# CHARACTERISING THE SOURCE: AN IMPORTANT ASPECT FOR MORE ACCURATE AND MEANINGFUL SPEECH INTELLIGIBILITY AND PRIVACY ASSESSMENT

FF Li Acoustics Research Centre, University of Salford, Greater Manchester, UK

## 1 INTRODUCTION

Speech intelligibility and privacy depend upon three distinctive elements, namely acoustic characteristics of sources, transmission channel effects and talker-listener matching. Transmission channel effects might be quantified by tailored physical properties of the channels and have been studied extensively in the past few decades: speech transmission index (STI) and articulation index (AI) are typical examples. Source characteristics have been neglected to some extent and were not systemically studied in the past. This paper presents some of the results and thoughts from a large scale study into statistical features of speech levels or more precisely vocal effort levels in anechoic conditions. In addition, some results from a pilot investigation into the relations between speech perception and the variations of acoustic phonetic components are discussed. And last but not least, the paper should shed a light on the appropriate speech levels when setting up speech intelligibility and privacy tests.

## 2 SOME PROBLEMS IN SPEECH INTELLIGIBILITY AND PRIVACY ASSESSMENTS

### 2.1 A Sandwich Model for Speech Comprehension

Speech intelligibility and privacy are important concerns in the design of built and human environments, such as classrooms and lecture theatres, where lecturers' voices need to be clearly delivered, transportation hubs where the clarity of Tannoy broadcast is important, and offices or meeting rooms in which intended speech communication should be intelligible but neighboring conversations often need to be kept private. Over the past few decades speech intelligibility and privacy have been one of the research foci of building and architectural acoustics, accumulating a good collection of assessment methods and a fairly large knowledgebase. Strictly speaking, intelligibility and privacy of speech should be referred to as the amount of information of speech of a talker that can be decoded by a human listener via an acoustic or electronic transmission channel. Under such a definition, clarity and intensity of the original articulation of the talker, quality of speech transmission channel and accent/dialect matching between the talker and the listener can all affect speech intelligibility or privacy, and hence complicating the case.

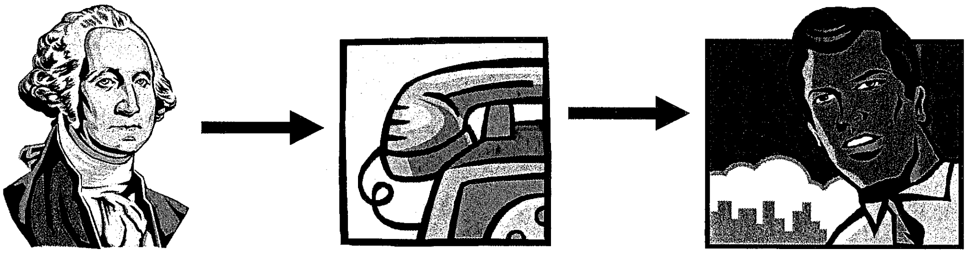


Figure 1. A sandwich model for speech intelligibility/privacy, Loudness and clarity of the source, Degradation caused by the transmission channel, and apprehension of listeners (including talker listener matching).

There are good reasons to single out a transmission channel and define its own "intelligibility" measure as a quality index. In building and architectural acoustics, intelligibility of a space or a system is usual deemed as physical or objective measures independent of the talker and listener, but best correlated to subjective intelligibility in a general sense. This is based upon certain assumptions, i.e. a typical speech source and source level, typical talker-listener matching. In addition to simplification of the complicated scenarios, there are other tangible benefits to do so: Physical parameters of transmission channels relate acoustics and intelligibility to building designs and facilitate the diagnoses of acoustic problems. In essence, all objective intelligibility/privacy assessment methods examine signal to noise ratios and channel distortion (frequency distortion and/or reverberation) in critical sub-bands. The underlying mechanism that makes intelligibility measures objective parameters is the assumption of a "normal" speech level and spectrum, and a typical talker-listener matching. So the two remaining issues here are proper standardization of the talker/source and accurately modeling speech perception process. The former is apparently more important for two distinctive reasons: (1) with limited knowledge and available tools to accurate model speech comprehension, the use of human listeners is still the ultimate solution for intelligibility tests. A good number of listeners are often used, but only a small number of talkers or recorded speech materials are involved in the testing. To set up an intelligibility test, the use of representative sources is crucial. (2) even if in the cases of objective assessment, e.g. setting up an STI measurement, source level and directivities need to be representative to make the measurement meaningful.

In telecommunications industry, there are also demands to assess usability or Quality of Service (QoS) of voice communications channels in terms of perceived speech quality or intelligibility. In parallel with acoustics research, a set of related but somehow diverse assessment regimes were developed. Perceptual Evaluation of Speech Quality (PESQ) is a typical example. At the first glance, it seems that the system gain can often be increased at the user end by tuning the volume up, nonetheless modern voice telecommunications systems are subject to digitization, non-linear codecs, acoustic and electronic noises. Vocal levels at the speech acquisition end do have a significant impact on the signal to noise ratio. In audiology and hearing aids research, it is also important to understand how loud people normally talk in various settings. In building sound insulation for speech privacy, how loud people normally talk and the variation of the loudness play an important role in the determination of necessary sound reduction.

All above examples justify the necessity of a knowledgebase about speech levels and directivities, their distribution and variation in various speech communication settings. Moreover as an engineering approach to the problems, a standardized "artificial talker" would perhaps make measurement easier and more repeatable and reliable.

This paper will identify what are available in the existing knowledgebase, present some results from the author's recent work and suggest further research in this field to fully establish a "standard talker" for speech intelligibility and privacy assessment purpose.

## 2.2 Acoustic Characteristics of Human Talkers in the Literature

Vocal Effort Level (VEL) also quoted as speech intensity level is often used to quantify how loud a talker talks in a particular communication setting. It is defined as an A-weighted or un-weighted equivalent continuous sound pressure level (SPL) of speech. In this text, the vocal effort level is more specifically defined as the on-axis A-weighted sound pressure level, or un-weighted 1/3 octave band SPL measured at 1 metre from the lips of a human speaker under anechoic conditions. The VEL is a key variable for the prediction of intelligibility of speech communications systems. It is also a critical reference value for the acoustical design to achieve desired speech intelligibility and/or privacy. In the light of its importance, the interest of quantifying VELs in various speech communication settings started from the era when communications systems emerged and measurement techniques became available. Early establishment of a small knowledgebase about the speech level was based on a series of scattered research activities taken place from the 20s to 50s. Further studies in 60s and early 70s made some enrichment to the body of knowledge. These early studies suffered from small number of samples, limited measurement techniques and less well-defined measurement conditions.

Crandall and Mackenzie made the first endeavor in defining "normal" speech level", but free-field microphone calibration was not available in the 1922 study[1]. In 1940 Dunn and White established the "normal" vocal effort level dataset in terms of long-term RMS and 1/8-second peak sound pressure levels at 30 cm from lips under anechoic conditions[2]. But the experiments used only 6 male and 5 female subjects. In 1947 French and Steinberg[3], and Benson and Hirsch[4] in1953 replicated the findings by Dunn and White with a larger number of subjects. Vocal effort level specified in the classical text by Beranek published in 1947[5] was largely based on the above studies. Thus the early days' "standard" speech level of 60-65 dB (long term RMS) at 1 metre from a male speaker's lips was established. Brandt et al. quantified the relation between changes in loudness and speech effort in 1969[6]. In 1976, Brown further argued that "comfortable effort level", as often instructed in speech related experiments, was not sufficiently constant[7]. All these authors called for a more detailed knowledgebase about vocal effort level distribution under stipulated efforts (e.g. casual, soft, normal, loud and shout) and more reliable statistical results from a larger number of subjects. Alongside the study of the "standard" or "normal" vocal levels, efforts were made to quantify Lombard effect – the phenomenon first described by Lombard in 1911 that speakers tend to increase their voice levels when the ambient noise increases[8]. Klumpp[9] and Gardner[10] made important contributions in this area prior to the publication of the Pearsons report in 1977.

Pearsons, Bennett, and Fidel were commissioned a large-scale research into the VELs under both controlled laboratory conditions and in real-life settings[11]. This is an important milestone. The report published in 1977 is often deemed as the "definitive" reference for English language vocal effort levels under anechoic conditions to date. No other anechoic chamber based study into the vocal effort level (English language) of a similar scale was documented in the literature. Given its importance, 21 year later Olsen published a summary of the report in 1998 as a journal paper[12]. (One most recent large-scale study in 2004 by Corthals[13] obtained speech levels of 400 normal subjects reading the "Dutch rainbow passage". However the measurements were not carried out in an anechoic chamber and results may not represent English language speech levels.) Given the larger number of subjects used, the better controlled and calibrated laboratory conditions and more up to date equipment, statistical results of the anechoic chamber measurements from Pearsons report should override the ones published prior to 1977.

The post 1977 era has seen some more research confirming the results from Pearsons' study and enriching the body of knowledge by adding more data and details, for example, directivity information. Studies were gradually directed to VEL adaptation and variation. Major contributions include detailed directivities of sound field around human talkers, vocal effort levels from more field measurements, refined coefficients of Lombard effect, and adaptation of vocal effort levels due to other variables such as communication distance. Chu and Warnock (2002) reported detailed directivity information of human speakers. Bozzoli and Farina measured speech levels in cars[15, 16]. Warnock[17], Bradley[18], Gover and Bradley[19], Gover and Bradley[20], Bradley and Gover[21] presented results from large-scale studies into speech levels in offices and meeting rooms. Navarra and Pimentel[22] and Hodgson et al. [23] measured speech levels in food courts and dining spaces. Variation of speech levels due to age, communication distance and ambient noise levels (Lombard effect) were studied by many authors. Hodge et al.[24], Huber et al.[25], Lienard and Benedetto[26], Brungart and Scott[27], Giguere et al.[28], and Pick et al.[29] all made important contributions.

## 2.3 Limitation of Existing Knowledgebase

Literature review showed continuous efforts over the past 90 years to characterise speech sources in terms of its intensity and directivities. Individual studies were reported by researchers from diverse fields, from dissimilar acoustic environments and with different purposes. This makes the comparison of their datasets difficult. There is no single standard definition for the vocal effort level, speech level, or speech intensity. The author advocates the use of the SPL at 1 meter on axis from talker's lips corrected to the sound pressure level obtained in anechoic conditions. The usefulness of the anechoic equivalent data is that they can be used to predict speech levels in various acoustic conditions taking into account the acoustic conditions and appropriate Lombard effect correction. It is known that auditory feedback can significant affect speech levels, measurements taken in an anechoic chamber can be used as a reference level, but they are acquired in an unnatural acoustic condition. It is therefore important to take into account vocal effort variations in environments.

The most comprehensive study of vocal effort level is probably the one carried out by Pearsons et al. The study was carried out in the United States 30 years ago. Although the controlled laboratory conditions are unlikely to change over time, the real-life settings 30 years ago in America may not represent the current reality in the UK. For example, the change of sizes of public venues may cause changes to background noise levels. Modern public transportation vehicles and road conditions in the UK are not identical to those in the US 30 years ago, and noise profiles can be different. Moreover, the stipulated vocal efforts used in Pearsons experiments are not clearly explained, it is speculated that subjects may interpret them differently.

Subjects involved in the Pearsons study speak American English. It is unknown whether there is a vocal effort discrepancy between British and American accents. Directivity data and vocal effort levels in offices published by National Research Council Canada were measured from a population of circa 90% English speakers and 10% French speakers. Whether there is a vocal effort level difference between English and French speakers is again unknown.

Lombard effect and Lombard slope is a useful and arguably robust prediction tool for vocal effort levels in noisy environment. However, there is a relatively large divergence in Lombard coefficients reported by different authors. Human perception and adaptation to environments typically feature a certain level of non-linearity. A simplex linear or piece-wise linear model for Lombard effect seems coarse and inadequate. More subtle non-linear models might be beneficial.

Several authors implied that the “normal speech levels” assumed by the current ANSI and ISO standards for speech intelligibility were too high to represent actual speech levels in certain real-life settings. For example, a low speech level of 50.2 dB(A) was suggested by Bradley (2003) for the assessment of intelligibility in open offices. Private conversations might have speech levels even lower than those of casual conversations. The data of such speech levels are crucial in the assessment of the viability of certain speech transmission systems. Unfortunately, statistical data about speech levels below casual conversations are not available from the literature.

Directivities of talkers again require more work. Several authors reported somewhat different data. There are also papers reporting large directivity discrepancies amongst the commercially available HATS and real human talkers: Responses of artificial mouth simulators of B&K HATS 4128, B&K 4227 and Head Acoustics HMS II.3 were compared against human talkers and non trivial discrepancies have been noted[30].

# 3 RECENT WORK

## 3.1 Vocal effort level of British English Speakers in anechoic conditions

To verify Pearsons anechoic vocal effort levels, similar experiments were carried out with 50 native British English speakers. The experiments aimed to (1) identify if American and British accents would affect the vocal effort levels, (2) extend the database to include “hushed” speech levels, and (3) mitigate the deviations in Pearsons dataset by giving clearer descriptions with examples.

Recordings were made in an anechoic chamber to determine the average vocal effort levels and spectra of adult males and females, using a 01dB-Metravib NetdB 12 kit, which allows for the simultaneous recording of multiple microphones. Five Omni-directional microphones (G.R.A.S. Type 26CA) were used: (1) at a 1m distance in front of the talker, (2) at 0.5m in front, (3) at 1m to the left, (4) at 1m to the right and (5) at 1m behind. All microphones were placed at the same height as the subjects’ mouths. Voices from 50 subjects with an average age of 30 years were recorded. No subjects reported any hearing or speech impairments.

Subjects were instructed to repeat the sentence 'Joe took father's shoe bench out, she was waiting at my lawn' three times with five different vocal efforts, namely hushed, normal, raised, loud and shouted. This particular phonetically balanced short sentence was chosen as it was used in the work by Pearsons, which would make the comparison between the two studies straightforward and robust. The second reason is due to its short length, as higher vocal efforts can be difficult to sustain for a long time and can risk damaging the vocal cords. In Pearsons' experiments, only very brief descriptions of the vocal efforts were given, with no examples of typical scenarios attached to each of the stipulated vocal effort labels. For example, the entire description for 'shout' was simply 'speak at a shouted level'. It was speculated that the lack of detailed descriptions could lead to ambiguous interpretations and subsequently a lager spread of data. In the current study, detailed descriptions (as shown in Table 1.) of each vocal effort were given with typical scenarios as examples to the subjects prior to starting the recording.

Table 1: Descriptions of vocal effort labels

| | |
|---|---|
| Hushed | This is the quietest level of voiced speech – just louder than whispering. Typically this speech level would be used in intimate situations where privacy is an issue; for example talking in a library so as not to disturb others, or talking in a doctor's waiting room. |
| Normal | This is a normal, everyday conversational speech level. Typically this speech level would be used in small quiet room with no more than two or three people involved in the conversation. |
| Raised | This speech level would typically be used when addressing multiple people in a medium sized room, or when in the presence of background noise such as a car or train. |
| Loud | This speech level would typically be used when issuing commands or attracting attention, expressing anger or assertiveness. A situation where this speech level would be used is when addressing a large number of people in a very large room without the aid of amplification. |
| Shout | This is the loudest possible speech level one can manage, without straining or hurting the vocal cords. |

Average vocal effort levels from the 1m microphone are presented in Table 2. All results are rounded to the nearest decibel. For a comparison purpose, Pearsons' results are shown in Table 3.

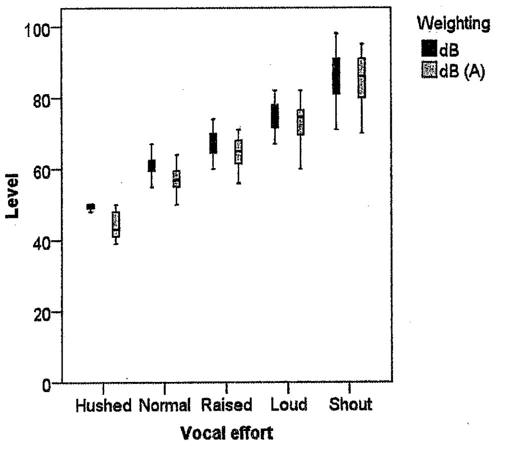
Table 2: Mean vocal effort levels in dB and dB(A) in anechoic conditions, measured at 1m. Unweighted levels in [ ], Standard deviations in ( ).

| | ***Hushed*** | ***Normal*** | ***Raised*** | ***Loud*** | ***Shout*** |
|---|---|---|---|---|---|
| **Males** | 47 [52] (2) | 58 [62] (3) | 67 [69] (5) | 76 [77] (6) | 89 [89] (6) |
| **Female** | 46 [49] (2) | 56 [58] (3) | 64 [66] (4) | 70 [71] (4) | 82 [82] (3) |

Table 3: Results from Pearsons *et al.* Format of results as is in Table 2. Format of results as is in Table 2.

| | **Casual** | **Normal** | **Raised** | **Loud** | **Shout** |
|---|---|---|---|---|---|
| **Males** | 52 [56] (4) | 58 [61] (4) | 65 [68] (5) | 76 [77] (6) | 89 [89] (7) |
| **Females** | 50 [54] (4) | 55 [58] (4) | 63 [65] (4) | 71 [72] (6) | 82 [82] (7) |

Figures 2 and 3 show the statistical distribution of each vocal effort level in terms of un-weighted and A-weighted sound pressure levels for male and female talkers. Both groups show an increase in between-subject variation as vocal effort increases, apart from the female shouted levels which decrease in standard deviation. Comparison between ours and Pearsons' results shows that the standard deviations are lower from our results, which might be attributed to the more detailed vocal effort labels used.



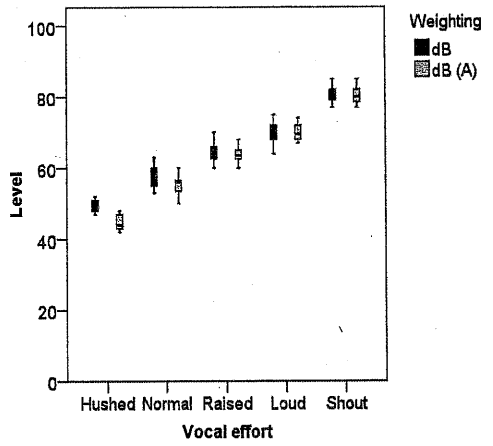




Figure 2: Average vocal effort levels for males Figure 3: Average vocal effort levels for females

For males and females respectively, the difference between A-weighted and unweighted levels is 5 dB and 3 dB for hushed speech, 4 dB and 2 dB for normal speech, 2 dB for raised speech, 1 dB for loud speech and 0 dB for shouted speech. Male speech is consistently louder than female speech, and the difference increases from 1 dB(A) to 7 dB(A) as vocal effort increases from hushed through to shouted speech. As expected, hushed speech shows consistently the lowest average level, approximately 11 dB(A) lower than normal speech for both male and female talkers. Raised speech is 9 dB(A) and 8 dB(A) more intense than normal speech for the male and female groups respectively. Loud speech is 9 dB(A) higher than raised speech for males and 6 dB(A) for females. Shouting speech gives the highest levels, with a 13 dB(A) increase from raised speech for males, and a 12 dB(A) increase for females. Between the two extreme ends of the vocal effort scale, the hushed and the shouted, there is a 42 dB(A) dynamic range for males and a 36 dB(A) for females.

Figures 4 and 5 show averaged one-third octave band speech spectra for male and female talker groups at each different vocal effort level, from the 1m microphone. More details about speech levels and spectra around the talkers can be found in the refrence[31].

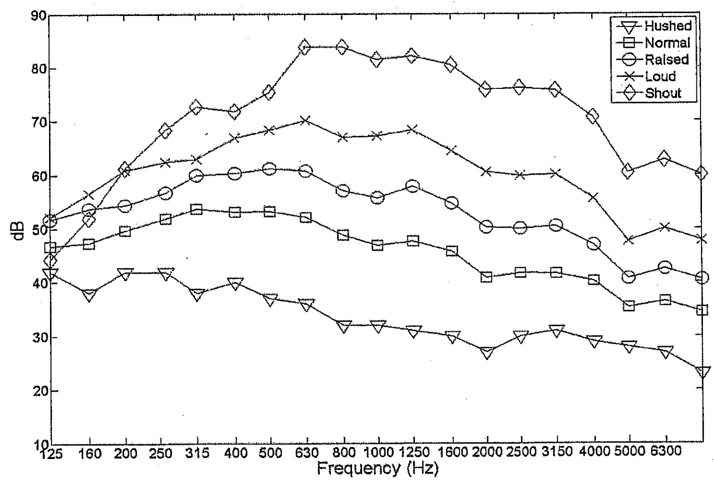



Figure 4: Speech spectra for male talkers at different vocal effort levels

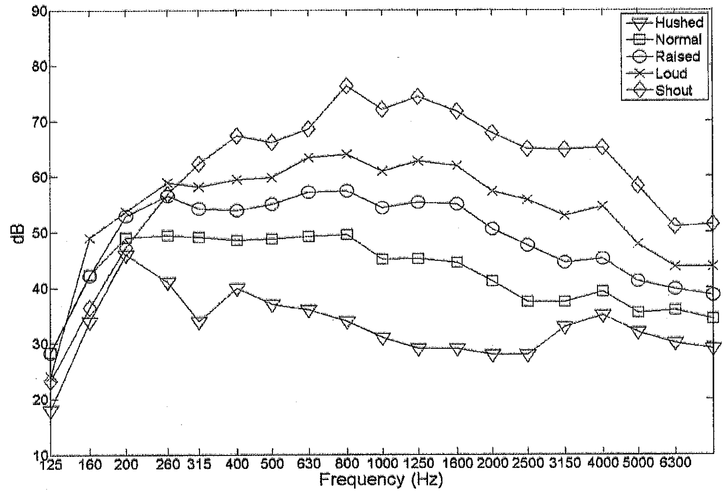



Figure 5: Speech spectra for female talkers at different vocal effort levels

## 3.2 Acoustic phonetic feature variations under diverse vocal efforts

Vocal effort variation results in changes of prosody and acoustic phonetic features of speech, not just overall energy levels. Simply increasing the volume of a whispered recording does not make it sound like a shouted one. Pilot studies were carried out to identify how vocal effort was perceived and the relations between vocal effort and clarity of speech by listening testing.

Subjects were asked to read five phonetically balanced nonsense wordlists at five specified vocal efforts detailed in Table 1. A total of 4340 uttered words were collected to form a corpus: 816 hushed, 864 normal, 882 raised, 882 loud and 896 shouted words. A hundred isolated CVC words selected randomly from the corpus were used for the listening tests. The overall sound level was equalised across all stimuli to the same $L_{eq}$. This was to remove the variance in volume and ensure that subjects only use the subtle phonetic aspects of the speech to complete the task. Stimuli were presented over an Apple Macintosh computer via reference headphones using Praat experiment-mfc software. During the first 5 trial tests, listeners were able to adjust the volume to a comfortable level, which then remained unchanged throughout the testing. The experiment interface is shown in Figure 6. The top row collected the responses for 'perceived vocal effort level, the bottom one collected the perceived clarity of articulation. A 'replay' button was positioned under the bottom row. The number of available replays was set to be unlimited. Subjects were given detailed descriptions of each vocal effort label. The following description of clarity of articulation was given: 'how well the speaker enunciates the word; how defined and clear the articulation is. Good clarity of articulation is where each individual speech sound is easily heard and recognised, whereas poor clarity of articulation would be if the speaker mumbles or it is difficult to recognise what they are saying'. Twenty-five normal hearing native English-speaking subjects participated in the listening experiments; 14 males and 11 females. The average age of subjects was 35. No subjects had any experience in speech transcription or similar work. Listening tests took place in a quiet environment free from any potential disruptive background noise. A total number of 2500 responses were collected for further analysis.

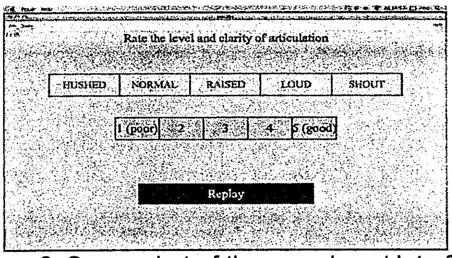



Figure 6. Screenshot of the experiment interface

The relationship between actual vocal efforts (the stipulated level for the original recordings) and the perceived vocal efforts (the responses from the listening tests) was investigated and statistical results across all listeners are shown in Figure 7. In the boxplot, the thick black lines represent the mean score, the filled box represents the interquartile range and the horizontal lines at the end of the 'whiskers' represent the minimum and maximum values. Asterisks represent suspect outliers in the hushed and normal cases. Results clearly show for relatively loud speech, the actual vocal effort can be appreciated from subtle changes in acoustic phonetic features, regardless of the playback volume. To some extent, this result confirms turning the volume of normal speech up does not mimic shouted vocal effort.

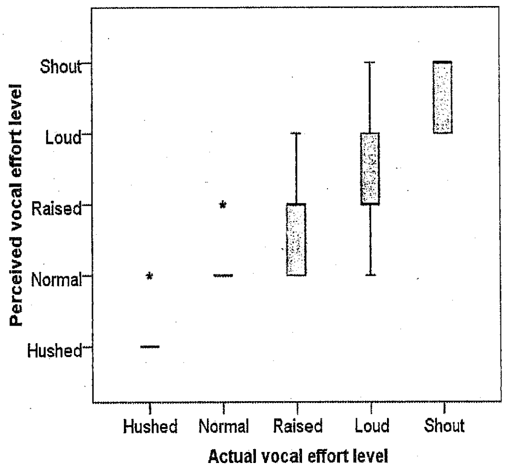



Figure 7. Relations between actual and perceived vocal effort.

Comparisons were also made to analyse the correlation between vocal effort and perceived clarity of articulation. The results are shown in Figure 8.

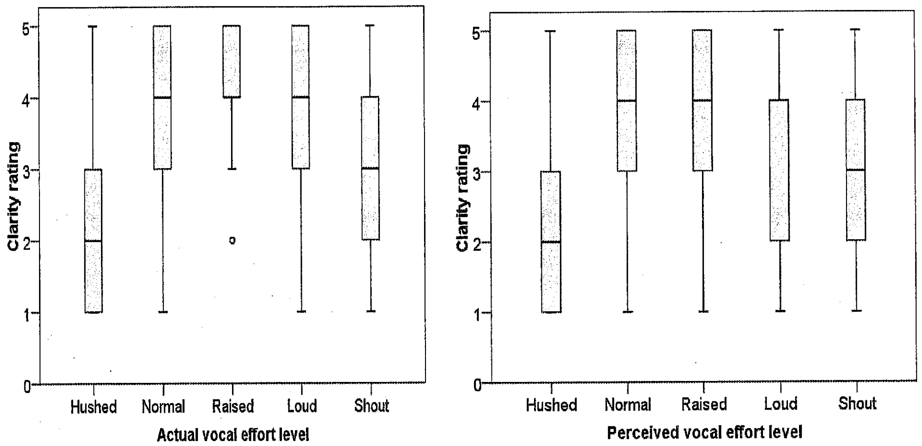



Figure 8. Relations between vocal effort and clarity.

Figure 8 shows the average clarity ratings against the actual vocal efforts (the ones stipulated during the recordings, left) and perceived vocal efforts (right). Mean clarity ratings for each stipulated vocal effort were: hushed = 2.19, normal = 4.04, raised = 4.12, loud = 3.65 and shout = 2.90. Standard deviations were: hushed = 1.18, normal = 1.01, raised = 0.87, loud = 1.12 and shout = 1.21. Post-hoc tests (Tukey's) revealed that normal and raised were the only pair not to be significantly different from one another ($p < 0.05$). Female speakers were rated slightly higher in clarity than male speakers. Females scored an average rating of 3.71 whereas males scored 3.16. The results show that the words said at normal and raised levels were deemed to have the best perceived clarity of articulation. The words said at a hushed level were rated the lowest. The pattern of the graph suggests that the 'ideal' speech level for the best clarity of articulation is around the normal to raised level. As vocal effort increases from the raised level, perceived clarity of articulation decreases. This implies that excessively raising one's voice does not necessarily result in an increase in clarity. Whereas raising voice is typically associated with a desire to improve intelligibility (against background noise), extreme vocal efforts such as loud and shout can actually have the opposite effect by decreasing perceived clarity.

# 4 CONCLUDING REMARKS AND FUTURE WORK

More accurate characterisation of speech sources is a step towards more reliable subjective and objective assessments of speech intelligibility and privacy. The speech level reported here can be used as a reference when setting up a speech intelligibility or privacy test. A device that can completely replicate the speech from a typical human talker in terms of its spectrum, dynamic range and directivity does not exist so far, partly because of the lack of statistical data, and partly because of technical challenges to reproduce phonetically dependent directivity patterns of a real speech source. More research is needed to fully establish a knowledgebase of statistical distribution of vocal effort levels, their variations in diverse communication settings, directivities and more reliable Lombard coefficients. With the knowledge, beamforming techniques and DSP algorithms, an artificial talk might be possible to completely simulate a human talker.

# 5 REFERENCES

1. Crandall, I. B. and MacKenzie, D. "Analysis of the Energy Distribution in Speech," Physical Review, Vol. 19, Issue 3, 221-232, 1992.
2. Dunn, H. K. and White S. D. "statistical measurements on conversational speech," J. Acoust. Soc. Am. Vol. 11, 278-288, 1940

3. French, N. R. and Steinberg J. C. "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. Vol. 19, 90-119, 1947
4. Benson, R.W. and Hirsh I. J., "Some variables in audio spectrometry", J. Acoust. Soc. Am. Vol. 25, 499-505. 1953.
5. Beranek, L. L., Acoustics, McGraw-Hill, 1954.
6. Brandt, J. F., Ruder, K. F. and Shipp, T. Jr., "Vocal Loudness and Effort in Continuous Speech," J. Acoust. Soc. Am. Vol. 46 (6B), 1543-1548,1969.
7. Brown, W. S. Jr., Murry, T. and Hughes D., "Comfortable effort level: An experimental variable," J. Acoust. Soc. Am. Vol. 60 (3), 696-699, 1976.
8. Lombard, E., "Le signe de l'elevation de la voix," Ann. Maladie Oreille Larynx Nez Pharynx, vol. 37, 101–119, 1911.
9. Klumpp, R. G. and Webster, J. C., "Physical measurements of equally speech-interfering Navy noises," J. Acoust. Soc. Am. 35, 1328-1338., 1963.
10. Gardner, M. B., "Effect of Noise, System Gain, and Assigned Task on Talking Levels in Loudspeaker Communication," J. Acoust. Soc. Am. Vol. 40 (5), 955-965, 1966.
11. Pearsons, K.S., Bennett, R.L., and Fidel, S., "Speech Levels in Various Noise Environments,' EPA-6001-77-025, U. S. Environmental Protection Agency, May 1977, Available from National Technical Information Services as No. PB-270 053, 1977.
12. Wayne O. Olsen, "Average Speech Levels and Spectra in Various Speaking/Listening Conditions- A Summary of the Pearson, Bennett, & Fidell (1977) Report," American Journal of Audiology, Vol.7, 21-25, 1998.
13. Paul Corthals, "Sound pressure level of running speech: percentile level statistics and equivalent continuous sound level," Folia Phoniatrica et Logopaedica, Vol. 56 No. 3, pp. 170-181, 2004.
14. Chu, W. T. and Warnock, A.C.C., "Detailed Directivity of Sound Fields Around Human Talkers," report number IRC-RR-104, National Research Council Canada, January 2002.
15. Bozzoli, F. and Farina, A., "Measurement of Speech Transmission Index Inside Cars Using Throat-Activated Microphone and Analysis of Its Correlation with Drivers' Impression" 117th AES convention, Preprint Number: 6307, San Francisco, USA, 2004.
16. Bozzoli, F. and Farina, A., "Measurement of Active Speech Level Inside Cars using Throat-activated Microphone" 116th AES convention, Preprint Number: 6004, Berlin, Germany, 2004.
17. Warnock, C. C., "Acoustical privacy in the landscaped office," J. Acoust. Soc. Am. 53, 1535–1543, 1973
18. Bradley, J. S., "A Renewed look at open office acoustical design," Report NRCC-46399 ,A version of this document is also published in Inter-Noise 2003, Seogwipo, Korea, pp. 1-8, 2003.
19. Bradley, J. S. and Gover, B. N., "Speech and Noise Levels Associated with Meeting Rooms, IRC, National Research Council Canada, Research Report, IRC RR-170, March, 2004.
20. Gover, B. N. and Bradley, J. S., "Measurement of Architectural Speech Security of Closed Offices and Meeting Rooms," National Research Council Canada (NRCC) technical report 48206, a version has also been published at 119th Audio Engineering Society Convention, New York, NY., Preprint Number: 6529, Oct. 7, 2005.
21. Bradley, J.S., Gover, B. N., "Measurement of Sound Transmission from Meeting Rooms," IRC Research Report, IRC RR-220 (Institute of Research in Construction, National Research council Canada), March, 2006.
22. Navarra, M. P. N. and Pimentel R.L., "Speech interference in food courts of shopping centres," Applied Acoustics 2007, 68(12), 364-375, 2007
23. Hodgson M., Steininger, G. and Razavi, Z., "Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments," J. Acoust. Soc. Am. 121 (4), 2023-2033, 2007.
24. Hodge, F. S. Colton, R. H. and Kelley, R. T., "Vocal Intensity Characteristics in Normal and Elderly Speakers," Journal of Voice Vol. 15, No. 4, pp. 503–511, 2001.
25. Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A. and Johnson, A., "Formants of children, women, and men: The effects of vocal intensity variation," J. Acoust. Soc. Am. 106 (3), Pt. 1,1352-1542, 1999.

26. Lienard, J. S. and Benedetto M. G. D., "Effect of vocal effort on spectral properties of vowels," J. Acoust. Soc Am 106, 411-422, 1999.
27. Brungart D. S. and Scott, K. R., "The effect of production and presentation level on auditory distance perception of speech," J. Acoust. Soc Am 110 (1), 425-440, 2001.
28. Giguere, C. et al., "Quantifying the Lombard effect in different background noises" J. Acoust. Soc. Am., Vol. 120 (5), 3378, 2006.
29. Pick, H. L. Jr., Siegel, G.M. and Fox, P.W., "Inhibiting the Lombard effect" J. Acoust Soc. Am. 8 5 (2), 894-900, 1989.
30. Halkosaari, T. and Vaalgamaa, M., "Directivity of human and artificial speech," Workshop on Wideband Speech Quality in Terminals and Networks: Assessment and Prediction, 8th and 9th June 2004 - Mainz, Germany, 2004.
31. Cushing I. R., Li F. F., Cox T. J., Worrall K., Jackson T., "Vocal effort levels in anechoic conditions," Applied Acoustics Vol. 72 pp. 695–701, 2011
32. Montgomery, A.A., Prosek, R.A., Walden, B.E & Cord, M.T., The effects of increasing consonant/vowel intensity ratio on speech loudness. Journal of Rehabilitation Research and Development 24 (4): 221 – 228, 1987.
33. Freyman, R.L., Nerbonne, G.P., The importance of consonant-vowel intensity ratio in the intelligibility of voiceless consonants. Journal of Speech and Hearing Research 32 (3): 524 – 535, 1989.
34. Shoba, N.H., Thomas, T.G & Subbarao, K., Influence of consonant-vowel intensity ratio on speech perception for hearing impaired listeners. Image and Signal Processing 5: 342 – 346, 2008.