question was investigated. Repeated measures ANOVA is a statistical analysis method that is applied when the same participants attend to all conditions of an experiment (Field, 2009). Similar to the previously mentioned ANOVA tests, the confidence interval was set to 95%, and $p < 0.05$ indicates that null hypothesis was not valid (alternative hypothesis is valid). Prior to the repeated measures ANOVA calculations, sphericity of the data set needs to be tested for the homogeneity of variance across conditions (Field, 2009). Sphericity was tested by Mauchly's test, which SPSS software produces when a repeated measures ANOVA test is conducted. A significant Mauchly's test means that there are significant variances in between conditions; therefore, it’s sphericity is not met. In that case, F-ratios are corrected by Greenhouse-Geisser corrections, which are used to calculate the p values (Field, 2009). In the second phase of the study, the three independent variables were the language (four levels: English, Polish, Arabic, and Mandarin), the room acoustic conditions (four levels: STI=0.6, STI=0.5, and STI=0.4), and the environments (the airport, the hospital, and the café). Repeated measures ANOVA was applied to the data set to test three null hypotheses; these null hypotheses were 'the change of language does not affect the participants’ subjective ratings of the acoustic environments', 'the change of room acoustic conditions does not affect the participants’ subjective ratings of the acoustic environments', and 'there is no interaction between the languages, the room acoustic conditions, and the participants’ subjective ratings of the acoustic environments'. The alternative hypotheses are the opposite of the null hypothesis; these were 'the change of language affects the participants’ subjective ratings on the acoustic environments', 'the change of room acoustic conditions affects the participants’ subjective ratings of the acoustic environments', and 'there is an interaction between the languages, the room acoustic conditions, and the participants’ subjective ratings of the acoustic environments'.

In the second phase of the study the interaction between the semantic attributes was investigated by using the PCA. The aim of the analysis was to extract meaningful factors (correlated groups of attributes) that can explain the relations between socio-cultural backgrounds of the participants and speech intelligibility. Detailed information on the PCA was presented in Section 3.3.4.

The results of the statistical analysis are presented along with the test results in Chapter 5.

3.5 Conclusions

In this chapter the methodology used in both phases of the study was presented. First, the selection process of the languages was presented. For the first phase of the study, the followings were described: the preparation process of the word and sentence lists, recording process of the lists, information on the speakers and listeners, post-processing of the audio recordings, equipment and facilities that were used, the acoustic conditions that were tested, the listening test procedure, and the statistical analysis methods of the results were described. For the second phase of the study the followings were described: the selection process of the tested acoustic environments, the preparation of the sentence lists, the recording and post-processing process, preparation of the visual material, equipment and facilities that were used for both recording and listening tests, listening test procedure, and the statistical analysis methods applied to examine results.

Results of the first phase of the study that investigates the relations between room acoustic properties (reverberation time and signal-to-noise ratio) and linguistic properties of the languages can be found in Chapter 4, and results of the second phase that investigates the relations between speech intelligibility and socio-cultural background of the participants are presented in Chapter 5.

CHAPTER 4

Room acoustics and speech intelligibility of English, Polish, Arabic, and Mandarin

---

4.1 Introduction

This chapter discusses comparisons of the subjective listening test results and the objective speech transmission index (STI) results for four languages (English, Polish, Mandarin, and Arabic) in order to understand relations between language specific effects and speech intelligibility, as well as relations between room acoustic properties and speech intelligibility of the different languages.

In this chapter, results of the speech intelligibility tests are presented and analysed. Three types of subjective listening test results were carried out. The first one was the Diagnostic Rhyme Test (DRT), the second one was the Phonemically Balanced word test (PB), and the last one was the Phonemically Balanced sentence test. DRT and PB tests were used to examine word intelligibility, whilst Phonemically Balanced sentence tests were used for the analysis of sentence intelligibility. It should be noted that PB word tests were only used for Polish because of the lack of DRT material in Polish; however, the results are still comparable as explained in Chapter 3.

It is important to point out that both word and sentence tests have some limitations with regard to comparisons between languages. For example, Kang (1998) 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 (Beranek, 1949), i.e. less suited 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.

All the statistical analysis presented in this thesis has been made using Rationalized Arcsine Units (RAU) (Studebaker, 1985) (i.e., rationalized arcsine transformed data), to ensure that the homogeneity assumption of ANOVA was not violated. Furthermore, the *p* values given have not been corrected for multiple comparisons. All the statistical analysis has been carried out using the Statistical Package for Social Sciences (SPSS), and all the results given in figures include standard errors of the mean and logarithmic regressions.

Subjects' consistency across all tests presented in this section (word scores, distinctive features' scores and sentence scores) was analysed using the Intra-Class Correlation Coefficient (ICC). The absolute agreement average measures ICC analysis with the two-way mixed model revealed that the answers of participants agree with each other for English (ICC = 0.973), Mandarin (ICC = 0.948), Arabic (ICC = 0.925), and Polish (ICC = 0.991), where ICC > 0.720 is usually considered as an acceptable value for social sciences (Shrout and Fleiss, 1979). This confirms that the use of only 6 subjects per language was appropriate and that the results presented are reliable. Actually, further statistical analysis indicated that only 2 subjects are needed to achieve an ICC just above 0.720. This lower number of participants could then be used in future work.

4.2 Word intelligibility tests

4.2.1 Overall intelligibility scores

One of the most common ways to assess subjective speech intelligibility is the use of word intelligibility tests. In the current research the test materials used were isolated words, and listeners were asked to select from presented monosyllabic words or write down the whole word that was presented throughout a loudspeaker (for details refer to Chapter 3). In this section, the relation between subjective overall speech intelligibility scores and the objective Speech Transmission Index (STI) measured under four room acoustic conditions is examined. The results are presented in Figure 4.1, where the horizontal axis shows the STI results, and the vertical axis shows the word intelligibility test results for all languages. As stated previously, DRT scores are shown for English, Arabic, and Mandarin, whilst PB scores are shown for Polish.

To calculate the results of the DRT, Voiers (1983) suggested the formula given below, in order to eliminate the effects of guesswork.

$$P_c = 100\frac{N_r - N_w}{T} \tag{4.1}$$

where, N_r is the number of correct responses, N_w is the number of incorrect responses, T is the total number of the test items, and P_c is the percentage correct score. Phonemically balanced word test scores were converted into percentage correct scores, and the arithmetic average of all of the participants' results were computed.

Figure 4.1 illustrates that there are differences between subjective speech intelligibility scores of English, Mandarin, Polish, and Arabic. First of all, English is the most intelligible language under all acoustic conditions. For the STI = 0.2 condition (S/N = -5 dB, high reverberation time) the DRT score of English is 37% and for the STI = 0.8 condition (no artificial background noise, low reverberation time) the DRT score is above 90%. It is also observed that Mandarin is more intelligible than Arabic and Polish at the STI = 0.4 condition (S/N = +5 dB, high reverberation time), in which participants were first introduced to the artificial background noise. The word intelligibility score of Mandarin at the STI = 0.4 condition is 70%, which is approximately 25% higher than the word intelligibility scores of Arabic and Polish. It is also seen that Arabic and Polish are the most sensitive languages to artificial background noise. For Arabic, the difference of word intelligibility scores between the STI = 0.4 condition and the STI = 0.6 condition is 38%, and for Polish it is 46%. It is also apparent that the difference between intelligibility scores of languages become more conspicuous under poor acoustic conditions (STI = 0.4 and STI = 0.2). It is observed that there is an approximate difference between language scores of 9% for the STI = 0.8 condition; however, it increases to much larger differences of 33% for STI = 0.4, and 30% for STI=0.2. All of the differences listed above in terms of speech intelligibility might be caused by linguistic properties of each language, and this is discussed in Section 4.2.2.

The difference of word intelligibility scores between languages was analysed statistically by the Factorial Analysis of Variance (ANOVA) method. Similar to the one-way ANOVA method, the Factorial ANOVA method also compares mean values of different groups; however, it is also capable of analysing more than one independent variable (Field, 2009). In the present study, language and room acoustic condition were the two independent variables. Factorial ANOVA showed that there was a main effect ($p < 0.01$) of language [$F(3, 80) = 26.09$, $p = 0.000$] and a main effect ($p < 0.01$) of STI conditions [$F(3, 80) =$

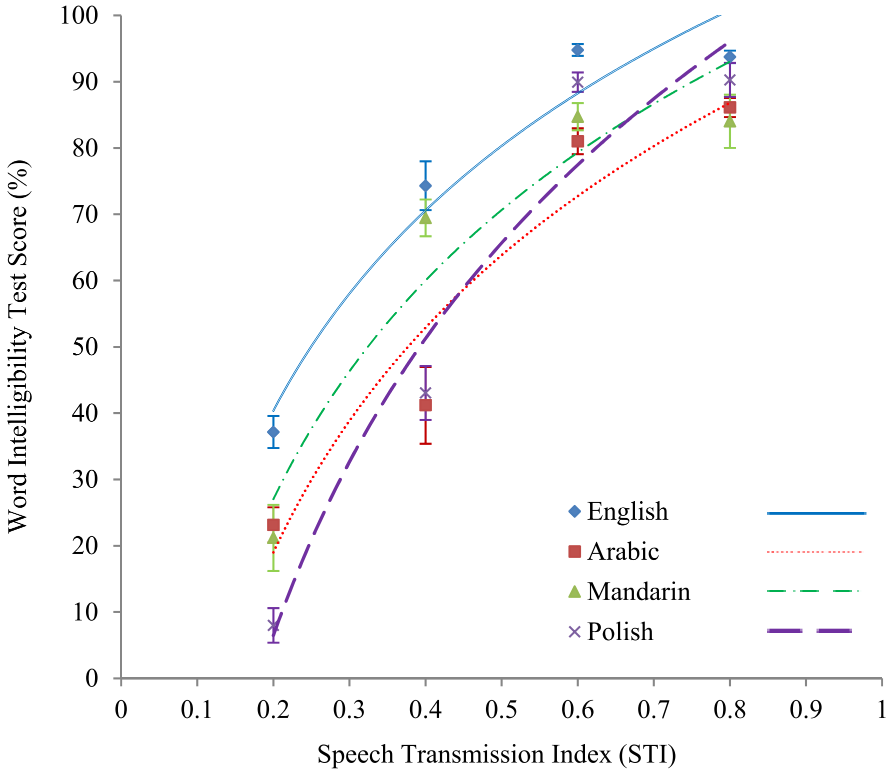


Figure 4.1 Word intelligibility scores and STI results for English, Arabic, Mandarin, and Polish. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

339.45, $p = 0.000$] on word intelligibility, as well as an interaction ($p < 0.01$) of language and STI conditions [$F(9, 80) = 6.55$, p = 0.000] on word intelligibility. These results indicate that both the variation of languages and the variation of room acoustic conditions (STI) affected the word intelligibility, and that the amount by which word intelligibility changes is governed by both of these factors and cannot be predicted by a single factor.

One-way ANOVA tests were also carried out for each STI condition, and these clarified that the word intelligibility scores of the four languages examined were significantly different ($p < 0.01$) at STI = 0.6 [$F(3, 20) = 16.35$, $p = 0.0000$], STI = 0.4 [$F(3, 20) = 16.38$, $p = 0.000$] and STI = 0.2 [$F(3, 20) = 11.45$, $p = 0.000$], whilst differences were not significant ($p > 0.05$) at STI = 0.8 [$F(3, 20) = 2.99$, $p = 0.055$]. In other words, word intelligibility of different languages is comparable under excellent room acoustic conditions, but is not comparable under all other conditions. PB Polish word scores were then removed from the statistical analysis, to check whether differences in test methods affected findings. Statistically significant differences ($p < 0.05$) between DRT scores were then found at all conditions: at STI = 0.8 [*F*(2, 15) = 4.67, *p* = 0.027], STI = 0.6 [*F*(2, 15) = 23.10, *p* = 0.000], STI = 0.4 [*F*(2, 15) = 16.67, *p* = 0.000] and STI = 0.2 [*F*(2, 15) = 4.75, *p* = 0.025]. This confirms that the main findings are not affected by the different word test used for Polish.

Speech intelligibility research (Peutz, 1971) suggests that the consonants are more sensitive to room acoustic conditions, such as background noise and signal-to-noise ratio, compared to vowels. Therefore, the consonant-to-vowel ratios of languages was one of the most important criterion for the selection of the tested languages. The four selected languages, Polish, Arabic, Mandarin, and English have high, moderately high, average, and low consonant-to-vowel ratios respectively. Results of the word intelligibility tests suggest that the difference between the subjective test scores of the languages is highly correlated with the consonant-to-vowel ratios of the languages. As shown by the data of Figure 4.1, it is apparent that English is the most intelligible language among the others under all room acoustic conditions, and has the lowest consonant-to-vowel ratio. Also, Mandarin, which has the second lowest consonant-to-vowel ratio, is significantly more intelligible at the STI = 0.4 acoustic condition compared to Arabic and Polish, which have moderately high and high consonant to vowel ratios, respectively.

The word intelligibility score of Mandarin is approximately 70%, whereas for Arabic and Polish it is approximately 40%. It should be noted that STI = 0.4 is the condition in which listeners were first introduced to the artificial background noise. However, comparing the word intelligibility scores for STI = 0.6 and STI = 0.8 conditions, the correlation between consonant to vowel ratios and speech intelligibility is no longer obvious. Although English is still the most intelligible language for these conditions, the difference between intelligibility scores decrease to approximately 9% at STI = 0.8 and Polish is the second most intelligible language.

Results suggest that the word intelligibility scores of Arabic and Polish depend on whether the acoustic conditions are challenging or not. For STI = 0.6 and STI = 0.8, the word intelligibility scores for Arabic are approximately 80% and for Polish they are approximately 90%, which indicates that Polish is more intelligible under good acoustic conditions. However, for STI = 0.4 the scores of both languages are approximately 40% and for STI = 0.2 Arabic is more intelligible than Polish, with a score difference of approximately 15%. It can therefore be assumed that Arabic is more intelligible than

Polish under very challenging acoustic conditions. Statistical analysis clarified that there is a significant correlation between the consonant to vowel ratios of languages and the subjective speech intelligibility results in more challenging room acoustic conditions with high reverberation time and signal-to-noise ratio (STI = 0.2 and STI = 0.4). The correlation was tested by the Spearman's rho analysis and statistically significant negative correlations were found between the consonant-to-vowel ratios of languages and the word intelligibility results for the most challenging room acoustic conditions, Spearman's correlation analysis results being $\rho = -0.73$ ($p < 0.01$) for STI = 0.2, and $\rho = -0.76$ ($p < 0.01$) for STI = 0.4. The negative sign indicates that word intelligibility decreased with increasing consonant-to-vowel ratio, as expected (Peutz, 1971).

One of the most related researches on comparing the intelligibility of different languages was conducted by Kang (1998). The intelligibility of English and Mandarin were compared in two spaces (a seminar room and a corridor) for three different room acoustic conditions. It was found that in the corridor, for a relatively high STI (high signal-to-noise ratio), the word intelligibility of Mandarin was better than English, and for a low STI, the intelligibility of English was better. However, in the seminar room the difference in word intelligibility of Mandarin and English was no longer obvious (almost no difference for STIs below 0.5 and only around +2% for Mandarin at STI = 0.6 and above). Contradicting the study conducted by Kang (1998), in the present study it is observed that English is the most intelligible language in all acoustic conditions. The difference between the two studies might be explained by the fact that the speech materials and room acoustic conditions were different. Both the PB word test and the DRT are reliable for testing the intelligibility of speech; however, when more than one language is being tested, the results might be affected by the linguistic properties of languages. Due to the fact that several linguistic properties of each language can be tested by the DRT, the DRT was used instead of the phonemically balanced word lists used by Kang (1998). Furthermore, in the previous study, the most challenging room acoustic condition had an S/N ratio of 2 dB, whereas in the present study for STI = 0.2 the S/N ratio was -5 dB.

Main findings:

- There is a statistically significant difference between the word intelligibility of the four languages tested, and this is observed for most of the acoustic conditions tested (STI = 0.2, 0.4, and 0.6). The difference is higher towards low intelligibility conditions, such as STI = 0.2 and STI = 0.4 where the difference between language scores is approximately 35%; however, for STI = 0.8 the difference is lower and approximately 10%.

- The results suggest that there is a correlation between consonant-to-vowel ratios of languages and subjective speech intelligibility scores under the low intelligibility conditions (STI = 0.2 and STI = 0.4). The Spearman’s rho analysis showed that the correlation is significant ($p = .000$) at STI = 0.2 and STI =0.4. Arabic and Polish, which have high consonant-to-vowel ratios, are significantly less intelligible than English and Mandarin, which have lower consonant-to-vowel ratios.
- Under noisy conditions Mandarin is more intelligible compared to Arabic and Polish. This might be caused by the linguistic properties of Mandarin, which are analysed in Section 4.2.2.

4.2.2 Language specific intelligibility scores

In this section, distinctive features’ scores of the Diagnostic Rhyme Tests (DRT) of English, Arabic, and Mandarin languages are analysed. Due to the fact that there were no DRT tests available in Polish at the time when the study was conducted, linguistic properties of Polish could not be examined and compared with English, Arabic, and Mandarin.

Each DRT includes scores of six different distinctive features depending on the tested language tested. The six distinctive features do not need to be identical across languages, as some distinctive features might be relevant in one language but irrelevant in another, and this is why different distinctive features might need to be considered to correctly represent a language. For English, voicing, nasality, sustention, sibilation, graveness, and compactness properties of consonants were tested (Voiers, 1977). For Arabic, tenseness, nasality, mellowness, flatness, graveness, and compactness were tested (Boudrea *et al.*, 2008). And for Mandarin, airflow, nasality, sustention, sibilation, graveness, and compactness were tested (Li *et al.*, 2000). In order to understand the effects of room acoustic properties on distinctive features and overall intelligibility of languages, DRT scores of each linguistic property are compared and analysed within each language (Figures 4.2, 4.3, and 4.4). Furthermore, the three shared distinctive features are compared and analysed in between languages, those being nasality, graveness, and compactness (Figures 4.5, 4.6, and 4.7).

4.2.2.1 Distinctive Features

The distinctive features used in the DRT were suggested by Jakobson *et al.* (1952). These discriminate consonantal properties between two sounds. Descriptions of each distinctive feature are listed below. An example list of the English DRT is also given in Table 4.1.

Table 4.1. An example list of the English DRT (Voiers, 1977).

Voicing		**Nasality**		**Sustention**	
Voiced	Unvoiced	Nasal	Oral	Continuant	Interrupted
Veal	Feel	Meat	Beat	Vee	Bee
Bean	Peen	Need	Deed	Sheet	Cheat
Gin	Chin	Mitt	Bit	Vill	Bill
Sibilation		**Graveness**		**Compactness**	
Sibilant	Non-sibilant	Grave	Acute	Compact	Diffuse
Zee	Thee	Weed	Reed	Yield	Wield
Cheep	Keep	Peak	Teak	Key	Tea
Jilt	Gilt	Bid	Did	Hit	Fit

- Voicing (voiced / unvoiced) is the distinctive feature between the voiced and the unvoiced sounds. While producing the voiced sounds vocal cords vibrate; however, the unvoiced sounds do not require a vocal cord vibration (i.e. *p* is an unvoiced sound, *b* is a voiced sound) (Jakobson *et al.*, 1952).
- Nasality (nasal / oral) differentiates among the nasal and the oral consonants. If air escapes from the mouth, it is called an oral sound. While producing the nasal sounds, air escapes from the nose (i.e. *b*, *w*, *v*, and *x* are oral sounds, *m* and *n* are nasal sounds) (Jakobson *et al.*, 1952).
- Sustention (continuant / interrupted) is the distinctive feature between the continuant and the interrupted consonants. The continuant sounds can be continued indefinitely, whilst the interrupted sounds show a sudden spread of energy (i.e. *f* is a continuant sound and *b* is an interrupted sound) (Jakobson *et al.*, 1952).
- Sibilation (sibilant / non-sibilant) is the distinctive feature between the sibilant and the non-sibilant sounds. The sibilant sounds can be described as hissing sounds such as the pronunciation of `*th*` in English (Jakobson *et al.*, 1952).
- Graveness (grave / acute) is the distinctive feature between the grave and the acute sounds. It is a subjective classification according to which the consonant sounds dull or sharp (i.e. *p* is a grave sound and *t* is an acute sound) (Jakobson *et al.*, 1952).
- Compactness (compact / diffuse) is the distinctive feature between the compact and the diffuse sounds. The compact consonants are generated at the front part of the oral

cavity, and the diffuse consonants are generated at the back (i.e. *k* and *g* are compact; *m* and *f* are diffuse consonants) (Jakobson *et al*., 1952).

- Tenseness (tense / lax) distinguishes between the tense and the lax consonants. The tense consonants are generated by using more effort compared to the lax consonants (i.e. *k* is a tense consonant, *g* is a lax consonant) (Jakobson *et al*., 1952).
- Mellowness (strident / mellow) is the distinctive feature between the strident and the mellow / non-strident consonants. Strident consonants are affricates and grooved fricatives, on the other hand, mellow or non-strident consonants are slit fricatives (i.e. *z* and *s* are strident consonants; *g* and *k* are mellow / non-strident consonants) (Jakobson *et al*., 1952).
- Flatness (flat / plain) is the distinctive feature between flat and plain vowels. The flat vowels are produced by lip rounding opposing to plain consonants. This feature is used only in the Arabic DRT, which considers vowels as well as consonants because of the taxonomic features of the language (Boudraa *et al*., 2008).
- Airflow (airflow / no-airflow) is a Mandarin only distinctive feature, which replaces voicing. It distinguishes between consonants with airflow (airflow the consonant gives) and non-airflow (airflow the consonant does not give) consonants (Li *et al*., 2000).

The scores of the DRT can be used for evaluating the overall speech intelligibility, and to evaluate the ease of discriminating consonantal properties. As stated above, each distinctive feature (voicing, nasality, sustention, sibilation, graveness, compactness, tenseness, mellowness, flatness, and airflow) consists of two consonantal properties. The score of a distinctive feature explains how well a listener can discriminate between two consonantal properties (Jakobson *et al*., 1952).

4.2.2.2 The results of language specific DRTs

This section initially compares the distinctive features' results within each language (English, Arabic, and Mandarin), and then across those languages. As discussed in section 4.2.1, English was the most intelligible language among the selected languages. It was hypothesized that the consonant-to-vowel ratio of a given language might have an impact on the overall intelligibility. The overall word test scores revealed that the language with lowest consonant-to-vowel ratio (English) was the most intelligible in all conditions. Additionally, by looking at Figure 4.2, it is seen that for the English language, nasal/oral consonants can be discriminated easily under all room acoustic conditions. It is surprising that even when STI = 0.2, the acoustic condition in which there is high artificial background noise (S/N = -5 dB), the nasality score is 85%. This is in line with a previous study of Voiers (1999), who found very high nasality scores even for challenging acoustic conditions. For most acoustic conditions graveness has the lowest DRT score compared to the other five properties. At the STI = 0.8 condition the DRT score for graveness is 79%, and for the STI = 0.2 condition it is 12%. The results also show that continuant/interrupted discrimination (sustention) is sensitive to the artificial background noise. At STI = 0.4 the sustention DRT score is 45%, and at STI = 0.6 it is 97%. The voiced/unvoiced discrimination (voicing) is also sensitive to the background noise. The voicing DRT scores show a 32% difference between STI = 0.4 and STI = 0.6. Additionally, sibilation is the second most intelligible distinctive feature at STI = 0.4, and it is as intelligible as nasality at STI = 0.6 and STI = 0.8. It can therefore be assumed that the low consonant-to-vowel ratio, and the intelligibility of nasal/oral (nasality) and sibilant/non-sibilant (sibilation) consonants increase the overall intelligibility of English under all room acoustic conditions.

Presumably because of the fact that the Arabic language has a moderately high consonant-to-vowel ratio, it was found to be one of the two least intelligible languages among the four participating languages; however, the results of the language specific DRT proves that the consonant-to-vowel ratio is not the sole reason of low speech intelligibility. In Figure 4.3, it is clear that there is a significant decrease of consonantal intelligibility between STI = 0.6 (no artificial background noise, low reverberation time) and STI = 0.4 (S/N = +5 dB, high reverberation time). For STI = 0.6 the DRT scores for all of the consonants vary between 70% and 95%. For STI = 0.4 the DRT scores decrease significantly to a range between 20% and 60%. It is also observed that the intelligibility of all types of consonants decrease under noisy conditions. Graveness is the most sensitive distinctive feature to background noise. The difference of DRT scores for graveness between STI = 0.4 and STI = 0.6 is 75%. Also, the discrimination between grave-acute consonants decreases by ~75%, and the discrimination between compact-diffuse consonants decrease by ~50%. According to independent sample *t*-tests, the changes in Arabic intelligibility between STI = 0.6 and STI = 0.4 were statistically significant ($p < 0.01$) for graveness [$t(10) = 8.62$, $p = 0.000$], compactness [$t(10) = 3.86$, $p = 0.003$] and mellowness [$t(10) = 4.50$, $p = 0.001$].

In section 4.2.1, it was stated that Mandarin was significantly more intelligible than Arabic and Polish, especially when artificial noise was introduced (STI = 0.4). The language specific DRT results reveal that for Mandarin, the discrimination between airflow/no airflow is high at most conditions (Figure 4.4). Even under noisy conditions such as STI = 0.4 (S/N = +5 dB), the intelligibility of these consonants is as high as 85%; however, this effect diminishes under very high background noise levels (STI = 0.2). Also, at STI = 0.8 the DRT score for airflow/non-airflow consonants is lower than at STI = 0.6, suggesting that too much absorption might reduce the intelligibility of airflow/non-airflow consonants. Figure 4.4 also illustrates that the nasal-oral DRT scores show a similar result to the airflow/non-airflow consonants, in terms of the inverse correlation of the STI and DRT scores. The nasality DRT scores is 66% at the STI = 0.8 condition, 68% at the STI = 0.6 condition and 75% at the STI = 0.4 condition. However, the effect can no longer be seen at STI = 0.2 because of the high artificial background noise. Another distinctive feature which shows an inverse correlation is sustention. Approximately 8% difference can be seen by comparing the sustention DRT results for the STI = 0.6 and STI = 0.8 conditions. However, independent sample *t*-tests showed that none of these changes are statistically significant ($p > 0.05$), so that no conclusions can be drawn from these unexpected variations.

The comparison figures of nasality, compactness, and graveness scores in-between languages are presented in Figure 4.5, Figure 4.6, and Figure 4.7, respectively. Figure 4.5 shows that the nasality scores of English are the highest under all acoustic conditions. The largest difference in-between languages is seen at STI = 0.2, where it is as large as ~80%. At STI = 0.8 the difference is about 35%. The nasality DRT scores of Arabic are better than Mandarin under most acoustic conditions (STI = 0.2, STI = 0.6 and STI = 0.8); however, Mandarin has higher nasality DRT scores at STI = 0.4. Comparison of the compactness DRT scores in-between languages can be seen in Figure 4.6. Compact-diffuse discrimination is better for English compared to Arabic and Mandarin under all room acoustic conditions. Another noticeable result is the large change in compactness DRT scores of Arabic between STI = 0.4 and STI = 0.6 (~45%), which shows that compact/diffuse discrimination in Arabic is sensitive to background noise. The largest difference between the compactness scores of different languages is 45%, which is observed at STI = 0.4. At STI = 0.8, the difference is approximately 10%. Figure 4.7 illustrates the comparison of graveness DRT scores in-between languages. At good room acoustic conditions, such as STI = 0.6 and STI = 0.8, Mandarin graveness DRT scores are the highest, and English are the lowest. The difference between English and Mandarin scores at STI = 0.8 is approximately 20%. The highest score difference can be seen at STI

= 0.4, which is ~40% between English and Arabic. Similar to the Arabic compactness DRT scores, the difference between STI = 0.4 and STI = 0.6 (75%) reveals that grave/acute discrimination in Arabic is also sensitive to background noise.

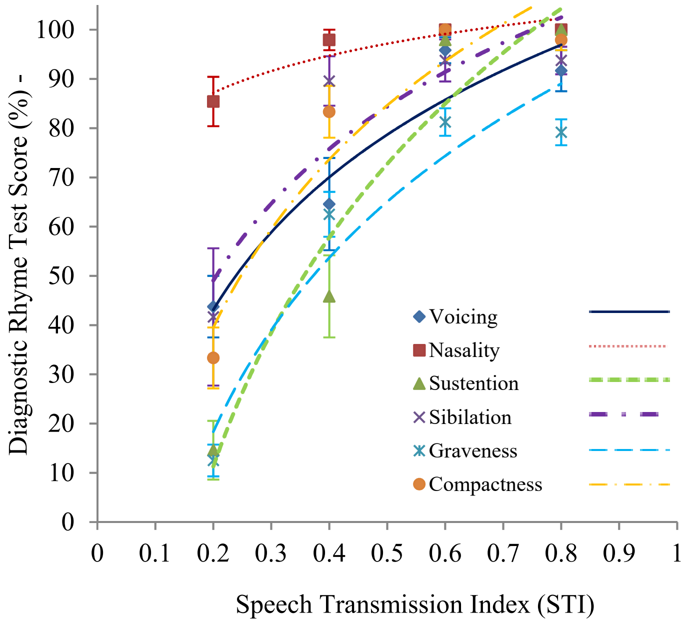


Figure 4.2 Comparison graph of consonantal property scores for English. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

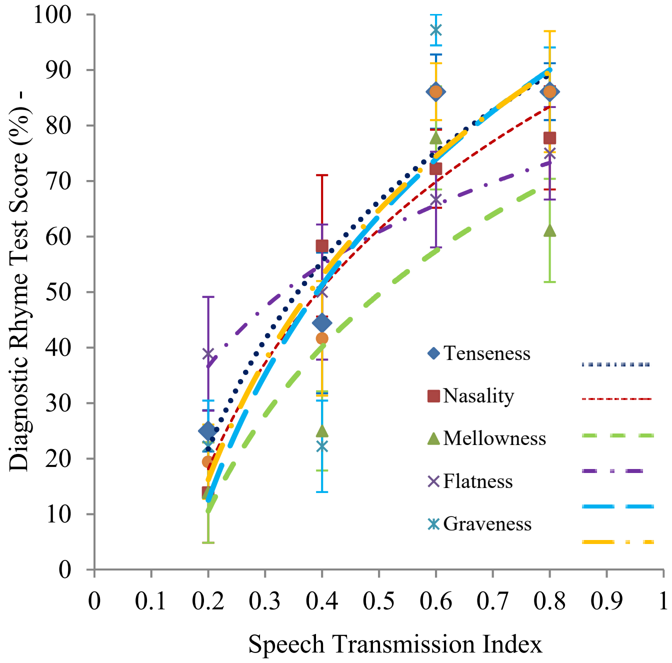


Figure 4.3 Comparison graph of consonantal property scores for Arabic. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

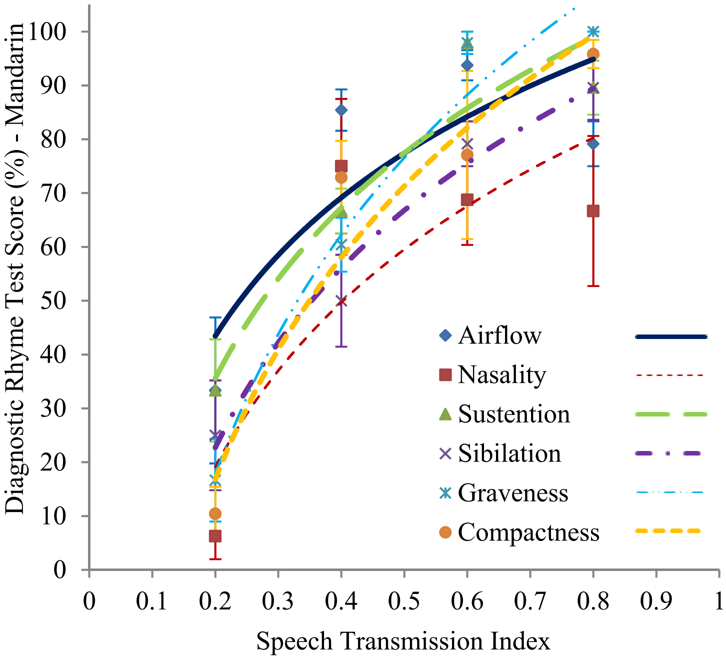


Figure 4.4 Comparison graph of consonantal property scores for Mandarin. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

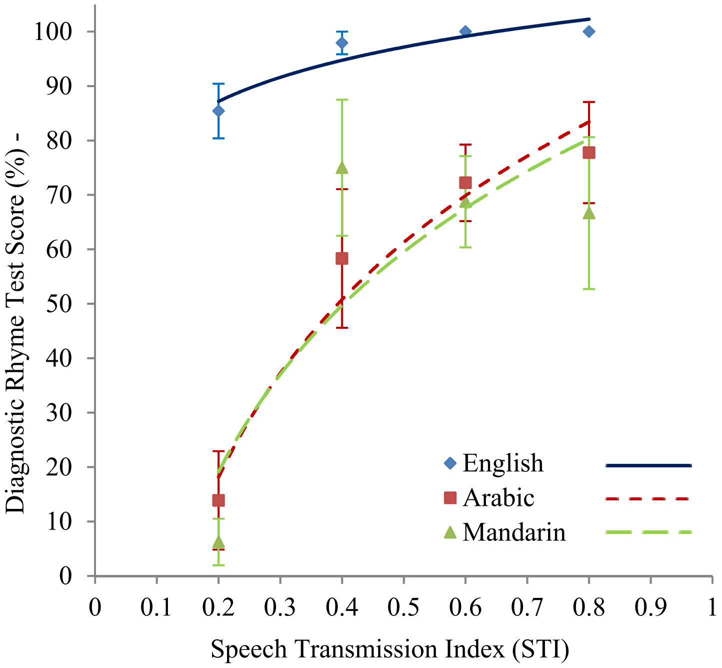


Figure 4.5 Comparison graph of nasality in between languages. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

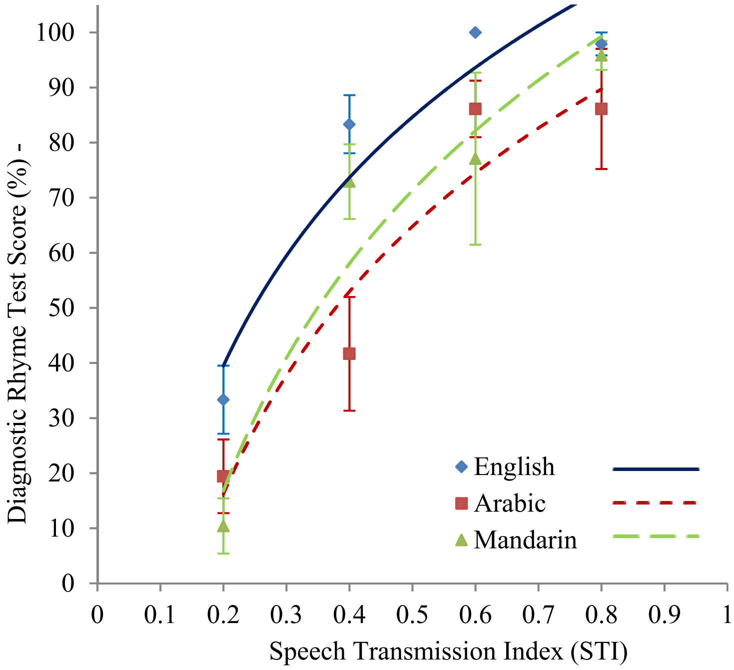


Figure 4.6 Comparison graph of compactness in between languages. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

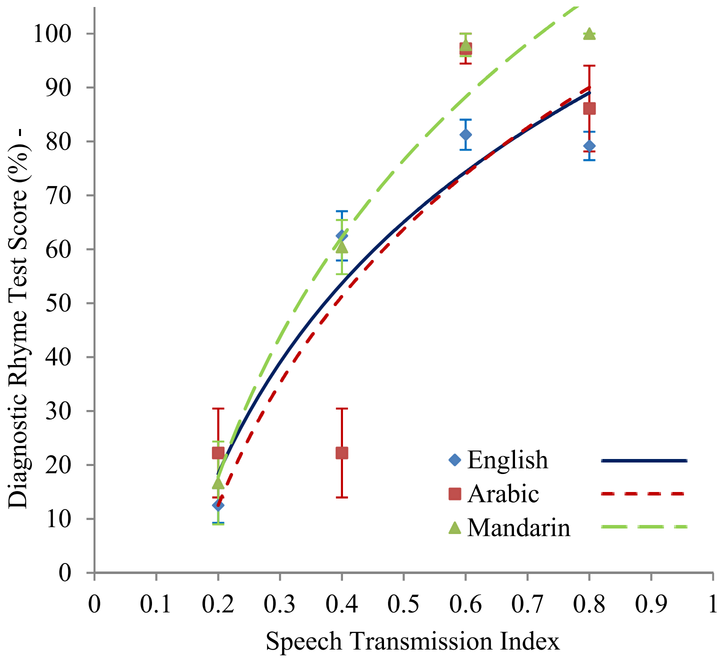


Figure 4.7 Comparison graph of graveness in between languages. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

4.2.2.3 Statistical analysis

For the purpose of statistical analysis, two separate methods have been used. First, in order to analyze the effects of STI and languages on the intelligibility of distinctive features, the factorial ANOVA method was used. By using factorial ANOVA, it has been possible to interpret the individual effects of language and STI, as well as the combined effects of language and STI on the intelligibility of consonants. In order to see if there were any combined effects that could be interpreted as factors, a principal component analysis was also conducted. Detailed descriptions of the statistical methods were given in Chapter 3.

Factorial ANOVA revealed that there was a main effect ($p < 0.01$) of language for nasality [$F(2, 60) = 34.85$, $p = 0.000$], graveness [$F(2, 60) = 6.39$, $p = 0.003$], and compactness [$F(2, 60) = 7.66$, $p = 0.001$], as well as a main effect ($p < 0.01$) of STI conditions for nasality [$F(3, 60) = 22.25$, $p = 0.000$], graveness [$F(3, 60) = 104.25$, $p = 0.000$], and compactness [$F(3, 60) = 47.07$, $p = 0.000$]. Furthermore, an interaction ($p < 0.05$) between languages and STI conditions was found for nasality [$F(6, 60) = 2.40$, $p = 0.038$] and graveness [$F(6, 60) = 6.12$, $p = 0.000$], but not for compactness [$F(6, 60) = 1.59$, $p = 0.166$]. The differences observed might be related to variations in the distinctive features' frequencies produced by the different languages, as well as by differences in the dynamic range of languages; however, at this stage this is just a hypothesis that will need to be verified by further research.

A principal component analysis (PCA) was also conducted. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, with a KMO = 0.82 and all KMO values for individual items were greater than 0.79, which is above the acceptable limit of 0.5 (Kaiser, 1974). An initial analysis was run to obtain eigenvalues for each component in the data. One component had eigenvalues over Kaiser's criterion of 1 and in combination explained 78% of the variance. Three components were retained in the final analysis, which were the three common distinctive features (graveness, nasality, and compactness). This indicates that these three components represent the intelligibility of speech and there is a relationship between these three distinctive features within each language.

4.2.2.4 Discussion

Looking solely at the language specific results would not be sufficient to analyse the overall effects of room acoustic properties on languages. The results should be analysed from a wider perspective by combining overall word intelligibility scores and language specific word intelligibility scores (i.e. distinctive features). The overall word intelligibility results revealed that English was the most intelligible language. It can be assumed that the intelligibility of nasal-oral and compact-diffuse consonants, and the low consonant-to-vowel ratio increased the overall intelligibility of English under all room acoustic conditions. For Mandarin, the intelligibility of airflow/no airflow consonants, and low consonant-to-vowel ratio, made Mandarin highly intelligible under most acoustic conditions. In particular, it was highly intelligible at a relatively low signal-to-noise ratio of +5 dB, but quickly became unintelligible under extreme room acoustic conditions (STI = 0.2). Arabic was the most sensitive language to background noise and more specifically the introduction of artificial noise (STI = 0.4 condition). The language specific DRT results support the overall DRT results. The difference between the STI = 0.4 and STI = 0.6 conditions for graveness scores was as high as 75% for Arabic, and 45% for compactness. Nasal-oral consonants showed the smallest changes in intelligibility for Arabic compared to other distinctive features, with a DRT score difference of ~10% between STI = 0.6 and STI = 0.4 conditions. The results of graveness and compactness suggested that Arabic was more sensitive to high background noise levels compared to the other languages.

To summarise, the results show that there was a significant difference between speech intelligibility of languages and this could be related to their distinctive features.

Main findings:

- The one-way ANOVA results revealed that the word intelligibility scores of languages vary significantly with the Speech Transmission Index (STI) (*p* = .000).
- English was the most intelligible language among the four languages tested not only because it had the lowest consonant-to-vowel ratio, but also because nasal/oral and compact/diffuse consonants were significantly more easy to discriminate in English.
- Mandarin was the second highest intelligible language among the four selected languages, because of the airflow properties of consonants which were significantly intelligible under most acoustic conditions, except STI = 0.2 (S/N = -5 dB).

- Arabic was highly sensitive to background noise. Under noisy room acoustic conditions (STI = 0.2 and STI = 0.4), the intelligibility of all types of consonants decreased significantly.

4.3 Phonemically balanced (PB) sentence test scores

Phonemically balanced (PB) sentence tests are more representative of a real-life situation compared to word intelligibility tests; however, the lack of sensitivity to room acoustic properties is the limitation of such tests (Beranek, 1949). In this section, phonemically balanced sentence test results for English, Arabic, Mandarin, and Polish are presented and analysed. The results of the PB sentence tests are also compared with the word intelligibility results that were presented in the previous sections to analyse the difference of sensitivity to room acoustic conditions. Sentence test scores have been converted into percentages correct scores, and the arithmetic average of all of the participants' results for each room acoustic condition was computed. A more detailed explanation of the scoring procedure was given in Chapter 3.

The number of words in each sentence varied with the language. In the English PB sentence list, there were a minimum of 6 and a maximum of 7 words in each sentence. In the Arabic PB sentence the minimum number of words was 3 and the maximum was 6. In the Mandarin list, all of the sentences had 7 words. Lastly, in the Polish PB list the minimum number of words was 5. A more detailed explanation of sentence lists was given in Chapter 3.

The comparison graph of the four languages' PB sentence test results is presented in Figure 4.8. Also, the comparison graphs of word and PB sentence scores for each language are presented in Figures 4.9, 4.10, 4.11, and 4.12. By looking at the trend lines created from individual PB sentences tests in Figure 4.8, it is seen that Arabic was noticeably less intelligible compared to the other three languages. Furthermore, unlike word scores, English was not noticeably more intelligible than all the other languages. This might be explained by the low predictability sentences used in English. It should also be noted that at STI = 0.4 (high reverberation time, S/N = +5 dB) the variance of intelligibility was the largest. The difference between highest and lowest intelligible language at that point is approximately 40%. As stated in the results of language specific interpretation of DRT results, Arabic had the highest sensitivity to background noise,

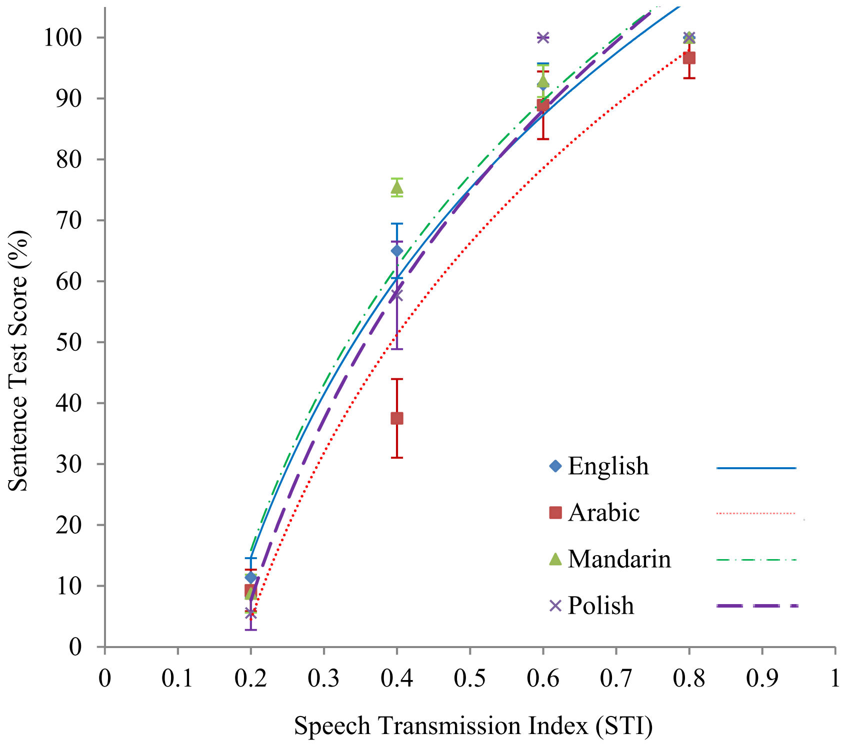


Figure 4.8 Comparison graph between sentence intelligibility scores. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

whereas Mandarin and English were the least sensitive languages to high background noise. At both ends of the trend lines, corresponding to STI = 0.8 and STI = 0.2, the intelligibility difference between languages is smaller than 10%. The difference between lowest and highest PB sentence test scores are larger at STI = 0.4 (~37%) and STI = 0.6 (~15%) compared to STI = 0.2 (~6%) and STI = 0.8 (~5%). Therefore, it can be stated that PB sentence tests are less accurate in identifying differences between languages when the acoustic condition is either very challenging (STI = 0.2), or very good (STI = 0.8).

Factorial ANOVA showed that there was a main effect ($p < 0.05$) of language [$F(3, 80) = 3.87$, $p = 0.012$] and a main effect ($p < 0.01$) of STI conditions [$F(3, 80) = 361.75$, $p = 0.000$] on sentence intelligibility, as well as an interaction ($p < 0.01$) of language and STI conditions [$F(9, 80) = 2.85$, p = 0.006] on sentence intelligibility. These results indicate that both the variation of languages and the variation of room acoustic conditions (STI) affected the sentence intelligibility, and that the amount by which sentence intelligibility changes is governed by both of these factors and cannot be predicted by a single factor. These factorial ANOVA findings are identical to those obtained in the analysis of word intelligibility scores.

However, one-way ANOVA tests carried out for each STI condition, indicated that the sentence intelligibility scores of the four languages examined were significantly different ($p < 0.01$) only at STI = 0.4 [$F(3, 20) = 6.99$, $p = 0.002$], whilst differences were not significant ($p > 0.05$) at STI = 0.8 [$F(3, 20) = 1.00$, $p = 0.413$], STI = 0.6 [$F(3, 20) = 2.07$, $p = 0.137$] and STI = 0.2 [$F(3, 20) = 5.71$, $p = 0.641$]. In other words, the sentence intelligibility of different languages was comparable under most conditions, with the exception of the poor room acoustic condition represented by STI = 0.4.

Further analysis of the sensitivity of PB sentence scores could be achieved by comparing the sentence and word intelligibility scores. The word intelligibility test – PB sentence score comparison graphs illustrate that there is a threshold where word and sentence intelligibility scores intercept (Figures 4.9, 4.10, 4.11, and 4.12). PB sentence tests tend to have higher intelligibility scores than the word tests above the threshold; however, below the threshold the word intelligibility scores are higher than the PB sentence test scores. The STI threshold value for the transition depends on the language. For instance, for English the threshold is STI ≈ 0.6, for Mandarin it is STI ≈ 0.35, for Arabic it is STI ≈ 0.45, and for Polish it is STI ≈ 0.25. The difference between word and sentence intelligibility scores depends on the distance from the threshold value. The threshold can be interpreted as the STI level where context becomes intelligible enough. When the context becomes intelligible, even if not all the words can be understood, context can be transferred from the talker to the listener, and the sentences become 100% intelligible. Below the threshold, the boundary between syllables and words tends to disappear due to the high reverberation time and low signal-to-noise ratio. The lack of word and syllable boundaries decrease the overall intelligibility of speech (Cutler and Butterfield, 1991). Mandarin and Polish have a lower threshold compared to Arabic and English. Because of the varying thresholds observed for different languages, this further suggests that there is no single optimum speech intelligibility level for all of the languages.

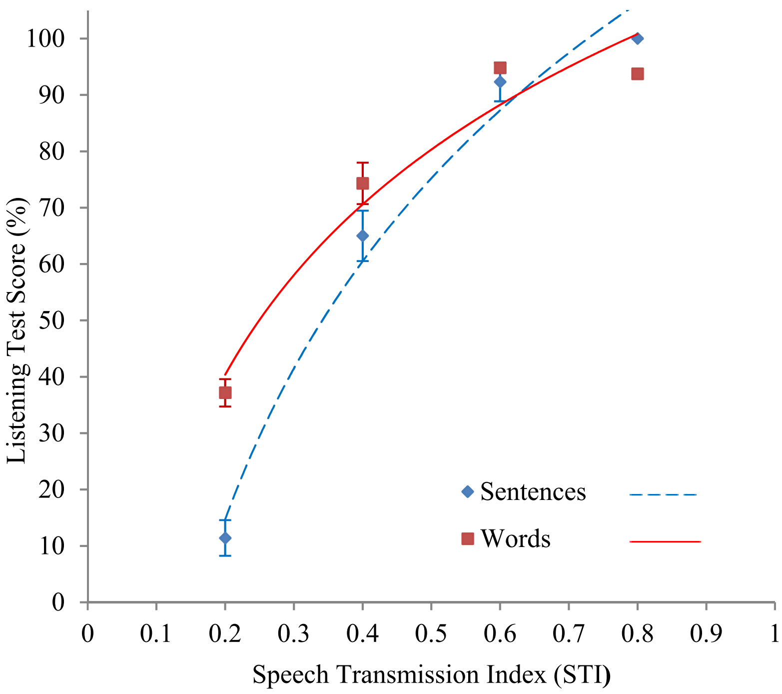


Figure 4.9 Comparison graph between sentence and word intelligibility scores for English. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

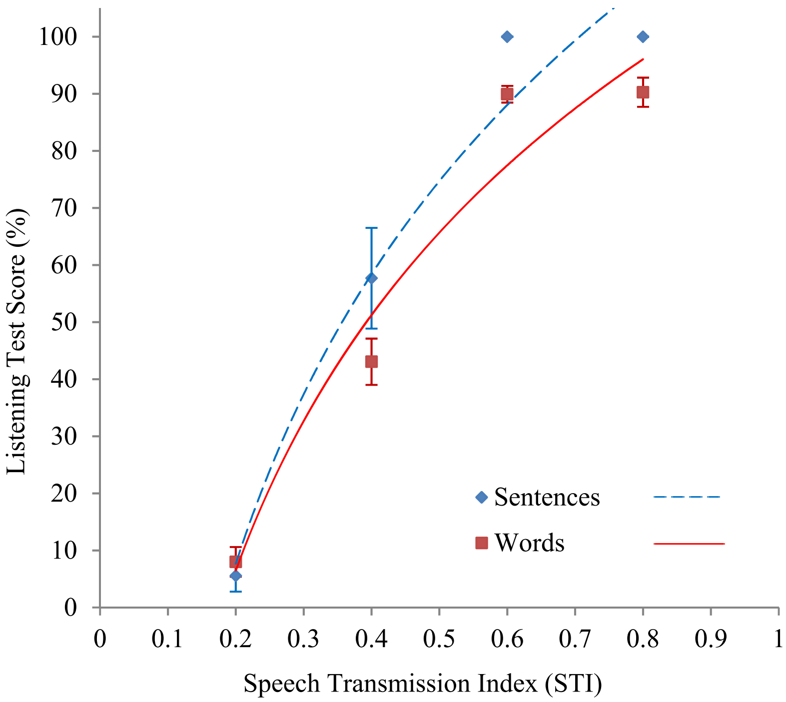


Figure 4.10 Comparison graph between sentence and word intelligibility scores for Polish. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

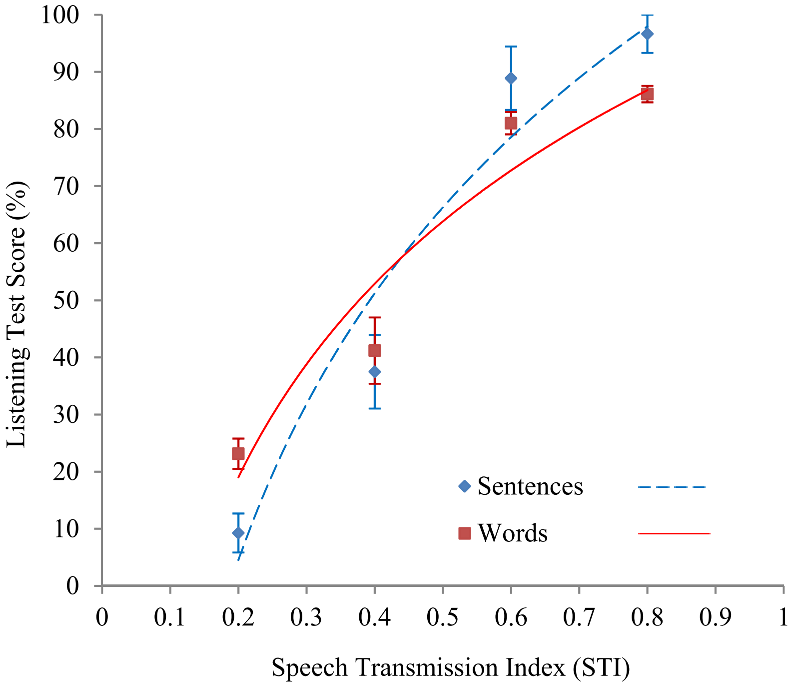


Figure 4.11 Comparison graph between sentence and word intelligibility scores for Arabic. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

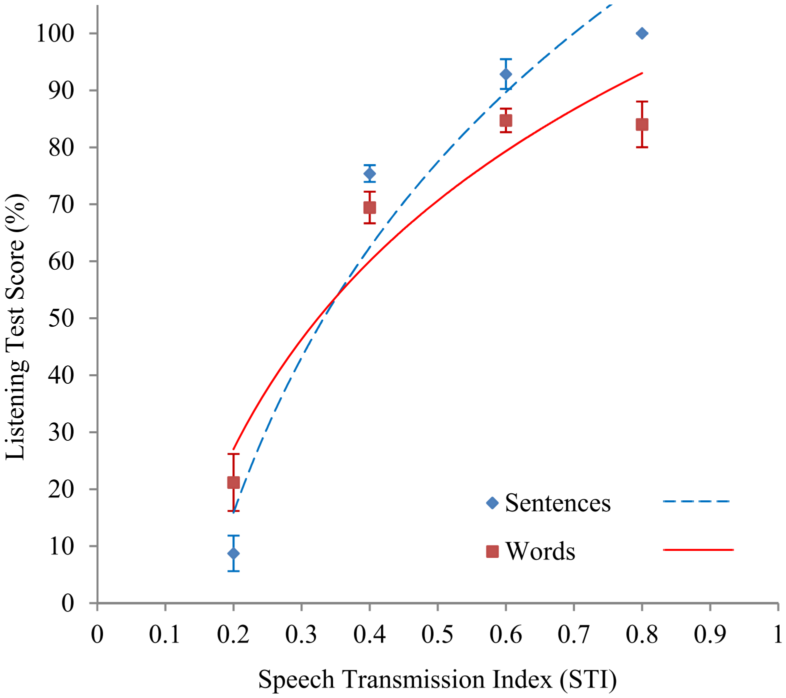


Figure 4.12 Comparison graph between sentence and word intelligibility scores for Mandarin. Actual data markers, standard errors of the means, and logarithmic regression lines are shown in the figure.

After analysing the word score vs. sentence score for English (Figure 4.9), it can be seen that there is a 25% difference at the STI = 0.2 point. This situation may be explained by the fact that the signal-to-noise ratio was too high, so that the significance of context was diminished. As it was stated in the previous section, especially the discrimination of nasal/oral (nasality) and compact/diffuse (compactness) consonants are relatively easy even at highest background noise levels. Additionally, after reaching a language specific threshold, in which the context becomes clear, sentence intelligibility scores tend to increase. At STI = 0.8, sentence intelligibility scores where higher than word intelligibility scores. The positive effects of the context on intelligibility can clearly be seen by looking at the intelligibility score difference of approximately 10%.

Among the four languages, Polish has the lowest threshold for word – sentence intelligibility score (Figure 4.10). At STI = 0.25 and above, sentence intelligibility scores are higher than word intelligibility scores. It is therefore likely that Polish can transmit the context more easily than the other three languages. Sentence intelligibility scores are 100% for both STI = 0.6 and STI = 0.8, and the difference between sentence and word intelligibility scores is ~15%. Similar to Arabic, which is a language with a moderately high consonant-to vowel ratio, there is a significant intelligibility loss when participants are exposed to the artificial background noise. Both sentence and word intelligibility scores drop by ~45% after the participants are exposed to the artificial background noise. At the STI = 0.2 condition, the word intelligibility score is slightly higher than the sentence intelligibility score (Figure 4.10), and both scores are significantly lower than other languages' intelligibility scores at the same room acoustic condition (Figure 4.1 and Figure 4.8).

Figure 4.11 illustrates the difference between word and sentence intelligibility scores for Arabic. At STI = 0.2 the difference between word and sentence intelligibility scores is approximately 15%. Assuming that the score difference occurs due to the lack of context, it can be seen that Arabic sentences rapidly become as intelligible as monosyllabic words at STI = 0.4 point. Additionally, Arabic's high sensitivity to background noise can be seen both in the sentence intelligibility scores, and the word intelligibility scores. The difference of sentence intelligibility scores between STI = 0.4 and STI = 0.6 conditions is as high as approximately 50%. That difference supports the idea of Arabic's high background noise sensitivity. As the context becomes more significant, the sentence intelligibility score increases to close to 100%, with a difference of approximately 10% with the word intelligibility score.

The comparison of Mandarin word and sentence intelligibility scores is presented in Figure 4.12. The threshold of word-sentence intelligibility score for Mandarin occurs at approximately STI = 0.35. Three out of four acoustic conditions tested in the current study are over that threshold, therefore for Mandarin, sentence intelligibility scores are higher than word intelligibility scores with the exception of the STI = 0.2 condition. The sentence intelligibility score is ~15% lower than word intelligibility score at that point. Similarly, at the STI = 0.8 condition, like other languages which have been discussed, sentence intelligibility score is 100%, ~15% higher than the word intelligibility score. Similar to the results of the word intelligibility scores, Mandarin language maintains its intelligibility under high background noise levels according to the sentence intelligibility scores. At the STI = 0.4 level, Mandarin has the highest sentence intelligibility score of 75%, among other participating languages (Figure 4.8). McLaughlin (2008) stated that each Mandarin character is a monosyllabic word, and the monosyllabic properties of the language might be an advantage in intelligibility testing. To some extent this may be interpreted as follows: the positive effects of context on Mandarin speech intelligibility can be seen even under noisy conditions.

In the literature, there are three main studies that compared sentence and word intelligibility test results. First, a comparison between PB word, CVC word, and sentence intelligibility scores was presented in EN 60268-16 (2003). Figure 4.13 shows that sentence intelligibility scores are close to 100% above STI ≈ 0.8 and always the highest above STI ≈ 0.5; however, below this threshold the sentence intelligibility scores rapidly decrease. The sentence intelligibility scores are the lowest below STI ≈ 0.4. It should also be noted that the PB word intelligibility scores are always higher than the CVC word intelligibility scores. Kinsler *et al*. (1962) also presented a comparison between word and sentence intelligibility test results as a function of the signal-to-noise ratio (Figure 4.14). The graph illustrates that the sentence intelligibility test scores are higher than the word intelligibility test scores under all S/N ratios, except S/N < -12 dB. The last study was conducted by Kang (1998). In this study, the sentence and word intelligibility test results for English and Mandarin as a function of STI were presented. The results of sentence intelligibility tests were higher than the results of word intelligibility tests for both languages. Also, the sentence intelligibility score difference between two languages was more conspicuous than the word intelligibility score difference. However, the room acoustic conditions of both past studies are different from the present study. The most challenging room acoustic condition of Kinsler *et al*. (1962) was S/N = -12 dB, which was the only variable. Reverberation time was not considered. Kang (1998) achieved a wide range of STI by using two room acoustic conditions and five seating positions. The first condition was S/N > 25 (the noise source was off), and the second condition was S/N = 2 dB (the noise source was on). Additionally, the early decay times (EDT) were between 0.6s and 1.3s in the corridor and between 0.6s and 0.8s in the seminar room (at 500 Hz - 1 kHz). It can therefore be argued that in both of the previous studies, the room acoustic conditions were not challenging enough to lose the context. In order to see the effects of context on the intelligibility of speech, the room acoustic conditions should be more challenging both in terms of low signal-to-noise ratio and high reverberation time.

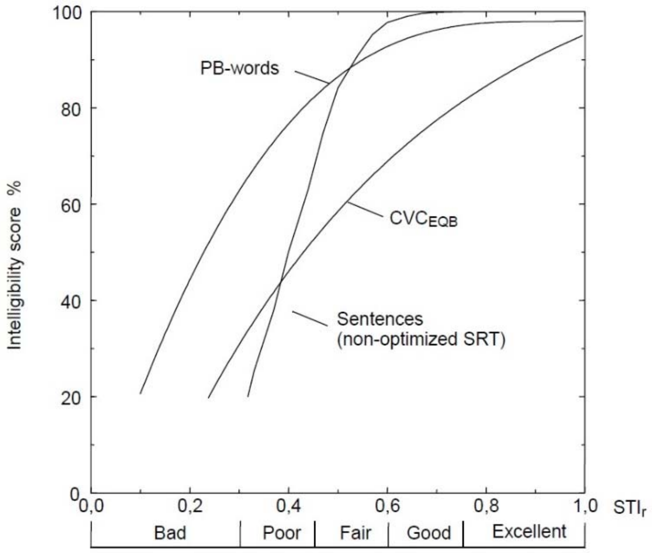


Figure 4.13 Comparison graph of PB-word, CVC word, and sentence intelligibility scores (EN 60268-16, 2003).

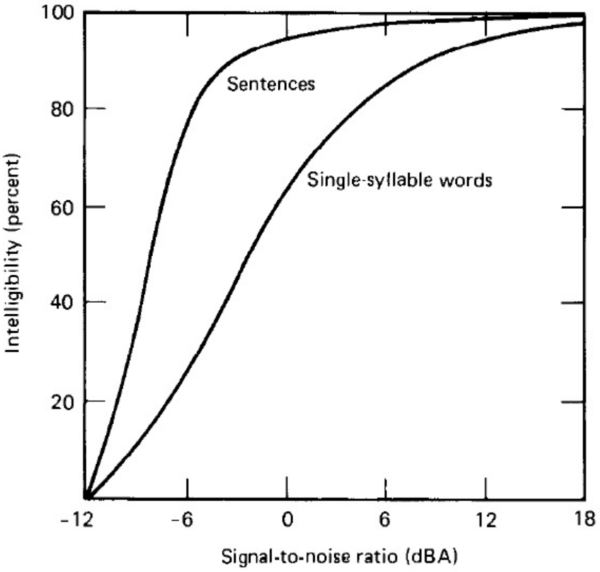


Figure 4.14 Comparison graph of sentence intelligibility and word intelligibility scores (Kinsler *et al.*, 1962).

To summarise, the sentence intelligibility results were more diverse at STI = 0.4 and at STI = 0.6, compared to STI = 0.2 (in which the context is not intelligible) and STI = 0.8 (in which the acoustic condition is very good). However, the relation between word and sentence intelligibility scores might be interpreted in a way that shows the effects of context on the intelligibility of speech. The STI threshold at which sentence intelligibility scores become better than word intelligibility scores can be seen as a point where context of the speech becomes effective on the intelligibility of speech. In this section, all four languages' word and sentence intelligibility scores were analysed and discussed from that perspective. The results of the study suggest that each of the examined languages is significantly different from the other in terms of speech intelligibility. In order to see the cumulative effects of room acoustics, linguistics, and context on multi-lingual speech intelligibility a variety of tools should be used, such as word intelligibility tests, sentence intelligibility tests, and objective speech intelligibility measures.

The smaller variations observed for sentence intelligibility compared to word intelligibility can be explained by the followings: 1) Sentence scores tend to be high under good acoustic conditions, regardless of language (Beranek, 1949); 2) Under very noisy and reverberant conditions the boundaries between syllables can disappear (Butler and Butterfield (1991) and sentence scores can then become very low across all languages. Smaller variations between languages are therefore to be expected for sentence intelligibility at either very good or very challenging room acoustic conditions (STI = 0.8 and STI = 0.2 respectively), justifying the fact that only the STI = 0.4 and STI = 0.6 conditions show comparable variations between the word and sentence scores.

Finally, it is worth pointing out again that sentence intelligibility is influenced by many factors that were not clearly defined in the sentence material used here. Therefore, sentence intelligibility comparisons between languages, as well as comparisons between word and sentence scores should be considered with caution. In particular, the accuracy of word-sentence thresholds is limited due to the unreliability of sentence materials and variations shown by the standard errors of the word intelligibility scores.

Main findings:

- The one-way Analysis of Variance (ANOVA) results suggested that there was a significant difference between the phonemically balanced sentence intelligibility test results and the objective speech transmission index (STI) measurements for each language (p = .000).

- For each language, there was an STI level threshold over which sentence intelligibility scores became better than word intelligibility scores. This threshold might be interpreted as the initial point where the context of a sentence has an effect on the intelligibility of speech.
- By analysing the difference between the STI = 0.6 and the STI = 0.4 levels it can be stated that Arabic was the most sensitive language to artificial background noise in terms of sentence intelligibility.
- Mandarin had the highest sentence intelligibility score when there were moderately high levels of background noise (STI = 0.4).
- Polish had the highest sentence intelligibility scores when there was no artificial background noise (STI = 0.6 and STI = 0.8).

4.4 Discussion and further analysis

This section examines possible reasons for the differences in intelligibility observed between languages. In section 4.2.1, correlations showed that consonant-to-vowel ratios can justify variations observed under poor room acoustic conditions, but not variations observed under good room acoustic conditions. Furthermore, distinctive features identified which types of phonemes are more easily discriminated across languages, but no explanation was given of potential reasons for such differences. Analysis of the spectral content and temporal variability of the speech signals are discussed in this section, to provide a further insight into the differences observed.

First of all, spectral analysis (Figure 4.15) of uninterrupted speech (word test materials used and all talkers included in the signals analysed) indicates that for an identical sound pressure level of 65 dBA, high-frequencies (and in particular 4 kHz and 8 kHz) are more pronounced for English (up to +5 dB). Such high frequencies contribute to the clarity of consonants and might justify the better consonantal discrimination observed for English. By contrast, Arabic has the lowest high frequency content. It should however be noted that spectral content only provides a limited insight into the acoustical properties of languages.

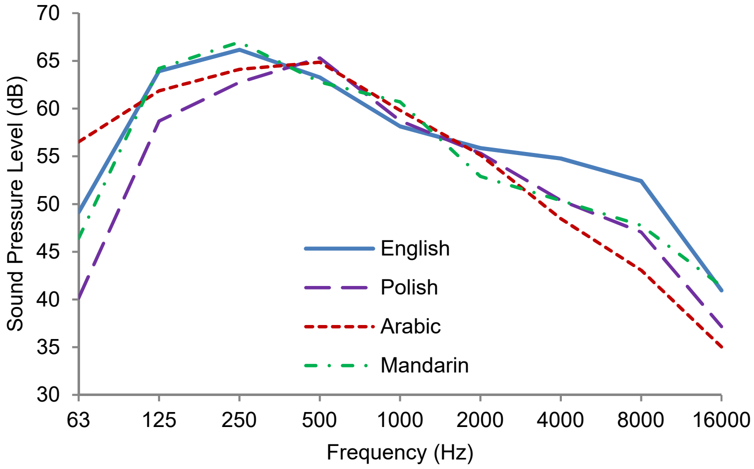
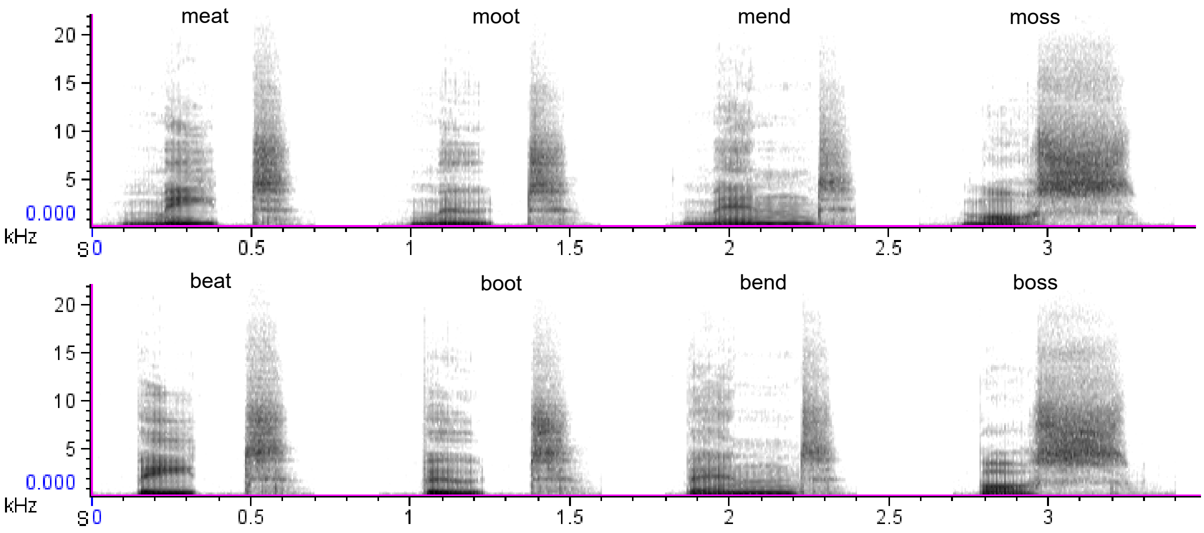
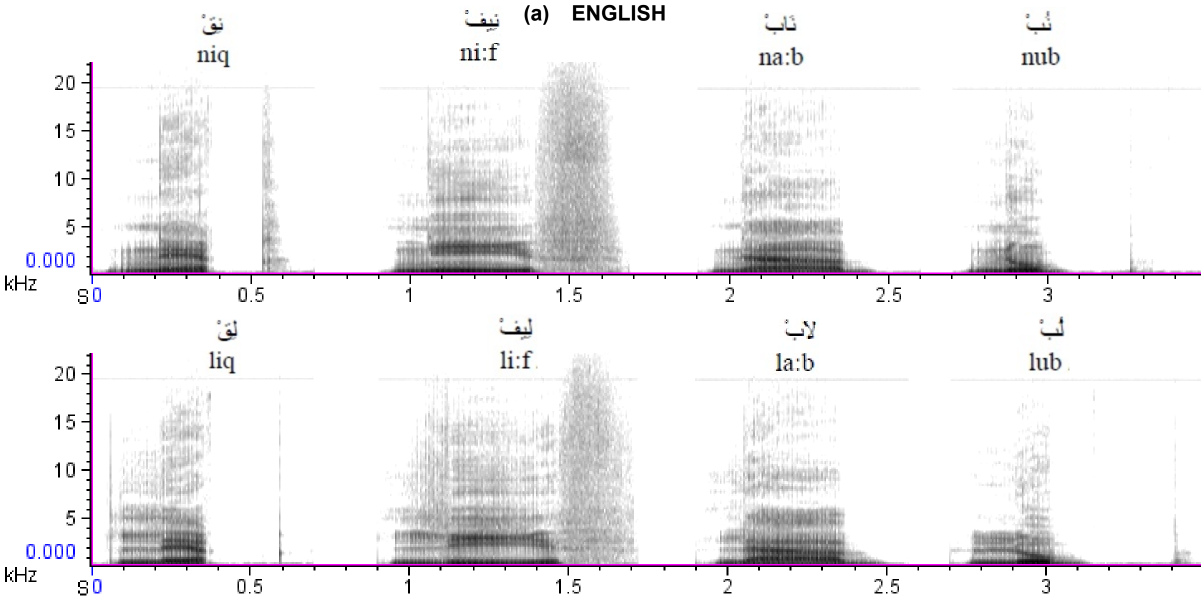


Figure 4.15 Spectra of the languages tested.

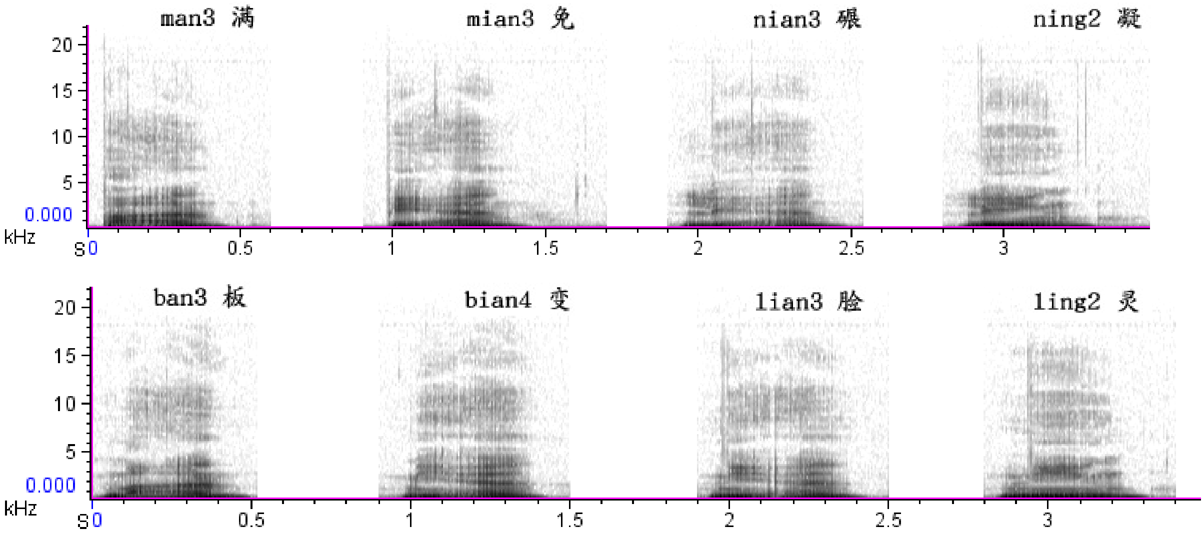
A more in depth analysis can be carried out using spectrograms, which allow examining frequency content, temporal variability and signal amplitude at the same time. This has been done here to compare nasal and oral words, in order to identify possible reasons for the excellent nasality scores observed for English. Spectrograms were produced using the software *RavenLite* and are shown in Figure 4.16., with four words displayed per graph. The words selected represent a wide range of nasal/oral sounds within each language (Voiers, 1977; Boudraa *et al.*, 2008; Li *et al.*, 2000). The spectrograms shown correspond to male speakers, although it can be noted that identical findings were found for female speakers. Most of the English monosyllabic words show a clear drop in high frequency amplitude between their initial and final parts (Figure 4.16(a)), unlike Arabic (Figure 4.16(b)) and Mandarin (Figure 4.16(c)), for which words show a fairly steady amplitude and frequency content. The drops observed in the English words correspond to vowel sounds contained between consonants (CVC sequences used), and could help better discriminate the initial consonants tested in the DRT method. Furthermore, English nasal consonants show an increase in the amplitude of high frequencies towards the beginning of the word, unlike oral consonants, as well as consistently longer durations. By contrast, nasal and oral words of Arabic and Mandarin show similar frequency contents and words' durations. The differences observed for English might help discriminate nasal vs. oral consonants, unlike Arabic and Mandarin that do not exhibit significant differences between the spectrograms of their nasal and oral words. Additional spectral analysis confirmed the spectrograms' findings, as it highlighted a larger temporal variability (quantified by $L_{10} - L_{90}$) for English nasality, especially for frequencies of 2 kHz and above (as much as +10-20 dB at 8 kHz compared to the other languages tested). It can also be noted that the temporal variability of high frequencies of Mandarin was on average



(a) ENGLISH



(b) ARABIC



(c) MANDARIN

Figure 4.16 Spectrograms of four nasal (top) and four oral (bottom) words, for English (a), Arabic (b) and Mandarin (c). The horizontal axis corresponds to time in seconds, while the vertical axis corresponds to frequency in kHz. The darker areas represent larger amplitude of the signal.

higher than what was found for Polish and Arabic (when taking into account all distinctive features). A larger temporal variability means a larger dynamic range, a property that can contribute to better intelligibility by picking up of the higher peaks (Kang, 1998). Spectrograms' analysis of other distinctive features showed that English consistently exhibits a drop in high frequency amplitude in the middle part of its words (vowel sounds), but the duration of this drop tends to be shorter than what was observed for nasality. Furthermore, words' durations were not consistently different in the other distinctive features of English.

To summarise, the better intelligibility of English appears to be justified by its low consonant-to-vowel ratio, its larger high frequency content, as well as its larger temporal variability and dynamic range at high frequencies. Mandarin can also take advantage of an average consonant-to-vowel ratio and fairly high temporal variability at high frequencies, previous work having also pointed out that tonality can improve its intelligibility (Zhang *et al.*, 1981). By contrast, the low word intelligibility of Arabic and Polish appears to be related to moderately high and high consonant-to-vowel ratios respectively, as well as low high frequency content and temporal variations. All of the above findings have been obtained from the acoustical analysis of word test materials used in the present work. In order to confirm these findings, further analysis will need to be carried out on additional test materials as well as on a larger number of speakers.

4.5 Conclusions

In this chapter, the impact of acoustic and linguistic factors on the speech intelligibility of four languages (English, Polish, Arabic and Mandarin) was examined. The study found that there was a significant difference between the word intelligibility scores of these languages. Under the same acoustic conditions (reverberation time and S/N ratio), the word intelligibility scores of each language differed between each other, depending on the linguistic and distinctive features' properties of the languages. For word intelligibility, the differences were found to be statistically significant for all conditions but the excellent room acoustic condition (STI = 0.8), indicating that the word intelligibility of different languages was comparable under excellent room acoustic conditions, but was not comparable under any other condition. The largest difference between word intelligibility scores (33%) was observed at STI = 0.4, in which the listeners were presented to increased reverberation time and artificial background noise. As the acoustic conditions improved, the difference decreased to 9% at STI = 0.8. It was found that distinctive features of the selected languages have an impact on the overall intelligibility, nasal/oral consonants being particularly intelligible in English. Furthermore, a significant correlation was found between the consonant-to-vowel ratios and the word intelligibility scores of languages at poor room acoustic conditions (STI = 0.2 and STI = 0.4). Further analysis suggested that the better intelligibility of English might be justified by its low consonant-to-vowel ratio, its larger high frequency content, as well as its larger temporal variability and dynamic range at high frequencies. Furthermore, it was found that Mandarin can take advantage of a fairly high temporal variability at high frequencies, previous work having also pointed out that tonality can improve its intelligibility (Zhang *et al.*, 1981). English, Arabic and Mandarin were tested using DRT lists, whilst Polish was assessed using PB words, because of the lack of DRT material in Polish. This is a limitation of the current study, although comparisons between DRT and PB words data are normally acceptable (as long as the effect of guesswork is removed from the calculation of DRT scores). It can also be noted that removing Polish from the analysis did not affect the main findings.

In contrast to word scores, sentence scores showed statistically significant differences between languages only at the STI = 0.4 condition, but this was justified by the lower sensitivity of sentence tests to either very good or very challenging room acoustic conditions. Additionally, the comparison between the word and the sentence intelligibility scores revealed that there is a language specific STI threshold over which the context of speech becomes intelligible, therefore increasing the intelligibility of sentences. This threshold was lower for Polish and Mandarin compared to English and Arabic.

Overall, the results of the study revealed that each language is affected differently by room acoustic properties, and these variations are due to differences between the linguistic properties of each language. As the STI is affected by reverberation time and signal-to-noise ratio only, a single STI value might then be insufficient for designing a multi-lingual environment, or even for designing the same type of space within different countries (as previously pointed out by Li et al. (2011)). This is discussed further in Chapter 6.

CHAPTER 5

Soundscape and speech intelligibility of English, Polish, Arabic, and Mandarin

5.1 Introduction

After analysing the room acoustic and linguistic effects on the speech intelligibility of four different languages in Chapter 4, the second phase of the study investigated how soundscape perception might affect speech intelligibility and communication. More specifically, this chapter presents and discusses how soundscape perception might affect the perceived speech intelligibility of English, Polish, Arabic, and Mandarin, by comparing the subjective assessment of three multi-lingual spaces (an airport, a hospital, and a café) tested under three room acoustic conditions (STI=0.4, STI=0.5, and STI=0.6). The airport check-in area was chosen as an example of a high reverberation time and high background noise acoustic environment, which is common in a majority of global cities. The hospital reception area was a medium sized enclosure leading to a medium to low reverberation time, with a relatively low and fairly steady background noise. The café environment was a medium to large sized space, with a moderately-high reverberation time and steady background noise.

15 native speakers per language (a total of 60) participated to semantic listening tests. The semantic descriptors were rated on a five-point scale and the questionnaires used are available in Appendix E. Further details about the recordings that were used, post-processing, the participants, and the test setup were presented in Chapter 3. In this chapter, results of the semantic differential analysis are presented and analysed. A principal component analysis was also conducted on the results of the study, and the results are given in Section 5.3. The results investigate the relationship between socio-cultural backgrounds of the native speakers of the four languages and perceived speech intelligibility in multi-lingual environments.

5.2 Semantic differential analysis

The acoustic environments were evaluated through semantic differential analysis. It is an evaluation survey technique that was developed by Osgood *et al.* (1957), and suggested as a useful method to evaluate sound environments (Kang, 2009). 11 semantic descriptors were suggested for the current study (3 descriptors to evaluate the speech (intelligibility, loudness, and pleasantness) and 8 descriptors to evaluate the acoustic environment (noisiness, annoyance, relaxation, comfort pleasantness, eventfulness, excitement, and familiarity)). All the semantic attributes' results presented in this chapter are based on a -2 to +2 range (e.g.: -2 = very unintelligible and +2 = very intelligible).

Table 5.1 (a) Spearman's correlation coefficient between the intelligibility attribute and the other attributes at the airport (** Correlation is significant at the 0.01 level, * Correlation is significant at the 0.05 level).

Airport	**English**			**Polish**			**Arabic**			**Mandarin**		
	STI=0.4	**STI=0.5**	**STI=0.6**	**STI=0.4**	**STI=0.5**	**STI=0.6**	**STI=0.4**	**STI=0.5**	**STI=0.6**	**STI=0.4**	**STI=0.5**	**STI=0.6**
Loudness	0.172	0.301	0.519*	-0.052	0.147	0.000	0.121	-0.069	0.044	0.205	0.203	0.038
S. Pleasantness	0.291	0.105	-0.212	-0.064	-0.153	0.384	-0.162	0.208	0.271	0.272	0.647**	0.574*
Noisiness	-0.071	0.231	-0.419	-0.015	-0.130	0.000	0.100	0.021	-0.362	0.193	-0.306	0.084
Annoyance	-0.117	-0.86	-0.280	-0.313	-0.154	0.348	0.402	-0.110	-0.299	-0.140	0.154	-0.129
Relaxation	0.331	0.421	0.134	0.44	0.353	0.000	-0.330	0.440	0.334	0.290	0.448*	0.222
Comfort	0.172	0.324	0.321	0.137	0.036	0.330	0.146	0.433	0.267	0.247	-0.116	0.176
E. Pleasantness	0.236	0.534*	0.519*	-0.256	0.311	-0.319	0.016	0.517*	-0.67	0.241	0.433	0.068
Eventfulness	-0.324	-0.049	-0.737**	-0.599**	-0.128	0.115	-0.597**	-0.249	0.277	0.018	-0.063	0.053
Excitement	0.054	-0.002	0.196	0.088	0.356	0.229	-0.226	-0.143	0.389	-0.22	-0.348	-0.119
Familiarity	-0.168	0.093	0.517*	-0.088	0.447*	0.204	-0.250	0.090	0.299	0.160	0.216	0.234

Table 5.1 (b) Spearman's correlation coefficient between the intelligibility attribute and the other attributes at the hospital (** Correlation is significant at the 0.01 level, * Correlation is significant at the 0.05 level).

Hospital	**English**			**Polish**			**Arabic**			**Mandarin**		
	STI=0.4	**STI=0.5**	**STI=0.6**	**STI=0.4**	**STI=0.5**	**STI=0.6**	**STI=0.4**	**STI=0.5**	**STI=0.6**	**STI=0.4**	**STI=0.5**	**STI=0.6**
Loudness	0.169	0,376	0,356	0.434	N/A	0,563*	0.238	0,598**	0,000	-0.201	0,488*	0,038
S. Pleasantness	0.505*	0,209	0,111	0.392	0,302	0,052	0.108	0,457*	0,225	0.258	0,514*	0,536*
Noisiness	-0.017	0,037	0,051	-0.071	0,062	-0,118	0.194	0,156	-0,124	0.169	-0,631**	-0,254
Annoyance	-0.296	-0,805**	-0,099	-0.292	0,000	-0,200	0.154	-0,032	0,216	-0.596**	-0,418	-0,707**
Relaxation	0.000	0,267	-0,040	-0.075	-0,169	N/A	-0.236	0,048	-0,316	0.308	0,518*	0,100
Comfort	-0.55	0,472*	-0,053	0.000	0,390	0,114	-0.637**	0,351	0,125	0.565*	0,375	0,533*
E. Pleasantness	0.051	0,690**	0,423	N/A	0,292	0,483*	-0.158	0,402	-0,152	0.565*	0,266	0,577*
Eventfulness	-0.434	0,114	-0,194	-0.294	-0,402	0,377	-0.129	-0,169	0,365	0.236	-0,228	0,144
Excitement	-0.013	0,250	0,086	-0.228	-0,442*	0,415	-0.129	0,347	0,191	0.289	0,480*	-0,094
Familiarity	0.090	-0,184	0,366	-0.183	-0,088	0,121	0.105	0,390	-0,044	-0.048	-0,034	0,314

Table 5.1 (c) Spearman’s correlation coefficients between the intelligibility attribute and the other attributes at the café (** Correlation is significant at the 0.01 level, * Correlation is significant at the 0.05 level).

Café	**English**			**Polish**			**Arabic**			**Mandarin**		
	STI=0.4	**STI=0.5**	**STI=0.6**	**STI=0.4**	**STI=0.5**	**STI=0.6**	**STI=0.4**	**STI=0.5**	**STI=0.6**	**STI=0.4**	**STI=0.5**	**STI=0.6**
Loudness	0,236	0,438	0,297	0,000	0,038	N/A	0,319	-0,033	-0,103	0,265	-0,282	0,000
S. Pleasantness	0,078	0,288	0,164	0,240	0,052	-0,115	0,076	0,296	0,511*	0,347	0,577*	0,358
Noisiness	-0,236	-0,219	0,000	0,354	0,000	-0,357	0,072	0,494*	0,062	-0,064	-0,140	-0,226
Annoyance	-0,255	-0,477*	0,022	0,189	0,000	-0,293	0,117	0,409	0,266	-0,714**	-0,340	-0,530*
Relaxation	0,166	0,175	0,083	-0,157	0,386	0,422	-0,187	-0,500*	0,000	0,231	0,297	0,000
Comfort	-0,007	0,232	0,022	0,270	0,392	0,218	-0,214	-0,423	0,094	0,437	0,290	0,355
E. Pleasantness	0,047	0,411	0,212	-0,270	0,542*	0,328	0,027	0,000	0,438	0,626**	0,290	0,316
Eventfulness	-0,424	-0,195	0,061	-0,037	-0,134	-0,148	-0,234	0,197	0,234	0,264	0,069	-0,122
Excitement	-0,484*	0,380	-0,083	-0,039	-0,248	0,250	0,176	0,000	0,210	0,396	0,222	0,173
Familiarity	0,028	0,018	0,176	0,054	-0,286	0,006	0,407	0,423	0,004	-0,031	0,402	0,546*

Three types of statistical analysis methods were applied to the listening test results: repeated measures ANOVA, one-way ANOVA, and Spearman’s correlation analysis. Detailed information on the statistical methods were explained in Chapter 3. The repeated measures ANOVA and one-way ANOVA were used in order to test the null hypothesis for the languages, the room acoustic conditions, and the environments, i.e. that these have no effect on the results of the attributes tested. The Spearman’s correlation analyses (one-tailed) were conducted to understand the relationship between the attributes tested and perceived speech intelligibility. The one-tailed correlation analysis is commonly used when a hypothesis is directional (i.e. the direction of change is expected) (Field, 2009). The results of the ANOVAs are presented under each attributes sub-section, and the summaries of the correlation analysis are given upfront in Table 5.1 and Table 5.2.

Table 5.2 Overall Spearman’s correlation coefficient between the intelligibility attribute and the other attributes across all 9 conditions (3 environments x 3 STIs) (** Correlation is significant at the 0.01 level, * Correlation is significant at the 0.05 level).

	English	**Polish**	**Arabic**	**Mandarin**
Loudness	0,522**	0,244**	0,297**	0,376**
S. Pleasantness	0,325**	0,190*	0,355**	0,509**
Noisiness	-0,321**	-0,195*	-0,084	-0,300**
Annoyance	-0,551**	-0,265**	-0,065	-0,461**
Relaxation	0,487**	0,289**	0,110	0,421**
Comfort	0,457**	0,310**	0,182*	0,439**
E. Pleasantness	0,486**	0,167*	0,245**	0,487**
Eventfulness	-0,417**	-0,188*	-0,084	-0,058
Excitement	0,028	0,73	0,017	0,102
Familiarity	0,062	0,43	0,072	0,306**

Prior to the correlation analysis, between subjects reliability was checked by computing the Intra-Class Correlation Coefficient (ICC) for the participants of each language. The average measures ICC analysis revealed that the answers of participants agree with each other for English (ICC = 0.924), Mandarin (ICC = 0.898), Arabic (ICC = 0.912), and Polish (ICC = 0.881), where ICC > 0.720 is usually considered as an acceptable value for social sciences (Shrout *et al.*, 1979).

5.2.1 Intelligibility

Fifteen participants per language (i.e. a total of sixty) were asked to subjectively evaluate intelligibility by answering a five-point semantic scale (from very unintelligible to very intelligible), under three room acoustic conditions, in three digitally simulated multi-lingual environments (i.e., nine cases were rated by each participant). Figure 5.1(a), 5.1(b), and 5.1(c) show the relationship between the intelligibility attribute scores and the STI levels at the airport, the hospital, and the café. It is found out that there are differences between the perceived intelligibility scores of English, Mandarin, Polish, and Arabic. The figures show that the scores vary both between the STI conditions, between the environments and between the languages. Results also show that the perceived intelligibility scores tend to increase as the STI increases.

The largest difference between the languages is observed at the airport and at the café, at STI=0.6. In both cases, the language perceived to be the most intelligible is English (airport=+1.48, café=+1.8) and the language perceived to be the least intelligible is Polish (airport=+0.53, café=+0.8). The smallest difference between scores is observed at the hospital, at STI=0.5, in which English and Arabic are the most intelligible languages with an average score of +1.06, while Polish is the least intelligible language with an average score of +0.8.

Table 5.3 Differences between the lowest and the highest average intelligibility attribute scores of English, Polish, Arabic, and Mandarin at the airport, the hospital, and the café.

	English	Polish	Arabic	Mandarin
Airport	1,52	0,86	1,12	0,92
Hospital	1,59	0,99	0,40	0,93
Café	1	0,47	0,13	0,54

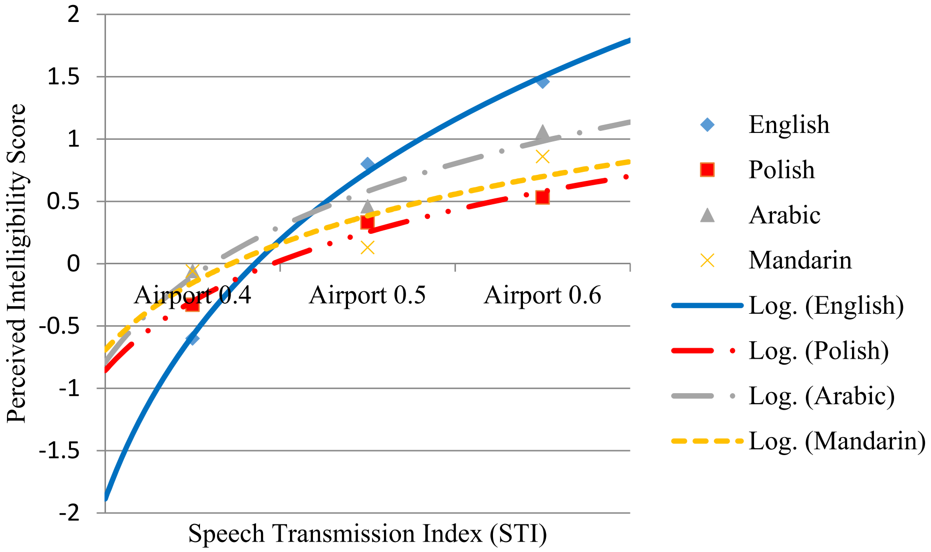


Figure 5.1 (a) Intelligibility attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

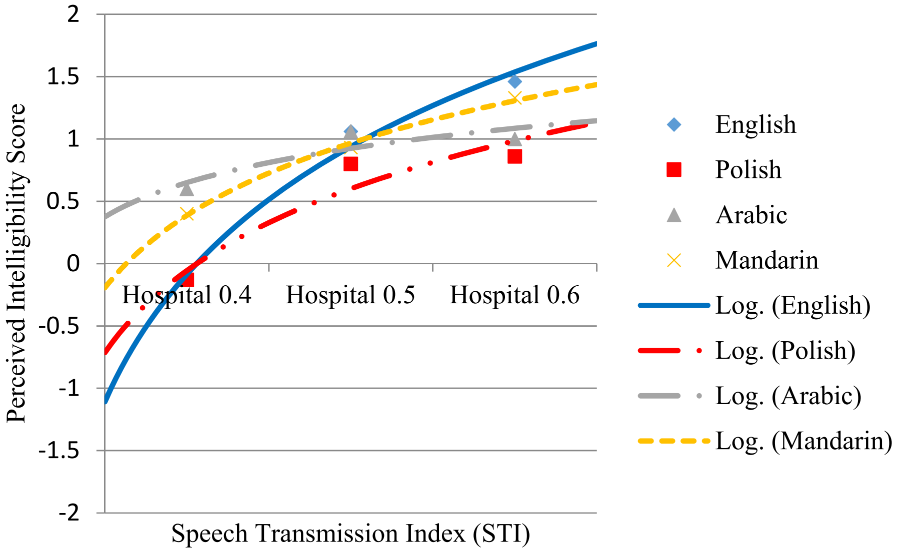


Figure 5.1 (b) Intelligibility attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

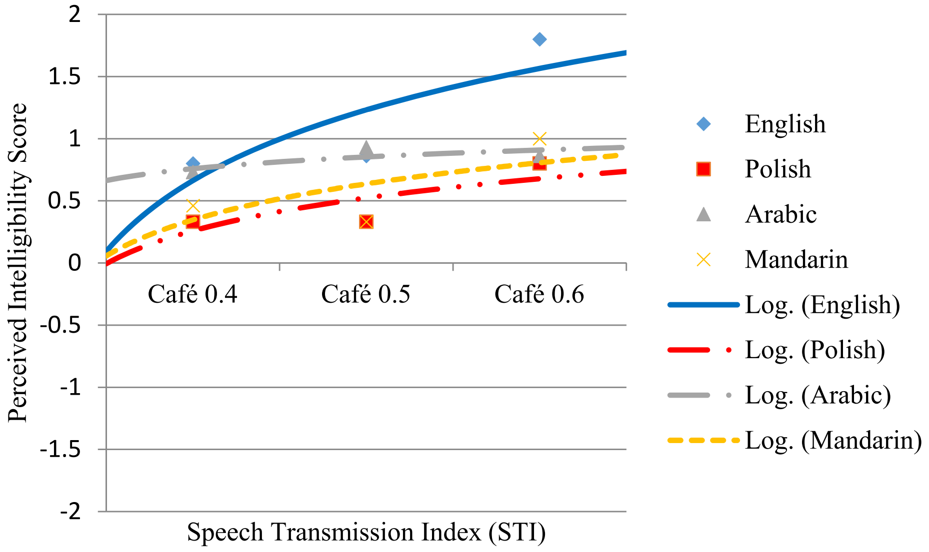


Figure 5.1 (c) Intelligibility attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

The most unexpected result is that English was perceived to be the least intelligible language at STI=0.4 in 2 out of 3 cases (airport=-0.6, hospital=-0.13), but the most intelligible language at the café (+0.8). It is also apparent that English was the most intelligible language at STI=0.5 and STI=0.6 in 5 out of 6 cases, with the exception of the café at STI=0.5 (+0.86). When these results are compared with the results of the first phase of the study, contradictions are observed. For instance, the sentence scores of English from the first phase of the study showed that English was the second most intelligible language at STI=0.4. The word intelligibility scores also revealed that it was the most intelligible language under all the acoustic conditions tested. Table 5.3 presents the differences between the highest and the lowest average intelligibility attribute scores of each language at the airport, the hospital and the café. It is seen that English showed the largest variance for all the environments.

The differences might have occurred due to the fact that the background noise samples that were used in the second phase of the study were representative of a real multi-lingual environment, containing specific distractive noise sources (i.e. public announcement in the airport and phone ringing in the hospital) that are particularly noticeable at STI = 0.4 (i.e. when they are louder). Additionally, the background noise in the airport environment contained public announcements that are in English, which might have been more distracting for native English speakers. Moreover, the sound environment of the café was more relaxing, therefore no specific distractive noise sources were present. In summary, these results might suggest that English is more sensitive to meaningful and distractive sound events that could be caused by a socio-cultural reaction, justifying why the scores at STI=0.4 at the airport and the hospital are lower than at the café. The results at the café do comply with the results of the first phase of the study, arguably because of the steady background noise sample used in that environment (i.e. background noise was more comparable to the white noise used in phase 1), showing that it was the most intelligible language at STI=0.4.

The analysis of the intelligibility attribute scores of Polish revealed that it was the least intelligible language in 7 out of 9 cases, except at the airport at STI=0.4 (-0.33) and STI=0.5 (+0.33), in which it was one of the two least intelligible languages. The average intelligibility attribute scores at the café were the lowest at STI=0.4 (+0.33) and at STI=0.5 (+0.80), and lowest at the hospital at STI=0.6 (+0.86). Furthermore, at STI=0.6, Polish was the least intelligible language at all the environments (airport=+0.53, hospital=+0.86, and café=+0.8). When compared to the results of the first phase of the study, contradictions are again observed. The sentence intelligibility scores from the first phase of the study revealed that Polish was the most intelligible language at STI=0.6; and according to the word intelligibility scores of the first phase of the study, Polish was the second most intelligible language at STI=0.6. However, in the second phase of the study it was the least intelligible language at STI=0.6 for the three environments tested.

The analysis of the intelligibility attribute scores of Arabic revealed that it had the highest average scores at STI=0.4 (+0.73) at the café and at STI=0.4 (+0.60) and STI=0.5 (+1.06) at the hospital. Arabic also had contradictory intelligibility attribute scores when compared with the first phase word and sentence intelligibility scores. The first phase word and sentence intelligibility scores of Arabic were the lowest at STI=0.4 and STI=0.6; however, the intelligibility attribute scores of Arabic in the second phase were the highest in 2 out of 3 cases at STI=0.4 (airport=-0.06, hospital=+0.6), with the exception of the café (+0.73), in which it had the second highest intelligibility attribute score. The rankings were low at STI=0.6, but not the lowest, as Arabic had the second highest intelligibility attribute scores at the airport (+1.06), and the second lowest intelligibility attribute scores at the hospital (+1.0) and at the café (+0.86). Similar to the scores at STI=0.4, it had the highest intelligibility attribute scores in 2 out of 3 cases at STI=0.5 (hospital=+1.06, café=+0.93), with the exception of the airport (+0.46). Table

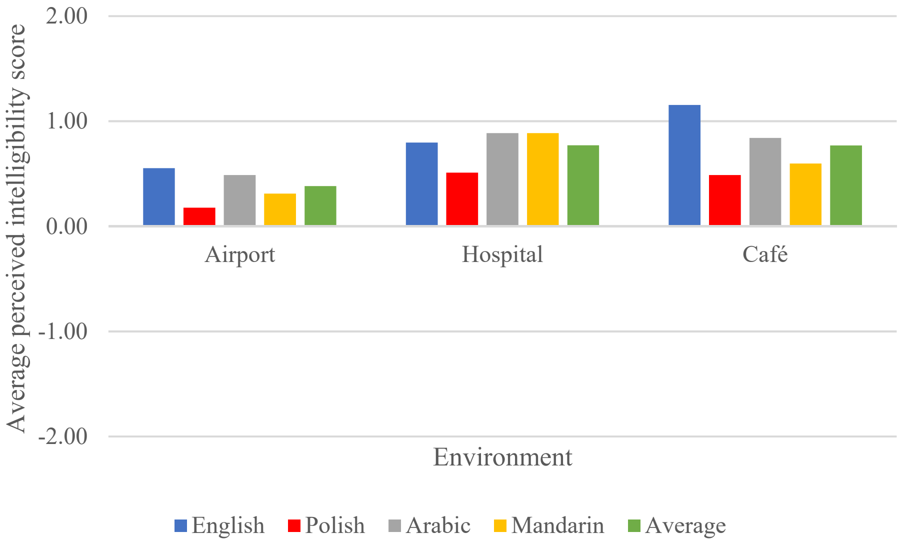


Figure 5.2 The average perceived intelligibility attribute scores of the 3 room acoustic conditions for each environment and language.

5.3 shows that Arabic had the lowest variance of the intelligibility attribute scores at the hospital (0.4) and the café (0.13); therefore, it can be assumed that the change in the room acoustic conditions did not affect significantly the subjective ratings of speech intelligibility as it did in the first phase of the study, where it was found to be much less intelligible at STI=0.4.

Analysis of the average intelligibility attribute scores of Mandarin revealed that it had the lowest average scores at STI=0.5 at the airport (+0.13), and at STI=0.5 at the café (+0.33). It was the most intelligible language only at STI=0.4 at the airport (-0.06). The largest difference between average scores was observed at the hospital (0.93), and the smallest difference was observed at the café (0.54) (Table 5.3). In the first phase of the study, Mandarin had the highest sentence intelligibility scores and the second highest word intelligibility scores at STI=0.4, which complies with the average intelligibility attribute scores obtained for the airport.

Figure 5.2 shows the average perceived intelligibility scores of the 3 room acoustic conditions for each environment and language. It is seen that the language with the highest average perceived intelligibility scores at the airport and the café was English (airport=0.55 and café=1.15). The average scores of Polish were always the lowest between languages (airport=0.18, hospital=0.51, and café=0.49). Additionally, Figure 5.2

shows that the average perceived intelligibility score of Polish was the lowest at the airport (0.38).

The repeated measures ANOVA revealed that the effects of speech transmission index (STI) [$F(1.7,56.5) = 82.35$, $p = 0.000$] and environment [$F(1.9,108.3) = 17.24$, $p = 0.000$] on the intelligibility attribute were statistically significant. Additionally, marginal significance [$F(3,56) = 2,75$, $p = 0.051$] was observed for variations between the speech intelligibility of different languages. Furthermore, combined effects of STI and language [$F(1,56) = 4490.73$, $p = 0.000$], environment and STI [$F(3.41,191.08) = 6.71$, $p = 0.003$], and environment, language, and STI [$F(10.23,191.08) = 2.30$, $p = 0.013$] on the scores of the intelligibility attribute were statistically significant.

The differences between the subjective scores of intelligibility of the four languages were statistically analysed for each room acoustic condition by using the one-way Analysis of Variance (ANOVA) method. The confidence interval was set to 95% for the one-way ANOVA. Three out of nine conditions showed significant differences ($p < 0.05$) between languages. These were the airport – STI=0.6 [$F(3,59) = 5.29$, $p = 0.003$], the café – STI=0.5 [$F(3,59) = 3.80$, $p = 0.015$], and the café – STI=0.6 [$F(3,59) = 6.31$, $p = 0.001$].

Overall, the analysis of the intelligibility attribute scores revealed that the subjective evaluation of speech intelligibility of each language varies, depending on the type of the environment, the type of the background noise, reverberation time, and signal-to-noise ratio. Perceived intelligibility of English appeared to be mostly influenced by the information carried in the background noise, for instance public announcement systems.

5.2.2 Loudness (Speech)

The next semantic attribute to be analysed is the loudness of speech. The participants were asked to evaluate loudness of speech by answering a five-point semantic scale (from very high speech level to very low speech level), under three room acoustic conditions, in three digitally simulated multi-lingual environments (i.e., nine cases were rated by each participant. Figure 5.3(a), 5.3(b), and 5.3(c) show the relationship between the loudness attribute scores and the STI levels at the airport, the hospital, and the café. The largest difference between the loudness attribute scores of the four languages was observed at the hospital at STI=0.6 (0.67) and the smallest difference was observed at the hospital at

STI=0.4 (0.02) (Figure 5.3(b)). It should be stressed that a large variation was also observed at the airport at STI=0.4 (0.60). The smallest variation at the airport was seen at STI=0.6 (0.19) (Figure 5.3(a)). Results show that the speech loudness scores tend to increase as the STI increases.

When the score variations are compared between the airport and the hospital, it can be seen that these include conditions that shows the largest and the smallest variation changes. Perceived loudness at the airport varies more at STI=0.4 (0.6); however, at the hospital the largest variance was observed at STI=0.6 (0.67). The main differences between the two environments were the reverberation times and the signal-to-noise ratios; therefore, this suggests that there might be a relationship between the perceived speech loudness of each language and the room acoustic properties: the perceived speech loudness of different languages might vary depending on the room acoustic properties (i.e. reverberation times and signal-to-noise ratios) at a given STI condition. The results suggest that attention should be given to individual room acoustic properties in order to achieve homogeneous levels of perceived speech loudness in multi-lingual environments.

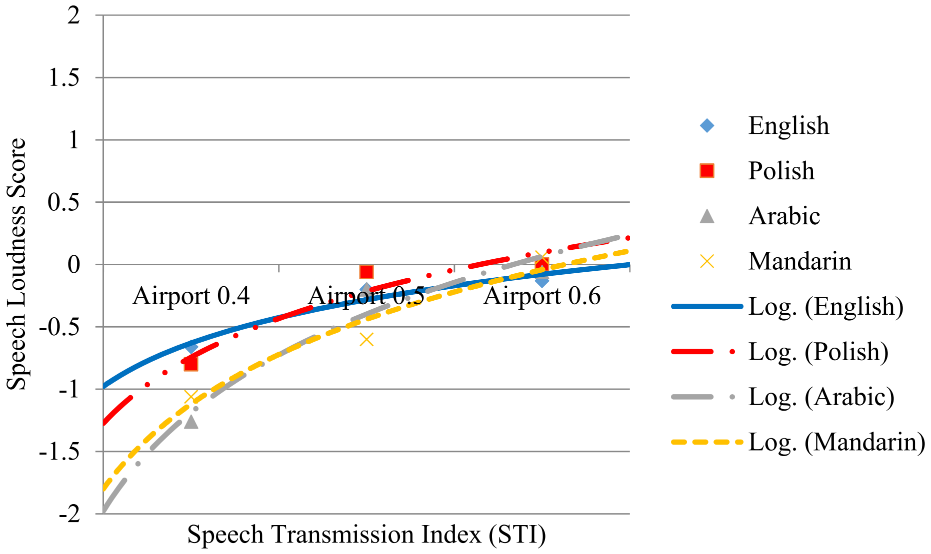


Figure 5.3 (a) Loudness attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

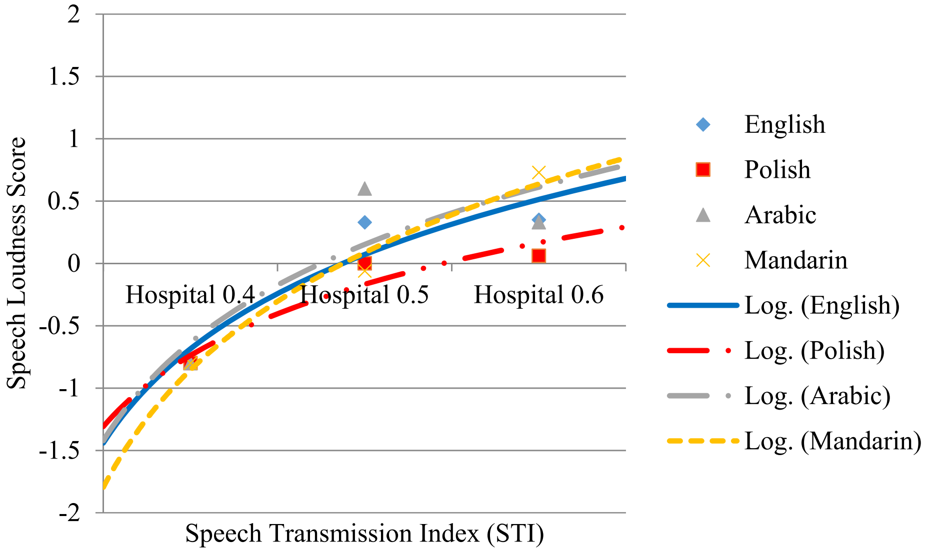


Figure 5.3 (b) Loudness attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

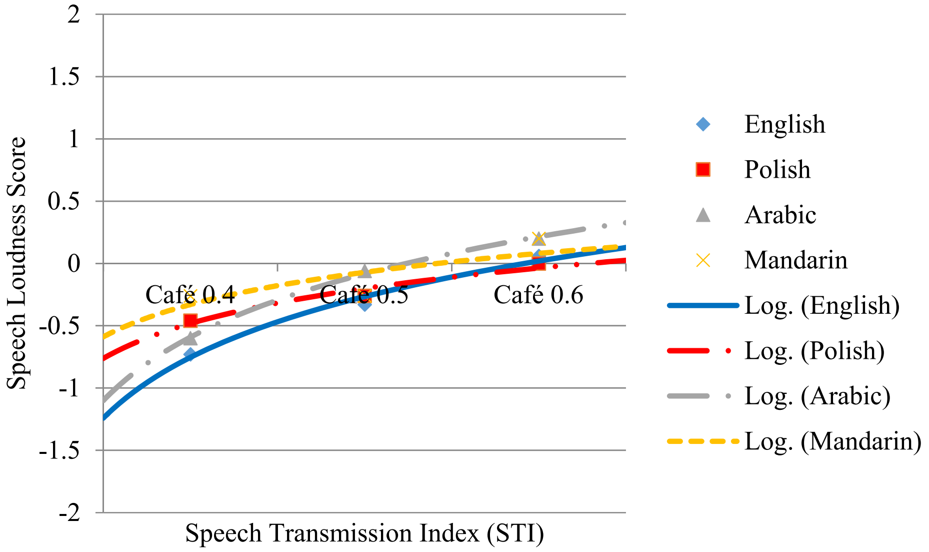


Figure 5.3 (c) Loudness attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

Figure 5.4 illustrates the average loudness attribute scores of the 3 room acoustic conditions for each environment and language, where it can be seen that values tend to be negative (i.e. speech loudness tended to be rated below a normal level). It is seen that the average loudness attribute scores of English were the highest at the hospital (-0.03), and the lowest at the café and the airport (-0.33). The Polish scores were the highest at the hospital and the café (-0.24), and the lowest at the airport (-0.28). The Arabic scores were the highest at the hospital (+0.04), and the lowest at the airport (-0.50). Lastly, similar to the Arabic scores, the Mandarin scores were the highest at the hospital (-0.04), and the lowest at the airport (-0.53). It should be noted that the variance of the average scores of Arabic and Mandarin in between the environments were larger when compared to the English and Polish scores. The difference between the lowest average scores of Arabic was 0.54 and Mandarin was 0.49, whereas the difference was 0.3 for English and 0.04 for Polish. These results suggest that the perceived speech loudness varied with the environment for Arabic and Mandarin more than for English and in particular Polish.

The repeated measures ANOVA revealed that the effects of speech transmission index (STI) [$F(1.78,100.00) = 96.48$, $p = 0.000$] and environment [$F(1.92,107.81) = 17.08$, $p = 0.000$] on the speech loudness attribute were statistically significant. Additionally, combined effects of environment and language [$F(5.77,107.81) = 2.84$, $p = 0.014$], and environment and STI [$F(3.76,210.54) = 5.28$, $p = 0.001$] were statistically significant on perceived speech loudness.

The differences between perceived speech loudness of the four languages were statistically analysed for each room acoustic condition by using the one-way ANOVA method, for which the confidence interval was set to 95%. Only two out of nine conditions showed significant differences ($p < 0.05$) between languages. These were the airport – STI=0.5 [$F(3,59) = 3.05$, $p = 0.036$] and the hospital – STI=0.5 [$F(3,59) = 4.05$, $p = 0.011$].

The correlation analysis given in Table 5.2 indicates that the loudness attribute scores were significantly correlated with the perceived intelligibility attribute scores of all four languages tested ($p < 0.01$). This suggests that perceived speech loudness directly affects perceived intelligibility, as expected. Additionally, the correlation analysis was conducted for each individual room acoustic condition (Table 5.1). The correlation analysis results for English indicate that there was a significant positive correlation at the airport at STI=0.6 ($p < 0.05$), for Polish at the hospital at STI=0.6 ($p < 0.05$), and for Arabic ($p < 0.01$) and Mandarin ($p < 0.05$) at hospital at STI=0.5. It should be noted that there were no significant correlations between the intelligibility attribute and the speech loudness attribute at STI=0.4, as well as at the café. Due to the fact that the correlations between these two attributes were significant only in 4 out of 36 cases, the significant correlations might be false positives.

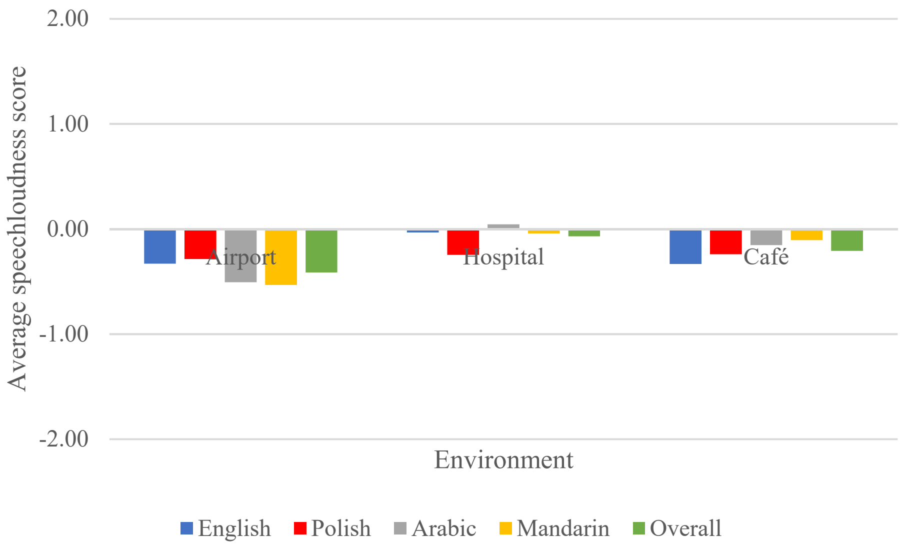


Figure 5.4 The average loudness attribute scores of the 3 room acoustic conditions for each environment and language.

The analysis of the speech loudness attribute scores showed that there is a difference between the perception of speech loudness of the four languages, especially at the hospital at STI=0.5 (maximum difference of 0.66) and at STI=0.6 (maximum difference of 0.67). The one-way ANOVA between the scores was statistically significant at STI=0.5 at the airport and hospital ($p < 0.05$). Another outcome of the analysis was that the variance between the perception of speech loudness in between languages were largest at STI=0.4 and decreased as the STI increased at the airport. However, at the hospital the difference between the languages was the lowest at STI=0.4 and increased as the STI increased. At the café, the score difference was approximately the same across the STI conditions tested. This might be caused by the different types of background noise samples used in the recordings. As mentioned in Chapter 3, the background noise sample that was used in the café environment was free of any intelligible speech samples and other distractive noise sources; however, there were PA samples at the airport and phone ringing samples at the hospital. Consequently, the scores of the café, the environment with the steadiest background noise, showed the least variance of score difference in between the STI conditions, while the results of the airport and the hospital showed variable scores with different trends. Additionally, the variance between the average loudness attribute scores of the three STI conditions for each environment were larger for Arabic and Mandarin than for English and especially Polish, revealing that the type of the environment has a more pronounced effect on the perceived speech loudness for the former two languages.

5.2.3 Pleasantness (Speech)

The next semantic attribute to be analysed is the pleasantness of speech. The participants were asked to subjectively evaluate speech pleasantness by answering a five-point semantic scale (from very pleasant to very unpleasant), under three room acoustic conditions, in three digitally simulated multi-lingual environments (i.e., nine cases were rated by each participant). Figure 5.5(a), 5.5(b), and 5.5(c) show the relationship between the speech pleasantness attribute scores and the STI levels at the airport, the hospital, and the café. Results show that the speech pleasantness scores tend to increase as the STI increases. The largest difference between the speech pleasantness attribute scores of the four languages was observed at the airport at STI=0.5 (0.92) (Figure 5.5(a)), in which English was the highest (+0.26) and Mandarin was the lowest (-0.66). The smallest difference was observed at the café at STI=0.4 (0.00) (Figure 5.5(c)), in which the scores of all four languages were equal (+0.13).

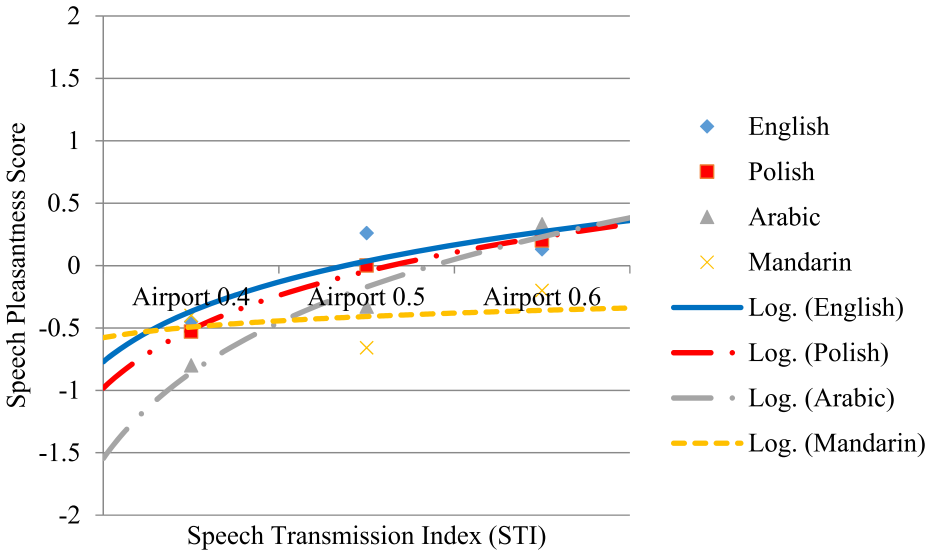


Figure 5.5 (a) Speech pleasantness attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

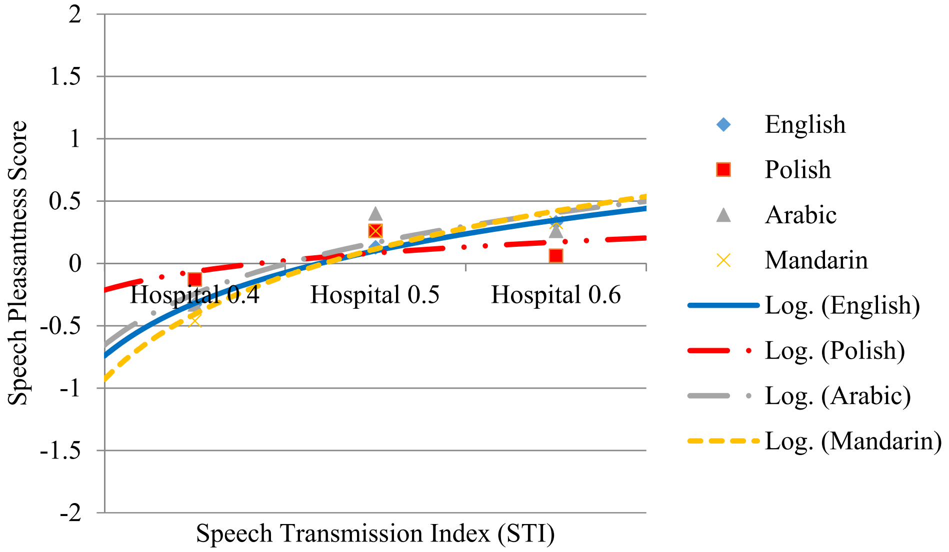


Figure 5.5 (b) Speech pleasantness attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

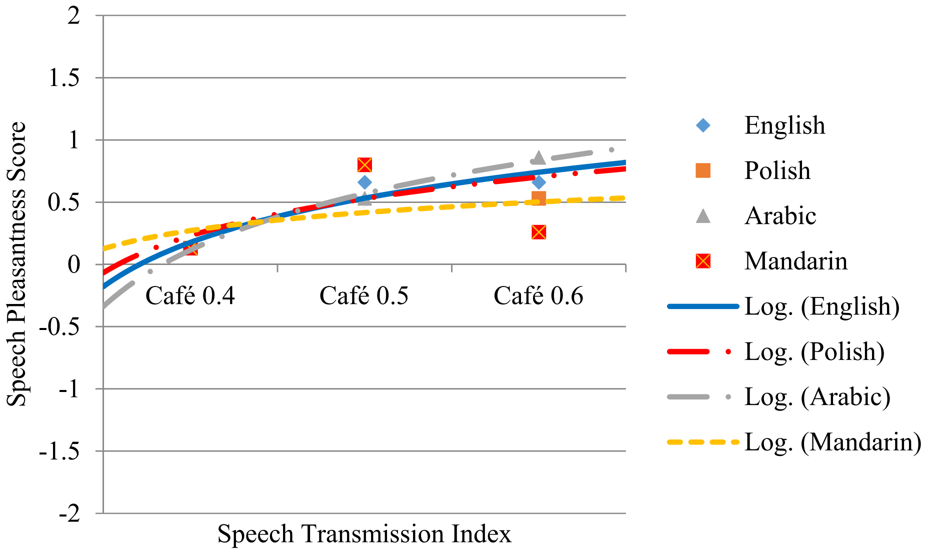


Figure 5.5 (c) Speech pleasantness attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

The variance between the scores did not show any clear pattern, suggesting that the effects of languages are not significant for speech pleasantness. In order to examine the effects of the environments and the context of speech, the averages of all three STI conditions were calculated and are presented in Figure 5.6.

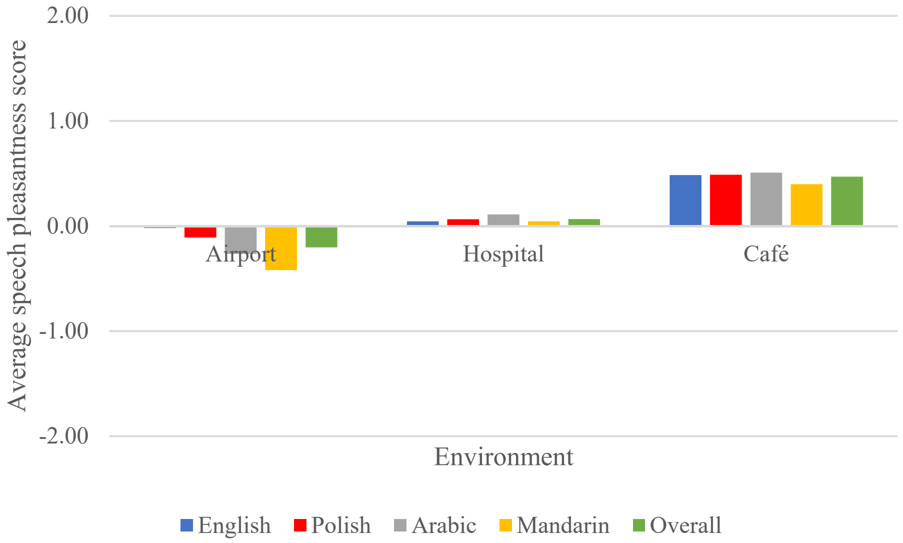


Figure 5.6 The average speech pleasantness attribute scores of the 3 room acoustic conditions for each environment and language.

The results show that native speakers of all languages rated the speech at the café as the most pleasant (English=+0.48, Polish=+0.48, Arabic=+0.50, Mandarin=+0.39), followed by the hospital (English=+0.04, Polish=+0.06, Arabic=+0.11, Mandarin=+0.04), and the airport (English=-0.02, Polish=-0.11, Arabic=-0.26, Mandarin=-0.42), respectively. The results do not suggest any relationship between the languages and the speech pleasantness attribute scores, indicating that speech pleasantness was mainly affected by the environment and the context of speech.

The repeated measures ANOVA revealed that the effects of speech transmission index (STI) [$F(1.50,84.29) = 28.23$, $p = 0.000$] and environment [$F(1.60,90.10) = 21.11$, $p = 0.000$] on the speech pleasantness attribute were statistically significant; however, neither the effects of the languages, nor the combined effects of the three variables were statistically significant. Additionally, according to the one-way ANOVA results, the differences between the four languages' speech pleasantness attribute scores were only significant at the airport at STI=0.5 [$F(3,59) = 4.10$, $p = 0.011$], in which the observed score variation was the largest. English had the highest average speech pleasantness attribute score (+0.26), followed by Polish (0.00), Arabic (-0.33), and Mandarin (-0.66), respectively.

The correlations between the intelligibility attribute and the speech pleasantness attribute were statistically analysed by running Spearman's correlation analysis. The correlation analysis results indicate that the speech pleasantness attribute scores were significantly correlated with the perceived speech intelligibility attribute scores of all four languages tested (English: $p < 0.01$, Polish: $p < 0.05$, Arabic: $p < 0.01$, and Mandarin: $p < 0.01$) (Table 5.2). Additionally, the correlation analysis was conducted for each individual room acoustic condition. The results show that there was a significant positive correlation for Arabic at the hospital at STI=0.5 ($p < 0.05$) and at the café at STI=0.6 ($p < 0.05$), and for Mandarin at the airport at STI=0.5 ($p < 0.01$) and STI=0.6 ($p < 0.05$), at the hospital at STI=0.5 ($p < 0.05$) and at STI=0.6 ($p < 0.05$), and at the café at STI=0.5 ($p < 0.05$) (Table 5.1). Although the total number of correlations between the intelligibility attribute and the speech pleasantness attribute was only 8 out of 36, a difference between the languages could still be observed. The number of significant correlations was 5 for Mandarin, whereas Arabic had 2, English had 1, and Polish had no significant correlations.

Overall, the speech pleasantness attribute was not affected by the languages; however, it was affected by the type of the environment, which suggested that there is a relationship between the context of speech and the speech pleasantness attribute scores.

5.2.4 Noisiness

The next semantic attribute to be analysed is noisiness. The participants were asked to subjectively evaluate the noisiness of the acoustic environment on a 5-point scale (from very noisy to very quiet). Figure 5.7(a), 5.7(b), and 5.7(c) show the relationship between the noisiness attribute scores and the STI levels at the airport, the hospital, and the café. Results indicate that the noisiness scores tend to decrease as the STI increases. The largest difference between the noisiness attribute scores of the four languages was observed at the airport at STI=0.6 (0.87) (Figure 5.7(a)), in which Polish was the highest (+1.00) and Arabic was the lowest (+0.13). The smallest difference was observed at the hospital at STI=0.5 (0.13) (Figure 5.7(b)), in which English was the highest (+0.73) and Arabic was the lowest (+0.60).

First, Figure 5.7 shows that there was a difference between the noisiness attribute scores of the four languages. For English, the scores of the noisiness attribute was the highest in six out of nine cases, with the exceptions of the airport – STI=0.5 (+0.93) and STI =0.6

(+0.86), and the hospital – STI=0.6 (+0.40). However, the English intelligibility attribute scores at these conditions were the highest among all four languages, indicating that the noisiness attribute scores and the intelligibility attribute scores of English were contradicting, as lower noisiness would be expected for high intelligibility scores. In section 5.2.1 it was revealed that English was the least intelligible language in 2 out of 3 cases at STI=0.4 and the most intelligible language in 5 out of 6 cases at STI=0.5 and STI=0.6 combined; however, the scores of the noisiness attribute were the highest in 6 out of 9 cases. It can therefore be suggested that perceived noisiness is not negatively related with perceived intelligibility for native English speakers.

It is also important to note that the perceived noisiness was lowest in all cases for Arabic listeners with no exceptions, indicating that Arabic participants were more tolerant to background noise. As mentioned in Section 5.2.1, Arabic had contradictory results with the results of the first phase of the study (see Chapter 4), in which the Arabic word and sentence intelligibility scores decreased by approximately 40% and 50%, respectively, when the artificial background noise was first introduced to the listeners. Therefore, it can be assumed that the difference between artificial and realistic noise sources is important in such intelligibility tests. The contradiction between the first phase and second phase results suggest that Arabic listeners were not sensitive to the realistic noise sources as much as the artificial white noise that was used in the first phase of the study. This might be caused by the socio-cultural habits of the Arabic listeners, due to the fact that realistic background noise samples were more familiar to the listeners compared to the artificial background noise, eventually increasing the perceived intelligibility of speech under noisy conditions.

The average noisiness attribute scores across the three STI conditions were calculated for each environment and language (Figure 5.8). The differences between the lowest and highest average scores of English was 0.29, Polish was 0.35, Arabic was 0.11, and Mandarin was 0.31. Figure 5.8 shows that the average noisiness attribute scores of all four languages were the lowest at the hospital. The airport was considered the noisiest by Polish, Arabic, and Mandarin listeners, while the café was considered the noisiest by the English listeners.

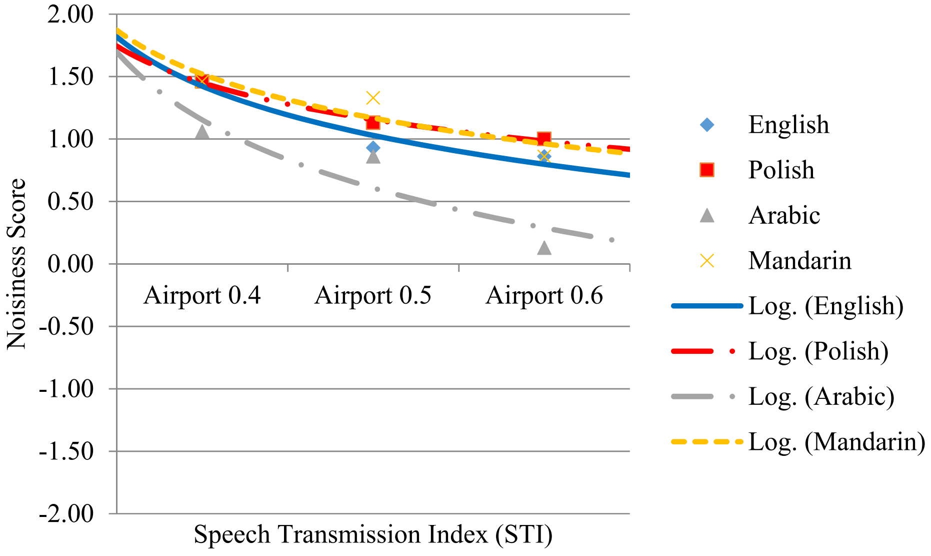


Figure 5.7 (a) Noisiness attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

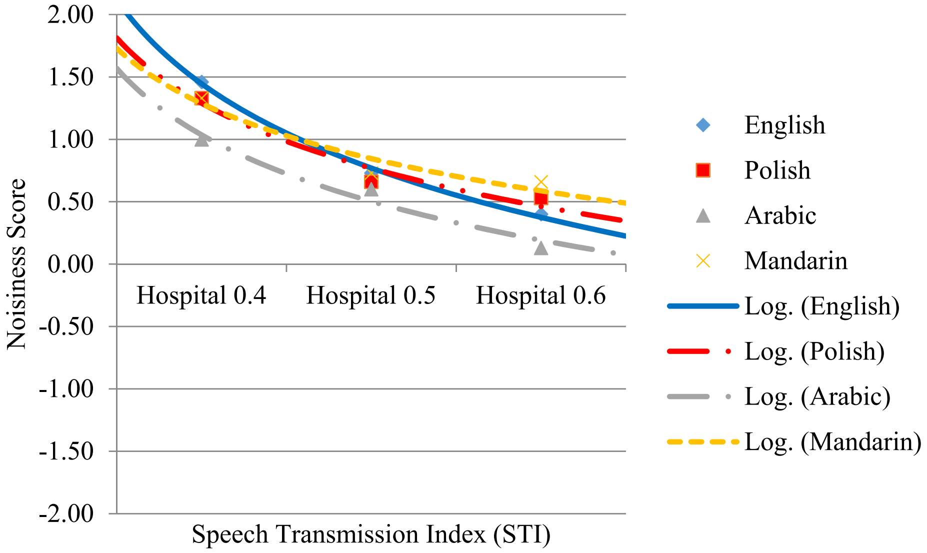


Figure 5.7 (b) Noisiness attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

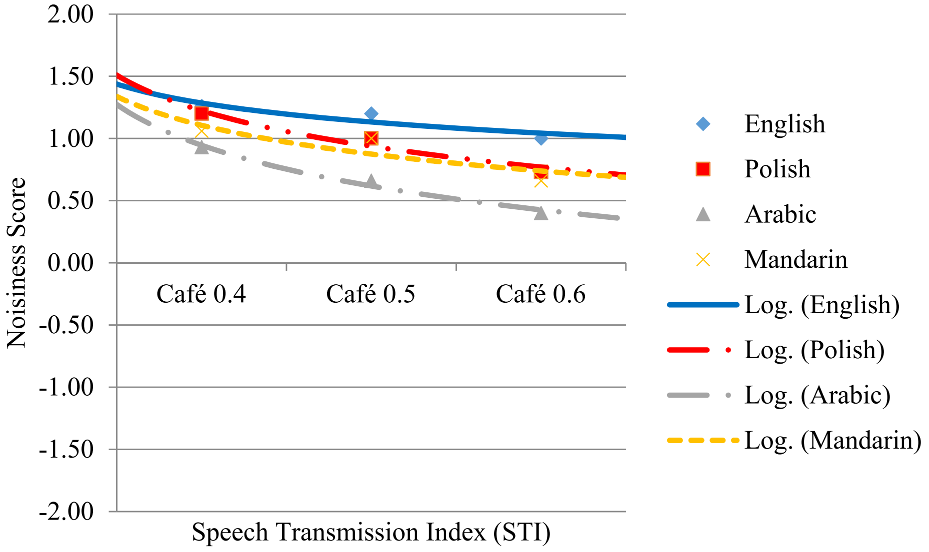


Figure 5.7 (c) Noisiness attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

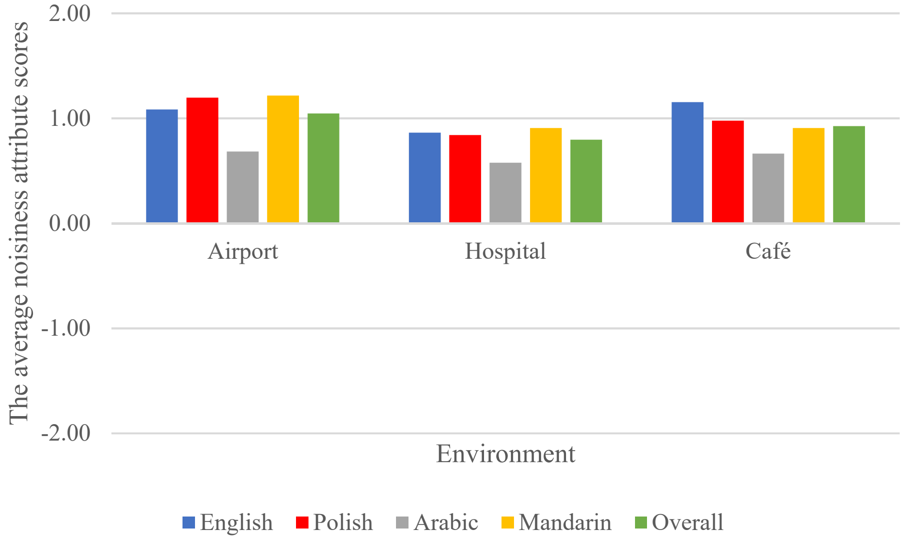


Figure 5.8 The average noisiness attribute scores of the 3 room acoustic conditions for each environment and language.

The repeated measures ANOVA revealed that the effects of languages on the noisiness attribute scores was statistically significant [*F*(3,56) = 2.82, *p*=0.047], along with the effects of speech transmission index (STI) [*F*(1.86,104.49) = 79.98, *p*=0.000] and the environments [*F*(1.97,110.71) = 5.40, *p*=0.006]. According to the one-way ANOVA results, the differences between the four languages' noisiness attribute scores were only significant at the airport at STI=0.6 [*F*(3,59) = 5.61, *p*=0.002], in which the observed score variation was the largest. In this case Polish had the highest average noisiness attribute score (+1.00), followed by English and Mandarin (+0.86), and Arabic (+0.13), respectively.

The correlations between the intelligibility attribute and the noisiness attribute were statistically analysed by running Spearman's correlation analysis (one-tailed). The analysis revealed that the correlations between the noisiness attribute scores and the perceived intelligibility attribute scores were statistically significant for English ($p < 0.01$), Polish ($p < 0.05$), and Mandarin ($p < 0.01$). The correlation was not statistically significant for Arabic (Table 5.2). This result further supports the hypothesis of the Arabic listeners' resilience to realistic background noise sources. Furthermore, the correlation analysis for each acoustic condition revealed that there were only 2 significant correlations out of 36 cases; a negative correlation was found for Mandarin at the hospital at STI=0.5 ($p < 0.01$), and a positive correlation was found for Arabic at the café at STI=0.5 ($p < 0.05$) (Table 5.1).

An interesting result of the study is that, although the differences between the four languages' noisiness attribute scores were only significant in 1 out of 9 cases, the repeated measures ANOVA revealed that the effect of languages on the noisiness attribute scores was statistically significant, along with the effects of speech transmission index (STI) and the environments. This can be explained by the fact that the variation between the scores of the four languages were not consistent. For instance, the variance between the results of the Polish and the Mandarin listeners was not as large as the variance between the English and Arabic listeners. The repeated measures ANOVA revealed that there was an effect of languages on the variation of the noisiness attribute scores, and the one-way ANOVA results showed that the effect was not consistent among the four languages.

The analysis of the noisiness attribute scores revealed that English listeners were the most sensitive to the background noise. The noisiness attribute scores of the Arabic listeners were the lowest, contradicting the interpretations of the first phase results (i.e. the large decrease of Arabic word and sentence intelligibility scores, when the listeners were first introduced to the artificial background noise) (see Chapter 4). The correlation analysis of the nosiness attribute scores and the perceived intelligibility attribute scores further support the hypothesis of the Arabic listeners' resilience to realistic background noise.

5.2.5 Annoyance

The next semantic attribute to be analysed is annoyance. The participants were asked to subjectively evaluate the annoyance of the acoustic environment on a 5-point scale (from very annoying to very favorable). Figure 5.9(a), 5.9(b), and 5.9(c) show the relationship between the annoyance attribute scores and the STI levels at the airport, the hospital, and the café. Results indicate that the annoyance scores tend to decrease as the STI increases. The largest difference between the annoyance attribute scores of the four languages was observed at the café at STI=0.4 (1.13) (Figure 5.9(c)), in which the scores of English was +1.46, Arabic was +0.66, Polish was +0.53, and Mandarin was +0.33. The smallest difference was observed at the hospital at STI=0.6 (0.20) (Figure 5.9(b)), in which the average annoyance attribute score of English was -0.06, Polish was 0.00, and Arabic and Mandarin was -0.20.

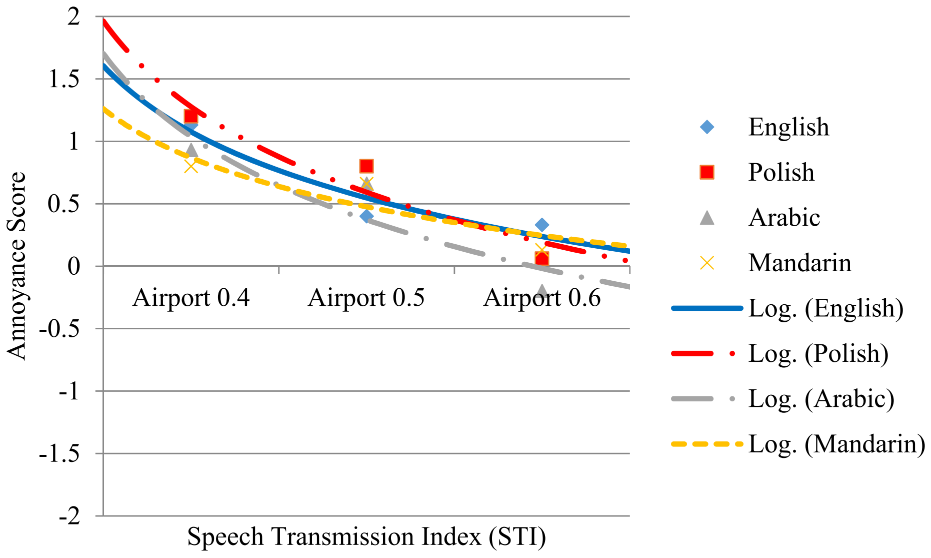


Figure 5.9 (a) Annoyance attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

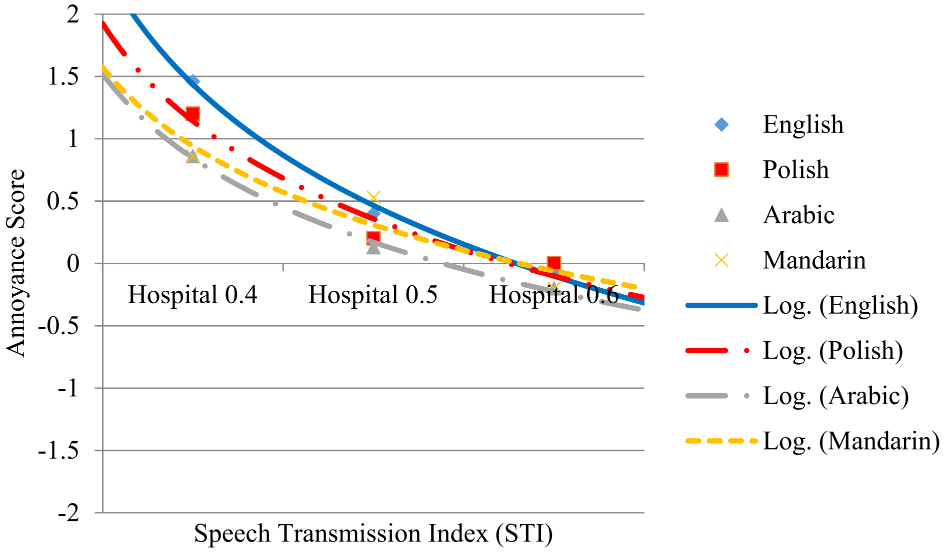


Figure 5.9 (b) Annoyance attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

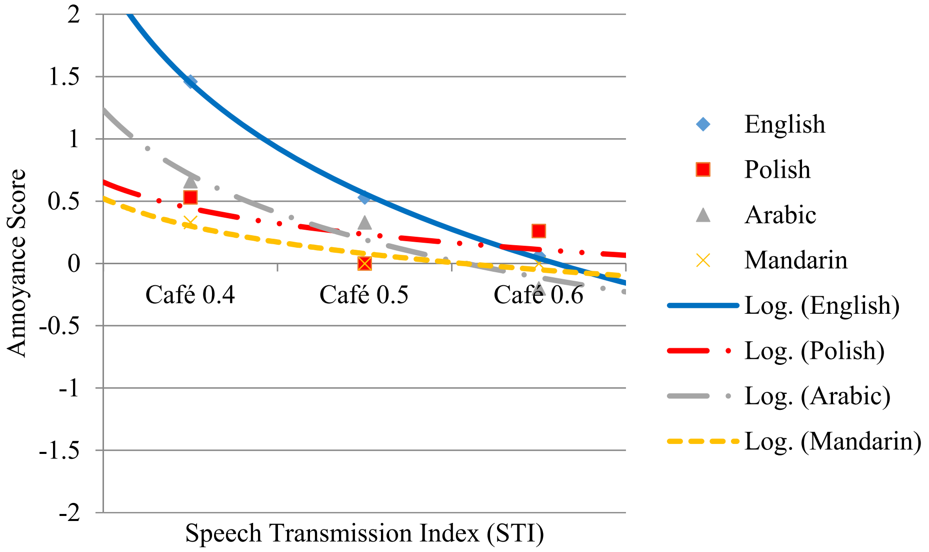


Figure 5.9 (c) Annoyance attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

After inspecting Figure 5.9, it is observed that English and Polish annoyance attribute scores were the highest in 4 out of 9 cases, whereas Mandarin and Arabic scores were the lowest in 5 out of 9 cases. These results can be interpreted as the English and Polish listeners being more annoyed by the overall acoustic conditions of the environments compared to the Mandarin and Arabic Listeners.

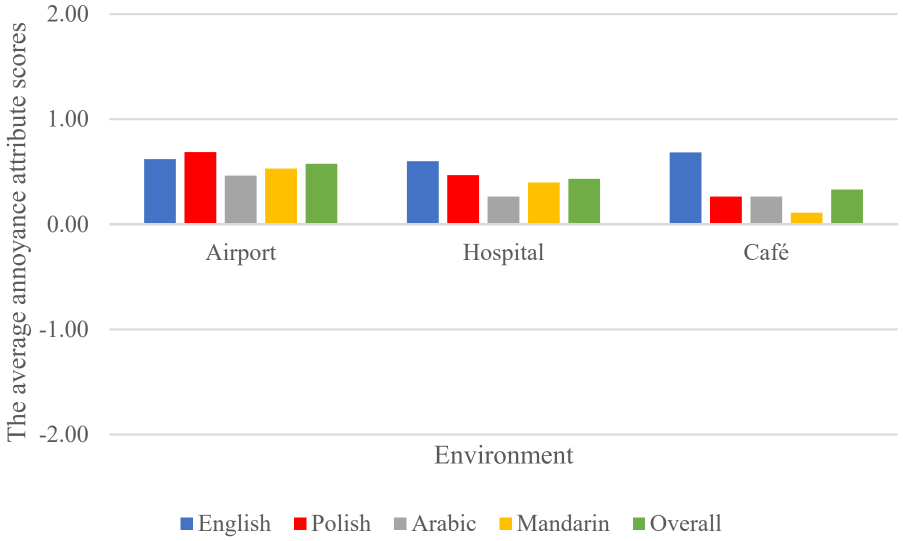


Figure 5.10 The average annoyance attribute scores of the 3 room acoustic conditions for each environment and language.

It should be noted that the annoyance attribute might be affected by both the room acoustic conditions (reverberation time and signal-to-noise ratio) and the context of speech, which is in turn related to the type of environment. In order to eliminate the effects of room acoustic conditions and observe the effects of the environment only, the average annoyance scores of the 3 room acoustic conditions were calculated for each environment and language and are presented in Figure 5.10. The results suggest that the airport was the most annoying environment for Polish, Arabic, and Mandarin participants, followed by the hospital, and the café; however, the results were different for English participants. Given the fact that the café environment had the most relaxed context of speech (see Chapter 3 for details), the Polish, Arabic, and Mandarin results are not surprising. However, the café was the most annoying environment for English, followed by the airport and the hospital, respectively; but the differences between the average annoyance scores of English (0.08) did not vary as much as the other languages (Polish=0.42, Arabic=0.20, Mandarin=0.42). Therefore, it can be assumed that the English listeners were equally annoyed by the acoustic environment and the context of speech of all of the environments.

The repeated measures ANOVA revealed that the effects of speech transmission index (STI) [$F(1.98,110.91) = 77.23$, $p=0.000$] and environment [$F(1.93,108.53) = 6.63$, $p=0.002$] on the annoyance attribute were statistically significant. Additionally, according to the one-way ANOVA results, the differences between the four languages' annoyance attribute scores were only significant at the hospital at STI=0.4 [$F(3,59) = 4.54$, $p=0.006$].

Correlations between the intelligibility attribute and the annoyance attribute were statistically analysed by running Spearman's correlation analysis. The analysis revealed that the negative correlations between the annoyance attribute scores and the perceived intelligibility attribute scores were statistically significant for English ($p < 0.01$), Polish ($p < 0.01$), and Mandarin ($p < 0.01$). Similar to the noisiness attribute results, Arabic was the only language that did not show any correlations between the two attributes (Table 5.2). The correlation analysis for each acoustic condition revealed 5 significant negative correlations out of 36 cases. These were found at the hospital ($p < 0.01$) and at the café ($p < 0.05$) at STI = 0.5 for English, at the hospital at STI = 0.6 ($p < 0.01$), and at the café at STI = 0.4 ($p < 0.01$) and STI = 0.6 ($p < 0.05$) for Mandarin. No significant correlations were found for Polish and Arabic (Table 5.1).

The analysis of the annoyance attribute scores revealed that the airport was the most annoying environment for Polish, Arabic, and Mandarin listeners, while English listeners were equally annoyed at all of the 3 acoustic environments. These results indicate that socio-cultural backgrounds might have an effect on the annoyance attribute. It can be hypothesised that Arabic and Mandarin listeners are used to noisier environments with poor acoustic conditions, consequently decreasing the annoyance attribute scores. This is suggested by previous research that showed how certain cultures can be more tolerant to noisy conditions (Yang and Kang, 2005), although noisiness and annoyance results obtained here were not similar and showed different trends.

5.2.6 Relaxation

The next semantic attribute to be analysed is relaxation. The participants were asked to subjectively evaluate how relaxing was the acoustic environment on a 5-point scale (from very relaxing to very stressful). Figure 5.11(a), 5.11(b), and 5.11(c) show the relationship between the relaxation attribute scores and the STI levels at the airport, the hospital, and the café. Results show that the relaxation scores tend to increase as the STI increases. The

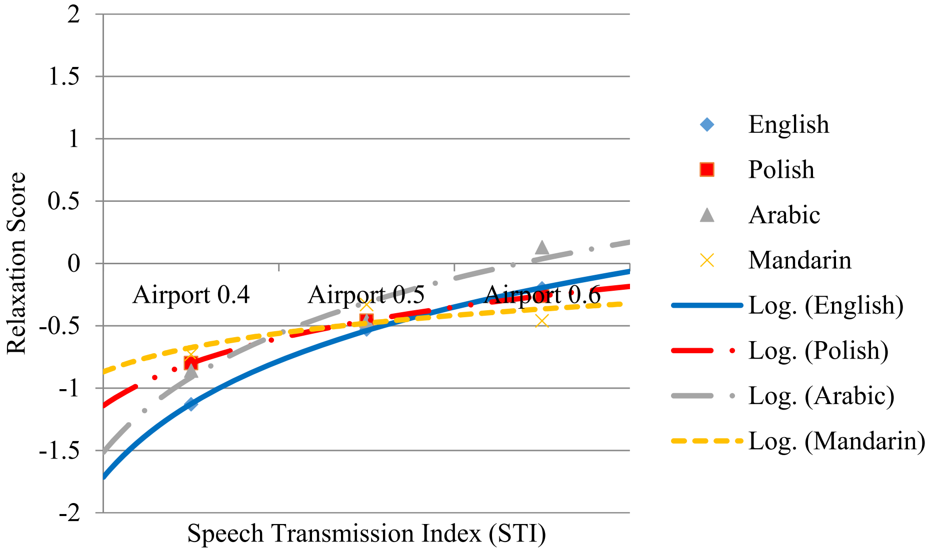


Figure 5.11 (a) Relaxation attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

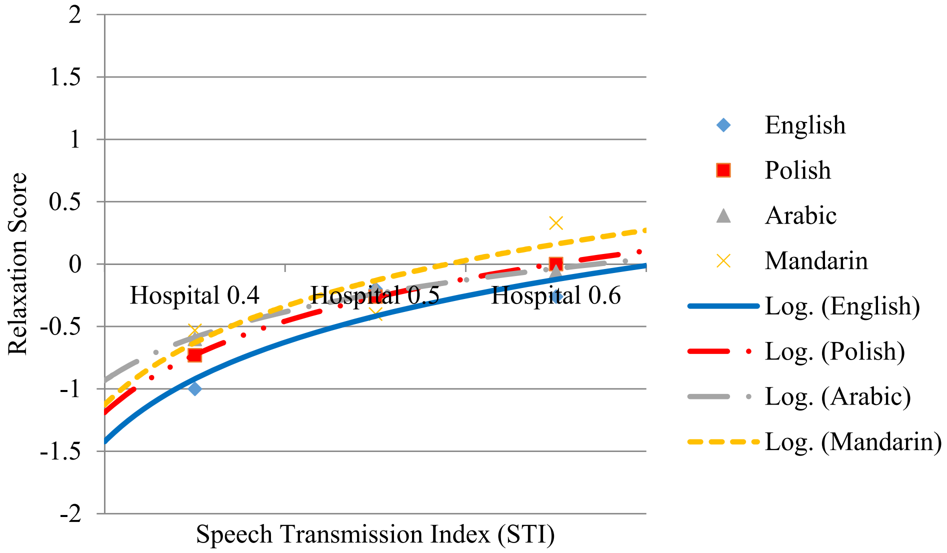


Figure 5.11 (b) Relaxation attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

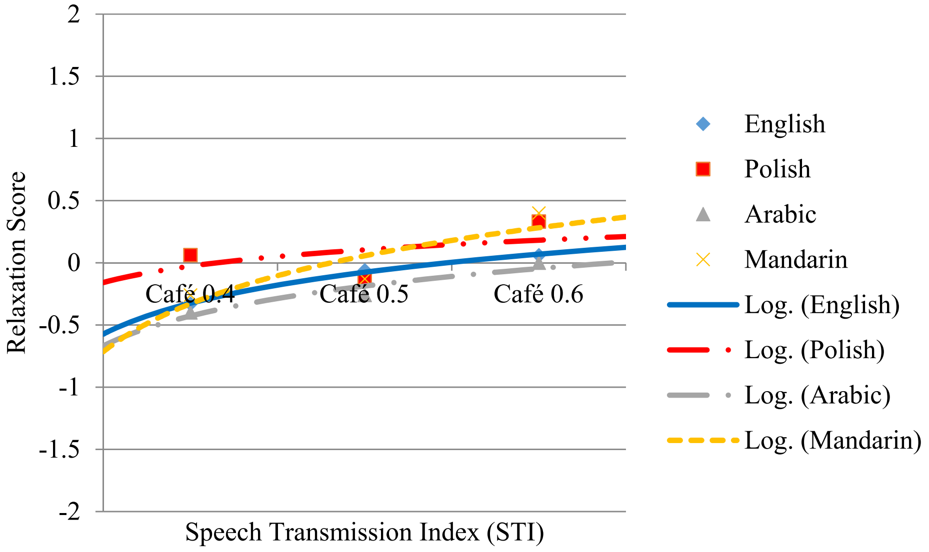


Figure 5.11 (c) Relaxation attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

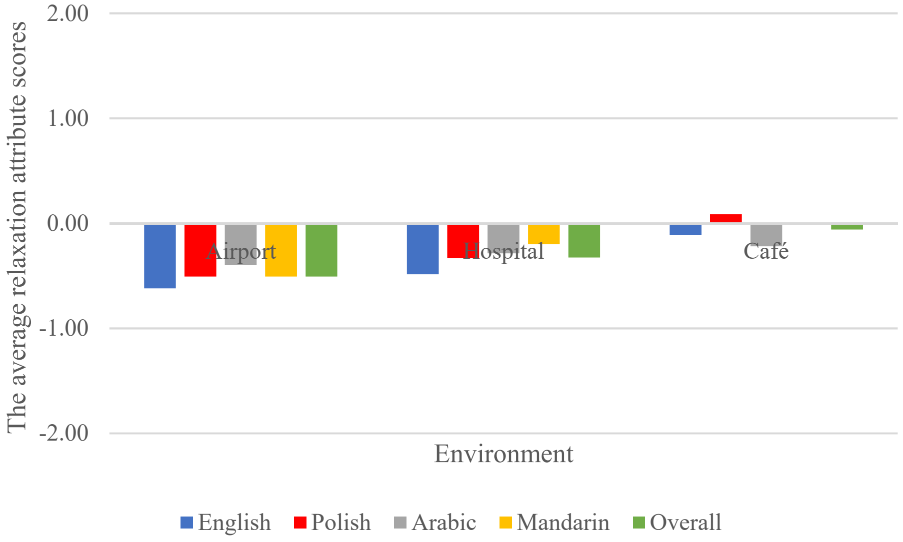


Figure 5.12 The average relaxation attribute scores of the 3 room acoustic conditions for each environment and language.

largest difference between the relaxation attribute scores of the four languages was observed at the hospital and the airport at STI=0.6 (0.59) (Figure 5.11(a) and 5.11(b)) and the smallest difference was observed at the airport, the café, and the hospital at STI=0.5 (0.20) (Figure 5.11(a), 5.11(b), and 5.11(c)). It was also found that the variance between the four languages' relaxation attribute scores increased at STI=0.4 and STI=0.6, and decreased at STI=0.5. The scores were always the highest at STI = 0.6, followed by STI

= 0.5 and STI = 0.4, respectively, with the exceptions of the Polish scores at the café and the Mandarin scores at the airport, for which the scores were the highest at STI = 0.5.

As expected, in all cases, relaxation attribute scores increased as the STI increased (Figure 5.11). Similar to the annoyance attribute scores, the café had the highest and the airport had the lowest average relaxation attribute scores across all languages (Figure 5.12). Figure 5.12 shows that the differences between the lowest and the highest average relaxation scores were similar between English (0.51), Polish (0.58), and Mandarin (0.50); however, Arabic showed a smaller difference across the environments (0.17). The less diverse results of Arabic suggest a resilience to stress sources of the environments, given the fact that Arabic had the lowest noisiness attribute scores, as well.

The repeated measures ANOVA revealed that the effects of speech transmission index (STI) [$F(1.97,110.74) = 47.70$, $p=0.000$] and environment [$F(1.73,97.39) = 23.82$, $p=0.000$] on the annoyance attribute were statistically significant. Additionally, according to the one-way ANOVA results, the differences between the four languages' relaxation attribute scores were only significant at the hospital at STI=0.6 [$F(3,59) = 4.66$, $p=0.006$].

The correlations between the intelligibility attribute and the relaxation attribute were statistically analysed by running Spearman's correlation analysis. The analysis revealed that the correlations between the two attributes are statistically significant ($p < 0.01$) for English, Polish, and Mandarin. However, the results of Arabic do not show any significant correlations between the relaxation attribute and the perceived intelligibility attribute (Table 5.2). The correlation analysis for each acoustic condition revealed 3 significant correlations out of 36 cases. Two significant positive correlations were found at the airport at STI = 0.5 ($p < 0.05$) and the hospital at STI = 0.5 ($p < 0.05$) for Mandarin, and one negative significant correlation was found at the café at STI = 0.5 ($p < 0.05$) for Arabic (Table 5.1).

According to the results of the relaxation attribute, there was a significant difference between environments and STI conditions. Additionally, Arabic listeners seemed to be less affected by the change in room acoustic conditions, when compared to English, Polish, and Mandarin participants, as the relaxation scores were fairly consistent across the environments.

5.2.7 Comfort

The next semantic attribute to be analysed is comfort. The participants were asked to subjectively evaluate the comfort of the acoustic environment on a 5-point scale (from very comfortable to very uncomfortable). Figures 5.13(a), 5.13(b), and 5.13(c) show the relationship between the comfort attribute scores and the STI levels at the airport, the hospital, and the café. Results show that the comfort scores tend to increase as the STI increases. The largest difference between the comfort attribute scores of the four languages was observed at the airport at STI=0.6 (0.79) (Figure 5.13(a)), in which the average score of English was +0.13, Polish was -0.40, Arabic was +0.26, and Mandarin was -0.53. The smallest difference was observed at the hospital at STI=0.5 (0.13) (Figure 5.13(b)), in which the average score of English was 0.00, Polish was -0.06, and Arabic and Mandarin were -0.13.

A comparison between the four languages' comfort attribute scores was made by marking the lowest and highest scoring languages for each of the 9 cases. English had the highest scores in 4 out of 9 cases, Polish had the highest and lowest scores in 3 out of 9 cases, Arabic had the lowest scores in 4 out of 9 cases and highest scores in 1 out of 9 cases, and lastly, Mandarin had the lowest scores in 5 out of 9 cases and highest scores in 2 out of 9 cases. This can be interpreted as, under similar acoustic conditions, English listeners felt more comfortable compared to Polish, Arabic and Mandarin listeners. This contradicts the noisiness scores that tended to be rated higher by English listeners. Another interesting result is that Arabic listeners' comfort attribute scores tended to be lower at the hospital and at the café when compared to the other languages; however, at the airport, the scores of Arabic listeners were closer to the English listeners' scores. As previously suggested for the noisiness attribute in Section 5.2.4, this might be caused by the socio-cultural backgrounds of the listeners.

The average comfort attribute scores of the 3 room acoustic conditions for each environment and language is given in Figure 5.14. It demonstrates that the café was the most comfortable environment, followed by the hospital and the airport, respectively. It was also observed that the Arabic scores show little variation (-0.07) across the environments.

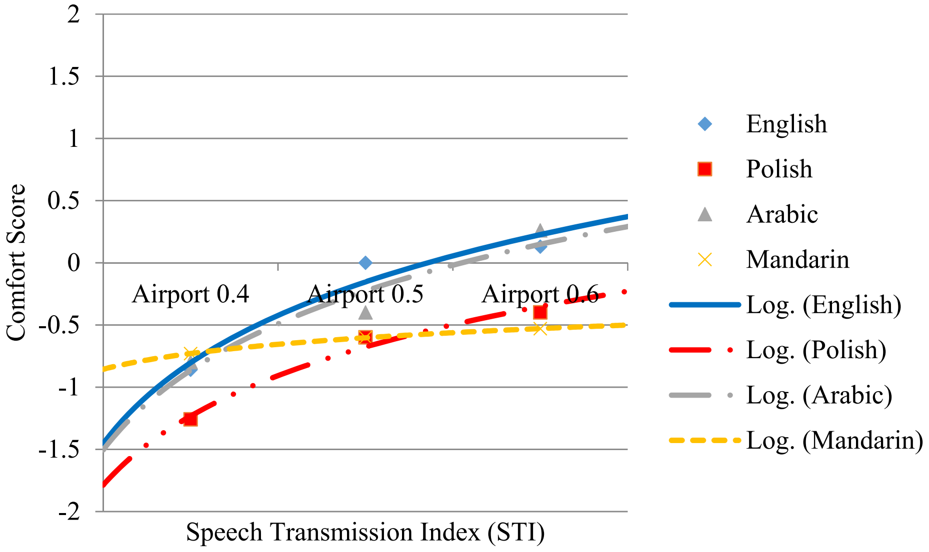


Figure 5.13 (a) Comfort attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

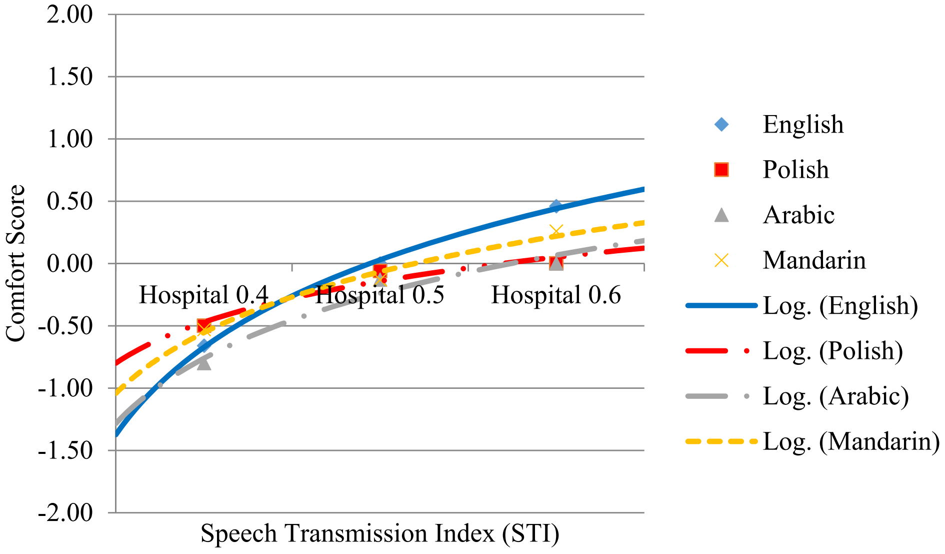


Figure 5.13 (b) Comfort attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

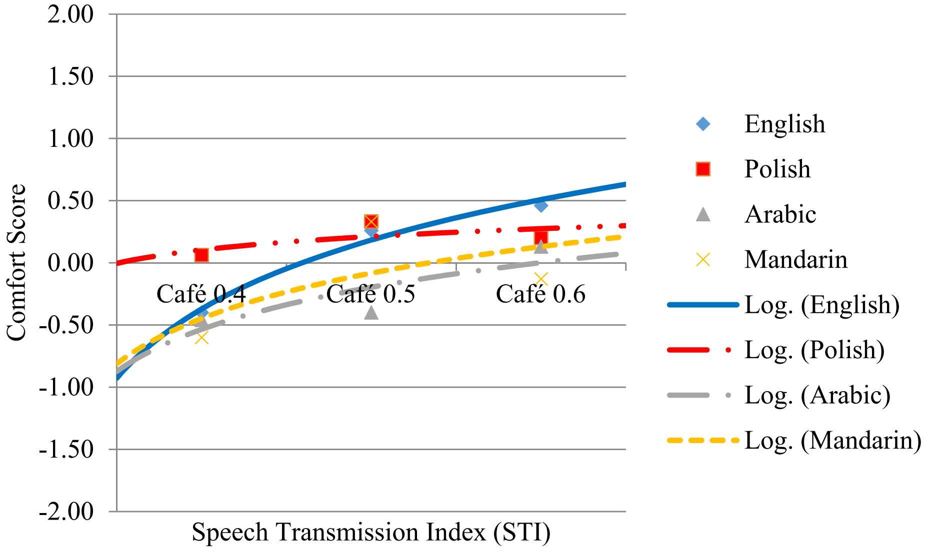


Figure 5.13 (c) Comfort attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

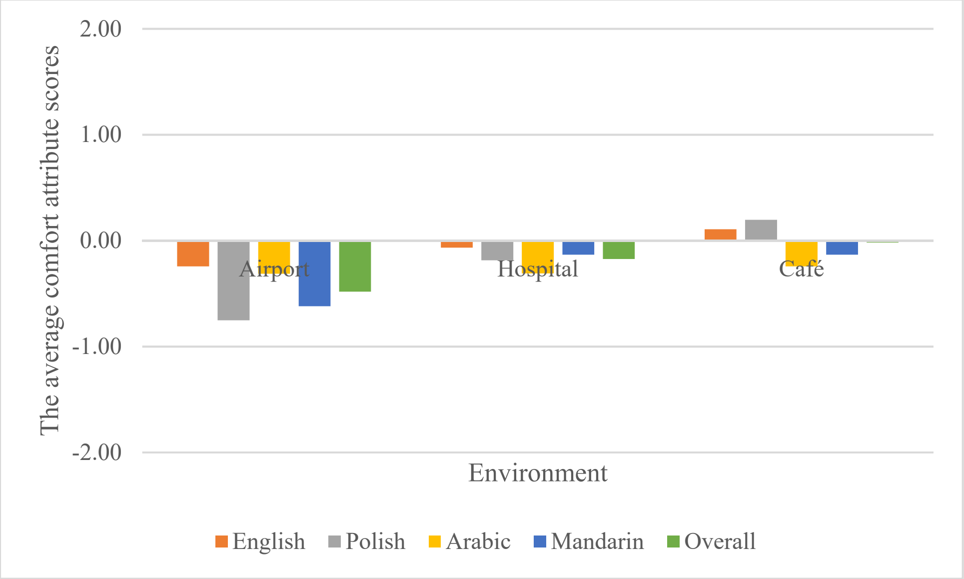


Figure 5.14 The average comfort attribute scores of the 3 room acoustic conditions for each environment and language.

The repeated measures ANOVA revealed that the effects of speech transmission index (STI) [*F*(1.92,107.65) = 62.31, *p*=0.000] and environment [*F*(1.98,111.14) = 13.44, *p*=0.000] on the comfort attribute were statistically significant. It also revealed that the combined effects of environment and language [*F*(5.95,111.14) = 2.91, *p*=0.011], and STI and language [*F*(5.76,107.65) = 2.43, *p*=0.032] on the comfort attribute were statistically significant. Additionally, according to the one-way ANOVA results, the differences between the four languages' comfort attribute scores were significant at the airport at STI=0.5 [$F(3,59) = 3.23$, $p=0.029$] and STI=0.6 [$F(3,59) = 4.31$, $p=0.008$].

Correlations between the intelligibility attribute and the comfort attribute were statistically analysed by running Spearman's correlation analysis. The analysis revealed that the two attributes are significantly correlated for the 4 languages tested (English: $p < 0.01$, Polish: $p < 0.01$, Arabic: $p < 0.05$, and Mandarin: $p < 0.01$) (Table 5.2). Additionally, the correlation analysis for each room acoustic condition was conducted. The analysis revealed 4 significant correlations out of 36 cases. Three significant positive correlations were found at the hospital at STI = 0.4 ($p < 0.05$) and STI = 0.6 ($p < 0.05$) for Mandarin, and the hospital at STI = 0.5 ($p < 0.05$) for English. A significant negative correlation ($p < 0.01$) was found at the hospital at STI = 0.4 for Arabic (Table 5.1).

The results suggest that the variance between comfort attribute scores of the four languages was dependent on the STI and the environment. Furthermore, English listeners tended to have higher comfort attribute scores in 4 out of 9 cases, and the combined average scores of all three STI conditions showed that English listeners have the highest scores at the hospital and at the café (Figure 5.14). Additionally, Arabic scores showed little variation (-0.07) across the environments.

5.2.8 Pleasantness (Environment)

The next semantic attribute to be analysed is the environmental pleasantness. The participants were asked to subjectively evaluate the pleasantness of the acoustic environment on a 5-point scale (from very unpleasant to very pleasant). Figure 5.15(a), 5.15(b), and 5.15(c) show the relationship between the environmental pleasantness attribute scores and the STI levels at the airport, the hospital, and the café. Results show that the environmental pleasantness scores tend to increase as the STI increases, with a minor exception for Mandarin at the airport. The largest difference between the environmental pleasantness attribute scores of the four languages was observed at the hospital at STI=0.6 (1.06) (Figure 5.15(b)), in which the average score of English was 0.00, Polish was -0.66, Arabic was +0.33, and Mandarin was +0.40. The smallest difference was observed at the hospital at STI=0.5 (0.07) (Figure 5.15(b)), in which the average score of English was -0.20, and Polish, Arabic, and Mandarin were -0.13.

The analysis of Figure 5.15 shows that there was a difference between the environmental pleasantness attribute scores of the four languages. It was observed that the Polish listeners' scores were the lowest in 7 out of 9 cases and the Mandarin listeners' scores were the highest in 5 out of 9 cases. It should be noted that the highest and the lowest scores varied depending on the environment; for instance, Polish scores were lowest in 3 out of 3 cases at the airport and the Mandarin listeners' scores were the highest in 3 out of 3 cases at the hospital. English and Arabic scores did not show such consistency. It should also be stated that the Mandarin environmental pleasantness scores were the highest at STI=0.4 at the airport (-0.46), the hospital (-0.53), and the café (-0.20). The results suggest that the Mandarin listeners felt more pleasant in most of the acoustic conditions, including poor acoustic conditions as seen at STI=0.4, when compared to the other 3 languages. Additionally, it was seen that Polish listeners felt more unpleasant in most of the acoustic conditions.

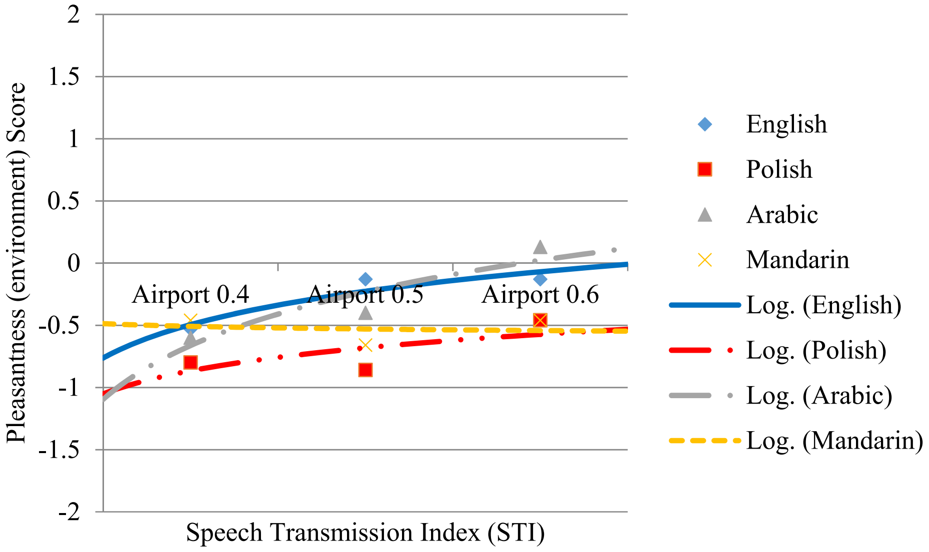


Figure 5.15 (a) Environmental pleasantness attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

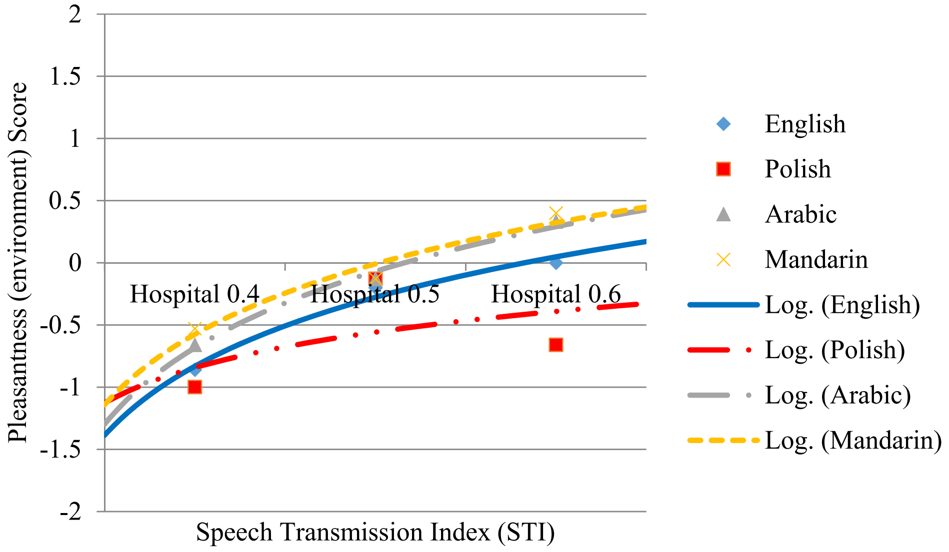


Figure 5.15 (b) Environmental pleasantness attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

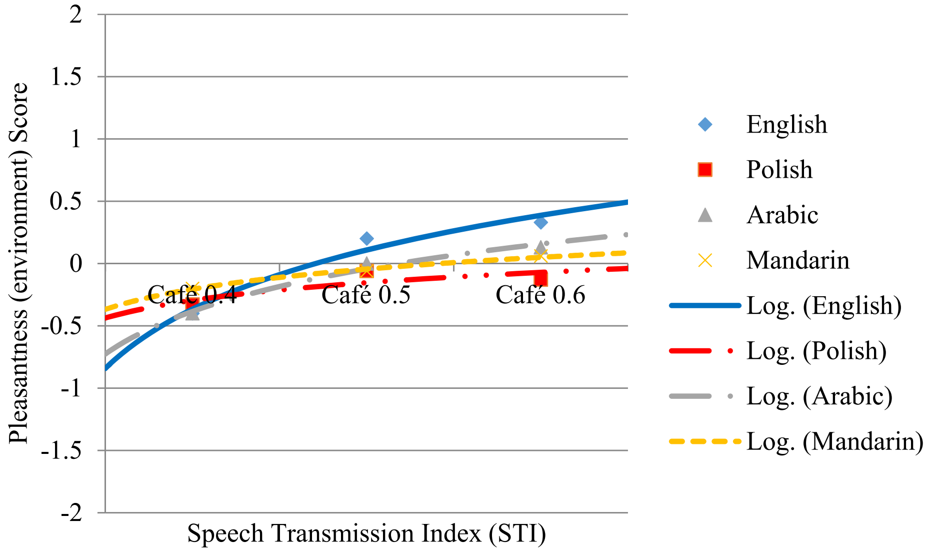


Figure 5.15 (c) Environmental pleasantness attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

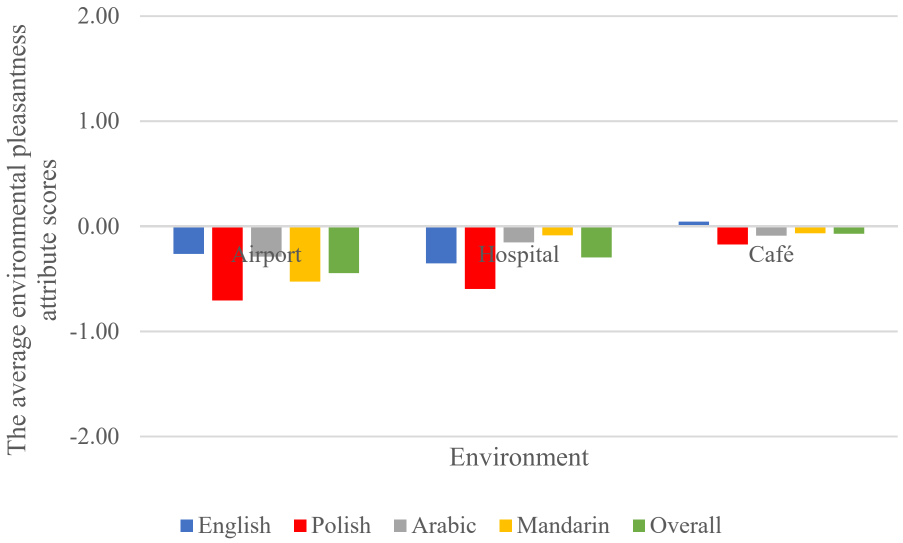


Figure 5.16 The average environmental pleasantness attribute scores of the 3 room acoustic conditions for each environment and language.

Figure 5.16 shows the average environmental pleasantness scores for each environment and language. It illustrates that the Polish, Arabic, and Mandarin listeners felt most comfortable at the café, followed by the hospital and the airport, respectively. The English listeners felt most comfortable at the café, as well; however, they found the airport more pleasant compared to the hospital, unlike the other languages results.

The repeated measures ANOVA revealed that the effects of speech transmission index (STI) [$F(1.89,105.91) = 37.21$, $p=0.000$] and environment [$F(1.83,103.00) = 11.13$, $p=0.000$] on the environmental pleasantness attribute were statistically significant. It also revealed that the combined effects of STI and language [$F(5.67,105.91) = 2.58$, $p=0.024$], and environment and STI [$F(3.82,214.10) = 5.07$, $p=0.001$] on the environmental pleasantness attribute were statistically significant. Additionally, according to the one-way ANOVA results, the differences between the four languages' environmental pleasantness attribute scores were significant at the airport at STI=0.5 [$F(3,59) = 5.21$, $p=0.003$] and STI=0.6 [$F(3,59) = 3.74$, $p=0.016$], as well as at the hospital at STI=0.6 [$F(3,59) = 8.24$, $p=0.000$].

Correlations between the intelligibility attribute and the environmental pleasantness attribute were statistically analysed by running Spearman's correlation analysis. The two attributes were significantly correlated for the four languages tested (English: $p < 0.01$, Polish: $p < 0.01$, Arabic: $p < 0.05$, and Mandarin: $p < 0.01$) (Table 5.2).

The correlation analysis was also conducted for each room acoustic condition separately. The analysis revealed 9 significant positive correlations out of 36. The average scores of English were significantly correlated with the intelligibility scores at the airport at STI = 0.5 ($p < 0.05$) and STI = 0.6 ($p < 0.05$), and at the hospital at STI = 0.5 ($p < 0.01$). For Polish, the significant correlations were found at the hospital at STI = 0.6 ($p < 0.05$) and at the café at STI = 0.5 ($p < 0.05$), for Arabic at the airport at STI = 0.5 ($p < 0.05$), and lastly for Mandarin at the hospital at STI = 0.4 ($p < 0.05$) and STI = 0.6 ($p < 0.05$), and at the café at STI = 0.4 ($p < 0.01$). Compared to the other 9 semantic attributes that were tested, environmental pleasantness had the most correlations with the intelligibility attribute; therefore, highlighting its importance in terms of multi-lingual speech intelligibility (Table 5.1).

The results of the analysis of the environmental pleasantness attribute scores revealed that there were significant differences between the environments and the STI conditions, as well as the languages. Mandarin listeners tended to rate environmental pleasantness higher in most of the acoustic conditions, while Polish listeners rated it lower compared to the other languages.

5.2.9 Eventfulness

The next semantic attribute to be analysed is the eventfulness. The participants were asked to subjectively evaluate the eventfulness of the acoustic environment on a 5-point scale (from very eventful to very uneventful). Figure 5.17(a), 5.17(b), and 5.17(c) show the relationship between the eventfulness attribute scores and the STI levels at the airport, the hospital, and the café. Results show that the eventfulness scores tend to decrease as the STI increases, with the exceptions of Polish and Arabic at the café. The largest difference between the eventfulness attribute scores of the four languages was observed at the hospital at STI=0.4 (1.00) (Figure 5.17(b)) and the smallest difference was observed at the airport at STI=0.6 (0.20), and at the café at STI=0.5 and STI=0.6 (Figure 5.17(c)).

The main finding of the analysis of the eventfulness attribute scores was that English scores were the highest in 6 out of 9 cases. At the airport, it was the highest in 3 out of 3 STI conditions. Additionally, the Arabic scores were the lowest in 5 out of 9 cases, and the Mandarin scores were the lowest in 4 out of 9 cases. These results can be interpreted as English listeners focus on the distractive noise sources in the environment more than the Arabic and Mandarin listeners, especially at the airport, which had the most distractive sound events compared to the hospital and the café. It should be noted that the public announcement in the airport background noise was in English, a factor that might have affected the native English participants' scores.

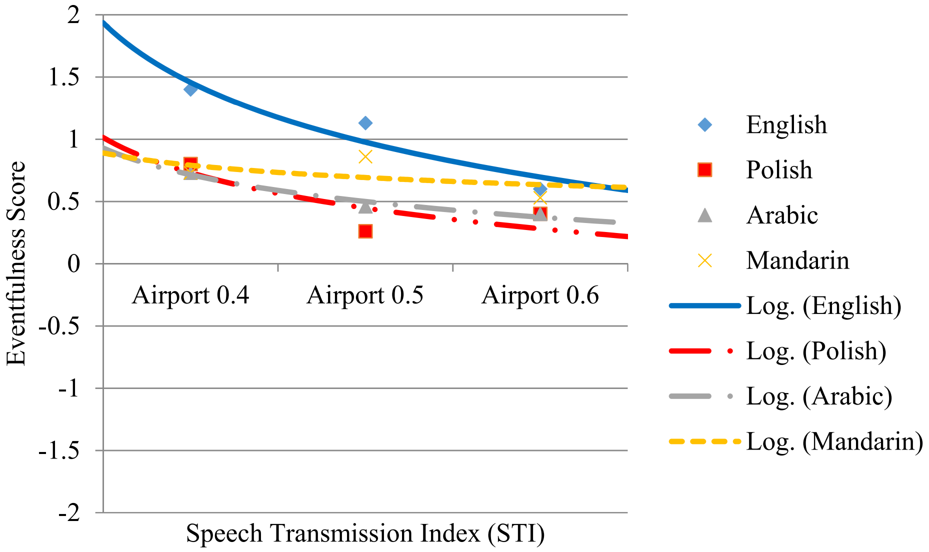


Figure 5.17 (a) Eventfulness attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

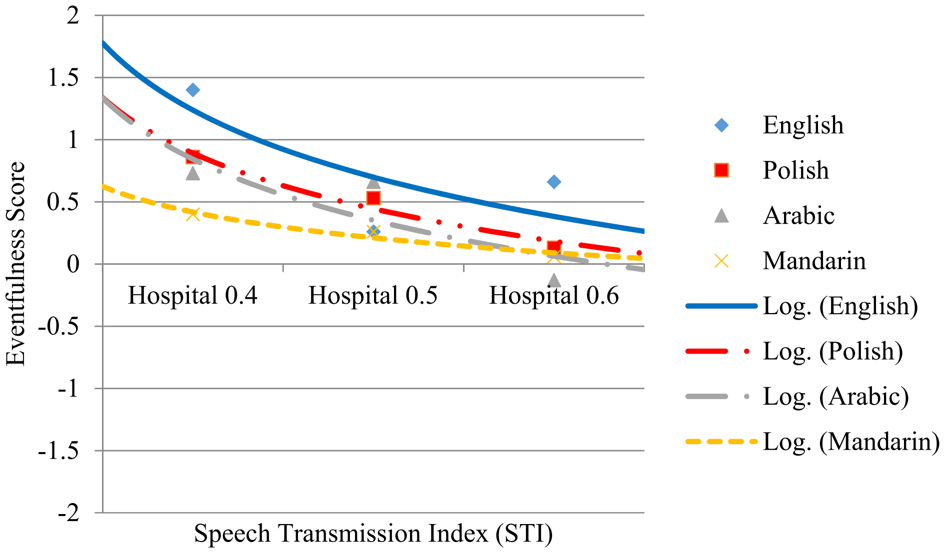


Figure 5.17 (b) Eventfulness attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

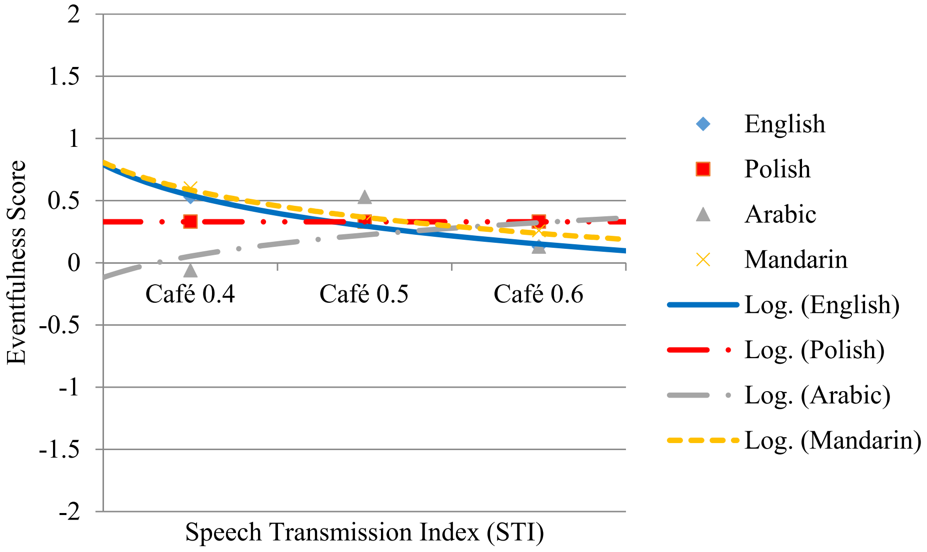


Figure 5.17 (c) Eventfulness attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

The average eventfulness attribute scores obtained for each environment and language is presented in Figure 5.18. It shows that the eventfulness attribute scores of English and Arabic were the highest at the airport, followed by the hospital, and the café, respectively. The Polish participants rated the hospital as the most eventful environment, followed by the airport and the café. Lastly, similar to the English and Arabic eventfulness attribute scores, the Mandarin scores were the highest at the airport; however, the score at the café was higher than at the hospital for Mandarin. Therefore, it can be suggested that the

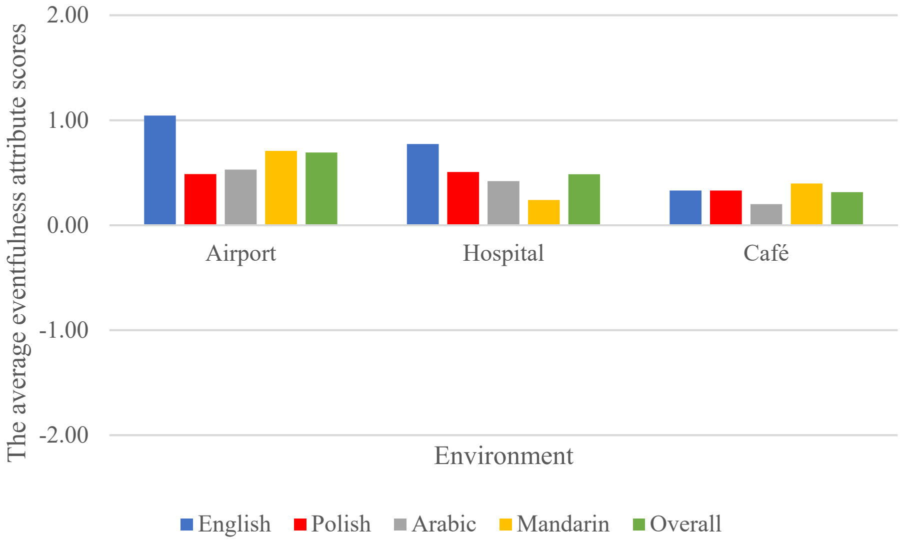


Figure 5.18 The average eventfulness attribute scores of the 3 room acoustic conditions for each environment and language.

eventfulness attribute scores depend on the language and the environment, which suggests that the native speakers of different languages might focus differently on sound sources in a given environment. This might eventually affect the intelligibility of speech in a multi-lingual environment.

The repeated measures ANOVA revealed that the effects of speech transmission index (STI) [$F(1.96,110.04) = 20.74$, $p=0.000$] and environment [$F(1.71,95.83) = 6.19$, $p=0.005$] on the eventfulness attribute were statistically significant. It also revealed that the combined effects of STI and language [$F(5.89,110.04) = 2.73$, $p=0.017$], STI and environment [$F(3.56,199.56) = 4.26$, $p=0.004$] on the eventfulness attribute were statistically significant. Additionally, according to the one-way ANOVA results, the differences between the four languages' eventfulness attribute scores were significant at the airport at STI=0.4 [$F(3,59) = 2.94$, $p=0.041$] and STI=0.5 [$F(3,59) = 3.74$, $p=0.016$], as well as at hospital at STI=0.4 [$F(3,59) = 4.29$, $p=0.009$].

Correlations between the intelligibility attribute and the eventfulness attribute were statistically analysed by running Spearman's correlation analysis. The analysis revealed that the two attributes were significantly correlated (negatively) only for English ($p < 0.01$) and Polish ($p < 0.05$). The correlations between the eventfulness attribute scores and the perceived intelligibility attribute scores for Arabic and Mandarin were not statistically significant (Table 5.2). Furthermore, the correlation analysis for each room acoustic condition revealed 2 significant correlations out of 36 cases. The average scores of English showed a significant positive correlation with the intelligibility scores at the airport at STI = 0.6 ($p < 0.01$) For Polish, a significant negative correlation was found at the airport at STI = 0.4 ($p < 0.01$) (Table 5.1).

Overall, the results suggest that the native English speakers focus on distractive sound events in an environment more than the other three languages. It was also observed that the scores vary depending on the language, the STI, and the environment. Additionally, the relationships between the combination of these three variables and the eventfulness attribute were statistically significant according to the repeated measures ANOVA results.

5.2.10 Excitement

The next semantic attribute to be analysed is the excitement. The participants were asked to subjectively evaluate the excitement of the acoustic environment on a 5-point scale (from very exciting to very boring). Figure 5.19(a), 5.19(b), and 5.19(c) show the relationship between the excitement attribute scores and the STI levels at the airport, the hospital, and the café. The largest difference between the scores of the four languages was observed at the airport at STI=0.6 (0.66) (Figure 5.19(a)) and the smallest difference was observed at the cafe at STI=0.5 (0.19) (Figure 5.19(c)).

Figure 5.19 shows that the excitement attribute scores of the four languages largely vary in between both STI conditions and environments, with no consistent pattern. The only generalisation can be made for Mandarin, which had the highest scores at the café in 3 out of 3 cases. These results suggest that excitement in an environment is largely personal rather than being caused by socio-cultural properties of the native speakers of a language.

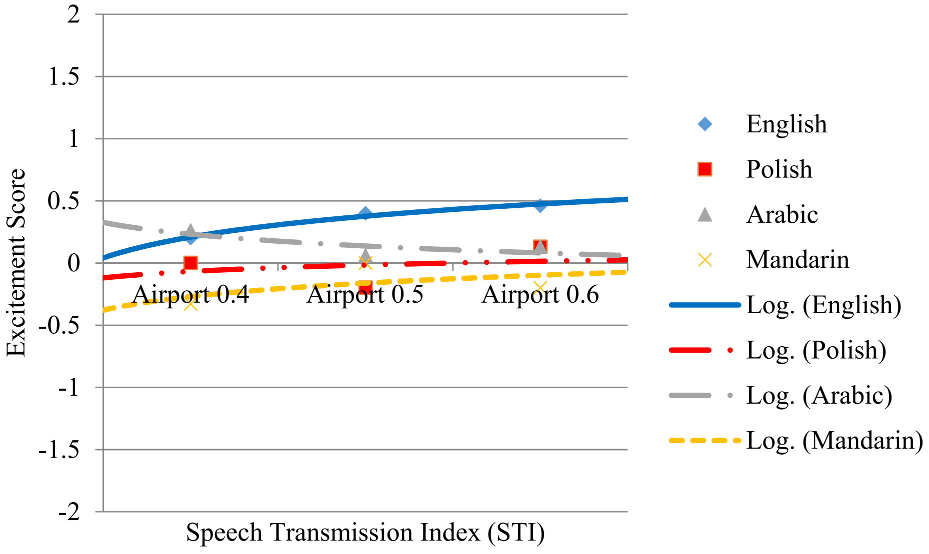


Figure 5.19 (a) Excitement attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

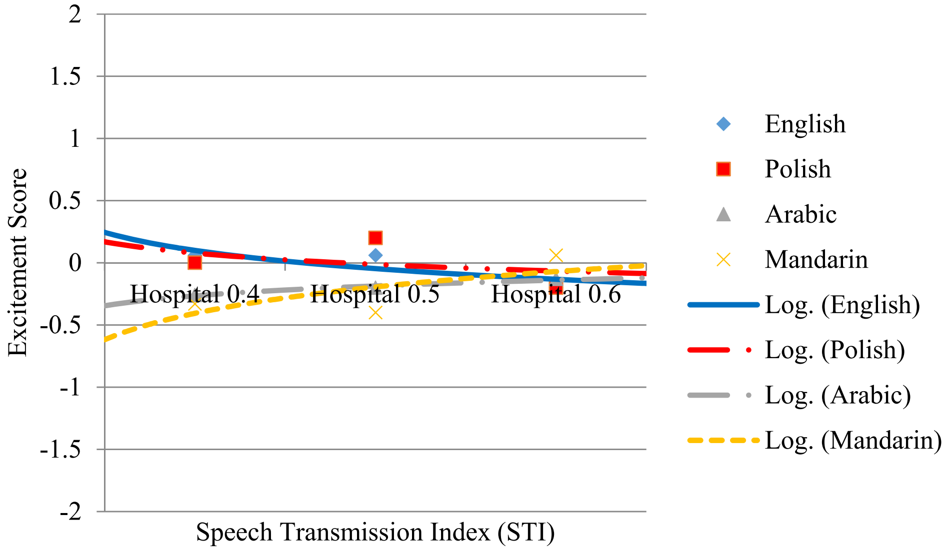


Figure 5.19 (b) Excitement attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

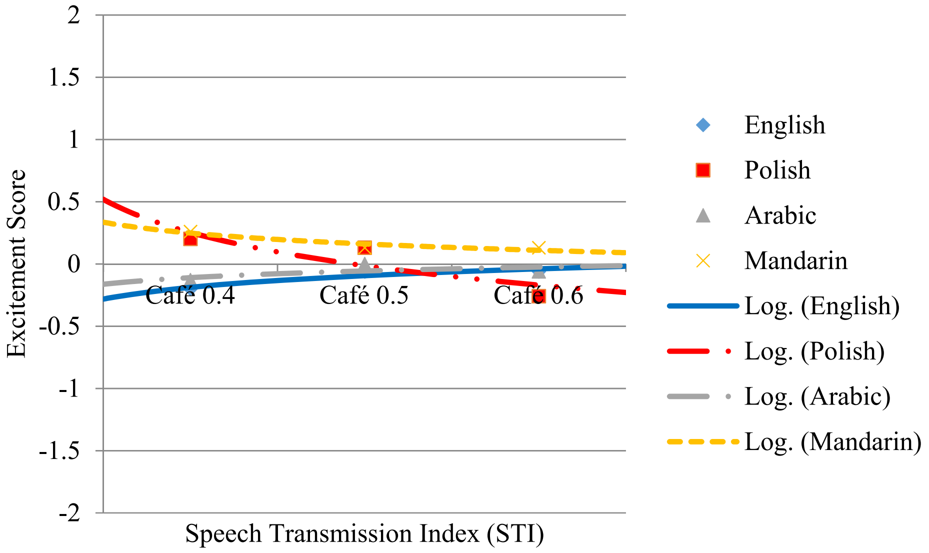


Figure 5.19 (c) Excitement attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

Figure 5.20 shows the average excitement attribute scores for each environment and language. Similar to the findings that were observed by analyzing Figure 5.19, the average scores do not show a particular pattern. The differences between the highest and lowest scores of English was 0.45, Polish was 0.04, Arabic was 0.34, and Mandarin was 0.34.

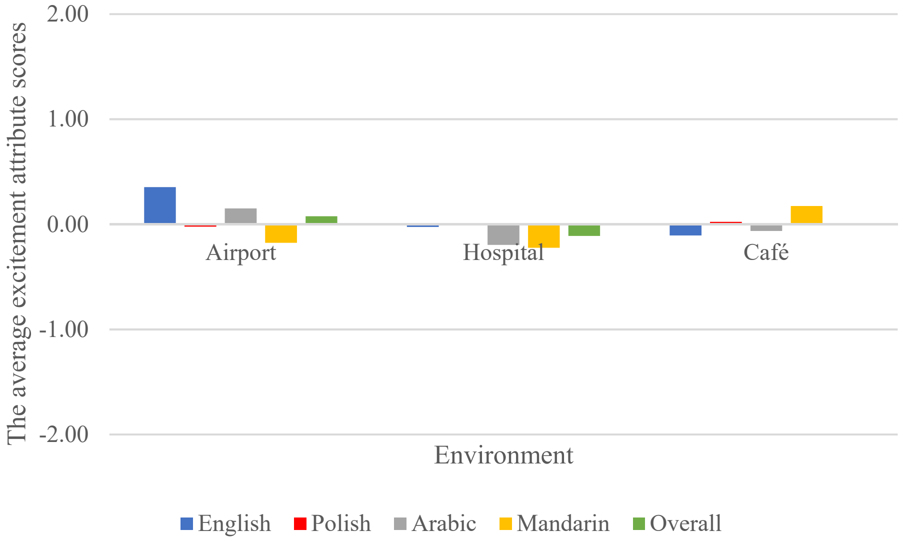


Figure 5.20 The average excitement attribute scores of the 3 room acoustic conditions for each environment and language.

The most exciting environment was the airport for the native speakers of English and Arabic and the café for the native speakers of Polish and Mandarin. The most boring environment was the café for the native speakers of English, the airport for the native speakers of Polish, and the hospital for the native speakers of Arabic and Mandarin.

The repeated measures ANOVA revealed that the effects of speech transmission index (STI) and environment on the excitement attribute were not statistically significant. It also revealed that the combined effects of environment and language [$F(5.85,109.30) = 2.67$, $p=0.019$], and environment, STI and language [$F(11.12,207.71) = 3.02$, $p=0.001$] on the excitement attribute were statistically significant. Additionally, according to the one-way ANOVA results, the differences between the four languages' excitement attribute scores were only significant at the hospital at STI=0.5 [$F(3,59) = 3.13$, $p=0.032$].

Correlations between the intelligibility attribute and the excitement attribute were statistically analysed by running Spearman's correlation analysis. The analysis revealed that there were no statistically significant correlations between the two attributes (Table 5.2). The correlation analysis for each room acoustic condition revealed 3 significant correlations out of 36 cases. The average scores of English showed significant negative correlations with the intelligibility scores at the café at STI = 0.4 ($p < 0.05$). A significant negative correlation was found for Polish ($p < 0.05$) and a significant positive correlation was found for Mandarin ($p < 0.05$) at the hospital at STI = 0.5 (Table 5.1).

Excitement was the only attribute for which the effects of neither the environment nor the STI were statistically significant. The variance between the results did not show any consistent pattern, eventually suggesting that excitement is largely a factor of personal feelings, rather than being clearly related with the environment, the room acoustic conditions, the languages, or the socio-cultural background of the participants.

5.2.11 Familiarity

The last semantic attribute to be analysed is the familiarity. The participants were asked to subjectively evaluate the familiarity of the acoustic environment on a 5-point scale (from very familiar to very unfamiliar). Figure 5.21(a), 5.21(b), and 5.21(c) show the relationship between the familiarity attribute scores and the STI levels at the airport, the hospital, and the café. The largest difference between the excitement attribute scores of the four languages was observed at the airport and at STI=0.4 (Figure 5.21(a)) and at the café at STI = 0.4 (0.67) (Figure 5.21(c)) and the smallest difference was observed at the café at STI=0.6 (0.20) (Figure 5.21(c)).

Figure 5.21 illustrates that the score variation between the STI conditions were irregular. At the airport, the English and Polish familiarity attribute scores were the highest at STI =0.4, followed by STI = 0.5, and STI = 0.6, respectively; however, for Arabic and Mandarin, the scores were the highest at STI = 0.6, followed by STI = 0.5, and STI = 0.4, respectively. At the café and the hospital, the results did not show any observable pattern; therefore, suggesting that the familiarity attribute is not a factor of the STI. In order to examine the effects of the environments independently from the STI conditions, average scores of all three STI conditions were calculated for each environment and language (Figure 5.22).

Figure 5.22 illustrates the average familiarity attribute scores of the 3 room acoustic conditions for each environment and language. It is observed that the café was the most familiar environment and the hospital was the most unfamiliar environment for all of the listeners, regardless of the native language.

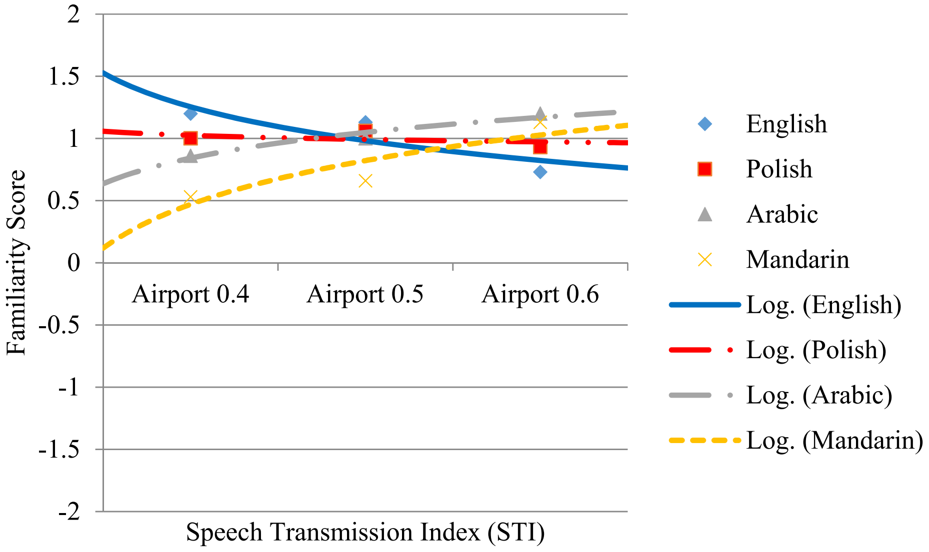


Figure 5.21 (a) Familiarity attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

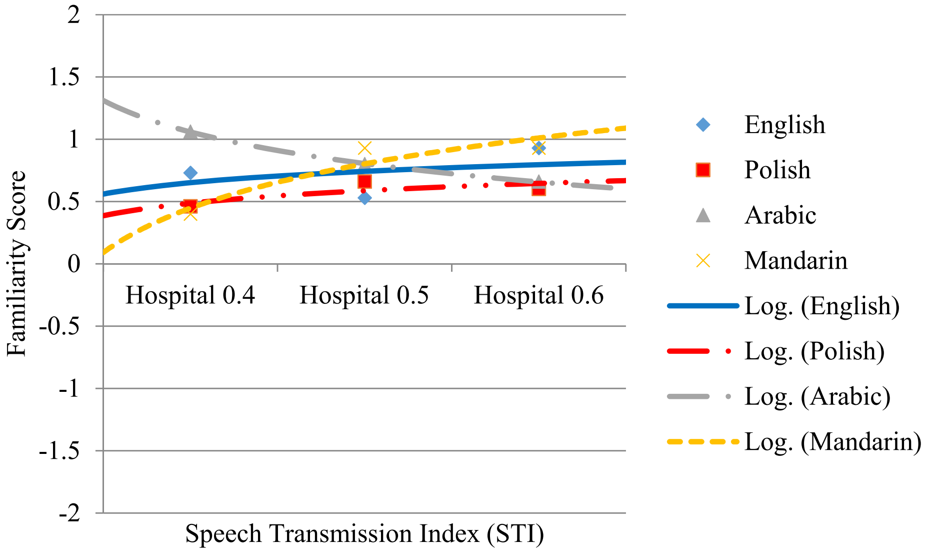


Figure 5.21 (b) Familiarity attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

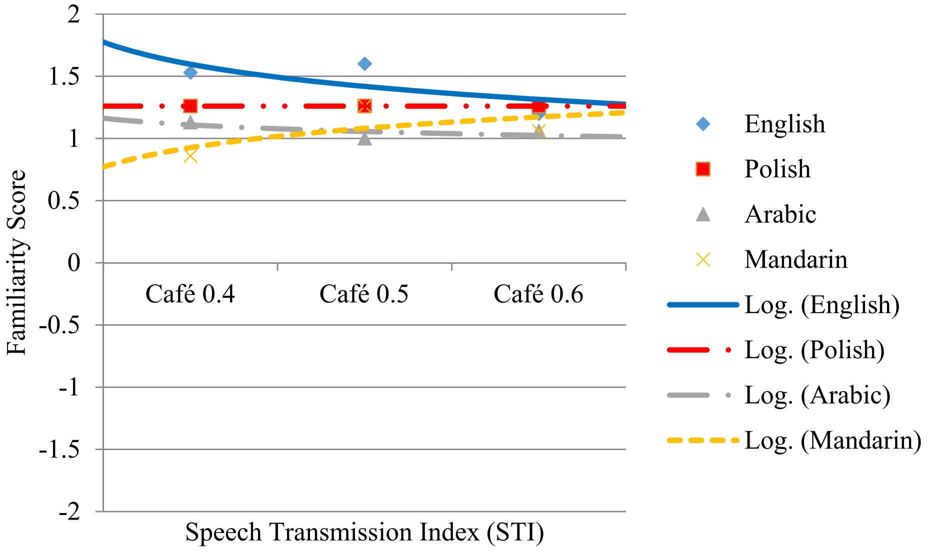


Figure 5.21 (c) Familiarity attribute scores of English, Polish, Arabic, and Mandarin at the café. Actual data markers and logarithmic regression lines are shown in the figure.

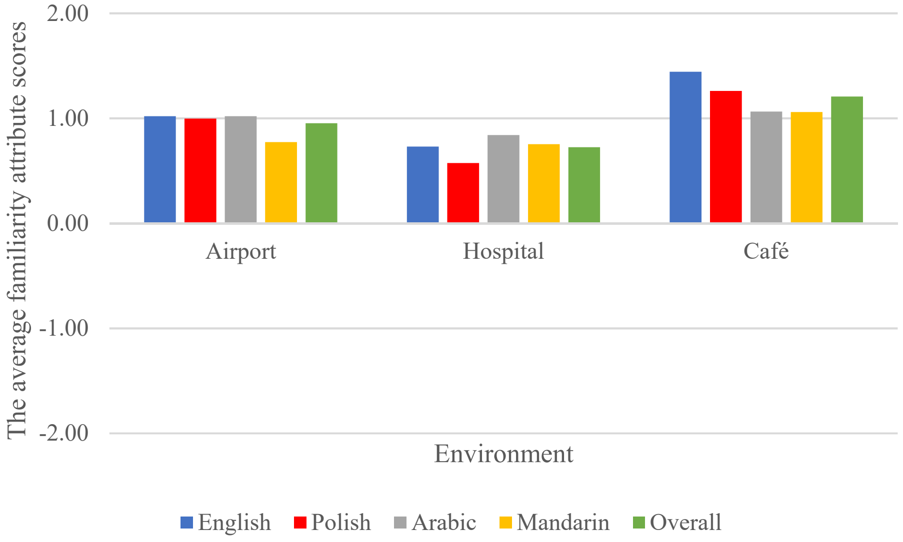


Figure 5.22 The average familiarity attribute scores of the 3 room acoustic conditions for each environment and language.

The repeated measures ANOVA revealed that familiarity, similar to the excitement attribute, did not show any significant effects of the STI on the attribute scores; however, the effect of the environment on the familiarity attribute was statistically significant [$F(1.99,111.65) = 14.76$, p=0.000]. It also revealed that the combined effects of STI and language [$F(5.88,109.90) = 2.89$, p=0.012], and environment, STI and language

[$F(9,168.02) = 2.03$, $p=0.038$] on the familiarity attribute were statistically significant. Additionally, according to the one-way ANOVA results, the differences between the four languages' familiarity attribute scores were significant at the café at STI=0.4 [$F(3,59) = 2.92$, $p=0.042$] and STI=0.5 [$F(3,59) = 4.41$, $p=0.007$].

The correlations between the intelligibility attribute and the familiarity attribute were statistically analysed by running Spearman's correlation analysis. The analysis revealed that the correlations between the two attributes were statistically significant for Mandarin only ($p < 0.01$). The correlations between the attributes were not statistically significant for English, Polish, and Arabic (Table 5.2). The correlation analysis for each room acoustic condition revealed 3 significant correlations out of 36 cases. English showed significant positive correlations with the intelligibility scores at the airport at STI = 0.6 ($p < 0.05$). A significant positive correlation was also found for Polish at the airport at STI = 0.5 ($p < 0.05$) and for Mandarin at the café at STI = 0.6 ($p < 0.05$) (Table 5.1).

The analysis of results revealed that the familiarity attribute is solely affected by the type of environment. The effects of STI and native language of the listeners did not have any significant effect on the familiarity attribute.

5.2.12 Summary of semantic differential analysis findings

Attribute scores showed fairly variable trends across languages, with only few noticeable findings: 1) Variations between languages were significant only for a minority of cases; 2) Speech pleasantness and environmental pleasantness were highly correlated to perceived intelligibility for all the languages considered (i.e., pleasantness improves speech intelligibility); 3) Similarly, comfort and speech loudness were highly correlated to perceived intelligibility for all the languages considered; 4) English participants tended to rate noisiness higher, and a significant negative correlation was found between noisiness and perceived speech intelligibility of English; Arabic participants tended to rate noisiness lower, but this did not correlate significantly with their perceived intelligibility; 5) Arabic showed the lowest number of correlations between semantic attributes and perceived intelligibility, suggesting that its perceived speech intelligibility is less affected by the overall soundscape perception compared to other language groups.

Detailed semantic differential analysis of the attributes revealed that the room acoustic conditions and the type of environments have varying effects on the perceived intelligibility of the four languages tested. It can therefore be stated that justifying speech intelligibility by room acoustic parameters only might mislead the designers and planners, especially while designing a multi-lingual environment.

Findings obtained from the semantic differential analysis are listed below.

Intelligibility

- Perceived intelligibility tended to increase as the STI increased.
- Perceived intelligibility of each language varied depending on the environment, background noise, reverberation time, and signal-to-noise ratio.
- Perceived intelligibility of English was mostly influenced by distractive noise sources present in the background noise (i.e. public announcements in the airport and phone ringing in the hospital).

Loudness (speech)

- Speech loudness scores tended to increase as the STI increased.
- Variance of the speech loudness scores depended on the type of background noise.
- The type of environment had a more pronounced effect on the perceived speech loudness of Arabic and Mandarin, compared to English and especially Polish.
- The correlations between perceived speech loudness and the perceived intelligibility attribute scores were statistically significant for the four languages tested ($p < 0.01$).

Pleasantness (speech)

- The pleasantness scores tended to increase as the STI increased.
- The correlations between the speech pleasantness attribute scores and the perceived intelligibility attribute scores were statistically significant for the four languages tested (English: $p < 0.01$, Polish: $p < 0.05$, Arabic: $p < 0.01$, and Mandarin: $p < 0.01$).
- Speech pleasantness was affected by the type of the environment.

Noisiness

- The noisiness scores tended to decrease as the STI increased.

- The correlations between the noisiness scores and the perceived intelligibility scores are statistically significant for English ($p < 0.01$), Polish ($p < 0.05$), and Mandarin ($p < 0.01$), but not statistically significant for Arabic.
- English listeners are most sensitive to the realistic background noise, whereas Arabic listeners are most resilient.

Annoyance

- The annoyance scores tended to decrease as the STI increased.
- The correlations between annoyance scores and perceived intelligibility scores were statistically significant for English ($p < 0.01$), Polish ($p < 0.01$), and Mandarin ($p < 0.01$), but not significant for Arabic.
- The results suggest that English and Polish listeners were more annoyed by the acoustic environments compared to the Arabic and Mandarin listeners.
- The airport was the most annoying acoustic environment for the Polish, Arabic, and Mandarin listeners, whereas English listeners were equally annoyed by all of the acoustic environments.

Relaxation

- The relaxation scores tended to increase as the STI increased.
- The correlations between the relaxation scores and the perceived intelligibility scores were statistically significant for English ($p < 0.01$), Polish ($p < 0.01$), and Mandarin ($p < 0.01$), but not significant for Arabic.
- Arabic listeners were less affected by the change in room acoustic conditions compared to English, Polish, and Mandarin listeners.

Comfort

- The comfort scores tended to increase as the STI increased.
- The correlations between the comfort attribute scores and the perceived intelligibility attribute scores were statistically significant for the four languages tested (English: $p < 0.01$, Polish: $p < 0.01$, Arabic: $p < 0.05$, and Mandarin: $p < 0.01$).
- Average scores of the three STI conditions show that English listeners achieved the highest scores at all of the acoustic environments.

Pleasantness (environment)

- The environments' pleasantness scores tended to increase as the STI increased.
- The correlations between the environments' pleasantness scores and the perceived intelligibility scores were statistically significant for the four languages tested (English: $p < 0.01$, Polish: $p < 0.01$, Arabic: $p < 0.05$, and Mandarin: $p < 0.01$).
- Environments' pleasantness had the most correlations with the perceived intelligibility attribute scores, across the all 11 attributes tested.

Eventfulness

- The eventfulness scores tended to decrease as the STI increased.
- The correlations between the eventfulness scores and the perceived intelligibility scores were statistically significant for English ($p < 0.01$) and Polish ($p < 0.05$).
- The results suggest that English speakers focused on distractive sound events in an environment more than the other three languages tested.

Excitement

- The excitement scores of the four languages varied in between both STI conditions and environments, with no consistent pattern, which suggests that excitement in an environment is largely personal rather than being caused by socio-cultural properties of native speakers of a language.
- The correlations between the excitement scores and the perceived intelligibility scores were not statistically significant.

Familiarity

- The relationship between the familiarity scores and the STI was irregular.
- The correlations between the familiarity scores and the perceived intelligibility scores were statistically significant for Mandarin only ($p < 0.01$).
- The familiarity attribute was solely affected by the environments. It was observed that the café was the most familiar environment and the hospital was the most unfamiliar environment for all of the listeners, regardless of the native languages.

5.3 Principal component analysis (PCA)

The interaction between the attributes were investigated by using the Principal Component Analysis (PCA). The aim of the analysis was to extract meaningful factors (correlated groups of attributes) that can explain the relations between soundscape perception of the participants and speech intelligibility. Detailed information on the PCA was presented in Chapter 3.

PCA with a varimax rotation of 11 semantic attributes and the native language variable was conducted on data gathered from 60 participants for each of the 9 cases (3 environments x 3 STI conditions).

5.3.1 Airport

Each STI condition is analysed separately for the airport environment. The common attributes that form a component across all STI conditions are sought and discussed at the end of the section.

<u>STI = 0.4</u>

The results of an orthogonal rotation of the Airport – STI = 0.4 case are shown in Table 5.4. When loadings less than 0.60 were excluded, the analysis yielded a four-component solution with a simple structure (component loadings =>0.60) (Field, 2009).

An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was factorable (KMO = 0.568) (Field, 2009). Kaiser (1974) suggests that values greater than 0.5 are acceptable.

Five attributes loaded onto Component 1, which are annoyance, relaxation, environmental pleasantness, comfort, and noisiness. These four attributes are all related to the overall acoustic environment. This component explains 24.3% of the total variance. Three attributes loaded onto Component 2, which are eventfulness, intelligibility, and language. This component explains 17% of the total variance.

Table 5.4 Rotation component matrix for the case: Airport – STI = 0.4.

	Component			
	1	2	3	4
Annoyance	**-,817**	-,144	-,167	,182
Relaxation	**,795**	,304	-,050	,070
Pleasantness (Environment)	**,724**	-,022	,246	,207
Comfort	**,718**	,100	-,096	,228
Noisiness	**-,673**	,051	,081	,213
Eventfulness	-,107	**-,803**	-,134	,144
Intelligibility	,059	**,708**	,144	,036
Language	,109	**,638**	-,205	,132
Loudness	-,133	-,047	**,777**	,038
Pleasantness (Speech)	,279	,151	**,765**	-,035
Familiarity	-,064	-,433	,235	**-,663**
Excitement	-,010	-,394	,342	**,633**

Two attributes loaded onto Component 3, which are loudness and speech pleasantness. This component explains 12.9% of the total variance. Lastly, familiarity and excitement loaded onto Component 4, which explains 8.8% of the total variance.

STI = 0.5

The results of an orthogonal rotation of the Airport – STI = 0.5 case are shown in Table 5.5. When loadings less than 0.60 were excluded, the analysis yielded a four-component solution with a simple structure (component loadings =>0.60) (Field, 2009).

An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was factorable (KMO = 0.644) (Field, 2009).

Four attributes loaded onto Component 1, which are annoyance, comfort, noisiness, and relaxation, which are the same attributes of Component 1 of the airport – STI = 0.4 case, except the environmental pleasantness. This component explains 19.8% of the total variance. Two attributes loaded onto Component 2, which are language and loudness. Similar to the component 3 of the airport – STI = 0.4 case, both attributes are related to the speech. This component explains 15.1% of the total variance. Component 3 has only one attribute, which is intelligibility. It explains 13.5% of the total variance. Lastly, Component 4 has two attributes loaded, which are eventfulness and excitement, and the component explains 12% of the total variance.

Table 5.5 Rotation component matrix for the case: Airport – STI = 0.5.

	Component			
	1	2	3	4
Annoyance	**-,816**	-,063	,062	-,094
Comfort	**,674**	,148	,252	,113
Noisiness	**-,662**	-,224	,015	,145
Relaxation	**,647**	-,213	,450	-,196
Language	-,080	**-,738**	-,119	-,202
Loudness	,102	**,683**	,056	-,273
Pleasantness (Speech)	,363	,574	,332	-,082
Familiarity	,003	,541	-,019	,178
Intelligibility	,041	,222	**,875**	-,042
Pleasantness (Environment)	,497	,070	,589	,399
Eventfulness	,019	,086	-,232	**,729**
Excitement	-,032	,025	,238	**,723**

STI = 0.6

The results of an orthogonal rotation of the Airport – STI = 0.6 case are shown in Table 5.6. When loadings less than 0.60 were excluded, the analysis yielded a four-component solution with a simple structure (component loadings =>0.60) (Field, 2009).

An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was factorable (KMO = 0.655) (Field, 2009).

Five attributes loaded onto Component 1, which are annoyance, noisiness, environmental pleasantness, relaxation, and comfort, which are exactly the same attributes of Component 1 of the airport – STI = 0.4 case. This component explains 24.4% of the total variance. Two attributes loaded onto Component 2, which are excitement and language. This component explains 12.5% of the total variance. Component 3 has only one attributes as seen in the previous case of airport – STI = 05, however, the attribute is familiarity instead of intelligibility. It explains 13.5% of the total variance. Lastly, Component 4 has two attributes loaded, which are eventfulness and excitement, and explains 12.2% of the total variance.

Common attributes

After examining the components' loadings of all three STI conditions at the airport, 4 common attributes were observed in Component 1 (annoyance, noisiness, relaxation, and

Table 5.6 Rotation component matrix for the case: Airport – STI = 0.6.

	Component			
	1	2	3	4
Annoyance	**-,821**	,081	,150	-,164
Noisiness	**-,764**	,247	-,018	,021
Pleasantness (Environment)	**,733**	,256	,041	,061
Relaxation	**,731**	-,085	,218	-,116
Comfort	**,724**	,205	,352	,046
Excitement	,060	**,820**	-,123	,151
Language	,026	**-,676**	-,335	,412
Intelligibility	,082	,141	**,791**	,287
Pleasantness (Speech)	,111	-,118	**,692**	-,154
Familiarity	,122	-,002	-,098	**,736**
Loudness	,031	-,032	,153	,561
Eventfulness	-,198	,386	-,047	,472

comfort) and 1 common attribute was observed in Component 2 (language). No common attributes were observed in Component 3 and Component 4. When the two largest components were considered together it was seen that environmental pleasantness is also common in 2 out of 3 STI conditions.

The four common attributes loaded onto Component 1 are all semantic attributes that were used to analyse the perception of the overall acoustic properties of an environment, rather than evaluating the speech. From a subjective point of view, the attributes are mainly related with the perception of noise and the feelings derived from it. It is not surprising to see the noisiness attribute and the annoyance attribute on the same factor, which are both mainly related to the assessment of background noise sources. Similarly, relaxation and comfort are both semantic attributes that were used to investigate the feelings of the participants. According to the analysis of the four attributes, Component 1 was labelled “emotional effects”. Language is the only common attribute that was loaded onto Component 2 across the three STI conditions. Therefore, component 2 was labelled “language”.

5.3.2 Hospital

Each STI condition is analysed separately for the hospital environment. The common attributes that form a factor across all STI conditions are sought and discussed at the end of the section.

<u>STI = 0.4</u>

The results of an orthogonal rotation of the hospital – STI = 0.4 case are shown in Table 5.7. When loadings less than 0.60 were excluded, the analysis yielded a five-component solution with a simple structure (component loadings =>0.60) (Field, 2009).

An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was factorable (KMO = 0.624) (Field, 2009).

Three attributes loaded onto Component 1, which are annoyance, relaxation, and environmental pleasantness. This component explains 20.7% of the total variance. One attribute loaded onto Component 2, which is excitement. This component explains 15.75% of the total variance. Component 3 has only one attributes well, which is

Table 5.7 Rotation component matrix for the case: Hospital – STI = 0.4.

	Component				
	1	2	3	4	5
Annoyance	**-,825**	,173	,001	-,109	,138
Relaxation	**,717**	,302	-,278	-,140	,019
Pleasantness (Environment)	**,676**	,234	-,041	,135	,476
Comfort	,567	,427	-,207	-,372	,161
Noisiness	-,555	,067	,534	-,061	,399
Language	,545	-,509	,107	,118	,361
Excitement	-,097	**,728**	,033	,091	,070
Eventfulness	-,411	,587	-,196	,268	,259
Loudness	,195	,335	**,653**	,007	,058
Familiarity	-,192	,048	-,474	**,744**	-,024
Intelligibility	,411	-,108	,434	**,606**	,033
Pleasantness (Speech)	,326	,431	,332	,132	**-,628**

loudness. It explains 13.3% of the total variance. Two attributes loaded onto Component 4, which are familiarity and intelligibility. Lastly, Component 5 has one attribute loaded, which is the speech pleasantness that explains 10.5% of the total variance.

STI = 0.5

The results of an orthogonal rotation of the hospital – STI = 0.5 case are shown in Table 5.8. When loadings less than 0.60 were excluded, the analysis yielded a four-component solution with a simple structure (component loadings =>0.60) (Field, 2009).

An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was factorable (KMO = 0.733) (Field, 2009).

Three attributes loaded onto Component 1, which are noisiness, relaxation, and environmental pleasantness. This component explains 22.2% of the total variance. Two attributes loaded onto Component 2, which are intelligibility and loudness. This component explains 19.4% of the total variance. Component 3 has two attributes as well, which are excitement and language. It explains 11.5% of the total variance. Lastly, one attribute loaded onto Component 4, which is familiarity. It explains 10.6% of the total variance.

Table 5.8 Rotation component matrix for the case: Hospital – STI = 0.5.

	Component			
	1	2	3	4
Noisiness	**-,809**	,074	-,073	,067
Relaxation	**,798**	,217	,203	,109
Pleasantness (Environment)	**,676**	,382	,072	,171
Eventfulness	-,533	-,076	,361	,223
Comfort	,495	,423	-,011	,335
Intelligibility	,067	**,814**	,013	,049
Loudness	,046	**,773**	,119	-,166
Pleasantness (Speech)	,312	,591	,057	,377
Annoyance	-,506	-,558	-,046	,118
Excitement	,140	,161	**,825**	,145
Language	-,025	-,005	**-,701**	,358
Familiarity	-,024	-,050	-,061	**,854**

STI = 0.6

The results of an orthogonal rotation of the hospital – STI = 0.6 case are shown in Table 5.9. When loadings less than 0.60 were excluded, the analysis yielded a five-component solution with a simple structure (Component loadings =>0.60) (Field, 2009).

An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was factorable (KMO = 0.517) (Field, 2009).

Two attributes loaded onto Component 1, which are speech pleasantness and annoyance. This component explains 17.65% of the total variance. Three attributes loaded onto Component 2, which are language, relaxation, and eventfulness. This component explains 14.7% of the total variance. Component 3 has only one attribute loaded, which is excitement. It explains 14% of the total variance. Two attributes loaded onto Component 4, which are loudness and intelligibility. Lastly, two attributes loaded onto Component 5, which are noisiness and familiarity. It explains 10.6% of the total variance.

Table 5.9 Rotation component matrix for the case: Hospital – STI = 0.6.

	Component				
	1	2	3	4	5
Pleasantness (Speech)	**,809**	,003	-,230	,006	,182
Annoyance	**-,720**	-,179	-,226	,118	,231
Pleasantness (Environment)	,509	,227	,336	,404	-,216
Language	-,022	**,836**	,095	,290	,127
Relaxation	,274	**,659**	,353	-,173	-,100
Eventfulness	-,003	**-,623**	,425	,190	,152
Excitement	-,041	,139	**,883**	-,016	,008
Comfort	,499	-,151	,537	,129	-,084
Loudness	-,203	,190	-,095	**,824**	-,006
Intelligibility	,270	-,281	,158	**,652**	,034
Noisiness	-,268	,033	-,127	-,100	**,829**
Familiarity	,418	-,084	,171	,146	**,628**

Common attributes

After examining the components' loadings of all three STI conditions at the hospital, no common attributes were identified in Component 1, Component 2, Component 3 and Component 4. When the two largest components were considered together, relaxation is common across all three STI conditions. Additionally, annoyance and environmental pleasantness were common in 2 out of 3 STI conditions. These attributes were also loaded onto Component 1 of the airport environment; however, instead of comfort and noisiness, the environmental pleasantness attribute loaded onto Component 1 of the hospital environment. It is important to note that language was not a common attribute in a single component at the hospital environment; however, it was common in between STI = 0.6 Component 2 and STI = 0.5 Component 3. Therefore, it is assumed that it still explains some variance at higher STI conditions.

5.3.3 Café

Each STI condition is analysed separately for the café environment. The common attributes that form a component across all STI conditions are sought and discussed at the end of the section.

STI = 0.4

The results of an orthogonal rotation of the café – STI = 0.4 case are shown in Table 5.10. When loadings less than 0.60 were excluded, the analysis yielded a five-component solution with a simple structure (Component loadings =>.60) (Field, 2009).

An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was factorable (KMO = 0.649) (Field, 2009).

Four attributes loaded onto Component 1, which are environmental pleasantness, comfort, relaxation, and annoyance. This factor explains 22.3% of the total variance. Two attributes loaded onto Component 2, which are language and familiarity. This component explains 13% of the total variance.

Table 5.10 Rotation component matrix for the case: Café – STI = 0.4.

	Component				
	1	2	3	4	5
Pleasantness (Environment)	**,805**	,083	-,003	,003	,037
Comfort	**,762**	-,187	-,081	-,125	,040
Relaxation	**,744**	-,074	,191	,245	-,317
Annoyance	**-,672**	-,093	,377	-,080	-,341
Excitement	,562	,238	,153	,509	,028
Language	,000	**,860**	,047	,110	,085
Familiarity	,002	**-,679**	,293	,310	,099
Noisiness	-,229	-,199	**,820**	,041	-,071
Pleasantness (Speech)	,238	,071	**,660**	-,256	,249
Eventfulness	,024	-,087	-,168	**,835**	-,021
Intelligibility	,109	-,268	,227	-,130	**,737**
Loudness	-,087	,351	-,093	,119	**,616**

Component 3 has two attributes loaded, which are noisiness and speech pleasantness. It explains 12.4% of the total variance. Only one loaded onto Component 4, which is eventfulness. It explains 10.3% of the total variance. Lastly, two attributes loaded onto Component 5, which are intelligibility and loudness. It explains 10.2% of the total variance.

STI = 0.5

The results of an orthogonal rotation of the café – STI = 0.5 case are shown in Table 5.11. When loadings less than 0.60 were excluded, the analysis yielded a four-component solution with a simple structure (Component loadings =>0.60) (Field, 2009). An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was factorable (KMO = .712) (Field, 2009).

Four attributes loaded onto Component 1, which are annoyance, relaxation, comfort, and environmental pleasantness. This component explains 26.15% of the total variance. Three attributes loaded onto Component 2, which are language, familiarity, and noisiness. This factor explains 17.5% of the total variance. Component 3 has one attribute loaded, which is eventfulness. It explains 10.9% of the total variance. Lastly one attribute loaded onto the Component 4 as well, which is intelligibility. It explains 10.3% of the total variance.

Table 5.11 Rotation component matrix for the case: Café – STI = 0.5.

	Component			
	1	2	3	4
Annoyance	**-,799**	,147	-,045	,011
Relaxation	**,792**	-,166	,098	,140
Comfort	**,792**	,207	-,120	-,064
Pleasantness (Environment)	**,745**	,179	,050	,291
Excitement	,497	,044	,443	,132
Language	-,085	**-,769**	,246	,245
Familiarity	,191	**,720**	,009	,084
Noisiness	-,467	**,659**	,147	,155
Loudness	,284	-,516	-,206	,309
Eventfulness	-,055	-,065	**,886**	-,098
Pleasantness (Speech)	,304	,396	,422	,352
Intelligibility	,076	-,092	-,022	**,891**

STI = 0.6

The results of an orthogonal rotation of the café – STI = 0.6 case are shown in Table 5.12. When loadings less than 0.60 were excluded, the analysis yielded a five-component solution with a simple structure (Component loadings =>0.60) (Field, 2009).

An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the sample was factorable (KMO = 0.568) (Field, 2009).

Two attributes loaded onto Component 1, which are excitement and familiarity. This component explains 18.2% of the total variance. Two attributes loaded onto Component 2, which are relaxation and environmental pleasantness. This component explains 16.5% of the total variance. Component 3 has two attributes loaded, which are language and intelligibility. It explains 14.1% of the total variance. One attribute loaded onto Component 4, which is eventfulness. It explains 12.6% of the total variance. Lastly one attribute loaded onto Component 5 as well, which is loudness. It explains 9.9% of the total variance.

Table 5.12 Rotation component matrix for the case: Café – STI = 0.6.

	Component				
	1	2	3	4	5
Excitement	**,807**	,220	-,030	-,103	,174
Familiarity	**,803**	-,143	,049	,039	-,170
Pleasantness (Speech)	,544	,211	,293	,405	-,195
Relaxation	,097	**,772**	-,098	,048	-,122
Pleasantness (Environment)	,199	**,736**	,380	,044	,251
Noisiness	,420	-,593	,209	,274	-,052
Language	,074	,196	**-,797**	,045	,232
Intelligibility	,110	,117	**,763**	,003	,178
Eventfulness	,104	,065	-,050	**,836**	,083
Annoyance	-,507	-,380	,065	,554	-,165
Comfort	,293	,417	,427	-,496	-,211
Loudness	-,046	-,004	-,018	,052	**,918**

Common attributes

After examining the components' loadings of all three STI conditions at the café, similar to the hospital environment, no common items were identified in Component 1; however, when Component 1 and Component 2 of the three environments were inspected together, the environmental pleasantness, the familiarity, and the relaxation attributes were common at all three STI conditions. The environmental pleasantness and the relaxation attributes loaded onto Component 1 at STI = 0.4 and STI = 0.5, and loaded onto Component 2 at STI = 0.6. The familiarity attribute loaded onto Component 2 at STI=0.4 and STI=0.5, and onto Component 1 at STI=0.6. The familiarity attributes loaded onto Component 1 at STI = 0.6 largely explains the total variation. The three identified attributes (environmental pleasantness, familiarity, and relaxation) are then considered as a common component. No common attributes were observed in Component 2, Component 3 and Component 4.

Additionally, contradicting the airport and the hospital results, the annoyance attribute was not common across the three STI conditions at the café. Environmental pleasantness, relaxation, and comfort were identified as semantic attributes that were used to subjectively evaluate an acoustic environment. It should be noted that language was not a common item in a single factor at the café environment; however, it was common in between Component 2 (STI = 0.4 and STI =0.5) and Component 3 (STI = 0.6), showing that it still explains some variation of the scores at the café.

5.3.4 Summary

According to the results of the principal component analysis (Table 5.13), 3 attributes are common in the first two components in minimum 6 (2 STI conditions x 3 environments) out of 9 acoustic conditions (3 STI conditions x 3 environments). These semantic attributes are relaxation, annoyance, and environmental pleasantness, which are considered as the main attributes that explain a large variance of the test results. It is important to note that relaxation is the only components' attribute that is common in all of the STI conditions and environments.

Overall these findings indicate that soundscape perception of the three environments tested was mainly affected by relaxation, annoyance and environmental pleasantness; relaxation being the only semantic attribute included in a component for all the cases considered (9 cases = 3 STI conditions x 3 environments). In addition to providing good speech intelligibility where communication is crucial, good acoustic design should then take into account these factors for improving soundscape quality. Furthermore, such factors can also be beneficial in improving speech intelligibility (as pointed out above for pleasantness).

Table 5.13 Summary table of the semantic attributes and components these attributes are loaded in.

	Airport			Hospital			Café		
	STI=0.4	STI=0.5	STI=0.6	STI=0.4	STI=0.5	STI=0.6	STI=0.4	STI=0.5	STI=0.6
Intelligibility	2	3	3	4	2	4	5	4	3
Loudness (Speech)	3								
Pleasantness (Speech)	3		3			1	3		
Noisiness	1	1	1		1		3	2	
Annoyance	1	1	1	1		1	1	1	
Relaxation	1	1	1	1	1	2	1	1	2
Comfort	1	1	1				1	1	
Pleasantness (Env.)	1		1	1	1		1	1	2
Eventfulness	2	4				2	4	3	4
Excitement	4	4	2	2	3	3			1
Familiarity	4		4	4	4		2	2	1
Loudness		2		3	2	4			
Language	*2*	*2*	*2*		3	2	2	2	3

- Relaxation was the only components' attribute that was common across all 3 STI conditions and environments.
- The airport had the most common attributes loaded onto the same components. Annoyance, relaxation, comfort, and noisiness attributes were loaded onto Component 1 under all 3 STI conditions.
- The airport was the only environment that had the language attribute loaded on a single component (component 2) as a common attribute across all 3 STI conditions.
- Familiarity was common across all 3 STI conditions only at the café.
- When the largest 2 components were considered together, relaxation, annoyance, and environmental pleasantness attributes explain a large variance across the 3 STI conditions and the 3 environments.
- Intelligibility attribute was always loaded onto a component, ranging from component 2 to component 5, across all 3 STI conditions and environments. It was loaded onto the same components with speech loudness (3 times), language (2 times), speech pleasantness (1 time), eventfulness (1 time), and familiarity (1 time) attributes.

5.4 Conclusions

This chapter presented and analysed the second phase of the study, which investigated how soundscape perception might affect the perceived speech intelligibility of English, Polish, Arabic, and Mandarin. First, semantic differential analyses of the results were investigated for 11 semantic attributes. A principal component analysis was also conducted for the 9 cases considered (3 environments and 3 STI conditions) in order to reveal the attributes that create a component, which consequently explains the total variations of the listening test scores.

The semantic differential analysis of the intelligibility attribute revealed that perceived speech intelligibility of each language varies with the type of environment, as well as the type of background noise, reverberation time, and signal-to-noise ratio. It should be noted that, perceived speech intelligibility and actual intelligibility can be different and are affected by multiple factors.

Perceived speech intelligibility of English appeared to be mostly affected negatively by the information content and distracting sounds present in the background noise. Public announcements made in English might have played a role in this, and further research could test whether public announcements made in different languages might lead to different findings. However, this might not necessarily be the case, as the distractive noise represented by a phone ringing in the hospital also appeared to have a greater negative effect on English participants. Furthermore, the contradictions between the two phases of the study also suggest that while English and Polish listeners were more sensitive to distractive noise sources, Arabic listeners were more resilient.

A principal component analysis showed that in addition to providing good speech intelligibility where communication is crucial, good acoustic design should then take into account these factors for improving soundscape quality. Furthermore, such factors can also be beneficial in improving speech intelligibility (as pointed out above for pleasantness).

The perceptual experiment of multi-lingual speech intelligibility presented in this chapter, aimed at highlighting the effects of language, environment, and room acoustic conditions on the total variance of the test results. Detailed semantic differential analysis and principal component analysis of the attributes revealed that the room acoustic conditions and the type of environments have varying effects on the perceived intelligibility of the four languages tested. It can therefore be stated that justifying speech intelligibility by room acoustic parameters only might mislead the designers and planners, especially while designing a multi-lingual environment. Design requirements of a multi-lingual environment are rather complex: the type of environment, type of background noise, reverberation time, and context can influence intelligibility of languages and more generally oral communication depending on the language an environment considered.

CHAPTER 6

Standards and design guidelines of spaces used for speech and their relation with multilingual intelligibility

---

6.1 Introduction

The chapter investigates the effectiveness of the current standards and design guidelines in terms of speech intelligibility in multi-lingual environments. This is done from the perspective of the outcomes of the present study, more specifically, the results of Chapter 4. These results showed that there was a significant difference between the word intelligibility scores of these languages. Under the same acoustic conditions (reverberation time and S/N ratio), the word intelligibility scores of each language differed between each other, depending on the linguistic and distinctive features' properties of the languages (see Chapter 4 for details). Therefore, in this chapter, implementation of such results to the current standards and design guidelines is aimed, by suggesting corrections to the STIs calculated based on the background noise (BN) and reverberation time (*T)* suggestions presented in the standards and design guidelines presented. Table 6.1 summarises the issues addressed in this chapter. The chapter initially reviews current standards and design guidelines, followed by a discussion looking at the application of STI corrections across languages.

Table 6.1 Issues to be addressed in relation to current design guidelines related to speech intelligibility.

	Current design guidelines	**Design guidelines for multilingual analysis**
Acoustic parameters	Background noise (BN) and reverberation time (*T*), with occasional specification of the speech transmission index (STI)	Calculate STI corresponding to each BN and *T* condition, in order to carry out multilingual comparisons based solely on the STI
Language effects	Languages effects ignored	Include languages effects by applying corrections to the recommended STI

There are 5 standards and 2 design guidelines that can be consulted in the process of designing various multilingual spaces: ‘sound system equipment – Part 16: objective rating of speech intelligibility by speech transmission index’ (BS EN 60268-16, 2011), ‘Ergonomics – assessment of speech communication’ (BS EN 9921, 2003) ‘Acoustical performance criteria, design requirements, and guidelines for schools’ (ANSI S12.60, 2002), ‘Sound insulation and noise reduction for buildings – Code of Practice’ (BS 8233, 1999), and ‘Acoustic design of schools: performance standards’ (BB93-PS, 2015). The guidelines presented are ‘Acoustic design of schools: a design guide’ (BB93-DG, 2015) and ‘Sound control for improved outcomes in healthcare settings’ (Joseph and Ulrich, 2007).

Each standard and design guideline is presented and discussed in terms of importance given to speech intelligibility, specifically to room acoustic parameters (i.e. reverberation time (T), signal-to-noise ratio (L_{SN}), and ultimately the speech transmission index (STI)), and multilingual communication. Very few documents presented in this section were focused on the STI, except for some sections of the standards and guidelines on open-plan environments. Due to the fact that the data presented in the current study is based on the STI, the signal-to-noise ratio and reverberation time information presented in such documents have been converted to the STI by using the modulation transfer function (MTF), in order to achieve comparable data.

The standards and the guidelines selected can be applied to multilingual spaces, and cover some of the spaces tested in the second phase of the present study (hospitals, and cafés). Additionally, standards and design guidelines also cover schools, due to the fact that such spaces have increasingly become multilingual in large-scale modern cities. It should also be noted that the airport terminal design guideline (FAA, 1988) is not specific to room acoustics; instead, it is an extensive guideline on most of the architectural and engineering aspects of terminal building design, and therefore has been excluded from the review.

The MTF forms the basis of the STI method and is typically determined from impulse responses (Rife, 1992), but can also be estimated from the reverberation time and signal-to-noise ratio present in the space (Houtgast *et al.*, 1980). The importance of being able to quantify the STI from simple room acoustic parameters lies in the fact that this allows determining a fundamental design parameter without the need of specialist equipment or software (e.g. maximum length sequence software or ray tracing software), as a simple spreadsheet can be used. This method can therefore be used by non-specialists for design purposes or acoustic assessments (Houtgast *et al.*, 1980).

Accuracy of the MTF based on simple room acoustic parameters was tested by Galbrun and Kitapci (2014), in order to define its applicability and limitations. This was achieved by comparing STI values obtained from the impulse response method based on maximum length sequence analysis (Rife, 1992) and for which accuracy is known, with STI values calculated from the reverberation time and signal-to-noise ratio (Houtgast *et al.*, 1980). Two rooms were tested under sixteen different acoustic conditions (different reverberation times and signal-to-noise ratios), allowing to examine a wide range of STI values (0.1–0.8) and carrying out a detailed analysis. Differences in STI between measured results and simple predictions based on T and L_{SN} were always lower than 0.1 (on a 0.0–1.0 scale), and on average always lower than 0.06. These differences were noticeable and therefore non-negligible, as a change in STI of 0.03 has been demonstrated as a just noticeable difference by Bradley *et al.* (1999). It should be noted that STI predictions based on T and L_{SN} tend to underestimate the STI.

In this chapter, the effectiveness of 5 standards and 2 design guidelines that can be consulted in the process of designing various multilingual spaces are discussed in relation with the results of the Chapter 4 of the thesis. The STIs are calculated based on the background noise (BN) and T suggestions presented in these standards and design guidelines, and a comparison of the STIs calculated is given in Section 6.3 in relation to the results of Chapter 4.

6.2 Standards and design guidelines

6.2.1 Speech communication standards

6.2.1.1 Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index (BS EN 60268-16:2011)

The standard ‘Sound system equipment – Part 16: objective rating of speech intelligibility by speech transmission index’ was published by the British Standards Institute in 2011. The document extensively discusses the theoretical background of STI and revised STI methods. In Annex A of the document, the modulation transfer function (MTF) and the calculation of the STI, including additional information on auditory masking, absolute speech reception thresholds, gender-specific octave band weighting as well as redundancy factors, and gender-specific spectra of STI test signals are explained. In Annex B, C, and D the alternative methods such as STIPA, STITEL, and the now obsolete RASTI are discussed. Annex E presents the comparison between the STI and other speech intelligibility measures, such as consonant-vowel-consonant (CVC) word tests, phonemically balanced (PB) word tests, and speech reception threshold (SRT) tests. The relationships between these measures are presented in Figure 6.1.

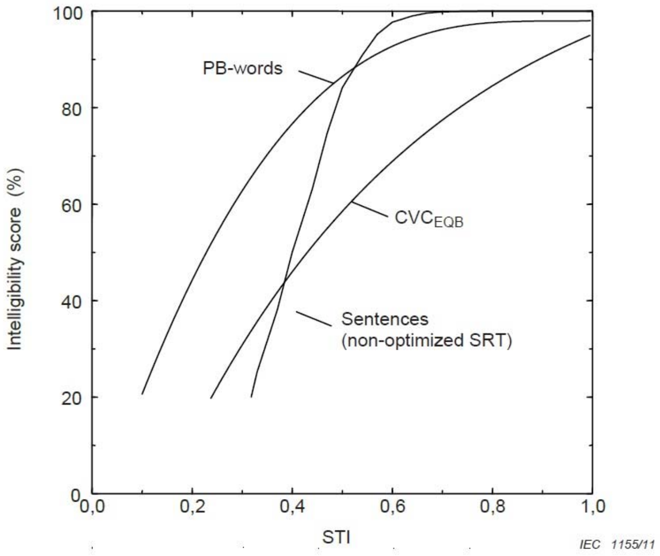


Figure 6.1 Relationships between some speech intelligibility measures (BS EN 60268-16, 2011).

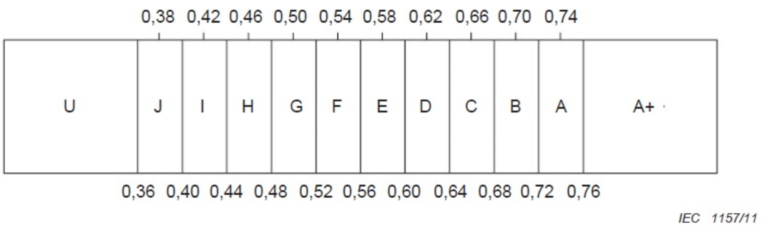


Figure 6.2 STI qualification bands (upper row of numbers represents the STI values at the centre of the bands, whereas lower row of numbers represents the STI values at the edges of the bands) (BS EN 60268-16, 2011).

Table 6.2 Adjusted intelligibility qualification tables for non-native listeners (BS EN 60268-16, 2011).

STI label range	Standard STI	Non-native category I experienced, daily second language use	Non-native category II intermediate experience, and level of second language use	Non-native category III new learner, infrequent second language use
bad - poor	0,30	0,33	0,38	0,44
poor - fair	0,45	0,50	0,60	0,74
fair - good	0,60	0,68	0,86	impossible
good - excellent	0,75	0,86	impossible	impossible

In Annex F and Annex G of the document, nominal qualification bands for the STI (Figure 6.2) and examples of typical applications are presented. The qualification scale is divided into several bands in order to provide flexibility for different applications.

Non-native speech intelligibility is discussed in Annex H of the document. It is suggested that non-native listeners require a 4 dB to 5 dB higher signal-to-noise ratio for similar intelligibility, compared to native listeners. Table 6.2 presents the adjusted intelligibility qualifications for three groups of non-native listeners based on the qualification rating presented in ISO 9921 (2003). The groups are defined depending on language experience, age of learning, and frequency of use of the second language. The concept of adjusted intelligibility is interesting in relation to multilingual communication, and this is discussed further in Section 6.3.

Overall, the document extensively presents the STI, alternative methods (e.g. STIPA and STITEL), qualification bands, and adjusted qualification tables based on the qualification rating presented in ISO 9921 (2003), which is discussed in the next section.

6.2.1.2 Ergonomics – Assessment of speech communication (BS EN ISO 9921:2003)

The standard 'Ergonomics – assessment of speech communication' was published by the British Standards Institute in 2003. The aim of the standard is to recommend required speech communication quality levels for various applications, such as warning and information messages, and general speech communication for work places, public areas, meeting rooms, and auditoria.

Table 6.3 Intelligibility rating and relations between various intelligibility indices (BS EN ISO 9921, 2003).

Intelligibility rating	**STI**	**Sentence score** %	**PB word score** %	**CVC word score** %
Excellent	> 0,75	100	> 98	> 81
Good	0,60 to 0,75	100	93 to 98	70 to 81
Fair	0,45 to 0,60	100	80 to 93	53 to 70
Poor	0,30 to 0,45	70 to 100	60 to 80	31 to 53
Bad	< 0,30	< 70	< 60	< 31

A total of 4 annexes out of 8 are found to be relevant to the present study. In Annex A of the standard, speaker and listener characteristics (e.g. vocal effort, distance between speaker and listener, and effect of non-native speakers and listeners) are presented. Next, in Annex B, fundamental information on CVC word tests, PB word tests, and sentence tests are given. It should be noted that the diagnostic rhyme test (DRT), which was used in the present study, is not mentioned in the standard. Basic information on the STI is presented in Annex C. Finally, in annex F of the document, an overview of subjective and objective test methods is given, and the relation between such intelligibility ratings are presented. Table 6.3 presents the intelligibility rating and the relations between various intelligibility indices.

Based on the intelligibility rating presented in Table 6.3, minimal performance rating recommendations are given in the standard, as well. It should also be noted that the intelligibility rating presented in Table 6.3 is mentioned throughout this chapter, in order to evaluate the speech intelligibility suggestions of the standards and design guidelines mentioned.

6.2.2 Design standards and guidelines

6.2.2.1 Sound insulation and noise reduction for buildings – Code of Practice (BS 8233:1999)

The standard ‘Sound insulation and noise reduction for buildings – code of practice’ was published by the British Standards Institute in 1999. Environmental noise limits in and around several types of buildings, including and not limited to dwellings, offices, schools, hospitals, and rooms for speech were defined, as well as design criteria, in order to accomplish satisfactory environments for most of the people on an objective and quantifiable basis. Although the standard focuses on background noise levels, reverberation time suggestions are also presented in most of the sections that are related to speech communication.

The document extensively presents information on measurement equipment, types of internal and external noise sources, and the procedure of planning and designing various building types in order to create a pleasing environment for most people under separate sections. In the first 6 sections, the limits of acceptable environmental noise levels and the techniques needed to achieve these levels are presented, but these are outside the scope of the thesis; however, information on environmental noise limits (i.e. background noise and signal-to-noise ratio) and reverberation time suggestions on previously mentioned building types are presented in section 7 - 'specific types of buildings' and more specifically in section 7.6.7 – 'rooms for speech'.

The standard suggests that indoor ambient noise should be controlled by limiting noise sources such as traffic and indoor mechanical services in order to achieve: reasonable working conditions, reasonable speech intelligibility, reasonable listening conditions, and reasonable resting/sleeping conditions. In terms of speech intelligibility, good and reasonable background noise levels for various types of rooms (Table 6.4) and suggested reverberation times for rooms for speech and rooms for music (Table 6.5) are presented. Ambient noise levels and reverberation times are considered separately, without mentioning the speech transmission index (STI), which is a combined function of both acoustic properties. The document also states several architectural design and planning decisions, such as separating rooms containing noise sources and rooms of work or resting.

In section 7.6.7.3 'design for good speech communication', it is suggested that the sound that arrives to the listener can be enhanced by improving the direct sound path and the early reflected sound that reaches listeners within the first 35 ms of the direct sound, and by controlling the reverberation times by aiming at the suggested values presented in Table 6.5. Additionally, general information on sound absorbent and sound insulating materials is given.

Table 6.4 Indoor ambient noise levels in spaces related to speech when they are unoccupied (BS 8233, 1999).

Criterion	Typical situations	Design range $L_{Aeq,T}$ dB	
		Good	**Reasonable**
Reasonable speech or telephone communications	Department store	50	55
	Cafeteria, canteen, kitchen	50	55
	Wash-room, toilet	45	55
	Corridor	45	55
Reasonable listening conditions	Classroom	35	40
	Church, lecture theatre, cinema	30	35
	Concert hall, theatre	25	30
	Recording studio	20	25

Table 6.5 Guide to reverberation time (*T)* at 500 Hz in unoccupied rooms used for speech and music BS 8233:1999 (1999).

Room volume (m^3)	**Reverberation time** *(T)*	
	Speech	**Music**
50	0.4	1.0
100	0.5	1.1
200	0.6	1.2
500	0.7	1.3
1000	0.9	1.5
2000	1.0	1.6

The STI values presented in Table 6.6 and Table 6.7 are based on the BN and the *T* suggestions for different types of rooms (i.e. type of activities and room volume in m^3) given in BS 8233 (1999), and computed by using the MTF method (Houtgast *et al.*, 1980). When calculating signal-to-noise ratios, speech signals were assumed to have a level of 65 dBA, which is in between the range of normal (~60 dBA) and raised (~70 dBA) voice levels measured at 1 m from a person speaking. It can also be noted that calculations assumed identical BN and *T* values across all frequencies, which is a limitation.

An initial investigation of the tables reveals that the STI varies between 0.53 (reasonable STI, 2000 m^3) and 0.78 (good STI for washrooms, toilets, and corridors, 50 m^3) for rooms for conversation, and 0.58 (good and reasonable STI, 2000 m^3) and 0.81 (good and reasonable STI, 50 m^3) for rooms for listening. As previously stated, STI predictions based on *T* and L_{SN} tend to underestimate the STI (average 0.06) (Galbrun and Kitapci, 2014). It is likely that in practice these values correspond to 'good' (STI

Table 6.6 Predicted speech transmission index (STI) values of various types of rooms ($50m^3$-$2000m^3$) based on the background noise (BN) and reverberation time (*T*) suggestions presented in BS 8233 (1999).

(a) $50m^3$ to $200m^3$

Typical situations	**STI**					
	$50m^3$		**$100m^3$**		**$200m^3$**	
	Good	**Reasonable**	**Good**	**Reasonable**	**Good**	**Reasonable**
Rooms for conversation						
Department store	0,74	0,67	0,7	0,64	0,66	0,61
Cafeteria, canteen, kitchen	0,74	0,67	0,7	0,64	0,66	0,61
Wash-room, toilet	0,78	0,67	0,73	0,64	0,69	0,61
Corridor	0,78	0,67	0,73	0,64	0,69	0,61
Rooms for listening						
Classroom	0,81	0,8	0,75	0,74	0,69	0,66
Lecture theatre, cinema	0,81	0,81	0,75	0,75	0,71	0,69
Concert hall, theatre	0,81	0,81	0,75	0,75	0,71	0,71
Recording studio	0,81	0,81	0,75	0,75	0,71	0,71

(b) $500m^3$ to $2000m^3$

Typical situations	**STI**					
	$500m^3$		**$1000m^3$**		**$2000m^3$**	
	Good	**Reasonable**	**Good**	**Reasonable**	**Good**	**Reasonable**
Rooms for conversation						
Department store	0,63	0,58	0,58	0,54	0,56	0,53
Cafeteria, canteen, kitchen	0,63	0,58	0,58	0,54	0,56	0,53
Wash-room, toilet	0,65	0,58	0,6	0,54	0,58	0,53
Corridor	0,65	0,58	0,6	0,54	0,58	0,53
Rooms for listening						
Classroom	0,67	0,66	0,61	0,61	0,58	0,58
Lecture theatre, cinema	0,67	0,67	0,61	0,61	0,58	0,58
Concert hall, theatre	0,67	0,67	0,61	0,61	0,58	0,58
Recording studio	0,67	0,67	0,61	0,61	0,58	0,58

0.6 - 0.75) and ‘excellent’ (STI 0.75 - 1.0) speech intelligibility, according to the qualification rating of ISO 9921 (2003), although the uncertainty of the predictions cannot guarantee that.

It is also interesting to see that the recommended STI decreases as the room volume increases. Additionally, the rooms for conversation require lower STI values compared to the rooms for listening, especially in smaller rooms. Although there is an observable difference between the STI recommendations of the two categories of rooms, the STI difference within each category is very small. The largest difference observed between the categories is 0.1, and the largest difference observed within the categories is 0.05. It should also be noted that, differences between good and reasonable values are also small, as these vary between 0.00 and 0.11.

Finally, it is worth noting that the document states the importance of considerable differences of sensitivity to noise between people. It points out that only physical characteristics of noise sources were taken into account, and that the difference between pleasant and unpleasant sounds, and psychological factors were not considered due to practical difficulties.

6.2.2.2 Acoustical performance criteria, design requirements, and guidelines for schools (ANSI S12.60-2002)

Acoustical performance criteria, design requirements, and guidelines for schools is an American national standard that aims to improve the quality of education by enhancing the acoustical conditions of learning spaces, and consequently increasing the effectiveness of communication between teachers and students, including those who have hearing, language, speech, attention deficit, or learning disabilities. It states that by improving room acoustic conditions, learning and teaching should be more effective and less stressful.

The standard mainly focuses on specifying acoustic design criteria based on *T* and BN. Although the aim of the standard is to achieve sufficient speech levels compared to the background noise levels in a classroom or learning space, the signal-to-noise ratio is not within the scope of the standard, as well as the noise generated by the occupants of such spaces (teachers and students), or the noise created by various activities within the learning spaces. It is suggested that the noise generated by the occupants or the activities can be actively prevented by appropriate controls by the teacher. It is also pointed out that lower reverberation times encourage the users of the space to lower the level of their voices, and increase the intelligibility of speech.

The background noise levels and reverberation times suggested in the document are presented in Table 6.7. Separate sections (C3.3 and C4) in Annex C (design guidelines for controlling reverberation in classrooms and other learning spaces) of the document present guidelines for ancillary spaces (i.e. corridors, gymnasia, and cafeterias) and minor information regarding learning spaces that are larger than 566 m^3. Absorbing material locations and mounting techniques of such materials are also presented for general type and lecture type classrooms separately.

The maximum reverberation time suggestions are made for three octave-band frequencies that are 500 Hz, 1000 Hz, and 2000 Hz. Two different reverberation times are suggested for two categories of classrooms and learning spaces: for spaces smaller than 283 m^3 it should be 0.6 seconds or less, and for spaces larger than 283 m^3 it should be 0.7 seconds or less.

In terms of background noise levels, the standard aimed to reach a +15 dB signal-to-noise ratio in classrooms and other learning spaces, according to the recommendations of the American Speech-Language-Hearing Association (ANSI S12.60, 2002). It is pointed out that a minimum speech level of 50 dBA would be guaranteed to have +15 dB signal-to-noise ratio when the background noise level does not exceed 35 dBA (Table 6.7). It is also stated that both reverberation time and background noise level criteria should be met in order to achieve a satisfying teaching environment.

Table 6.7 Maximum A-weighted steady background noise levels and maximum reverberation times in unoccupied, furnished learning spaces (ANSI S12.60, 2002).

Learning space	**Maximum one-hour-average A-weighted steady background noise level (dB)**	**Maximum reverberation time for sound pressure levels in octave bands with mid-band frequencies of 500, 1000, and 2000 Hz**
Core learning space with enclosed volume < 283 m^3 (< 10.000 ft^3)	35	0,6
Core learning space with enclosed volume > 283 m^3 and < 566 m^3 (> 10.000 ft^3 and < 20.000 ft^3)	35	0,7
Core learning space with enclosed volume > 566 m^3 (> 20.000 ft^3)	50	-

Table 6.8 Predicted speech transmission index (STI) values in core learning spaces based on the background noise (BN) and the reverberation time (*T*) suggestions presented in ANSI S12.60 (2002).

Learning space	**STI**
Core learning space with enclosed volume < 283 m^3 (< 10.000 ft^3)	0,70
Core learning space with enclosed volume > 283 m^3 and < 566 m^3 (> 10.000 ft^3 and < 20.000 ft^3)	0,67

The STI values presented in Table 6.8 are based on the BN and the *T* suggestions given in ANSI S12.60 (2002), and computed by using the MTF method (Houtgast *et al.*, 1980). The reverberation time (*T*) suggestion for core learning spaces larger than 566m^3 was not presented in the standard; therefore, its STI value could not be computed. While calculating signal-to-noise ratios, speech signals were assumed to have a level of 65 dBA.

The suggested BN values for both sizes of core learning spaces are 35 dBA, and the *T* values are 0.6 for spaces smaller than 283 m^3 and 0.7 for spaces larger than 283 m^3. Consequently, the STI values computed were similar: STI=0.70 for spaces smaller than 283 m^3, and STI=0.67 for spaces larger than 283 m^3. These values correspond to 'good' (0.60 - 0.75) speech intelligibility according to the qualification rating of ISO 9921 (2003).

Guidelines for ancillary and large learning spaces are based on the noise reduction coefficient (NRC), which is the arithmetic mean of the sound absorption coefficients at 250, 500, 1000, and 2000 Hz. It was suggested that a suspended ceiling should be constructed by the use of materials with an NRC of 0.7 or higher in cafeterias and large learning spaces with a ceiling height of 3.7 meters or less. Additionally, professional consultancy was suggested for areas that have a ceiling height of 3.7 meters and more.

Annex B and Annex D of the document focus on internal and external noise control. Extensive information on the noise control of HVAC, electrical equipment, plumbing systems, and instructional equipment are presented in Annex B. Further information on the isolation between learning spaces and between learning spaces and other interior or exterior spaces are given in Annex D. It should be noted that architectural design suggestions are also given in both sections in order to isolate spaces containing noise sources and classroom/learning spaces.

6.2.2.3 Acoustic design of schools: performance standards (Building bulletin 93)

The document ‘Acoustic design of schools: performance standards’ presents guidelines for the acoustic design of schools by suggesting standards for indoor ambient noise levels, reverberation time, and speech transmission index. Additionally, useful information on sound insulation is presented. Several room types, in which speech is crucial are investigated, such as teaching spaces (both regular and open-plan), lecture rooms, teaching spaces for students with special hearing and communication needs, meeting rooms, dining rooms, and offices. Other types of rooms, such as rooms for music (i.e. music classrooms, ensemble and recording rooms, and control rooms), sports halls, libraries, administration, and ancillary spaces are also considered in the document; however, such rooms are outside the scope of this thesis and are not considered in the following tables.

All the spaces mentioned above should meet the standards for indoor ambient noise level, sound insulation (airborne and impact), and reverberation time, while the spaces are finished, furnished, and unoccupied. Additionally, meeting the performance standards for speech transmission index (STI) in open-plan spaces is required.

In the document, background noise level in enclosures are referred as indoor ambient noise levels (IANL). It is stated that the aim of reaching acceptable noise levels is to achieve intelligible communication between students, and between students and teachers during learning and study activities in teaching environments. The noise sources taken into account are external sources (i.e. traffic noise and industrial/commercial structures), building services (i.e. HVAC systems, plant, and drainage), and actuator/damper noise. Noise generated by the teaching activities, equipment used in the space, and rain noise are excluded from the IANL, similar to other standards and guidelines that are presented in the current study. The suggestions for upper limits for IANL in rooms used for speech are presented in Table 6.9. It should be noted that the table presented in the original document is more extensive due to the fact that it includes spaces other than rooms for speech, such as rooms for music and sport halls.

Table 6.9 Noise activity and sensitivity levels and upper limits for indoor ambient noise levels in rooms used for speech (BB93, 2015).

Type of room	**Room classification for the purpose of airborne sound insulation**		**Upper limit for the indoor ambient noise level $L_{Aeq,30mins}$ dB**	
	Activity noise (Source room)	**Noise tolerance (Receiving room)**	**New build**	**Refurbishment**
Nursery school rooms, primary and secondary school teaching spaces	Average	Medium	35	40
Open-plan teaching spaces	Average	Medium	40	45
Lecture room	Average	Medium	35	40
Teaching space for students with special hearing and communication needs	Average	Low	30	35
Meeting room, interviewing/counselling room, video conference room	Low	Medium	40	45
Dining room	High	High	45	50
Office, staff room	Low	Medium	40	45

The next room acoustic parameter presented in the document is the reverberation time. The recommended reverberation times given in Table 6.10 are the arithmetic average of the reverberation times in the 500 Hz, 1 kHz, and 2 kHz octave bands, or the one-third octave bands from 400 Hz, to 2.5 kHz. Furthermore, for the rooms used by hearing impaired students or students with language deficiencies, the arithmetic average of the reverberation times in the 125 Hz to the 4 kHz octave bands, or the one-third octave bands from 100 Hz to 5 kHz are used.

It is important to clarify that the recommended STIs given in the document are only reported for open-plan teaching spaces, in order to comply with Requirement E4 of the Building Regulations of England and Wales. Attention should be given to the special needs of such spaces, mainly the need of speech privacy, as well as the speech intelligibility required for intelligible communication. Similar to the other standards discussed in the current study, specialist advice is recommended for open-plan teaching spaces. The recommended STIs are presented in Table 6.11.

The STI values presented in Table 6.12 are based on the BN and the *T* suggestions given in Building Bulletin 93 (2015) (Table 6.9 and Table 6.10), and computed by using the MTF method (Houtgast *et al.*, 1980). While calculating signal-to-noise ratios, speech signals were assumed to have a level of 65 dBA. The STI values were calculated for both new build and refurbishment conditions, and are presented in Table 6.12.

Table 6.10 Performance standards for reverberation time in rooms used for speech (BB93, 2015).

Type of room	**T_{mf} seconds**	
	New build	**Refurbishment**
Nursery school rooms, primary school teaching spaces	≤ 0.6	≤ 0.8
Secondary school teaching spaces	≤ 0.8	≤ 1.0
Open-plan teaching spaces	≤ 0.5	≤ 0.5
Lecture room (fewer than 50 people)	≤ 0.8	≤ 1.0
Lecture room (more than 50 people)	≤ 1.0	≤ 1.0
Teaching space for students with special hearing and communication needs	*T* ≤ 0.4 averaged from 125 Hz to 4 kHz octave band centre frequencies and T ≤ 0.6 in every octave band in this range	≤ 0.4
Meeting room, interviewing/counselling room, video conference room	≤ 0.8	≤ 0.8
Dining room	≤ 1.0	≤ 1.5
Office, staff room	≤ 1.0	≤ 1.2

Table 6.11 Performance standards for speech intelligibility and speech privacy in open-plan spaces – speech transmission index (STI) (BB93, 2015).

Condition	Speech transmission index (STI)
Instruction or critical listening activity - within group	≥ 0.6
Between groups (during critical listening activities)	≤ 0.3

Table 6.12 Predicted speech transmission index (STI) values in core learning spaces based on the background noise (BN) and the reverberation time (*T*) suggestions presented in the Building Bulletin 93 (2015).

Type of room	**STI**	
	New build	**Refurbishment**
Nursery school rooms, primary school teaching spaces	0,7	0,63
Secondary school teaching spaces	0,63	0,58
Open-plan teaching spaces	0,74	0,73
Lecture room (fewer than 50 people)	0,63	0,58
Lecture room (more than 50 people)	0,58	0,58
Teaching space for students with special hearing and communication needs	0,81	0,81
Meeting room, interviewing/counselling room, video conference room	0,63	0,62
Dining room	0,58	0,48
Office, staff room	0,58	0,54

The table illustrates that the lowest STI corresponds to the refurbished dining room (0.49), and the highest STI corresponds to both the new build and the refurbished teaching spaces for students with special communication needs (0.81). When the underestimation of the MTF method based on BN and *T* is taken into account (Galbrun and Kitapci, 2014), all of the STI values computed can be expected to be ‘good’ or ‘excellent’, except for the refurbished dining room, office, and staff room, which are considered to have ‘fair’ speech intelligibility (ISO 9921, 2003). Although conversations and other listening activities in such spaces are not as crucial as a listening activity in a lecture room, attention should also be given to improve speech communication in these spaces.

Prior to the acoustic design process of open-plan offices, preparing an activity management plan is also recommended. According to the activity management plan, STI calculations are needed in order to place appropriate absorption, diffusion, and screening to achieve intelligible performance. Additionally, computer modelling and simulations are highly recommended. The prediction software is suggested to be capable of creating a 3 dimensional model of the space, and comprising surface finishes with adequate absorption and scattering coefficients in relevant octave frequency bands (BB93, 2015).

6.2.2.4 Acoustics of schools: a design guide

The document 'Acoustics of schools: a design guide' is advised to be read in conjunction with the previously discussed document 'Acoustic design of schools: performance standards'. It is an extensive guideline that discusses acoustic performance specifications, noise control, sound insulation, and design of rooms for speech and music. This section exclusively focuses on the areas that are related to the intelligibility of speech between teachers and student, and students and students, which is presented in Section 4 (the design of rooms for speech) of the document. In this section, detailed information on acoustic and architectural design is given, such as indoor ambient noise levels, room size, shape/volume, reverberation times, type/location/distribution of the acoustic absorption, reflector and diffuser usage, and electronic sound reinforcement systems.

First of all, in terms of ambient noise levels, it is stated that the noises generated by teaching activities within the school premises and equipment used in the space, as well as rain noise, are excluded; however, such noise sources should be considered during the design process. High levels of noise that are generated by exceptional events can however be disregarded. It is also stressed that tonal and intermittent noises (e.g. noise sources from building services) are more annoying compared to other noises; therefore, such noises should be at least 5 dB below the given limits.

Additional information is provided in order to limit the ambient noise in teaching environments; such as airborne sound insulation between spaces (including insulation between circulation spaces and other spaces used by students), impact sound insulation (e.g. footsteps) of floors, and sound absorption in corridors, entrance halls, and stairways. It is stated that such improvements help reducing overall ambient noise levels in teaching environments and other crucial areas.

Section 4 of the document, ‘The design of rooms for speech’, extensively investigates the steps to follow in order to achieve the above mentioned criteria in teaching spaces, including but not limited to architectural design decisions, amount and placement of the sound absorbent/diffusing/reflective materials, and electronic sound reinforcement systems. The document initially presents the basics of speech acoustics, such as the frequency ranges of male and female voices, and the importance of recognition of consonants on the intelligibility of speech.

Guidelines on improving speech intelligibility in open-plan spaces are also presented in the document. Open-plan teaching environments are considered as complex acoustic spaces, due to high background noise levels caused by multiple groups of people working in the same environment. In order to maintain high flexibility in open-plan teaching spaces, it is advised to agree on an expected open-plan layout and activity plan with the client. The open-plan layout is suggested to include the position of the teacher during oral presentations, the seating plan of the students, and the learning base areas, and the activity plan is advised to include the estimated number of teachers and students engaged in a discussion at any given time, the number of people walking around the open-plan areas during teaching hours, and any machinery expected to be running in the open-plan areas. Based on the open-plan layout and the activity plan, creating a computer prediction model is suggested, by using a software capable of calculating the STI. The suggested STI levels for within groups should be equal to, or higher than 0.6, and between groups should be equal to, or lower than 0.3 (Table 6.11). The within groups value implies to the communication between the teacher and students, and the between groups value implies communication between adjacent teaching spaces, where the concern is speech privacy rather than speech intelligibility.

After mentioning the basics of speech acoustics, the importance of reverberation time, and equal distribution of sound absorbent materials, the impact of room geometry on direct sound paths is shortly discussed. It is important to note that speech privacy in open-plan teaching environments is also explained in Section 4 of the document. The suggestions for speech intelligibility and speech privacy in open-plan teaching environments include dimensions and materials of divider screens, and preparing appropriate time schedules for activities that occur in close proximity.

6.2.2.5 Sound control for improved outcomes in healthcare settings (The center for health design)

This document investigates possible effects of sound and noise on patients' health, work performance of staff, and speech communication in terms of ambient noise, speech intelligibility, speech privacy, and music impact by reviewing the previous literature of peer-reviewed journal articles, research reports, and books in the area of medicine, psychology, architecture, and acoustics.

The document claims that today's hospitals are exceeding recommended guidelines for noise levels and are extremely noisy; this negatively affects patients' health, work performance and health of hospital staff and visitors. The document presents the effects of sound/noise and suggestions for improvement in terms of ambient noise control, improving speech intelligibility and speech privacy, as well as the impacts of music in such environments.

Firstly, the effects of ambient noise in healthcare settings are discussed. It is stated that the threshold ambient noise levels in hospital patient rooms should not exceed 35 dBA during the day and 30 dBA during the night, according to the World Health Organization (WHO) guidelines (Berglund *et al.*, 1999); however, among 35 published research studies on hospital noise levels, no studies reported noise levels that were within the suggested limits of the WHO (Berglund *et al.*, 1999). Similarly, although the peak noise levels in hospital patient rooms during the night is recommended as 40 dBA (Berglund *et al.*, 1999), it usually exceeds 85 dBA in today's hospitals.

The effects are discussed both from patients and from hospital staffs' perspective. Patients experience annoyance, sleep disruption and awakening, health issues (e.g. decreased oxygen saturation, elevated blood pressure, increased heart rate and respiration rate, and decreased rate of wound healing). On the other hand, hospital staff experience increased perceived work stress, increased fatigue, emotional exhaustion and burnout, and difficulty in communication. Additionally, suggestions on decreasing ambient noise, such as installing sound absorbing ceiling tiles, designing all single-bed rooms, and reducing noise generated by medical equipment and staff conversations are provided in the document. It should be noted that most of the above effects are outside the scope of the present study.

Secondly, speech privacy and patient confidentiality in healthcare settings are discussed. The document states that patients require adequate speech privacy especially while sharing confidential information, specifically in open environments. The lack of speech privacy leads patients to withhold important information, such as private history. Furthermore, it is mandatory to avoid overhearing of private conversations in hospitals. The document recommends the design of special enclosed rooms for admitting areas, psychological testing areas, haematology labs, and examination areas. In addition to architectural design solutions, using a high-performance acoustical ceiling is also recommended.

Lastly, speech intelligibility in healthcare settings is investigated, which is also the focus point of the present study. It is stated that speech intelligibility and speech privacy are closely related; therefore, the aim should be maximising speech intelligibility while maintaining adequate levels of speech privacy. It is interesting that no further information is presented on background noise levels, reverberation times, or speech transmission index in order to justify the suggestions given in the document, such as installing sound absorbing materials or using high performance sound absorbing ceiling tiles. The document mainly focuses on solutions provided by architectural design decisions, with no quantifiable design parameters.

6.3 Discussion

In this section, the relationship between the outcomes of Chapter 4 and the STI values calculated from the data presented in the standards and design guidelines is investigated, in order to identify the difference between the intelligibility of various languages under the same room acoustic conditions that are suggested by widely used standards and design guidelines.

Table 6.13 presents the word and sentence intelligibility scores of English, Polish, Arabic, and Mandarin for a range of STI values covering the standards and design guidelines reviewed (STI = 0.5 - 0.8, Table 6.14). The highest intelligibility score differences between the languages are also given for each STI value. The word and sentence intelligibility scores below STI = 0.5 are not shown, since none of the documents discussed go below this value (STI results rounded to one decimal place).

Table 6.13 Comparison of the word and sentence intelligibility scores for a range of STI values based on the standards and design guidelines reviewed.

	Speech intelligibility scores							
STI	**Word scores (%)**				**Sentence scores (%)**			
	English	Polish	Arabic	Mandarin	English	Polish	Arabic	Mandarin
0.8	93.8	90.3	86.1	84.0	100.0	100.0	96.7	100.0
	Maximum difference: 9.8				Maximum difference: 3.3			
0.7	95.0	87.4	80.3	86.7	97.4	99.3	90.0	100.0
	Maximum difference: 14.7				Maximum difference: 10.0			
0.6	94.8	89.9	81.0	84.7	92.3	100.0	88.9	92.9
	Maximum difference: 13.8				Maximum difference:11.1			
0.5	80.3	65.7	63.8	70.6	75.2	74.7	66.3	77.4
	Maximum difference: 16.5				Maximum difference: 11.1			

Logarithmic regressions were used to calculate speech intelligibility scores at STI = 0.5 and 0.7.

English word scores	$y = 43.585 \ln x + 110.52$	English sentence scores	$y = 66.039 \ln x + 120.99$
Polish word scores	$y = 64.621 \ln x + 110.47$	Polish sentence scores	$y = 73.165 \ln x + 125.44$
Arabic word scores	$y = 48.91 \ln x + 97.729$	Arabic sentence scores	$y = 67.403 \ln x + 113.01$
Mandarin word scores	$y = 47.613 \ln x + 103.64$	Mandarin sentence scores	$y = 67.18 \ln x + 123.99$

Table 6.14 The STI values suggested for various environments by the standards and design guidelines reviewed.

	Environment
STI=0.8	Small (50 m^3) department store, cafeteria, canteen, kitchen, wash-room, toilet, corridor (BS 8233, 1999) Small/medium (100 m^3) classroom, lecture theatre, cinema, concert hall, theatre, recording studio (BS 8233, 1999) New build open-plan teaching spaces (BB93, 2015) New build / refurbished teaching space for students with special hearing and communication needs (BB93, 2015)
STI=0.7	Medium (200 m^3) and medium/large (500 m^3) department store, cafeteria, canteen, kitchen, wash-room, toilet, corridor, classroom, lecture theatre, cinema, concert hall, theatre, recording studio (BS 8233, 1999) Core learning space with enclose volume > 283 m^3 and < 566 m^3 (ANSI S12.60, 2002) Core learning space with enclose volume < 283 m^3 (ANSI S12.60, 2002) New build nursery school rooms, primary school teaching spaces (BB93, 2015) Refurbished open-plan teaching spaces (BB93, 2015)
STI=0.6	Large (1000 m^3) and very large (2000 m^3) department store, cafeteria, canteen, kitchen, wash-room, toilet, corridor, classroom, lecture theatre, cinema, concert hall, theatre, recording studio (BS 8233, 1999) Refurbished nursery school rooms, primary school teaching spaces (BB93, 2015) New build / refurbished secondary school teaching spaces (BB93, 2015) New build / refurbished lecture room (fewer than 50 people) (BB93, 2015) New build / refurbished lecture room (more than 50 people) (BB93, 2015) New build / refurbished meeting room, interviewing/counselling room, video conference room (BB93, 2015) New build dining room (BB93, 2015) New build office, staff room (BB93, 2015)
STI=0.5	Refurbished dining room (BB93, 2015) Refurbished office, staff room (BB93, 2015)

The spaces corresponding to each STI value are also listed in the table, and were identified from the STI predictions illustrated in the previous section. Most of the estimated STI values presented corresponds to either ‘good’ (0.60 - 0.75) or ‘excellent’ (0.75-1.00) intelligibility in the qualification rating of ISO 9921 (2003). It should be noted that most of the room acoustic parameters suggested by the Building Bulletin 93 – Performance standards (BB93-BS, 2015) lead to STI = 0.5 and STI = 0.6, except for open-plan teaching spaces (new build: 0.8, refurbished: 0.7), teaching spaces for students with special hearing and communication needs (0.8), and new build nursery school rooms, as well as primary school teaching spaces (0.7). Furthermore, STI values based on the suggestions of ‘Sound insulation and noise reduction for buildings – Code of Practice’ (1999) also lead to STI = 0.6 for spaces larger than 1000 m^3. The Estimated STI values of ANSI S12.60 (2002) for learning spaces and the estimated STI values for new build nursery school rooms, primary school teaching spaces and refurbished open-plan teaching spaces are 0.7 (BB93-BS, 2015). The STI values based on the suggestions of BS 8233 (1999) vary between STI = 0.8 and STI = 0.6 depending on the room size and tend to decrease as the room volume increases. Altogether, these results highlight the variability between guidelines and their lack of consistency, as comparable spaces (e.g. learning spaces and lecture rooms) show different values of recommended STI, depending on the document considered.

The largest difference between the word intelligibility of languages was seen at STI = 0.5 (16.5%), and the largest sentence intelligibility difference is seen at STI = 0.5 and STI = 0.6 (11.1%) (Table 6.13). Even at higher STIs such as STI = 0.7 and STI = 0.8, the maximum difference between the word intelligibility scores of languages is as high as 14.7% at STI = 0.7 and 9.8% at STI = 0.8, and the maximum difference between the sentence intelligibility scores is 10% at STI = 0.7 and 3.3% at STI = 0.8. These differences highlight the inappropriateness of relying on a single STI value when designing multilingual spaces.

The ISO standard 9921 provides relations between qualification ratings (‘bad’, ‘poor’, ‘fair’, ‘good’ and ‘excellent’) and STI values, as well as relations between intelligibility scores of English and STI values. Such details were discussed in section 6.2.1.2 and are given in Table 6.3, where PB scores are listed instead of DRT scores (although these are expected to be very similar for English (as shown in Section 3.3.1.1)). The word and sentence scores obtained for the four languages tested can be related to the qualification

Table 6.15 Relationship between intelligibility scores and qualification ratings according to BS EN ISO 9921 (based on English language data) and actual STI values obtained from the current study (Chapter 4 results).

ISO 9921 intelligibility scores and ratings	STI				
	ISO 9921	English	Polish	Arabic	Mandarin
PB word score of 93% (‘good’)	0.6	0..6	0.76	0.9	0.8
PB word score of 98% (‘excellent’)	0.75	0.75	0.82	1.00	0.88
Sentence score of 100% (‘good’ / ‘excellent’)	0.6	0.73	0.6	0.82	0.7

Table 6.16 Range of differences in STI values between languages.

ISO 9921 scores and ratings	ΔSTI range
PB word score of 93% (‘good’)	+0.16 - +0.30
PB word score of 98% (‘excellent’)	+0.07 - +0.25
Sentence score of 100% (‘good’ / ‘excellent’)	+0.10 - +0.22

ratings listed in Table 6.3, in view of identifying the STI limits of different languages corresponding to these. Of particular interest are the lower STI limits of the ‘good’ and ‘excellent’ ratings, as these cover the vast majority of spaces listed in Table 6.13 in terms of STI (STI = 0.6 and 0.75 respectively, according to ISO 9921).

The results of this analysis are listed in Table 6.15. Looking first at word scores and their relations with STI values, it can be seen that STI values need to be higher than 0.6 for all languages except English, in order to be in the ‘good’ rating category. The same applies for the STI value of 0.75, which corresponds to the ‘excellent’ rating category. The variations in STI needed to achieve such ratings are given in Table 6.16, where is can be seen that large increases in STI of +0.16 to +0.30 are needed to achieve the ‘good’ rating across the four languages, whilst smaller increases of +0.07 to +0.15 are needed to achieve the ‘excellent’ rating. In any case, all of these increases are larger than 0.03, which is stated as the just noticeable difference in STI by Bradley *et al.* (1999). In other words, none of those increases is negligible. Comparisons can also be made between sentence intelligibility scores, although the comparability of sentence scores is limited (see Chapter 4). For sentence scores, English does not achieve a sentence intelligibility of 100% at STI = 0.6, as expected from ISO 9921, whilst Polish does. Mandarin, English and Arabic achieve that only at higher STI values (0.70, 0.73

and 0.82 respectively), which correspond to increases of +0.10 to +0.22 (Table 6.16). The increases of Table 6.15 can be compared to the increases in STI needed for non-native speakers stated in BS EN 60268-16 (2011). Table 6.2 shows that increases in STI of 0.08 (non-native category I) and 0.28 (non-native category II) are needed to achieve a 'good' rating, and this rating is actually not achievable for the non-native category III. The 'excellent' rating can be achieved with an increase in STI of 0.09 for the non-native category I, whilst it is not achievable for non-native categories II and III. Some of these increases are comparable to those observed between languages, suggesting that STI data such as the one given in Table 6.15 and Table 6.16 could be used for design purposes. It is however important to note that additional tests to those presented in this thesis would need to be carried out across a wider range of STI conditions, in order to obtain more robust and reliable data for languages' intelligibility scores.

To summarise, the results of Table 6.15 and Table 6.16 quantify the changes in STI (and therefore in room acoustic conditions) needed to achieve either 'good' or 'excellent' intelligibility ratings across languages, and these highlight the fact that the STI values recommended in BS EN ISO 9921 (2003) might be appropriate for some languages (e.g. English, because of its higher intelligibility scores) but not for others, and could benefit from adjustments such as those made for non-native speakers in BS EN 60268-16 (2011).

6.4 Conclusions

This chapter overviewed 5 standards (BS EN 60268-16, 2011) (BS EN 9921, 2003) (ANSI S12.60, 2002) (BS 8233, 1999) (BB93-PS, 2015) and 2 design guidelines (BB93-DG, 2015) (Joseph and Ulrich, 2007) that can be consulted in the process of designing various multilingual spaces, from the perspective of the outcomes of the present study. The suggestions provided by the above mentioned documents were presented and discussed in terms of the importance given to speech intelligibility. STI suggestions were found only for open-plan learning environments, most documents providing reverberation time and background noise limits instead. In order to investigate the relationship between the provided upper limits for room acoustic parameters (BN and *T*) and the results of the first phase of the study (Chapter 4), the STI values were calculated for each condition by using the MTF method (Houtgast *et al.*, 1980). It is important to state that this method tends to underestimate the STI by

0.06 on average (in the range of 0.0-1.0) (Galbrun and Kitapci, 2014); therefore, the calculated STI values should be considered with caution.

The results of the first phase of the study suggested that there was a difference between word/sentence intelligibility scores of different languages under the same STI values, especially in low intelligibility conditions. Furthermore, the intelligibility of different languages varied depending on the specific room acoustic condition considered. It is important to note that most of the standards present physical parameter suggestions, except BS 8233 (1999), which is the only standard that states the importance of considerable difference of sensitivity to noise between people. This is in line with the findings of Chapter 5, which showed that justifying speech intelligibility by room acoustic parameters only might mislead the designers and planners, because of the importance of perceptual factors that might also play a role.

Analysis of speech intelligibility scores obtained in phase 1 of the research (Chapter 4) and guidance values based on the STI, or on the STI rating scale defined by BS EN ISO 9921 (2003), showed that such STI recommendations might be appropriate for some languages (e.g. English, because of its higher intelligibility scores) but not for others. Based on the results of the research, this appears to be particularly true for spaces that are expected to be more challenging in terms of intelligibility, e.g. open-plan spaces where excellent room acoustic conditions are difficult to achieve in practice. Furthermore, even under the ‘excellent’ STI = 0.8 condition, differences between word intelligibility scores can be still significant and non-negligible (~10%), suggesting that variability across languages should be considered anyway. This can be done by adjusting recommended STI values across languages, to achieve the appropriate speech intelligibility. Results showed that adjustments can vary between +0.16 to +0.30 when aiming to achieve a ‘good’ word intelligibility rating across all languages, and between +0.07 to +0.15 when aiming to achieve an ‘excellent’ word intelligibility rating. Adjustments were found to be comparatively smaller for ‘good’ sentence intelligibility (+0.10 to +0.22). These variations are not negligible and highlight the significance of differences observed between languages. Furthermore, it should be noted that the accuracy of these STI adjustments is limited, as additional tests will be needed in order to obtain more robust and reliable data for languages’ intelligibility scores across a wider range of room acoustic conditions.

CHAPTER 7

Conclusions

7.1 Introduction

This chapter illustrates the main findings obtained from the research. A summary of conclusions is given for each chapter, and this is followed by a description of the impact of the research. Suggestions for future work are described at the end of the chapter, together with limitations of the current work.

7.2 Findings

The main aim of the thesis was to find out possible relations between speech intelligibility and multi-lingual communication, in terms of acoustics, linguistics and perceptual factors. More specifically, the work focused on the impact of room acoustic conditions on the speech intelligibility of four languages representative of a wide range of linguistic properties (English, Polish, Arabic and Mandarin). Additionally, perceived speech intelligibility and soundscape perception associated to these languages were also analysed. Lastly, the study investigated several standards and design guidelines of spaces used for speech and their relation with multilingual intelligibility from the perspective of the outcomes of the study.

Main findings of the first experimental phase:

- The study found that there was a significant difference between the word intelligibility scores of the languages tested at most acoustic conditions.
- Distinctive features and acoustic properties of languages have an impact on the overall intelligibility.
- There is a significant correlation between consonant-to-vowel ratios and the word intelligibility scores at low STIs.
- There is a language specific threshold over which the context of speech becomes intelligible, therefore increasing the intelligibility of sentences.

Main findings of the second experimental phase:

- Perceived speech intelligibility of each language varies with the type of environment, type of background noise, reverberation time, and signal-to-noise ratio.
- Perceived and actual intelligibility can be different.
- Information content and distracting sounds affect perceived intelligibility differently for different languages.
- The soundscape of the three environments tested was mainly affected by relaxation, annoyance and environmental pleasantness.

Main findings of the overview of the standards and design guidelines:

- STI recommendations presented in the current standards and design guidelines might be appropriate for some languages but not for others.
- Variability across languages can be accounted for by adjusting recommended STI values across languages, to achieve the appropriate speech intelligibility.

7.3 Summary of Chapters

This section presents a summary of the chapters throughout the thesis.

Chapter 2 provided the background information required to carry out the thesis. Room acoustics, speech intelligibility, assessment of speech intelligibility, and the factors affecting speech intelligibility such as socio-lingual factors and soundscape theory were covered and previous literature was critically analyzed. It was observed that, predictors of speech intelligibility (AI, STI or RASTI) were tested for several languages (Houtgast and Steeneken, 1984; Kang, 1998). Additionally, Peng (2011) and Zhu *et al.* (2014) compared the results of Mandarin and English word tests and found that English tends to be slightly more intelligible than Mandarin under most room acoustic conditions, although some contradictions were observed between the findings of these studies, especially for either very poor or very good room acoustic conditions. However, a wider comparison between more languages was still needed to understand the effects of acoustical factors on the speech intelligibility of different languages, as most comparisons have been made between English and Mandarin. The literature showed that speech intelligibility test materials vary, such as rhyme tests (Fairbanks, 1958), PB word list and PB sentence lists (Beranek, 1949). The diagnostic rhyme test (DRT) was identified as the most appropriate listening test for the present study, because of its capability of differentiating several phonemic properties of various languages. The sensitivity to the phonemic properties of a language is an important aspect of a multilingual speech intelligibility study. Also, in order to observe the effects of social and psychological factors, the soundscape approach could be used to examine the perception of indoor spaces used for speech, by designing the surveys/questionnaires accordingly.

Chapter 3 presented the methodology of the two phases of the study (room acoustics and speech intelligibility, and soundscape perception). Initially, the selection process of the languages was described. For the first phase, a description was given on the word and sentence lists that were used, the recording and post processing of these word and sentence lists, together with details on the laboratory space used and the equipment used, as well as the listening test procedure. For the second phase, the followings were presented: selection process of the cases to be examined, preparation of the sentence lists, recording and post processing of the sentence lists and the background noise samples, preparation of the visual materials, details on the laboratory space and the equipment used, and information on semantic differential analysis. For both phases of the study, the statistical analysis methods used to analyse results were also described.

Chapter 4 discussed comparisons of the subjective listening test scores obtained for four languages (English, Polish, Mandarin, and Arabic), under different room acoustic conditions defined by their speech transmission index (STI=0.2, STI=0.4, STI=0.6, and STI=0.8). Overall intelligibility scores, language specific intelligibility scores of distinctive features, and sentence intelligibility scores were presented and analysed in order to understand relations between language specific effects and speech intelligibility, as well as relations between room acoustic properties and speech intelligibility of the different languages. The study found that there was a significant difference between the word intelligibility scores of these languages. Under the same acoustic conditions (reverberation time and S/N ratio), the word intelligibility scores of each language differed between each other, depending on the linguistic and distinctive features' properties of the languages. For word intelligibility, the differences were found to be statistically significant for all conditions but the excellent room acoustic condition (STI = 0.8), indicating that the word intelligibility of different languages was comparable under excellent room acoustic conditions, but was not comparable under any other condition. It was found that distinctive features of the selected languages have an impact on the overall intelligibility, nasal/oral consonants being particularly intelligible in English. Furthermore, a significant correlation was found between the consonant-to-vowel ratios and the word intelligibility scores of languages at poor room acoustic conditions (STI = 0.2 and STI = 0.4). In contrast to word scores, sentence scores showed statistically significant differences between languages only at the STI = 0.4 condition, but this was justified by the lower sensitivity of sentence tests to either very good or very challenging room acoustic conditions. Additionally, the comparison between the word and the sentence intelligibility scores revealed that there is a language specific STI threshold over which the context of speech becomes intelligible, therefore increasing the intelligibility of sentences. This threshold was lower for Polish and Mandarin compared to English and Arabic. Acoustical analysis of the languages also suggested that the better word intelligibility of English might be related to its greater high frequency content, as well as its larger temporal variability and dynamic range at high frequencies.

Chapter 5 presented and discussed how soundscape perception might affect the perceived speech intelligibility of English, Polish, Arabic, and Mandarin, by comparing the subjective assessment of three multi-lingual spaces (an airport, a hospital, and a café) tested under three room acoustic conditions (STI=0.4, STI=0.5, and STI=0.6). Results of the semantic differential analysis and principal component analysis were also given in this chapter. The semantic differential analysis of the intelligibility attribute revealed that perceived speech intelligibility of each language varies with the type of environment, as well as the type of background noise, reverberation time, and signal-to-noise ratio. Variations between the perceived speech intelligibility of the four languages were only marginally significant ($p = 0.051$), unlike word intelligibility tested during the first phase which showed significant variations between languages across all conditions but the excellent room acoustic condition. Perceived speech intelligibility of English appeared to be mostly affected negatively by the information content and distracting sounds present in the background noise. Attribute scores showed fairly variable trends across languages, with only few noticeable findings: 1) Variations between languages were significant only for a minority of cases; 2) Speech pleasantness and environmental pleasantness were highly correlated to perceived intelligibility for all the languages considered (i.e., pleasantness improves speech intelligibility); 3) Similarly, comfort and speech loudness were highly correlated to perceived intelligibility for all the languages considered; 4) English participants tended to rate noisiness higher, and a significant negative correlation was found between noisiness and perceived speech intelligibility of English; Arabic participants tended to rate noisiness lower, but this did not correlate significantly with their perceived intelligibility; 5) Arabic showed the lowest number of correlations between semantic attributes and perceived intelligibility, suggesting that its perceived speech intelligibility is less affected by the overall soundscape perception compared to other language groups. The contradictions between the two phases of the study also suggest that while English and Polish listeners were more sensitive to distractive noise sources, Arabic listeners were more resilient. A principal component analysis showed that soundscape perception of the three environments tested was mainly affected by relaxation, annoyance and environmental pleasantness, relaxation being the only semantic attribute included in a component for all the cases considered (9 cases = 3 STI conditions x 3 environments). Detailed semantic differential analysis and principal component analysis of the attributes revealed that the room acoustic conditions and the type of environments have varying effects on the perceived intelligibility of the four languages tested. It can therefore be stated that justifying speech intelligibility by room acoustic parameters only might mislead the designers and planners, especially while designing a multi-lingual environment. Design requirements of a multi-lingual environment are rather complex: the type of environment, type of background noise, reverberation time, and context can influence intelligibility of languages and more generally oral communication depending on the language an environment considered.

Chapter 6 presented an overview of 5 standards and 2 design guidelines that can be consulted in the process of designing various multilingual spaces, from the perspective of the outcomes of the present study, and more specifically, the results of Chapter 4. Each standard and design guideline were presented and discussed in terms of importance given to speech intelligibility, specifically to room acoustic parameters (i.e. reverberation time (T), signal-to-noise ratio (L_{SN}), and ultimately the speech transmission index (STI)), and multilingual communication. The signal-to-noise ratio and reverberation time information presented in such documents were converted to STI values by using the modulation transfer function (MTF), and a comparison of the STIs calculated were presented in Section 6.3 in relation to the results of Chapter 4. The chapter investigated the effectiveness of the current standards and design guidelines in terms of speech intelligibility in multi-lingual environments. Comparisons between speech intelligibility scores obtained in phase 1 of the research (Chapter 4) and guidance values based on the STI, or on the STI rating scale defined by BS EN ISO 9921 (2003), showed that such STI recommendations might be appropriate for some languages (e.g. English, because of its higher intelligibility scores) but not for others. Based on the results of the research, this appears to be particularly true for spaces that are expected to be more challenging in terms of intelligibility, e.g. open-plan spaces where excellent room acoustic conditions are difficult to achieve in practice. Variability across languages can be accounted for by adjusting recommended STI values across languages, to achieve the appropriate speech intelligibility. Results showed that adjustments can vary between +0.16 to +0.30 when aiming to achieve a ‘good’ word intelligibility rating across all languages, and between +0.07 to +0.15 when aiming to achieve an ‘excellent’ word intelligibility rating. Adjustments were found to be comparatively smaller for ‘good’ sentence intelligibility (+0.10 to +0.22). These variations were not negligible and highlighted the significance of differences observed between languages. Furthermore, it was noted that the accuracy of these STI adjustments is limited, as additional tests will be needed in order to obtain more robust and reliable data for languages’ intelligibility scores across a wider range of room acoustic conditions.

7.4 Impact of the research

The results obtained from both phases developed the knowledge and understanding of multilingual communication, also in relation to existing standards and design guidelines. Such information could be used by architects, service and product providers, and acoustic engineers in order to minimise communication problems between end users in multilingual environments.

More specifically, current guidelines might mislead architects, service and product providers, as well as acoustic engineers involved in the design of multilingual spaces. The findings obtained from the research could be used to encourage the development of guidelines that take into account the variability of speech intelligibility across languages. In that sense, the approach used in Chapter 6 (STI corrections applied to languages) could be considered for the development of such new guidelines. It is however clear that additional academic research will be needed to consolidate the speech intelligibility data of a variety of languages before robust guidelines can be implemented, especially taking into account the complexity of speech intelligibility tests and their limitations, some of which are discussed in the following section. The current work also provides a strong basis for research planning to further look at the relationship between the soundscape and speech intelligibility.

7.5 Suggestions for future research and limitations

In this section, suggestions for future research are illustrated, based on the findings highlighted in this work and previous research.

Word intelligibility of Polish was assessed using PB words, unlike other languages for which DRT lists were available. This represents an important limitation of the current study. English data suggested that variations between DRT and PB results are small and therefore acceptable, but this alone cannot guarantee the same conclusion for Polish. Further research could be conducted when a Polish DRT is published, in order to achieve more comparable results.

The multiple factors affecting sentence intelligibility varied across the languages used (e.g. context, familiarity, predictability, prosody and number of words), making sentence tests less comparable than word tests. To obtain a further insight into sentence intelligibility, future work could compare sentences translated across different languages. It might be difficult to obtain phonemically balanced material across all the languages tested, but this approach could at least maintain context and provide useful comparisons of real life scenarios.

Although word tests were more sensitive to room acoustic conditions than sentence tests, it is important to remember that representing a language through words only is a limitation, as the PB words used might only represent a fraction of the type of words available in a language (as this is for example the case for English, as opposed to Mandarin).

The work might have included monolingual vs. multilingual speakers' effects, which might be partly responsible for some of the variations observed, although these effects alone cannot justify the large differences observed between languages. The fixed order of STI conditions tested might also have been responsible for order effects that could have been excluded through randomisation.

In order to accurately quantify STI adjustments between languages (see Chapter 6), additional tests will be needed in order to obtain more robust and reliable data for languages' intelligibility scores across a wider range of room acoustic conditions.

Lastly, it should be noted that in the first phase of the study, white noise was used in all the tests for the STI = 0.2 and STI = 0.4 conditions, but research has shown that the type of background noise used can affect DRT scores. Kondo (2011) found that, for identical signal-to-noise ratios, white noise produced lower DRT scores (for Japanese) than pseudo-speech noise and babble noise, and Astolfi et al. (2012) also found variations in DRT scores of Italian for a variety of noise sources in primary school classrooms (traffic vs. babble vs. fan-coil vs. impact). Further research will therefore need to examine whether different types of background noise can also affect languages' intelligibility differently.

References

Abushariah, M., Ainon, R., Zainuddin, R., Elshafei, M., and Khalifa, O. (2011). *Phonetically rich and balanced text and speech corpora for Arabic language*. Language Resources and Evaluation, 46(4), 601-634, 2011.

AC 150/5360-13 (1988). *Planning and design guidelines for airport terminal facilities*. Federal Aviation Administrarion, Washington, USA.

Alias, F., and Trivino, M. P. (2007). *A phonetically balanced modified rhyme test for evaluating Catalan intelligibility*. International Congress of Phonetic Sciences, 2007, Germany.

Anderson, B. W., and Kalb J. T. (1987) *English verification of the STI method for estimating speech intelligibility of a communication channel*. J. Acoust. Soc. Am., **81(6)**, 1982-1985, 1987.

ANSI/ASA S3.2. (2009). *Method for measuring the intelligibility of speech over communication systems*. American National Standards Institute, 2009.

ANSI S12.60 (2002). *Acoustical performance criteria, design requirements, and guidelines for schools.* American National Standards Institute, 2009.

AudioCheck. (2015). Online Audiogram Hearing Test. (Available online at http://www.audiocheck.net/testtones_hearingtestaudiogram.php, Accessed on 16/06/2015).

Ball, M. J. (2010). *The routledge handbook of sociolinguistics around the world.* Oxon, Routledge.

Barnett, P. W., Knight, R. D. (1995). *The common intelligibility scale.* Proceedings of the Institute of Acoustics, **17(7)**, 199-204, 1995.

Beranek, L. L. (1949). Test Lists. *Acoustics Measurements.* New York: Published for the Acoustical Society of America by the American Institute of Physics.

Beranek, L. L. (2006). *Analysis of Sabine and Eyring equations and their application to concert hall audience and chair absorption*. J. Acoust. Soc. Am., **120(3)**, 1399-1410, 2006.

Berglund B., Lindvall T., Schwela D.H. (1999). *Guidelines for community noise*. London.

Berglund, B., and Nilsson, M. E. (2006). *On a tool for measuring soundscape quality in urban residential areas*. Acta Acustica United with Acustica, **92,** 938-944, 2006.

Blommaert, J., Collins, J., and Slembrouck, S. (2005). *Spaces of multilingualism*. Language & Communication, **25,** 197-216, 2005.

Boudraa, B., Boudraa, M., and Guerin, B. (2008). *Arabic diagnostic rhyme test using minimal pairs*. Acoustics, Paris.

Boudraa, M., and Boudraa B. (2000). Twenty lists of ten Arabic sentences for assassment. Acta Acustica United with Acustica, **86**, 870-882, 2000.

Bradley, J. S. (1986). *Predictors of speech intelligibility in rooms*. J. Acoust. Soc. Am., **80,** 837-845, 1986.

Brown, A. L., Kang, J., and Gjestland, T. (2011). *Towards standardization in soundscape preference assessment*. Applied Acoustics, **72,** 387-392, 2011.

BS 8233 (1999). *Sound insulation and noise reduction for buildings – Code of Practice.* British Standards Institution, London, UK.

BS EN 60268-16 (2011). *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index.* British Standards Institution, London, UK.

BS EN ISO 9921 (2003). *Ergonomics – Assessment of speech communication.* British Standards Institution, London, UK.

Building Bulletin 93 (2015). *Acoustic design of schools: performance standards.* Department of Education, London, UK.

Building Bulletin 93 (2015). *Acoustics of schools: a design guide.* Department of Education, London, UK.

Cain, R., Jennings, P., and Poxon, J. (2011). *The development and application of the emotional dimensions of a soundscape.* Applied Acoustics, **74(2)**, 232-239, 2011.

Carles, J., Bernaldez, F. & Lucio, J. D. (1992). *Audio-visual interactions and soundscape preferences*. Landscape research, **17,** 52-56, 1992.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge University Press, Cambridge.

Christiansen, C., Pedersen, M. S., and Dau, T. (2010). *Prediction of speech intelligibility based on an auditory preprocessing model*. Speech Communication, **52,** 678-692, 2010.

Cohen, M. F. (1965). *Effects of stimulus presentation rate upon intelligibility-test scores*. J. Acoust. Soc. Am., **37**, 1206, 1965.

Cutler, A., Butterfield, S. (1991). *Word boundary cues in clear speech: A supplementary report*. Speech Communication, **10(4)**, 335-353, 1991.

Davies, W. J., Cain, R., Carlyle, A., Hall, D. A., Hume, K. I., and Plack, C. J. (2009a). *The positive soundscape project: A sysnthesis of result from many disciplines.* Internoise, 2009, Ottawa.

Davies, W. J., Mahnken, P. Z. & Gamble, P. (2009b). *Measuring and Mapping Soundscape Speech Intelligibility*. Euronoise, 2009, Edinburgh, Scotland.

Diaz, C., and Velazquez, C. (1995). *A live evaluation of the RASTI method.* Applied Acoustics, **46(4)**, 363-372, 1995.

Dokmeci, P. N., and Kang, J. (2010). *Objective Parameters for Acoustic Comfort in Enclosed Spaces.* 20th International Congress on Acoustics, 2007, Sydney, Australia.

Egan, D. (1988). *Architectural acoustics*, McGraw-Hill.

Extra, G., and Yagmur, K. (2011). *Urban multilingualism in Europe: Mapping linguistic diversity in multicultural cities*. Journal of Pragmatics, **43,** 1173-1184, 2011.

Fairbanks, G. (1958). *Test of Phonemic Differentiation: The Rhyme Test.* J. Acoust. Soc. Am., **30,** 596-600, 1958.

Field, A. (2009). *Discovering statistics using SPSS.* Sage, London.

French, N. R., and Steinberg, J. C. (1947). *Factors Governing the Intelligibility of Speech Sounds*. J. Acoust. Soc. Am., **19,** 90-119, 1947.

Fu, Q., Zhu, M., Wang X. (2011) *Development and validation of the Mandarin speech perception test.* J. Acoust. Soc. Am., **129(6)**, EL267-EL273, 2011.

Galbrun, L. (2011). Architectural acoustics lecture notes, Heriot-Watt University.

Galbrun, L., and Kitapci, K. (2014). *Accuracy of speech transmission index predictions based on the reverberation time and signal-to-noise ratio*. Applied Acoustics, **81**, 1-14, 2014.

Garcia Lecumberri, M. L., and Cooke, M. (2006). *Effect of masker type on native and non-native consonant perception in noise.* J. Acoust. Soc. Am., **119(4)**, 2445-2454, 2006.

Garcia Lecumberri, M. L., Cooke, M., and Cutler A. (2010) *Non-native speech perception in adverse conditions: A review*. Speech Communication, **52**, 864-886, (2010).

Gardner-Chloros, P. (2009). *Code-switching,* Cambridge, Cambridge University Press.

Goodenough, W. H. (1957). *Cultural anthropology and linguistics,* Washington, DC, Georgetown University Press.

Heller, M. (2007). Bilingualism as ideolojy and practice. *Bilingualism: A social approach.* Houndmills: Palgrave.

Holmes, J. (1992). *An introduction to sociolinguistics,* London, Longman.

Houtgast, T., and Steeneken, H. J. M. (1973). *The modulation transfer function in room acoustics as a predictor of speech intelligibility*. Acustica, **28,** 66-73., 1973.

Houtgast, T., and Steeneken, H. J. M. (1980). *Predicting speech intelligibility in rooms from the modulation transfer function. 1. General room acoustics. Acustica,* **46,** 60-72, 1980.

Houtgast, T., and Steeneken, H. J. M. (1984). *A multi-language evaluation of the rasti-method for estimating speech intelligibility in auditoria.* Acustica, **54,** 185-199, (1984).

Houtgast, T., and Steeneken, H. J. M. (1985). *A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria.* J. Acoust. Soc. Am., **77,** 1069-1077, 1985.

Hudson, R. A. (1996). *Sociolinguistics,* New York, Cambridge University Press.

Hutcheson, G. D., and Sofroniou, N. (1999). *The Multivariate Social Scientist: an introduction to generalized linear models.* Sage Publications.

ISO/TR 4870 (1991). *The construction and calibration of speech intelligibility tests.* International Organization for Standardization, Genève, Switzerland.

ISO 12913-1 (2014). *Soundscape - Part 1: Definition and conceptual framework.* International Organization for Standardization, Genève, Switzerland.

Jakobson, R, Fant, G., and Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates.* Technical Report 13. Massachusetts: Acoustics Laboratory, MIT.

Jones, D. (1917). *An English Pronouncing Dictionary*. Dent, London.

Joseph, A., Ulrich, R. (2007). *Sound control for improved outcomes in healthcare settings*. The Center for Health Design, **4**, 2007.

Kaiser, H. F. (1974). *An index of factorial simplicity.* Psychometrika, **39(1)**, 31-36, 1974.

Kalikow, D. N., and Stevens, K. N. (1977). *Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability*. J. Acoust. Soc. Am., **61,** 1337, 1977.

Kang, J. (1998). *Comparison of Speech Intelligibility between English and Chinese.* J. Acoust. Soc. Am., **103,** 1213-1216, 1998.

Kang, J. (2010). *From understanding to designing soundscapes*. Front. Archit. Civ. Eng. China, **4,** 403-417, 2010.

Kang, J., and Zhang, M. (2010). *Semantic differential analysis of the soundscape in urban open public spaces.* Building and Environment, **45,** 150-157, 2010.

Kinsler, L. E., and Frey, A. R. (1962). *Fundamentals of acoustics.* New York: Wiley.

Kondo, K. (2012). The Japanese diagnostic rhyme test. *In:* KONDO, K. (ed.) *Subjective quality measurement of speech.* Berlin: Springer.

Kowal S., Weise, R., and O'Connell, D., (1983). *The use of time in storytelling*. Language and Speech, 26(4), 377-392.

Kryter, K. D. (1970). *The effects of noise on man*. New York: Academic Press.

L,i Z., Tan, E. C., Mcloughlin, I. and Teo, T. T. (2000). *Proposal of standards for* *intelligibility tests of Chinese speech*. Iee Proceedings-Communications, 147.

Long, M. (2006). *Architectural Acoustics*, Elsevier.

Maddieson, I. (2008). Consonant-Vowel Ratio. *In: M. S. Dryer and M. Haspelmath, Martin (Eds.) The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology, 2008. (Available online at http://wals.info/chapter/3, Accessed on 01/12/2012).

Maddieson, I. (2013). Tone. *In: M. S. Dryer and M. Haspelmath, Martin (Eds.). The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. (Available online at http://wals.info/chapter/13, Accessed on 16/06/2015).

Martin-Jones, M., Blackledge, A., and Creese, A. (2012). *The routledge handbook of multilingualism,* Oxon, Rourledge.

McLoughlin, I. (2008). *Subjective intelligibility testing of Chinese speech.* IEEE Transactions on Audio, Speech, and Language Processing, **16(1)**, 23-33, 2008.

Mesthrie, R., Swann, J., Deumert, A., and Leap, W. L. (2009). *Introducing sociolinguistics,* Edinburgh, Edinburgh university press.

Miller, G. (1951). Language and communication, McGraw-Hill, New York.

Osgood, C.E., Suci, G.J., Tannenbaum, P.H. (1957). *The measurement of meaning*. University of Illinois Press, Champaign (IL).

Ozimek, E., Dariusz, K., Sek, A., and Wicher, A. (2009). *Development and evaluation of Polish digit triplet test for auditory screening.* Speech Communication, **51**, 307-316, 2009.

Peng, J. (2005). *Feasibility of subjective speech intelligibility assessment based on auralization*. Applied Acoustics, **66,** 591-601, 2005.

Peng, J. (2007). *Relationship between Chinese speech intelligibility and speech* *transmission index using diotic listening.* Speech Communication*,* **49,** 933-936, 2007.

Peng, J. (2010). *Chinese speech intelligibility at different speech sound pressure levels and signal-to-noise ratios in simulated classrooms.* Applied Acoustics, **71,** 386-390, 2010.

Peng, J., Bei, C., and Sun, H. (2011). *Relationship between Chinese speech intelligibility and speech transmission index in rooms based on auralization.* Speech Communication*,* **53,** 986-990, 2011.

Peutz, V. M. A. (1971). *Articulation loss of consonants as a criterion for speech transmission in a room.* Journal of the Audio Engineering Society, **19(11)**, 915-919, 1971.

Pheasant, R. J., Fisher, M. N., Watts, G. R., Whitaker, D. J.,and Horoshenkov, K. V. (2010). *The importance of auditory-visual interaction in the construction of 'tranquil space'*. Journal of Environmental Psychology*,* **30,** 501-509, 2010.

Phonics. (2012). *Phonics Definition & Glossary* http://www.atozphonics.com/phonicsdefinition.html [Online]. A to Z Phonics. [Accessed 6th of December 2012].

Pimsleur, P., Hnacock, C., and Furey, P. (1977). *Speech rate and listening comprehension.*Viewpoints on English as a second language, 27-34, New York, Regents Publishing.

Raimbault, M. (2006). *Qualitative judgements of urban soundscapes: Questionning questionnaires and semantic scales*. Acta Acustica United with Acustica*,* **92,** 929-937, 2006.

Raimbault, M., and Dubois, D. (2005). *Urban soundscapes: Experiences and knowledge*. Cities*,* **22,** 339-350, 2005.

Sapie, E. (1921). *Language,* New York, Harcourt, Brace, World.

Schafer, M. (1977). *The Soundscape: Our Sonic Environment and the Tuning of the World,* New York.

Shrout, P., Fleiss, J. (1979). *Intraclass correlations: Uses in assessing rater reliability.* Psychological Bulletin, **86(2)**, 420-428, 1979.

Steeneken, H. J. M., and Houtgast, T. (1980). *A physical method for measuring speech-transmission quality*. J. Acoust. Soc. Am., **67,** 318-326, 1980.

Steeneken, H. J. M., and Houtgast, T. (2002). *Validation of the revised STIr method.* Speech Communication, **38,** 413-425, 2002.

Studebaker, G. A. (1985). *A "rationalized" arcsine transform*. Journal of Speech, Language, and Hearing Research, **28(3),** 455-462, 1985.

Van Engen, K. J., and Bradlow, A. R. (2007). *Sentence recognition in native- and foreign-language multi-talker background noise*. J. Acoust. Soc. Am., **121(1)**, 519-526, 2007.

Van Engen, K. J. (2010). *Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble.* Speech Communication, **52**, 943-953, 2010.

Van Wijngaarden, S. J., Bronkhorst, A. W., Houtgast, T. & Steeneken, H. J. M. (2004). *Using the Speech Transmission Index for predicting non-native speech intelligibility*. J. Acoust. Soc. Am., **115,** 1281-1291, 2004.

Voiers, W. D. (1965). *Performance evaluation of speech process devices, II. the role of individual difference*s. Bedford: United Stated Air Force, 1965.

Voiers, W. D. (1977). Diagnostic evaluation of speech intelligibility. *In:* HAWLEY, M. E. (ed.) *Speech intelligibility and speaker recognition.* Stroudsburgh: Dowden, Hutchinson and Ross, Inc.

Voiers, W. D. (1999). *Uses of the diagnostic rhyme test (English version) for predicting the effects of communicators' linguistic backgrounds on voice communications in English: An exploratory study*. Austin: Dynastat, Inc.

Voiers, W. D. (1983). *Evaluating Processed Speech Using the Diagnostic Rhyme Test.* Speech Technology, 0-9, 1983.

Wardhaugh, R. (2006). *An introduction to sociolinguistics,* Victoria, Blackwell.

Wodak, R., Kryzyzanowski, M., and Forchtner, B. (2012). *The interplay of language ideologies and contextual cues in multilingual interactions: Language choice and code-switching in European Union institutions*. Language in Society*,* **41,** 157-186, 2012.

Yu, L., and Kang, J. (2010). *Factors influencing the sound preference in urban open spaces*. Applied Acoustics*,* **71,** 622-633, 2010.

Zhu, P., Mo, F., Kang, J. (2014). *Relationship between Chinese speech intelligibility and speech transmission index under reproduced general room conditions.* Acta Acustica United with Acustica, **100**, 880-887, 2014.

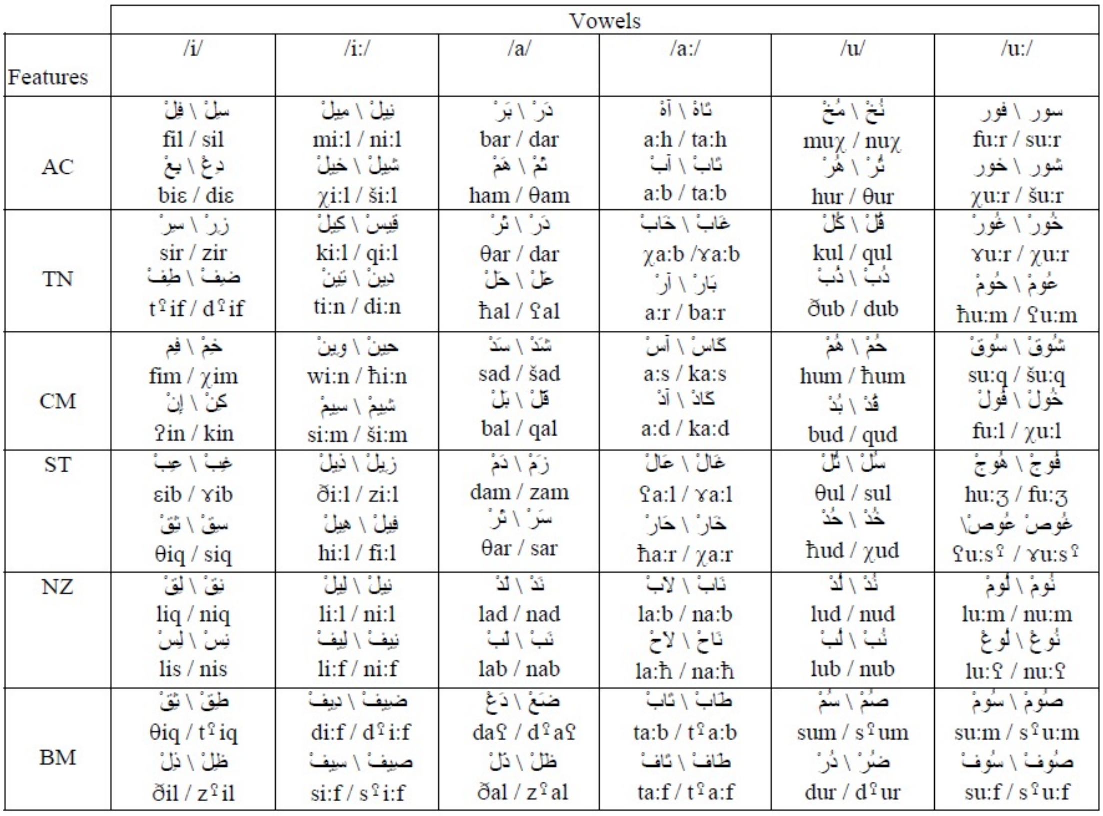
Appendix A: Diagnostic Rhyme Test – Word Lists

This Appendix illustrates the English (Voiers, 1977), Arabic (Boudraa *et al*, 2008), and Mandarin (Fu *et al*., 2011) diagnostic rhyme test (DRT) word lists used for the first phase listening tests.

English DRT

Voicing		**Nasality**		**Sustention**	
Veal	Feel	Meat	Beat	Vee	Bee
Bean	Peen	Need	Deed	Sheet	Cheat
Gin	Chin	Mitt	Bit	Vill	Bill
Dint	Tint	Nip	Dip	Thick	Tick
Zoo	Sue	Moot	Boot	Foo	Pooh
Dune	Tune	News	Dues	Shoes	Choose
Vole	Foal	Moan	Bone	Those	Doze
Goat	Coat	Note	Dote	Though	Dough
Zed	Said	Mend	Bend	Then	Den
Dense	Tense	Neck	Deck	Fence	Pence
Vast	Fast	Mad	Bad	Than	Dan
Gaff	Calf	Nab	Dab	Shad	Chad
Vault	Fault	Moss	Boss	Thong	Tong
Daunt	Taunt	Gnaw	Daw	Shaw	Chaw
Jock	Chock	Mom	Bomb	Von	Bon
Bond	Pond	Knock	Dock	Vox	Box
Sibilation		**Graveness**		**Compactness**	
Zee	Thee	Weed	Reed	Yield	Wield
Cheep	Keep	Peak	Teak	Key	Tea
Jilt	Gilt	Bid	Did	Hit	Fit
Sing	Thing	Fin	Thin	Gill	Dill
Juice	Goose	Moon	Noon	Coop	Poop
Chew	Coo	Pool	Tool	You	Rue
Joe	Go	Bowl	Dole	Ghost	Boast
Sole	Thole	Fore	Thor	Show	So
Jest	Guest	Met	Net	Keg	Peg
Chair	Care	Pent	Tent	Yen	Wren
Jab	Gab	Bank	Dank	Gat	Bat
Sank	Thank	Fad	Thad	Shag	Sag
Jwas	Gauze	Fought	Thought	Yawl	Wall
Saw	Thaw	Bong	Dong	Caught	Thought
Jot	Got	Wad	Rod	Hop	Fop
Chop	Cop	Pot	Tot	Got	Dot

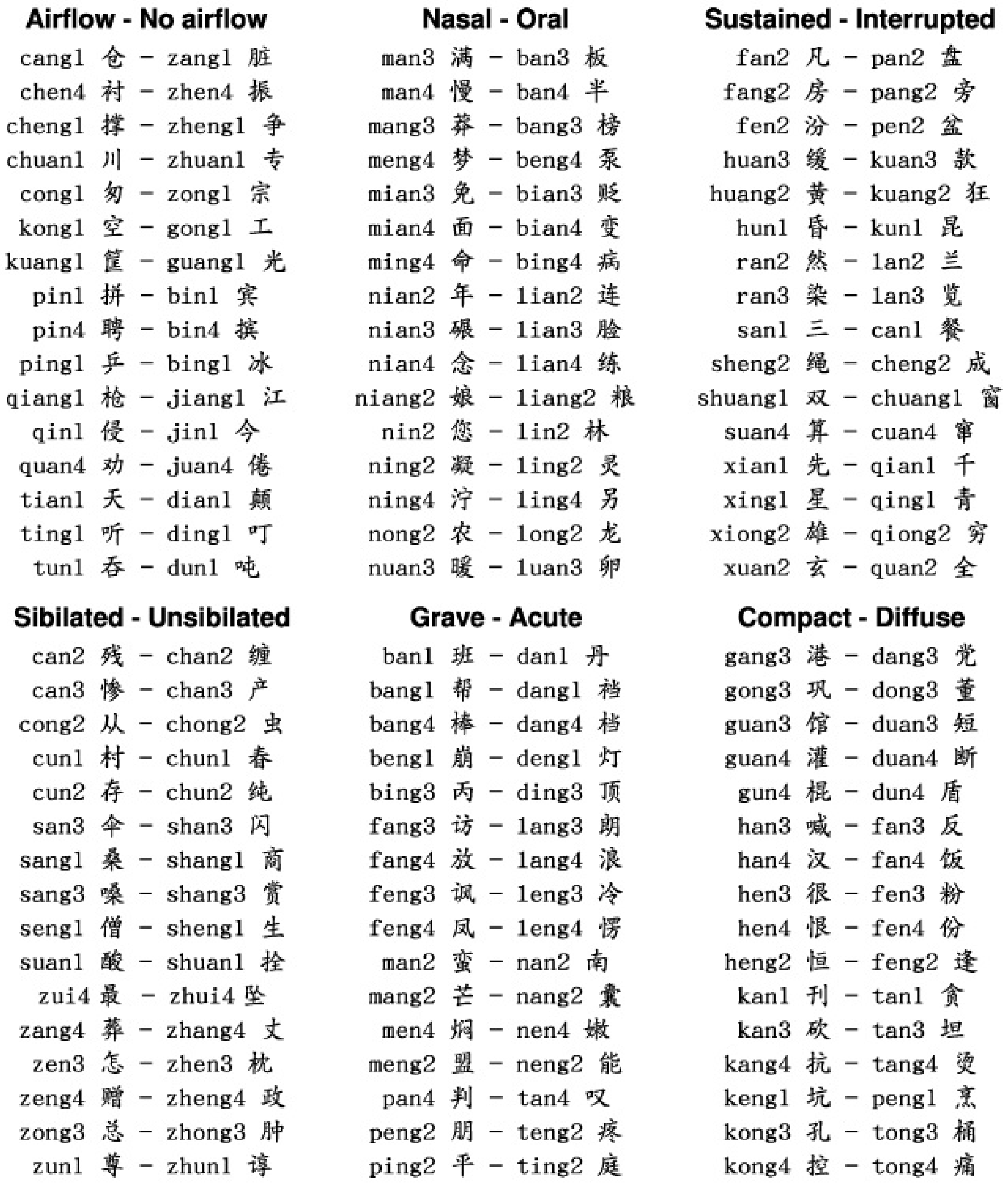
Arabic DRT



	Vowels					
Features	/i/	/i:/	/a/	/a:/	/u/	/u:/
AC	سِلْ \ فِلْ fil / sil دِعْ \ بِعْ biɛ / diɛ	نيلْ \ ميلْ mi:l / ni:l شيلْ \ خيلْ χi:l / ši:l	دَرْ \ بَرْ bar / dar ثَمْ \ هَمْ ham / θam	ثاهْ \ آهْ a:h / ta:h ثابْ \ آبْ a:b / ta:b	نُخْ \ مُخْ muχ / nuχ ثُرْ \ هُرْ hur / θur	سور \ فور fu:r / su:r شور \ خور χu:r / šu:r
TN	زِرْ \ سِرْ sir / zir ضِفْ \ طِفْ tˤif / dˤif	قيسْ \ كيلْ ki:l / qi:l دينْ \ تينْ ti:n / di:n	دَرْ \ ثَرْ θar / dar عَلْ \ حَلْ ħal / ʕal	غَابْ \ خَابْ χa:b /ɣa:b بَارْ \ آرْ a:r / ba:r	قُلْ \ كُلْ kul / qul دُبْ \ ذُبْ ðub / dub	خُورْ \ غُورْ ɣu:r / χu:r عُومْ \ حُومْ ħu:m / ʕu:m
CM	خِمْ \ فِم fim / χim كِنْ \ إِنْ ʔin / kin	حينْ \ وينْ wi:n / ħi:n شيمْ \ سيمْ si:m / ši:m	شَدْ \ سَدْ sad / šad قَلْ \ بَلْ bal / qal	كَاسْ \ آسْ a:s / ka:s كَادْ \ آدْ a:d / ka:d	حُمْ \ هُمْ hum / ħum قُدْ \ بُدْ bud / qud	شُوقْ \ سُوقْ su:q / šu:q خُولْ \ فُولْ fu:l / χu:l
ST	غِبْ \ عِبْ ɛib / ɣib سِقْ \ ثِقْ θiq / siq	زيلْ \ ذيلْ ði:l / zi:l فيلْ \ هيلْ hi:l / fi:l	زَمْ \ دَمْ dam / zam سَرْ \ ثَرْ θar / sar	غَالْ \ عَالْ ʕa:l / ɣa:l خَارْ \ حَارْ ħa:r / χa:r	سُلْ \ ثُلْ θul / sul خُدْ \ حُدْ ħud / χud	فُوجْ \ هُوجْ hu:ʒ / fu:ʒ غُوصْ عُوصْ\ ʕu:sˤ / ɣu:sˤ
NZ	نِقْ \ لِقْ liq / niq نِسْ \ لِسْ lis / nis	نيلْ \ ليلْ li:l / ni:l نيفْ \ ليفْ li:f / ni:f	نَدْ \ لَدْ lad / nad نَبْ \ لَبْ lab / nab	نَابْ \ لَابْ la:b / na:b نَاحْ \ لَاحْ la:ħ / na:ħ	نُدْ \ لُدْ lud / nud نُبْ \ لُبْ lub / nub	نُومْ \ لُومْ lu:m / nu:m نُوعْ \ لُوعْ lu:ʕ / nu:ʕ
BM	طِقْ \ ثِقْ θiq / tˤiq ظِلْ \ ذِلْ ðil / zˤil	ضيفْ \ ديفْ di:f / dˤi:f صيفْ \ سيفْ si:f / sˤi:f	ضَعْ \ دَعْ daʕ / dˤaʕ ظَلْ \ ذَلْ ðal / zˤal	طَابْ \ ثَابْ ta:b / tˤa:b طَافْ \ ثَافْ ta:f / tˤa:f	صُمْ \ سُمْ sum / sˤum ضُرْ \ دُرْ dur / dˤur	صُومْ \ سُومْ su:m / sˤu:m صُوفْ \ سُوفْ su:f / sˤu:f

AC: Graveness, TN: Tenseness, CM: Compactness, ST: Mellowness, NZ: Nasality, BM: Flatness

Mandarin DRT



Airflow - No airflow

cang1 仓 - zang1 脏
chen4 衬 - zhen4 振
cheng1 撑 - zheng1 争
chuan1 川 - zhuan1 专
cong1 匆 - zong1 宗
kong1 空 - gong1 工
kuang1 筐 - guang1 光
pin1 拼 - bin1 宾
pin4 聘 - bin4 摈
ping1 乒 - bing1 冰
qiang1 枪 - jiang1 江
qin1 侵 - jin1 今
quan4 劝 - juan4 倦
tian1 天 - dian1 颠
ting1 听 - ding1 叮
tun1 吞 - dun1 吨

Nasal - Oral

man3 满 - ban3 板
man4 慢 - ban4 半
mang3 莽 - bang3 榜
meng4 梦 - beng4 泵
mian3 免 - bian3 贬
mian4 面 - bian4 变
ming4 命 - bing4 病
nian2 年 - lian2 连
nian3 碾 - lian3 脸
nian4 念 - lian4 练
niang2 娘 - liang2 粮
nin2 您 - lin2 林
ning2 凝 - ling2 灵
ning4 泞 - ling4 另
nong2 农 - long2 龙
nuan3 暖 - luan3 卵

Sustained - Interrupted

fan2 凡 - pan2 盘
fang2 房 - pang2 旁
fen2 汾 - pen2 盆
huan3 缓 - kuan3 款
huang2 黄 - kuang2 狂
hun1 昏 - kun1 昆
ran2 然 - lan2 兰
ran3 染 - lan3 览
san1 三 - can1 餐
sheng2 绳 - cheng2 成
shuang1 双 - chuang1 窗
suan4 算 - cuan4 窜
xian1 先 - qian1 千
xing1 星 - qing1 青
xiong2 雄 - qiong2 穷
xuan2 玄 - quan2 全

Sibilated - Unsibilated

can2 残 - chan2 缠
can3 惨 - chan3 产
cong2 从 - chong2 虫
cun1 村 - chun1 春
cun2 存 - chun2 纯
san3 伞 - shan3 闪
sang1 桑 - shang1 商
sang3 嗓 - shang3 赏
seng1 僧 - sheng1 生
suan1 酸 - shuan1 拴
zui4 最 - zhui4 坠
zang4 葬 - zhang4 丈
zen3 怎 - zhen3 枕
zeng4 赠 - zheng4 政
zong3 总 - zhong3 肿
zun1 尊 - zhun1 谆

Grave - Acute

ban1 班 - dan1 丹
bang1 帮 - dang1 裆
bang4 棒 - dang4 档
beng1 崩 - deng1 灯
bing3 丙 - ding3 顶
fang3 访 - lang3 朗
fang4 放 - lang4 浪
feng3 讽 - leng3 冷
feng4 凤 - leng4 愣
man2 蛮 - nan2 南
mang2 芒 - nang2 囊
men4 焖 - nen4 嫩
meng2 盟 - neng2 能
pan4 判 - tan4 叹
peng2 朋 - teng2 疼
ping2 平 - ting2 庭

Compact - Diffuse

gang3 港 - dang3 党
gong3 巩 - dong3 董
guan3 馆 - duan3 短
guan4 灌 - duan4 断
gun4 棍 - dun4 盾
han3 喊 - fan3 反
han4 汉 - fan4 饭
hen3 很 - fen3 粉
hen4 恨 - fen4 份
heng2 恒 - feng2 逢
kan1 刊 - tan1 贪
kan3 砍 - tan3 坦
kang4 抗 - tang4 烫
keng1 坑 - peng1 烹
kong3 孔 - tong3 桶
kong4 控 - tong4 痛

Appendix B: Phonemically Balanced Word Lists

This Appendix illustrates the Polish phonemically balanced word lists (Ozimek *et al.*, 2007) used for the first phase listening tests.

List 1	**List 2**	**List 3**	**List 4**	**List 5**
plac	dres	kwas	płaz	twarz
zez	jaś	czas	sieć	rzecz
szyld	jacht	cyrk	czart	sierp
wał	mech	fach	leń	cham
skurcz	biust	ksiądz	zrost	złość
kat	tak	typ	kit	byt
grosz	głos	gnój	tchórz	plus
widz	wyż	wesz	maj	nić
pech	pył	dar	dym	bal
pierś	płaszcz	pieśń	wieprz	dreszcz
muł	woń	wór	wół	mur
grzyb	krzak	żbik	krzyk	bieg
cień	żal	dżem	dzień	syn
bank	kant	karp	pęk	kark
drań	krem	plan	kran	gmach
lódź	wódz	los	moc	nos
błąd	klops	sztorm	grunt	głąb
groch	broń	tłum	dłoń	król
rok	lud	lot	lok	łuk
tom	dom	gol	ból	ton
kwiat	zgryz	zwiad	stryj	wstyd
film	nerw	hełm	filc	walc
bez	bicz	bis	bas	paź
sens	liść	zięć	żerdż	rejs

List 6	**List 7**	**List 8**	**List 9**	**List 10**
klej	kraj	płeć	płacz	głaz
rzeź	dziś	dzicz	ciecz	jeż
cynk	sęp	sęk	żart	jęk
łan	rym	lew	leń	wir
spust	tłuszcz	złość	bluszcz	gwóźdź
bat	bak	gad	tik	byk
gruz	plusz	klosz	kłos	klucz
mecz	las	mysz	raj	nic
bar	tył	pan	dal	dach
wieść	wrzask	zjazd	wjazd	chrzest
łom	loch	muł	chór	rów
szpik	step	brzeg	styk	zbyt
zew	czyn	sen	cel	żar
garb	targ	park	pęd	tynk
płyn	gniew	gwar	krew	gram
mocz	noc	wóz	nóż	wuj
drąg	front	prąd	brąz	sport
dwór	plon	tron	grom	proch
lód	róg	ród	huk	łup
koń	dół	duch	puch	bon
strach	zb,eg	spryt	spływ	sklep
hymn	węch	chęć	myśl	marsz
kij	gaz	kac	gaj	pas
szewc	maść	sejm	zamsz	zysk

Appendix C: Phonemically Balanced Sentence Lists

This Appendix illustrates the English (Kalikow *et al.*, 1977), Polish (Ozimek, 2009), Arabic (Boudraa *et al.*, 2000), and Mandarin (Fu *et al.*, 2011) phonemically balanced sentence lists used for the first phase listening tests.

ENGLISH SENTENCE LIST

1. The watchdog gave a warning growl.
2. She made the bed with clean sheets.
3. The old man discussed the dive.
4. Bob heard Paul called about the strips.
5. I should have considered the map.
6. The old train was powered by steam.
7. He caught the fish in his net.
8. Miss Brown shouldn’t discuss the sand.
9. Close the window to stop the draft.
10. My T.V. has a twelve-inch screen.

POLISH SENTENCE LIST

1. Znowu ta winda nie działa
2. Najpierw zwabiło go światło
3. Wracam późno do hotelu
4. Taśma przesuwa się ciszej
5. Nie znamy wielu powodów
6. To były nasze pomysły
7. Dla ciebie była zawsze dobra
8. Na czele grupy stoi wódz
9. Ale potrzebny był następny
10. Były pewnie zbyt głęboko

ARABIC SENTENCE LIST

1 — ʒaḥfaẓu mina lʒarḍi More Conservative than the earth	أَحفَظُ مِنَ الأَرضِ
2 — ʒayna lmusa:firu:na ? Where are the passengers?	أَينَ المُسَافِرُونَ ?
3 — la: lam yastamtiɛ biṯamariha: No! he has not enjoyed its fruits	لَا لَم يَستَمتِع بِثَمَرِهَا
4 — sayuʒḏi:him zama:nuna: Our time will do harm to them	سَيُؤذِيهِم زَمَانُنَا
5 — kuntu qudwatan lahum I have been a model for them	كُنتُ قُدوَةً لَهُم
6 — ʒa:zara ṣa:ʒiman He rewards a fasting man	آزَرَ صَائِمًا
7 — ka:la wa ḡabaṭa lkabša He measured and tightened the sheep	كَالَ وَ غَبَطَ الكَبشَ
8 — hal laḏaɛathu biqawlin ? Has she offended him with words?	هَل لَذَعَتهُ بِقَولٍ ?
9 — ɛarafa wa:liyan wa qa:ʒidan He knew a governor and a commander	عَرَفَ وَالِيًا وَ قَائِدًا
10 — χala: ba:luna: minkuma: We did not pay attention to you	خَلَا بَالُنَا مِنكُمَا

MANDARIN SENTENCE LIST

1	今天的阳光真好	jīn tiān de yáng guāng zhēn hǎo
2	节假日不用门票	jié jiǎ rì bù yòng mén piào
3	晚上一块去跳舞	wǎn shàng yī kuài qù tiào wǔ
4	对面有两所高中	duì miàn yǒu liǎng suǒ gāo zhōng
5	这些衣服洗过吗	zhè xiē yī fú xǐ guò ma
6	**北京近来很寒冷**	běi jīng jìn lái hěn hán lěng
7	他家每年放鞭炮	tā jiā měi nián fàng biān pào
8	外孙出生在农村	wài sūn chū shēng zài nóng cūn
9	星期二别打篮球	xīng qī èr bié dǎa lán qiú
10	短裙长度正合适	duǎn qún cháng dù zhèng hé shì

Appendix D: 2^{nd} Phase Sentence Lists

This Appendix illustrates the English, Polish, Arabic, and Mandarin sentence lists used for the second phase listening tests.

ENGLISH SENTENCE LIST

Airport check-in area

I am afraid the luggage allowed on this flight is two pieces maximum, regardless of the maximum weight permitted. The charge per extra luggage is fifteen Euros, which you can pay at the airline's counter.

Unfortunately your luggage cannot go directly to New York, as you are flying with two different airlines. Once you arrive in London, please collect your luggage and check in again for your next flight.

I am sorry to inform you that, due to bad weather, the flight is delayed by at least one and a half hour. It will therefore not leave before two thirty. Please go to the airline's counter if you want to change your booking.

You have exceeded the baggage allowance by three kilos. You can remove some items, or you can choose to pay an excess baggage fee of thirty Euros. This fee's been calculated from a ten Euros charge per extra kilo.

Unfortunately we cannot proceed with the check-in: the luggage belt is broken at the moment. This is being fixed and the system should be working again soon. Please stay in the queue until I call you forward.

We have just been told that all flights are being cancelled or delayed because of the fog. Your flight has not been confirmed yet, so we cannot start the check-in. Please remain in the check-in area for further updates.

Hospital reception area

In order to book an appointment, I first need you to fill in this form and submit it to me when completed. Please write down your name, date of birth, phone number and health insurance number if available.

I will need to confirm the time of the appointment with the doctor, but I will give you a call once this has been done. Please remember that you will need to bring all your scans and medical reports from the previous two years.

Unfortunately there is no slot available this morning. I am aware that mornings are better for you, but we just had a cancellation for three pm today. Is there any chance you would be able to come then?

A doctor needs to review your medical history before we can register you at this health centre. This is an insurance requirement. Please fill in this standard health form and return it to me when finished.

The results of your blood tests should arrive by the end of this week. The laboratory still needs to confirm whether this will be on Thursday or Friday, but I will call you as soon as I receive this information.

We are very sorry, but your appointment is delayed by one hour, as the doctor has just been called on an emergency. If you cannot stay, we have a slot available tomorrow morning at nine. Is this ok?

Café

I am really looking forward to the weekend. Yesterday I spent some time planning a two hour hike in the mountains, as well as a short boat trip on the lake, if the weather is good. Would you be interested in coming with me?

I am thinking about organising a dinner party with all our friends this Saturday. We could have it at my place, and I could ask everybody to bring some homemade food. Do you think this is a good idea?

I really enjoy living in that neighbourhood. I am not far from the city centre, there are fantastic shops, and the rent is lower than what I used to pay. Why don’t you also consider moving into that area?

My brother is coming to visit next month. He doesn’t come here very often, so I am really looking forward to it. Maybe we could go out for dinner together when he is around and you would finally meet him! What do you think?

We could organise a night out during the week-end. First go to a restaurant and then watch a movie. There are so many good movies that came out recently, and I have not seen any of them. Would you like that?

Last year we travelled to the north and saw some amazing landscapes: beautiful mountains, lakes, forests... There was also a lot of wildlife. I am surprised you never went there. Why don’t you go there this summer?

Practice Sentence

I would like to thank you for taking part in this test. This practice session should clarify all the test procedures, but please do not hesitate to ask me any question if you need any further clarification.

POLISH SENTENCE LIST

Airport chack-in area

Obawiam sie, ze na ten lot dozwolone sa tylko dwie sztuki bagazu, niezaleznie od maksymalnej dozwolonej wagi. Oplata za dodatkowy bagaz, ktora mozesz uiscic w kasie linii lotniczych, wynosi pietnascie Euro.

Niestety twoj bagaz nie moze byc nadany bezposrednio do Nowego Jorku, jako ze lecisz dwoma roznymi liniami lotniczymi. Kiedy tylko przybedziesz do Londynu prosze odebrac bagaz i odprawic sie ponownie na nastepny lot.

Z przykroscia informuje, ze z powodu zlej pogody lot jest opozniony o co najmniej poltorej godziny. Zatem samolot nie wyleci przed druga trzydziesci. Jesli zyczylbys sobie zmiane rezerwacji prosze udac sie do kasy linii lotniczych.

Przekroczyles dozwolona wage bagazu o 3 kg. Mozesz wypakowac kilka rzeczy lub uiscic oplate za nadbagaz, ktora wynosi 30 Euro czyli 10 Euro za kazdy dodatkowy kilogram wagi.

Niestety nie mozemy przystapic do odprawy: tasma bagazowa.jest chwilowo w trakcie naprawy ale usterka powinna byc wkrotce naprawiona. Prosze czekac w kolejce do momentu wezwania.

Wlasnie sie dowiedzielismy,ze wszystkie loty zostaly odwolane lub opoznione ze wzgledu na mgle. Twoj lot nie zostal jeszcze potwierdzony, dlatego nie mozemy przystapic do odprawy. Prosze czekac w sali odlotow do czasu podania dalszych szczegolow.

Hospital reception area

Zeby umowic sie na wizyte musisz w pierwszej kolejnosci wypelnic ten formularz i oddac mi go wypelnionego. Prosze zapisac swoje imie, date urodzenia, numer telefonu i numer ubiezpieczenia zdrowotnego jesli jest dostepny.

Bede musial potwierdzic termin wizyty z doktorem ale zadzwonie do ciebie jak tylko to zrobie. Pamietaj prosze o przyniesieniu wszystkich swoich zdjec przeswietlen i raportow medycznych z poprzednich dwoch lat.

Niestety dzisiejszego ranka nie ma zadnych wolnych miejsc. Rozumiem, ze poranny termin jest dla ciebie dogodniejszy ale wlasnie zwolnilo sie miesjce na trzecia popoludniu. Czy jest szansa, ze bedziesz mogl przyjsc na ta wizyte?

Zanim zostaniesz zarejestrowany w tej przychodni lekarz bedzie musial przejrzec twoja historie medyczna. Takie sa wymogi ubezpieczenia. Prosze wypelnic standardowy formularz i zwrocic go do mnie kiedy bedzie wypelniony.

Wyniki badania krwi powinny byc gotowe pod koniec tego tygodnia. Laboratorium powinno potwierdzic czy bedzie to czwartek czy piatek. Zadzwonie jak tylko otrzymam informacje.

Bardzo mi przykro ale twoja wizyta lekarska jest opozniona o godzine poniewaz doktor zostal wezwany do naglego przypadku. Jesli nie mozesz poczekac dluzej mamy wolne miejsce jutro rano o 9. Odpowiada to tobie?

Café

Naprawde nie moge doczekac sie weekendu, wczoraj spedzilem czas na planowaniu dwugodzinnej wedrowki po gorach jak rowniez krotkiej wycieczki lodka po jeziorze, jezeli tylko pogoda bedzie dobra. bedziesz chcial pojechac ze mna?

Mysle nad zorganizowaniem kolacji w sobote ze wszytskimi naszymi przyjaciolmi. moglibysmy zorganizowac u mnie i moglabym poprosic wszystkich o przyniesienie jakiegos jedzienia.Myslisz, ze to dobry pomysl?

Naprawde dobrze mieszka mi sie w tym sasiedztwie. nie jest to daleko od centrum miasta, sa tu fantastyczne sklepy i wynajem jest tanszy niz placilam poprzednio. Moze takze rozwazysz przeprowadzke do tej dzielnicy?

Moj brat przyjezdza z wizyta w nastepnym miesiacu. on nie przyjezdza tutaj zbyt czesto dlatego z niecierpliwoscia oczekuje jego przyjazdu.Moglibysmy pojsc razem na kolacje jak on tu bedzie i wreszcie moglbys go poznasz.co on tym myslisz?

Moglibysmy zorganizowac wyjscie w weekend.najpierw pojsc do restauracji a potem do kina. Jest duzo dobrych filmow, ktore wyszly ostatnio a ja nie widzialam zednego z nich. chcialbys tak zrobic?

W zeszlym roku bylismy na polnocy i widzielismy wspaniale widoki, piekne gory, jeziora, lasy.... Bylo tam takze duzo dzikich zwierzat. jestem zdziwiona ze nigdy tam nie byles. moze wybierzesz sie ta tego lata?

ARABIC SENTENCE LIST

Airport check-in area

أخشى أن الأمتعة المسموح بها على هذه الرحلة هي حقيبتان بالحد الأقصى، بغض النظر عن الوزن الأقصى المسموح به. الغرامة لكل حقيبة اضافية هو خمسون يورو، والتي يمكنك دفعها على شباك شركة الطيران.

للاسف لا يمكن لحقيبتك الذهاب مباشرة الى نيويورك, لأنك تسافر مع شركتي طيران. حالما تصل إلى لندن، رجاءً استلم حقيبتك و قم باجراءات السفر مجدداً لرحلتك الثانية.

يؤسفني أن أعلمك أنه وبسبب الطقس السيئ فان الرحلة الجوية قد تأخرت على الأقل ساعة و نصف. لذلك فانها لن تغادر قبل الثانية و النصف. الرجاء الذهاب لشباك شركة الطيران اذا أردت تغيير حجزك.

لقد تجاوزت الحد المسموح للوزن بثلاث كيلو غرامات. يمكنك تفريغ بعض الاغراض أو يمكنك دفع اجر اضافي ثلاثون يورو للوزن الاضافي. هذا الاجر الاضافي تم حسابه بعشرة يورو لكل كيلو اضافي.

للاسف لن نتمكن من المتابعة باجراءات االسفر, حزام الأمتعة معطل حالياً. في هذه الأثناء يتم اصلاحه وينبغي أن يكون النظام جاهزاً للعمل في القريب العاجل. الرجاء البقاء في رتل الانتظار الى ان أطلب منك التقدم.

لقد أُعْلِمنا أن جميع الرحلات قد الغيت أو أجلت بسبب الضباب. لم يتم تأكيد موعد رحلتك بعد, لذلك لا يمكننا البدء باجراءات السفر. الرجاء البقاء بالقرب من منطقة الاجراءات لتبقى على علم باخر التطورات.

Hospital reception area

لتقوم بحجز موعد, أولا أريدك ان تقوم بملئ هذه الاستمارة وتسليمها لي حال الانتهاء من ذلك. الرجاء كتابة اسمك و تاريخ ميلادك ورقم هاتفك ورقم الضمان الصحي في حال توفره في الأسفل.

أريد تأكيد وقت الموعد مع الطبيب, و لكن سأتصل بك حال التاكد من ذلك. رجاءً تذكر أنه عليك احضار كل الصور الشعاعية و التقارير الطبية الخاصة بك للسنتين الماضيتين.

للأسف لا يتوفر أي وقت خالي هذا الصباح. أنا أعلم تماما أن المواعيد الصباحية هي أفضل لك و لكن للتو قد ألغينا حجز كان مقررا الساعة الثالثة ظهرا اليوم. هل من الممكن لك أن تكون قادرا على المجيئ في هذا الموعد؟

الطبيب بحاجة للاطلاع على سجلك الصحي قبل أن نتمكن من تسجيلك في المركز الصحي. هذا طلب التامين الصحي. الرجاء ملىء استمارتك الصحية و أعدها الي عند الانتهاء من ذلك.

نتائج فحص الدم المتعلقة بك سوف تصل في نهاية هذا الاسبوع. المخبر لم يؤكد بعد فيما لو ان ذلك سوف يتم يوم الخميس أو الجمعة, لكن سأتصل بك حال الحصول على هذه المعلومة.

نحن بغاية الاسف, ان موعدك قد تأجّل ساعة, كون الطبيب قد طلب بشكل اسعافي. اذا كنت لا تستطيع الانتظار، لدينا موعداً شاغراً صباح الغد في الساعة التاسعة. هل يناسبك ذلك؟

Café

أنا أنتظر بفارغ الصبر عطلة نهاية الأسبوع. البارحة قضيت بعض الوقت في التخطيط للتنزه ساعتين في الجبال، وكذلك رحلة قصيرة بالقارب في البحيرة ، إذا كان الطقس جيد. هل تود المجيئ معي؟

أنا أفكر في تنظيم حفل عشاء مع جميع أصدقائنا هذا السبت, يمكن أن نقيمه في منزلي، ومن الممكن أن أطلب من الجميع أن يجلبوا بعض الأطعمة المنزلية الصنع. هل تعتقد أن هذه فكرة جيدة؟

أنا أستمتع حقا بالعيش في ذاك الحي. أنا لست بعيد عن وسط المدينة، وهناك محلات رائعة، و تكاليف الإيجار أقل مما اعتدت على دفعه. لماذا لا تفكر أنت أيضا بالانتقال إلى تلك المنطقة ؟

أخي قادم لزيارتي الشهر القادم. هو لا ياتي الى هنا بشكل متكرر, لذلك انا أتطلع لزيارته بفارغ الصبر. من الممكن ان نخرج سوية للعشاء عندما يأتي و سوف تلتقي به أخيراً! ما رأيك؟

يمكننا تنظيم ليلة ساهرة نهاية هذا الاسبوع. في البداية نذهب الى المطعم ومن ثم نشاهدة فلم. هناك العديد من الأفلام الجيدة تعرض مؤخراً, و لم يتسنى لي مشاهدة أي منهم. هل ترغب بذلك؟

في السنة الماضية سافرنا إلى الشمال و رأينا بعض المناظر الطبيعة الخلّابة: جبال وبحيرات وغابات رائعة الجمال...واستمتعنا بمشاهدة الحياة البرية. أنا متفاجئ أنك لم تذهب إلى هناك مطلقاً. لماذا لا تذهب إلى هناك هذا الصيف؟

Practice sentence

أود أن أشكركم على المشاركة في هذا الاختبار. هذه الجلسة العملية يجب أن توضح جميع إجراءات الاختبار، ولكن من فضلك لا تتردد في سؤالي أي سؤال إذا كنت بحاجة إلى أية توضيحات اضافية.

MANDARIN SENTENCE LIST

Airport check-in area

尽管您的托运物没有超重，但按规定托运物最多两件，您超过两件的多余部分要收取额外的15欧元，您可以到相应的航空服务公司客服台前付费。

您的投运物不能直接运到纽约，因为您订的是两家不同的航空公司的机票。当您达到伦敦后，需要取您的托运行李，并且要为您的下一趟航班重新办理登记手续。

很遗憾告诉大家，由于天气状况不好，本次航班要推迟至少一个半小时，飞机在2:30之前不会起飞，如果您要更改预订的机票，请到相关航空公司柜台前办理。

您的托运行李超重了3**公斤。你可以取出一些**东西，或者你**也可以交**30**欧元超重**费。超重费是按每公斤10**欧元**计算出来的。

抱歉，由于**行李**传送带出故障现在不能办理登机手续。故障应该很快被排除。请继续排队等侯通知。

我们刚接到通知，大雾导致所有航班被取消或者晚点。您的班次有待确定，现在还不能办理登机手续。请在登机区等侯通知。

Hospital reception area

为完成您的预订，我需要您填写这个表，完成之后交给我。请填写您的姓名，出生日期，电话号码，如果有安全保险号的话也填写。

我需要您确认一下和医生的预约时间，我会在预约完成后给您打一个电话。请别忘了您需要带上您前两年的扫描片和医检报告

对不起，今早没有空余时间。我注意到早晨对您来说更好些，但我们今天只有下午三点有空，您看您这个时间能否有空过来?

因为保险的要求，您在本医疗诊所注册之前医生需要审查您的病史。请填写这份标准健康表格，完成之后交给我。

您的血检结果会在下周末前出来，实验室还没有告诉我们具体是下星期四还是星期五。我们收到通知后会立刻转告你。

很抱歉，由于医生急诊您的预约需要推后 1 小时。如果您没时间等，您可以明天早上 9 点再来。这样可以吗?

Café

我盼望着周末，昨天我花了点时间计划了 2 个小时的登山计划，如果天气好的话，还有一个短的湖上划艇，你有兴趣和我一起去吗？

我在考虑组织一次周六和朋友们一起度过的晚餐聚会，我们可以在我的住处举办，我希望每个人能带点自制的食物，你们觉得这是个好主意吗？

我非常喜欢周边环境，离市中心不远，有许多不错的超市，房租比我之前的要便宜。你也考虑搬过来。

我哥哥（弟弟）下个月来看我。因为他不经常来，我非常期望见到他。他来以后我们可以一起去吃顿饭，你们见见面。怎么样？

这个周末我们可以一起出去吃顿晚饭，然后看场电影。最近有许多好电影上演，我都还没有看。你想不想去？

去年到北方旅游，我们领略了很多漂亮的自然风光：有青山绿水和森林，还有许多野生动物。你居然没有去过。你干吗不今年夏天去一下呢？

Practice Sentence

我想请你参加这个试验。我会给你解释试验程序。你有不明白之处请问我。

Appendix E: Questionnaire

This Appendix illustrates the questionnaire used for the second phase listening tests, including listening test instructions.

SUBJECT 1

Gender Male ☐ Female ☐

Age (years) …………….

Do you have any hearing problems that you are aware of? Yes ☐ No ☐

Instructions

Thank you for accepting to take part in this listening test. The listening test will last approximately 20 minutes. If you become tired or uncomfortable, feel free to leave the room and/or take a break at any time.

The experiment consists of 9 different tests (combination of 3 different environments and 3 different speech recordings). Each test (1 slide) will combine a photograph of a specific environment and a speech recording related to it.

After listening to each test, you will be asked to evaluate various aspects of the speech and acoustic environment by filling the forms that have been given to you.

During the experiment, you are going to evaluate 3 different types of busy environments in a random order; an airport check-in area, a hospital reception area, and a café.

You will have to imagine that you are in the environment presented, and a person is in front of you and is talking to you. The acoustic conditions might change, so while listening you should pay particular attention to the speech and acoustic properties of the environment.

You can listen to each condition only one time.

In order to become familiar with the test procedures, you will now be presented a short practice speech sample.

LISTENING TEST

EVALUATION FORM

Please evaluate the speech by ticking a box under each question

- **How intelligible is the speech?**

Very intelligible	Intelligible	Neither intelligible nor unintelligible	Unintelligible	Very unintelligible
☐	☐	☐	☐	☐

- **How high is the speech level?**

Very high	High	Normal	Low	Very low
☐	☐	☐	☐	☐

- **How pleasant is the conversation?**

Very pleasant	Pleasant	Neither pleasant nor unpleasant	Unpleasant	Very unpleasant
☐	☐	☐	☐	☐

Please evaluate the acoustic environment by ticking a box under each question

- **How noisy is the acoustic environment?**

Very noisy	Noisy	Neither noisy nor quite	Quiet	Very quiet
☐	☐	☐	☐	☐

- **How annoying is the acoustic environment?**

Very annoying	Annoying	Neither annoying nor favourable	Favourable	Very favourable
☐	☐	☐	☐	☐

- **How relaxing is the acoustic environment?**

Very relaxing	Relaxing	Neither relaxing nor stressful	Stressful	Very stressful
☐	☐	☐	☐	☐

- **How comfortable is the acoustic environment?**

Very comfortable	Comfortable	Neither comfortable nor uncomfortable	Uncomfortable	Very uncomfortable
☐	☐	☐	☐	☐

- **How pleasant is the acoustic environment?**

Very pleasant	Pleasant	Neither pleasant nor unpleasant	Unpleasant	Very unpleasant
☐	☐	☐	☐	☐

- **How eventful is the acoustic environment?**

Very eventful	Eventful	Neither eventful nor uneventful	Uneventful	Very uneventful
☐	☐	☐	☐	☐

- **How exciting is the acoustic environment?**

Very exciting	Exciting	Neither exciting nor boring	Boring	Very boring
☐	☐	☐	☐	☐

- **How familiar is the acoustic environment?**

Very familiar	Familiar	Neither familiar nor unfamiliar	Unfamiliar	Very unfamiliar
☐	☐	☐	☐	☐