Subjective speech intelligibility and soundscape perception of English, Polish, Arabic and Mandarin

Subjective speech intelligibility and soundscape perception of English, Polish, Arabic and Mandarin

Kivanc Kitapci[a)]
Laurent Galbrun[b)]
Centre of Excellence in Sustainable Building Design, School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom

This paper analyses perceived speech intelligibility and, more generally, soundscape perception associated to four different languages. The research 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) (STI=0.4, STI=0.5, and STI=0.6). Correlation analysis was carried out between the eleven semantic attributes, the four languages, the three room acoustic conditions, and the three environments, in view of developing an understanding of how the use of different languages might influence perceived speech intelligibility and, more generally, oral communication. Overall, the statistical analysis showed that the rating of semantic attributes is affected by languages, suggesting that these influence subjective speech intelligibility and, more generally, soundscape perception associated with oral communication. The perceptual variations associated to different languages should then be taken into consideration in the design of multi-lingual environments.

1 INTRODUCTION

In a modern and globalised world, the interaction between multilingual and multicultural people in public, commercial and social spaces is gaining importance, and communication is at the centre of this interaction. In the current literature, there are multiple studies which are looking at communication between non-native speakers; however, only very few studies have been comparing objective and subjective differences in speech intelligibility for native speakers of varying languages. The aim of our project is to find out possible relations between speech intelligibility and multi-lingual communication, in terms of acoustics, linguistics, and socio-

[a)] email: kk220@hw.ac.uk
[b)] email: l.g.u.galbrun@hw.ac.uk

cultural factors. In order to investigate the multi-dimensional structure of the intelligibility of speech in multi-lingual spaces, the project has been divided into two main phases.

In the first phase of the project, the interaction of various room acoustic parameters with different languages was investigated [1]. The current paper presents results obtained from the second phase of the study, which investigates the role of soundscape perception on subjective speech intelligibility. The combination of the results obtained from both phases will 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.

Although there are many recent studies on socio-linguistics and multilingual communication, the number of studies relevant to the research presented here 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. However, the methodologies used in most of these studies are useful in informing the present research.

For example, a recent study examined urban multilingualism in Europe [2]. 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). The results of the research revealed the distribution and language vitality of immigrant minority languages at home across European cities. It was 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 languages other than the mainstream language at home [2].

Wodak et al. [3] conducted a study on language choice and code-switching in institutions of the European Union. The institutions examined were the European Parliament and the European Commission. It was claimed that these two institutions are representative of the European population, therefore reflecting the same multilingual characteristics. It was hypothesised that various contextual settings and different language ideologies affect multilingual communication. As a result, it was understood that various languages are being simultaneously used in the contexts investigated [3].

These socio-linguistic studies provide a wide perspective of European languages and their interaction in a multi-lingual setting, and also illustrate the significance of multi-lingual environments within Europe. Socio-linguistic studies often focus on the relationship between languages and demographic data. However, this is an aspect that has not been examined by the present research and is therefore of limited interest.

Socio-lingual factors are amongst the main factors affecting speech intelligibility in multilingual spaces. While dealing with socio-lingual factors and communication between people, the perception of the sound environment can 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 [4], which considers all the sound present within a space and the perception of that sound environment. The soundscape methodology is therefore a valuable approach that has been used in the present study to evaluate the multiple factors affecting multi-lingual communication.

After reviewing the literature on room acoustics and socio-linguistics, it was found that the number of studies that investigated the relationship between languages and perceived speech intelligibility is very limited. Therefore, the present study aims to bridge that gap with the help of the soundscape theory.

2 METHODOLOGY

This section presents the methodology used for the study, including the selection of environments, preparation of sound samples, and the semantic test undertaken. The study was carried out using several sample groups, in which the native language of each sample group was the variable. 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), Sino-Tibetan (e.g. Mandarin) and Altaic (e.g. Japanese) language families. The specific languages identified for the research were English, Polish, Arabic and Mandarin [1].

Fifteen participants per language (i.e. a total of sixty) were asked to subjectively evaluate the acoustic environments by answering eleven questions on a five-point semantic scale, under three room acoustic conditions, in three digitally simulated multi-lingual environments (i.e., nine cases were rated by each participant). The three multi-lingual environments chosen were an airport check-in area, a hospital reception area, and a café. The airport check-in area was chosen as representative of a high reverberation time and high background noise acoustic environment, the latter including a public announcement message. The hospital reception area was a medium sized enclosure leading to a medium to low reverberation time, with a relatively low background noise that included a phone ringing. The airport and hospital were chosen as environments where speech contains critical information that cannot be risked to be unintelligible. The café was a medium to large sized space, with a moderately-high reverberation time and fairly steady background noise, which was mostly composed of hubbub speech noise and did not contain any noticeable intermittent noise events. Furthermore, the café characterised a relaxed environment, as opposed to the fairly stressful environments represented by the airport check-in area and the hospital reception.

The speech samples were uniquely designed for each environment in order to achieve an appropriate context. Six sentences were created for each environment (and translated from English for Polish, Arabic and Mandarin), 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 sentences were developed as conversations in which participants were being talked to, in order to produce a sense of engagement. The three room acoustic conditions were created digitally by adding contextually appropriate background noise and reverberation to the speech recordings. These acoustic conditions were defined by the sound transmission index (STI=0.4, STI=0.5 and STI=0.6), which could be predicted from the reverberation time and signal-to-noise ratio present in the environment. The finalised audio material was then presented to the participants in combination with the visuals of the environments. The presentation was made in a random order to prevent order effects, and different sentences were used for different acoustic conditions of the same environment. Tests were carried out in the anechoic chamber of Heriot-Watt University, where subjects were asked to evaluate the audio-visual material through semantic differential analysis. The latter is an evaluation survey technique that was developed by Osgood et al. [5], and suggested as a useful method to evaluate sound environments [6]. Eleven semantic descriptors were identified for the current study (3 descriptors to evaluate speech (intelligibility, loudness, and pleasantness) and 8 descriptors to evaluate the acoustic environment (noisiness, annoyance, relaxation, comfort, pleasantness, eventfulness, excitement, and familiarity)). This paper mainly focuses on analysing results obtained for the intelligibility attribute. Additionally, a Spearman's rank-order correlation analysis was also conducted in order to examine the significant correlations between the intelligibility attribute and the other 10 attributes. The results of the experiment were statistically analysed using the Statistical Package for Social Sciences (SPSS) software.

3 SEMANTIC ANALYSIS OF SPEECH INTELLIGIBILITY

This section examines the semantic differential analysis of the speech intelligibility attribute. The participants were asked to subjectively evaluate the presented speech sample on a 5-point scale (from unintelligible to intelligible). Figs. 1-3 show the relationship between the perceived speech intelligibility attribute scores (on a -2 to +2 scale) and the STI levels at the airport (Fig. 1), the hospital (Fig. 2), and the café (Fig. 3) for the four languages. These figures indicate that there are differences between the speech intelligibility attribute scores of English, Polish, Arabic and Mandarin, and that the scores vary between the STI conditions, between the environments and between the languages. As expected, results also show that the perceived intelligibility scores tend to increase as the STI increases.

The repeated measures ANOVA revealed that the effects of the speech transmission index (STI) and the environment on the speech intelligibility attribute are statistically significant (p = .000). Additionally, marginal significance (p = 0.051) was observed for variations between the perceived speech intelligibility of different languages. Furthermore, combined effects of STI and language (p = .000), environment and STI (p = .000), and environment, language, and STI (p = .013), were statistically significant for the scores of the speech intelligibility attribute.

The difference between the intelligibility scores obtained from the four languages were statistically analysed for each room acoustic condition by using the one-way Analysis of Variance (ANOVA) method, with a confidence interval set to 95%. The conditions that showed statistically significant differences (p<.05) between languages were the airport – STI=0.6 [F(3,59) = 5.29, p=.003], the café – STI=0.5 [F(3,59) = 3.80, p=.015], and the café – STI=0.6 [F(3,59) = 6.31, p=.001].

The largest differences between the languages were observed at the airport (Fig. 1) and at the café (Fig. 3), at STI=0.6. In both cases, the language perceived to be most intelligible was English (airport=+1.48, café=+1.8) and the language perceived to be least intelligible was Polish (airport=+0.53, café=+0.8). The smallest difference between scores was observed at the hospital at STI=0.5, in which English and Arabic were the most intelligible languages with an average score of +1.06, while Polish was the least intelligible language with an average score of +0.8 (Fig. 2).

English was perceived to be 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). However, 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). When these results are compared with the results of the first phase of the study [1], contradictions are observed. For instance, the sentence scores of English from the first phase showed that English was the most intelligible language at STI=0.4. The word intelligibility scores also revealed that it was the most intelligible language for all the acoustic conditions tested (STI=0.2, STI=0.4, STI=0.6 and STI=0.8). It is worth noting that, unlike what was done for the tests presented here, artificial white noise was used as background noise in the first phase of the study, which might be responsible for these discrepancies.

Table 1 presents the differences between the highest and the lowest speech intelligibility attribute scores of each language at the airport, the hospital and the café. The differences between the spaces might have occurred due to the fact that the background noise samples used in the second phase of the study were variable and representative of complex real environments.

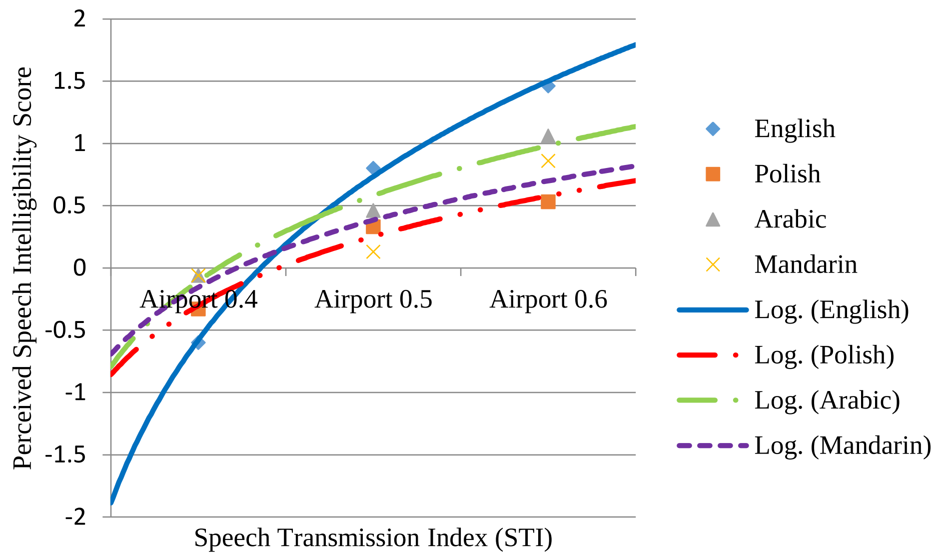

Fig. 1 - Comparison graph of the speech intelligibility attribute scores of English, Polish, Arabic, and Mandarin at the airport. Actual data markers and logarithmic regression lines are shown in the figure.

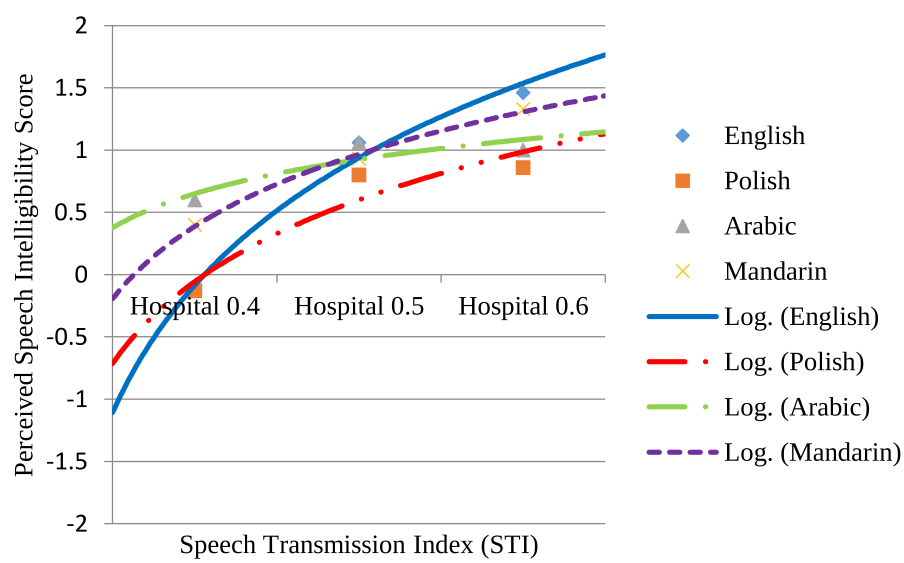

Fig. 2 - Comparison graph of the speech intelligibility attribute scores of English, Polish, Arabic, and Mandarin at the hospital. Actual data markers and logarithmic regression lines are shown in the figure.

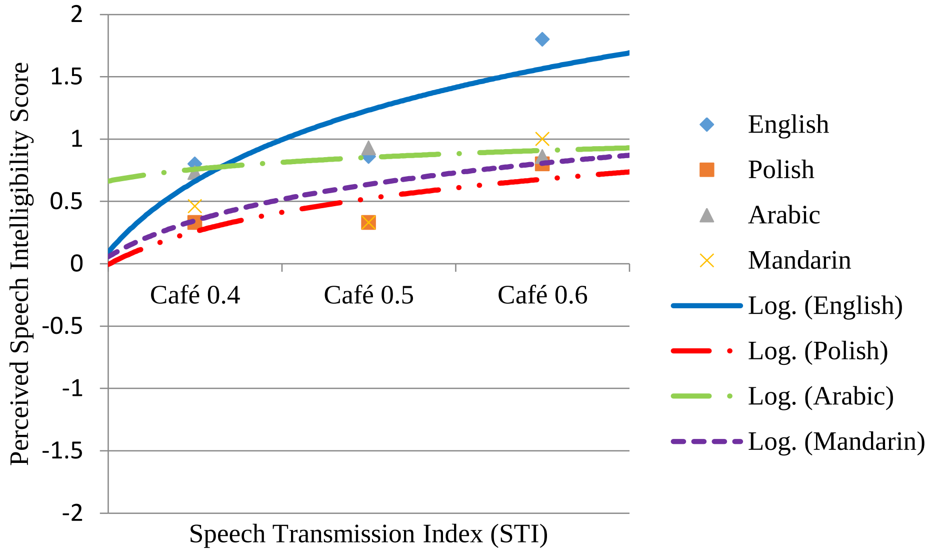

Fig. 3 - Comparison graph of the speech intelligibility attribute scores of English, Polish, Arabic, and Mandarin at the cafe. Actual data markers and logarithmic regression lines are shown in the figure.

For example, the airport and hospital environments contained distracting noise sources (public announcement in the airport and phone ringing in the hospital), whilst the sound environment of the café was designed to be more relaxing, with no noticeable intermittent noise events. The data shown in Table 1 also indicates that English showed the largest variance in all the environments. In that respect, it should be noted that the background noise in the airport environment contained a public announcement that was in English, and which might have been more distracting for native English speakers. In fact, results suggest that English is more sensitive to meaningful and distracting sound events, as lower scores are observed at the airport and the hospital rather than at the café, for the STI=0.4 condition where the level of background noise is higher. The results at the café do comply with the results of the first phase of the study [1], arguably because of the fairly steady background noise used in that environment.

The analysis of the perceived intelligibility scores of Polish revealed that it was the least intelligible language in 7 out of 9 cases, except at the airport for STI=0.4 (-0.33) and STI=0.5 (+0.33), in which it was one of the two least intelligible languages. It should also be noted that Polish was the least intelligible language across all the environments at STI=0.6 (airport=+0.53, hospital=+0.86, and café=+0.8). When compared to the results of the first phase of the study [1], contradictions are again observed. The sentence intelligibility scores obtained in the first phase

Table 1 - Differences between the highest and lowest perceived speech intelligibility 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	1.00	0.47	0.13	0.54

revealed that Polish was the most intelligible language at STI=0.6; and according to the word intelligibility scores of the first phase, Polish was the second most intelligible language at STI=0.6. However, in the results presented here it was the least intelligible language at STI=0.6 for the three environments tested, these results suggesting that the type of background noise might not justify such discrepancies.

The analysis of the perceived speech intelligibility scores of Arabic are also contradictory when compared to the first phase intelligibility scores. In fact, the first phase word and sentence intelligibility scores of Arabic were the lowest at STI=0.4 and STI=0.6 [1]. However, the perceived speech intelligibility scores of Arabic in phase 2 were the highest in 2 out of 3 cases at STI=0.4 (airport=-0.06, hospital=+0.6), whilst Arabic had the second highest perceived intelligibility score at the café (+0.73). The rankings were lower at STI=0.6, but were still not the lowest, as Arabic had the second highest intelligibility attribute scores in the airport (+1.06), and the second lowest intelligibility attribute score in the hospital (+1.0) and in the café (+0.86). Similar to the scores obtained 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), whilst it had the second highest score at the airport (+0.46). Table 1 shows that Arabic had the lowest variance of the perceived speech intelligibility scores at the hospital (0.4) and the café (0.13). Therefore, it can be assumed that the change in 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 [1]. In particular, it is worth noting that the large drop in word intelligibility of Arabic observed in phase 1 between STI=0.6 and STI=0.4, appeared to be justified by the introduction of artificial noise for the STI=0.4 condition. In that respect, the experimental procedures of phase 1 and phase 2 are not comparable, as identical STI values do not mean identical acoustic conditions.

Analysis of the perceived speech intelligibility scores of Mandarin revealed that it had the lowest scores at STI=0.5 at the airport (+0.13), and at STI=0.5 at the café (+0.33), and that it was the most intelligible language only at STI=0.4 at the airport (-0.06) (same result as Arabic). The largest difference between scores was observed at the hospital (0.93), and the smallest difference was observed at the café (0.54) (Table 1). 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 [1], which complies with the perceived intelligibility scores obtained here at the airport.

Overall, the analysis of the intelligibility attribute scores indicate that the subjective evaluation of speech intelligibility of each language varies, depending on the type of environment, the type of background noise, reverberation time, and signal-to-noise ratio. Perceived intelligibility of English appeared to be mostly affected by the information content (e.g. public announcement) and distracting sounds (e.g. phone ringing) carried in the background noise. Furthermore, some discrepancies were observed between the perceived intelligibility scores presented here and the objective intelligibility scores obtained from phase 1 of the study [1], which might be due to the complexity and variability of the cases considered here (e.g. differences in background noise, effect of distracting noise sources, variable environments and contexts).

4 CORRELATION ANALYSIS

In order to examine the relationship between the intelligibility attribute and the other 10 attributes tested (loudness, speech pleasantness, noisiness, annoyance, relaxation, comfort, environmental pleasantness, eventfulness, excitement, and familiarity), correlations were statistically analysed by running Spearman’s correlation analysis (one-tailed). 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 = .924), Mandarin (ICC = .898), Arabic (ICC = .912), and Polish (ICC = .881), where ICC > .720 is usually considered as an acceptable value for social sciences [7].

First, the Spearman's rank-order correlation analysis was conducted by including all the data, without subdividing it into groups of languages, environments, or STI conditions. The analysis revealed that 9 out of 10 attributes were strongly correlated with the speech intelligibility attribute, the exception being the excitement attribute. Loudness (r_s = .356, p = .000), speech pleasantness (r_s = .350, p = .000), relaxation (r_s = .321, p = .000), comfort (r_s = .356, p = .000), environmental pleasantness (r_s = .377, p = .000), and familiarity (r_s = .129, p = .001) were positively correlated, whilst noisiness (r_s = -.235, p = .000), annoyance (r_s = -.329, p = .000), and eventfulness (r_s = -.181, p = .000) were negatively correlated with perceived speech intelligibility.

Next, in order to examine the variance in terms of significant correlations in between languages, the analysis was repeated by dividing the data into groups of languages. The Spearman’s rank-order correlation analysis of English revealed that 8 out of 10 attributes were significantly correlated with the speech intelligibility attribute: loudness (r_s = .522, p = .000), speech pleasantness (r_s = .325, p = .000), relaxation (r_s = .487, p = .000), comfort (r_s = .457, p = .000), and environmental pleasantness (r_s = .486, p = .000) were positively correlated, whilst noisiness (r_s = -.321, p = .000), annoyance (r_s = -.551, p = .000), and eventfulness (r_s = -.417, p = .000) were negatively correlated with perceived speech intelligibility. The Spearman’s rank-order correlation analysis of Polish revealed that 8 out of 10 attributes were significantly correlated with the speech intelligibility attribute: loudness (r_s = .244, p = .002), speech pleasantness (r_s = .190, p = .014), relaxation (r_s = .289, p = .000), comfort (r_s = .310, p = .000), and environmental pleasantness (r_s = .167, p = .027) were positively correlated, whilst noisiness (r_s = -.195, p = .012), annoyance (r_s = -.265, p = .001), and eventfulness (r_s = -.188, p = .014) were negatively correlated. The Spearman’s rank-order correlation analysis of Arabic revealed that only 4 out of 10 attributes were significantly correlated with the speech intelligibility attribute: loudness (r_s = .297, p = .002), speech pleasantness (r_s = .355, p = .000), comfort (r_s = .182, p = .017) and environmental pleasantness (r_s = .245, p = .002) were positively correlated. It should be noted that no significant negative correlations were found. For Mandarin, the correlation analysis revealed that 8 out of 10 attributes were strongly correlated with the speech intelligibility attribute. The attributes that were positively correlated were loudness (r_s = .376, p = .000), speech pleasantness (r_s = .509, p = .000), relaxation (r_s = .421, p = .000), comfort (r_s = .439, p = .000), environmental pleasantness (r_s = .487, p = .000), and familiarity (r_s = .306, p = .000), whilst noisiness (r_s = -.300, p = .000) and annoyance (r_s = -.461, p = .000) were negatively correlated with perceived speech intelligibility. It is interesting to note that Mandarin was the only language that showed a significant correlation between the comfort attribute and the intelligibility attribute.

Overall, it was observed that the attributes correlated with perceived speech intelligibility were the same for English and Polish. Another interesting finding was that Arabic showed no negative correlations with the speech intelligibility attribute, suggesting that Arabic listeners might be more resilient to poor acoustic conditions. Additionally, Mandarin was the only language that showed a significant positive correlation between the familiarity and the speech intelligibility attributes. These results indicate that perceived intelligibility can show different correlations with perceptual attributes across different languages, confirming that multi-lingual communication is not solely affected by room acoustic conditions.

5 CONCLUSIONS

This paper analysed of the subjective speech intelligibility and soundscape perception of English, Polish, Arabic, and Mandarin. The semantic differential analysis of the speech intelligibility attribute revealed that the subjective evaluation of speech intelligibility of each language varies, depending on the type of environment, as well as the type of background noise, reverberation time, and signal-to-noise ratio. The perceived speech intelligibility of English appeared to be mostly affected by the information content and distracting sounds present in the background noise. It was also observed that correlations between perceived speech intelligibility and the other attributes vary with languages. The results presented here, together with those obtained from the first phase of the project [1], will ultimately lead to design recommendations that take into account language effects in view of improving speech intelligibility.

6 ACKNOWLEDGEMENTS

The authors would like to thank Dr Bernadette O'Rourke and Professor Graham H. Turner for the valuable help provided in the establishment and development of the research.

7 REFERENCES

1. K. Kitapci and L. Galbrun, "Comparison of speech intelligibility between English, Polish, Arabic and Mandarin", *Proceedings of Forum Acusticum 2014*, Krakow (2014).

2. G. Extra and K. Yağmur, "Urban multilingualism in Europe: Mapping linguistic diversity in multicultural cities", *Journal of Pragmatics,* **43**, (2011).

3. R. Wodak, M. Kryzyzanowski, and B. Forchtner, "The interplay of language ideologies and contextual cues in multilingual interactions: Language choice and code-switching in European Union institutions", *Language in Society,* **41**, (2012).

4. M. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*. New York, (1977).

5. C.E. Osgood, G. Suci and P. Tannenbaum, *The measurement of meaning,* Urbana, IL: University of Illinois Press, (1957).

6. J. Kang and M. Zhang, "Semantic differential analysis of the soundscape in urban open public spaces", *Building and Environment*, **45(1)**, (2010).

7. P. Shrout and J. Fleiss, "Intraclass correlations: Uses in assessing rater reliability", *Psychological Bulletin*, 86 (2), 420-428, (1979).