Speech intelligibility in multilingual spaces

Speech intelligibility in multilingual spaces

Kivanc Kitapci[1], Laurent Galbrun[2], Bernadette O'Rourke[3], and Graham H. Turner [4]

[1,2] Heriot-Watt University, School of Built Environment

Edinburgh, EH14 4AS, United Kingdom

[3,4] Heriot-Watt University, School of Management and Languages

Edinburgh, EH14 4AS, United Kingdom

ABSTRACT

The cultural and social diversity of modern urban environments, can affect oral communication within spaces where multiple languages are used. In such environments, the intelligibility of speech is influenced not only by room acoustic parameters, but also by languages' characteristics and socio-cultural factors related to the individuals using those languages. The aim of the present study is to examine how physical, linguistic and socio-cultural factors affect communication of multilingual environments. More specifically, the study considers four languages (English, Polish, Chinese and Arabic), and investigates the relationship between speech intelligibility parameters measured from physical tests and listening tests. Speech intelligibility comparisons focus on results obtained from the speech transmission index (STI), word lists based upon the diagnostic rhyme test (DRT), and phonemically balanced sentence lists. In its second stage, the research will look at how socio-cultural factors affect speech intelligibility and communication.

Keywords: Room acoustics, Speech intelligibility, Soundscape

1. 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 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 the current study is to find out possible relations between speech intelligibility and multi-lingual communication, in terms of acoustics, linguistics and social factors. This 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.

Houtgast and Steeneken [1] investigated the correlation between various languages and speech intelligibility. They carried out a study using 11 western languages (English, Finnish, French, German, Hungarian, Italian, Dutch, Maori, Polish, Swedish and Slovak) in 16 acoustic conditions. As a result, it

[1] kk220@hw.ac.uk
[2] l.g.u.galbrun@hw.ac.uk
[3] b.m.a.o'rourke@hw.ac.uk
[4] g.h.turner@hw.ac.uk

was found out that the differences among intelligibility tests may be caused by several effects, and that two of these effects are talker specific effects and phoneme or language specific effects [1].

One of the most relevant works on comparing the speech intelligibility between various languages focused on the differences between Mandarin and English [2]. 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 found that the word intelligibility of Mandarin is better than the word intelligibility of English in the high speech transmission index (STI) condition that corresponds to a high signal-to-noise ratio (S/N). The reason for this could be that some English consonants are not intelligible under reverberant conditions; however, tones in Mandarin language are helpful for increasing the word intelligibility. Furthermore, the word intelligibility of English is better than that of Mandarin at lower STI values caused by decreased S/N ratio. It is hypothesized that because of the wider sound pressure level dynamic range of the English language, some English words could be more intelligible by picking up only the high peaks [2].

Peng *et al.* [3] compared Chinese speech intelligibility scores obtained from physical measurement and computer simulations. The study was based on a computer model of an enclosure and used the auralization technique, under several acoustic conditions. The research was also replicated for different enclosures, including classrooms [4-5]. These studies showed that the auralization technique can be used to conduct speech intelligibility tests, but the enclosures should be modelled very carefully and in detail to achieve viable results.

Although there are many recent studies on socio-linguistics and multilingual communication, the number of studies relevant to the research presented 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 [6]. 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 [6].

Wodak *et al.* [7] 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 [7].

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 will not be examined by the present research and is therefore of limited interest.

Socio-lingual factors and room acoustic conditions are the two main factors affecting speech intelligibility in multilingual spaces. While dealing with socio-lingual factors and communication between people, the perception of the sound environment becomes 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 [8], 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 will be 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 has been found out that the number of studies that investigated the relationship between languages and speech intelligibility is limited. Except the studies which compared two languages in terms of room acoustics, no research has examined the links present between speech intelligibility and social, lingual, and cultural aspects. Therefore, the present study aims to bridge that gap with the help of the soundscape theory.

2. METHODOLOGY

The present study on acoustics and speech intelligibility in multi-lingual spaces consists of two main phases. The first phase of the study focuses on the effects of room acoustic factors on languages and speech intelligibility, and the second phase of the study focuses on the correlation of socio-cultural effects and the intelligibility of speech.

The study first focuses on the comparison between subjective and objective speech intelligibility obtained for a variety of language and acoustic conditions. Initially, speech intelligibility tests will be carried out in a physical space. Objective speech intelligibility will be obtained from measurements of the Speech Transmission Index (STI), which is a function of room acoustic properties and is based on the Modulation Transfer Function (MTF) method [9]. Subjective speech intelligibility will be obtained from conducting Diagnostic Rhyme Tests (DRT), which are typically based on the proportion of words correctly understood in a word list, as well as using lists of sentences. Comparison of these two results will identify the correlations between subjective and objective speech intelligibility scores.

The study is carried out using several sample groups, in which the native language of each sample group is the variable. Languages representative of a wide range of linguistic properties has been 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. Chinese) and Altaic (e.g. Japanese) language families. The specific languages identified for the research are English, Chinese, Arabic, and Polish.

The majority of participants have been selected from undergraduate and post-graduate students of Heriot-Watt University, and the male/female ratio has been kept equal in each sample group. Speech intelligibility tests have been held in a reverberant room in the acoustics laboratory of Heriot-Watt University, where tests have been repeated using different room acoustic conditions in terms of reverberation and background noise (i.e. different reverberation times and speech to noise ratios).

In addition to speech intelligibility, the qualitative perception of communication is influenced by a variety of factors such as speech privacy, density of occupation and space's layout. The rating of these factors is expected to vary within different socio-cultural groups and this will be analysed in a second stage, using both quantitative (e.g. speech privacy related to signal to noise ratio; density of occupation) and qualitative measures (e.g. questionnaires based on preference scales). The latter will be based on a number of simulations modelled for a variety of communication interactions.

Ultimately, the correlations between subjective and objective speech intelligibility scores and the qualitative analysis of communication will allow developing design guidelines that take into account multilingualism and socio-cultural variations. This will be done by providing guidelines in terms of reverberation time and speech to noise levels which should not be exceeded, maximum density of occupation and recommended layouts.

2.1 Selecting the Languages

The selection process of languages had various steps. First of all, a comparison table was prepared to identify the differences between some of the major languages used in the world (Table 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 is presented in the comparison table.

Table 1 – The comparison table of the common languages of the world

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)

It should be noted that the selected language group was required to represent a western multilingual environment. Another important criterion used in the selection of the languages was the consonant-to vowel ratio, as the speech intelligibility is affected by the loss of consonants. Therefore, it was hypothesized that the languages that have a high consonant-to-vowel ratio might be more sensitive to

the room acoustic conditions in terms of speech intelligibility. Another linguistic factor considered was the tonal properties of the languages. To examine the effects of the tonal system of a language on the speech intelligibility, at least one tonal language had to be selected. 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 speakers' populations were selected. The availability of native speakers who have a neutral accent in 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.

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

2.2 Room Acoustic Phase

The objective of the first phase of the study is to compare the subjective and objective speech intelligibility of the selected languages (English, Chinese, Arabic, and Polish) under various acoustic conditions. The gathered data will be used for comparing subjective and objective speech intelligibility both within each language, and between the languages. The results will lead to an understanding of the relations present between the linguistics properties and the acoustic properties of a room.

To assess the objective speech intelligibility, the speech transmission index (STI) method is used. To assess the subjective speech intelligibility, Diagnostic Rhyme Tests (DRT) and sentence lists are used. DRT is a listening test consisting of 192 words arranged in 96 pairs. The words are common, monosyllabic words, and most of them have three sounds ordered in a consonant-vowel-consonant sequence. The word pairs differ only in their initial consonants. The word lists [10] [11] [12] [13] have been recorded using professional native speakers of the languages (three males and three females), in an anechoic chamber [10].

The words and sentences have then been randomized in order to use the same lists several times for various acoustic conditions [10]. Before the actual recordings, a practice list was recorded by each talker, to make them familiar with the process.

The listening tests have been conducted in the reverberant chamber of the Heriot-Watt University acoustics laboratory. Three male and three female listeners have been selected from native speakers of each language. The recorded word lists were presented through a loudspeaker. Listeners' had to identify the spoken words and sentences within the visually presented material. The listening test was repeated for four different acoustic conditions, by changing the reverberation time and signal-to-noise ratio. The reverberation time was controlled by mounting sound absorber panels on the walls, and the signal-to-noise ratio was controlled by digitally adding an artificial noise signal to the speech signal by using a noise generator. The participants were advised to take a break after each test.

The objective evaluation of the speech intelligibility was measured using the commercial Maximum Length Sequence System Analyzer (MLSSA) software. The gathered data from MLSSA calculations will be compared to the subjective speech intelligibility scores to see any correlations between linguistic properties of a language, the acoustic properties of a room, and the speech intelligibility. The work is on-going at the time of writing, and provisional results will be presented at the conference.

2.3 Soundscape Phase

To emphasize human communication in multi-lingual spaces, the connection between social factors, cultural factors and emotional descriptors should also be established. In the second phase of the study, socio-lingual and cultural effects on the intelligibility of the speech will be investigated by using computer modelling, auralization, and soundscape theory.

Gathering the subjective data derived from emotional descriptors is possible by conducting a soundscape study and a semantic differential analysis. Davies *et al* [14] 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 the two main dimensions of the theory. The word "sound" was related with the description of physical properties of the sound itself, and the word "scape" was used for explaining the emotional interaction between the listeners and the environment (Figure 1) [14].

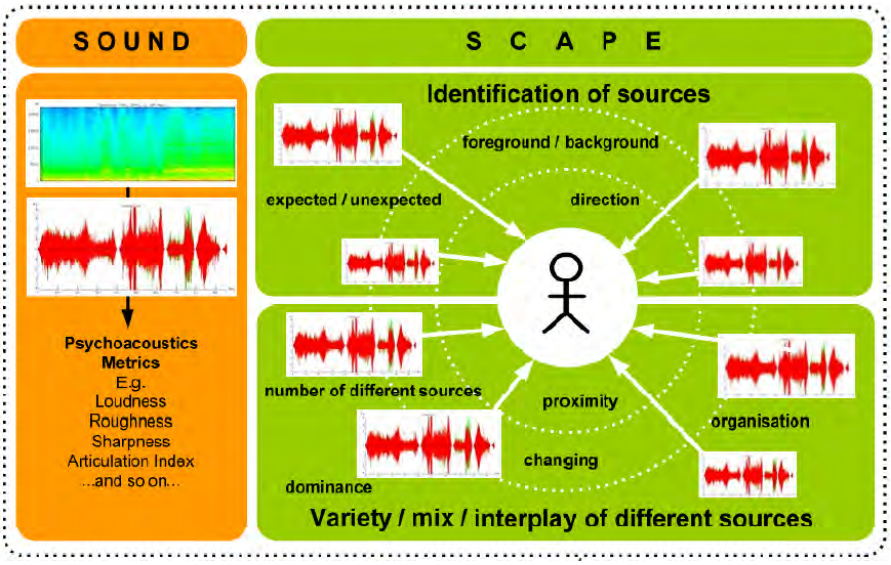

Figure 1 – The graphical explanation of the sound-scape framework [14]

In the present study, multi-lingual communication will be simulated in several multi-lingual environments where speech intelligibility is critical, such as public courts, hospitals, police stations, train stations, or airports. One of the most feasible ways for achieving this is through computer modelling and auralization techniques, where auralization is the acoustic equivalent of visualization.

The selected environments will be modelled digitally both for visualization and auralization purposes. The realistic images rendered from the computer generated models, and the simulated sonic environment will be presented to the participants. The participants will be required to subjectively assess both the intelligibility of the speech samples, and the overall sonic environment using semantic scales. The semantic differential technique uses descriptive listening and identifies the emotional meaning of words. An example response form is given in Figure 2.

	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 – An example response form of semantic differential analysis [15]

The collected data set will be used to investigate any relations between socio-lingual, cultural, and emotional aspects with the intelligibility of speech, in various multi-lingual environments. By correlating the data from both the first and the second phases of the study, an extensive database of subjective and objective measures of the speech intelligibility will be established.

3. Current Progress

In the current state of the study, the diagnostic rhyme test (DRT) word lists [10] [11] [12] and the phonemically balanced sentence lists [16] [17] [18-19] of the selected languages (English, Arabic, Chinese, and Polish) have been collected from various resources. It is important to note that no Polish DRT word list was found in the literature, therefore the Polish Digit Triplets Test (PDTT) has been used [13]. The lists have been edited to a standardised and consistent format.

The participants of the first phase of the study (the room acoustic phase) have been selected among students of Heriot-Watt University, by applying the criteria mentioned in the methodology section. For the selection of speakers, each participant was evaluated by a native speaker of that language to check that their accent is as neutral as possible. The recordings have been held in the anechoic chamber of Heriot-Watt University. A Brüel and Kjaer microphone, power supply, and a Zoom handheld sound recorder have been used to record the DRT word lists and the phonemically balanced sentence lists (Figure 3).

It is foreseen that the room acoustic phase of the study, including the data analysis, should be completed by the date of the Inter-noise 2013 conference. The results of the first phase of the study will therefore be presented at the conference.

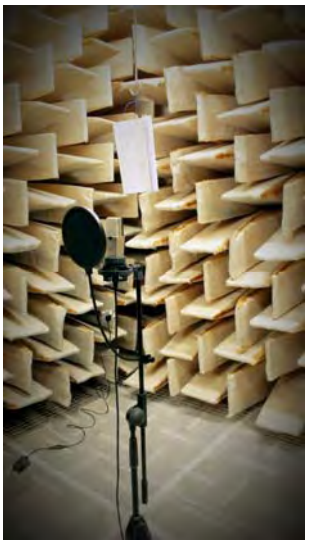

Figure 3 – The setup for voice recording in the anechoic chamber

4. Discussion

This article presented a review of the current literature on speech intelligibility and multilingual communication from two perspectives that are the room acoustics point of view and the socio-lingual point of view. It was found that there is a missing connection in the literature between these two approaches, which needs to be examined in order to have the full picture of speech intelligibility in multi-lingual spaces. The present study aims to fill this research gap.

The review of previous work suggests that room acoustics methodologies and socio-linguistics methodologies are not sufficient, on their own, for the analysis of speech intelligibility in multi-lingual spaces. Soundscape theory will therefore be used as an appropriate tool to assess social, cultural and behavioural effects on the intelligibility of speech. The room acoustic and the socio-linguistic aspects of the study will be connected by using the soundscape theory, as described in the methodology section. One of the limitations of the present study will be the use of laboratory experiments based on computer simulations and the auralization technique, rather than field experiments. Although the control of conditions will be higher in the laboratory, field experiments might be more representative of the complex interaction between social, cultural, and psychological phenomena. Future research will therefore have to consider field studies in view of complementing the present research.

Another factor which will have to be considered carefully in the present study is the sample size, especially for the soundscape phase. For the room acoustic phase, the sample size of six listeners per language is appropriate according to ANSI/ASAS3.2 [10]; however, much larger numbers will be needed for the soundscape study.

5. Conclusions

It is hypothesized that the social, cultural, lingual, and psychological background of a speaker/listener pair directly affects the intelligibility of speech. Therefore, in multilingual spaces, and especially in those where the communication is crucial (e.g. court rooms and hospitals), the intelligibility of speech should be assessed from a larger perspective than classical room acoustics. The present study aims to be a starting point for research assessing the intelligibility of speech by combining engineering methods with social sciences, from a multi-lingual perspective.

REFERENCES

[1] T. Houtgast and H. J. M. Steeneken, "A multi-language evaluation of the rasti-method for estimating speech intelligibility in auditoria," Acustica, 54, 185-199 (1984).

[2] J. Kang, "Comparison of Speech Intelligibility between English and Chinese," J. Acoust. Soc. Am., 103, 1213-1216 (1998).

[3] J. Peng, "Feasibility of subjective speech intelligibility assessment based on auralization," Applied Acoustics, 66, 591-601 (2005).

[4] J. Peng, "Relationship between Chinese speech intelligibility and speech transmission index using diotic listening," Speech Communication, 49, 933-936 (2007).

[5] J. Peng, "Chinese speech intelligibility at different speech sound pressure levels and signal-to-noise ratios in simulated classrooms," Applied Acoustics, 71, 386-390 (2010).

[6] G. Extra and K. Yağmur, "Urban multilingualism in Europe: Mapping linguistic diversity in multicultural cities," Journal of Pragmatics, 43, 1173-1184 (2011).

[7] 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, 157-186 (2012).

[8] M. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World* (Destiny Books, Rochester, VT, 1994).

[9] T. Houtgast and H. J. M. Steeneken, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," Acustica, 28, 66-73 (1973).

[10] ANSI/ASAS3.2, "Method for measuring the intelligibility of speech over communication systems," American National Standards Institute (2009).

[11] B. Boudraa, M. Boudraa, and B. Guerin, "Arabic diagnostic rhyme test using minimal pairs," Proceedings of Acoustics'08, Paris, France, 2329-2333 (2008).

[12] Z. Li, E. C. Tan, I. McLoughlin, and T. T. Teo, "Proposal of standards for intelligibility tests of Chinese speech," IEE Proceedings Vision Image Signal Processing, 147(3), 254-260 (2000).

[13] E. Ozimek, D. Kutzner, P. Libiszewski, A. Warzybok, and J. Kocinski, "The new Polish tests for speech intelligibility measurements," Signal Processing Algorithms, Architectures, Arrangements, and Applications Conference Proceedings (SPA), 163-168 (2009).

[14] W. J. Davies, R. Cain, A. Carlyle, D. A. Hall, K. I. Hume, and C. J. Plack, "The positive soundscape project: A sysnthesis of result from many disciplines," Proceedings of Inte-rnoise 2009, Ottawa, Canada (2009).

[15] J. Kang and M. Zhang, "Semantic differential analysis of the soundscape in urban open public spaces," Building and Environment, 45, 150-157 (2010).

[16] D. N. Kalikow and K. N. Stevens, "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am., 61(5), 1337-1351 (1977).

[17] M. Boudraa, B. Boudraa, and B. Guerin, "Twenty lists of ten Arabic sentences for assessment," Acustica, 86, 870-882 (2000).

[18] Q. J. Fu, M. M. Zhu, and X. S. Wang, "Development and validation of the Mandarin speech perception test," J. Acoust. Soc. Am., 129, EL267-EL273 (2011).

[19] E. Ozimek, "Speech recognition tests for two additional languages: Polish and French," FP6–004171 HEARCOM, Hearing in the Communication Society, Deliverable D7-1-b (2007).