Edinburgh, Scotland
EURONOISE 2009
October 26-28

Measuring and mapping soundscape speech intelligibility

William J Davies[a], Peter Z Mahnken, Phill Gamble
Acoustics Research Centre, University of Salford, M5 4WT, United Kingdom

Chris Plack
Human Communication & Deafness Division, Manchester University, United Kingdom

ABSTRACT

More than one factor analysis of soundscape perception has concluded that speech communication is a significant contributor to the overall perceived quality of urban soundscapes. While not everything that is perceptually significant in real soundscapes can be quantified it seems likely that speech intelligibility can. There is a large literature on intelligibility focussed on indoor settings. This paper reports on an attempt to use extended speech intelligibility index (ESII) to characterise time-varying speech intelligibility in real outdoor soundscapes. The experiment arose out of the Positive Soundscape Project, a large interdisciplinary project. The relationships between SII, subjective speech intelligibility and subjective speech quality will be explored as functions of signal-to-noise ratio. Using the ESII technique to map soundscapes as one possible sound quality indicator will be demonstrated. Potential implications for the rational planning of soundscapes will be outlined.

1. INTRODUCTION

The current paradigm of soundscape research in the acoustics community is centred on listening. The earliest soundscape works in acoustic ecology urged us to pay more attention to our sound environment, to listen, capture and preserve our soundscapes[1]. Useful ideas about listening have also been drawn from room acoustics, where the experience of listening to music in a concert hall was characterised as a multidimensional psychological process as early as 1971.[2] Soundscape theorists have proposed different types of listening,[3] though these have been rather under-researched to date. A large number of soundscape research reports are based on the soundwalk, where participants are led on a silent walk through the soundscape, listening intently.[4] Much has been learnt from the listener paradigm, and there is now some agreement on the most important perceptual and cognitive factors involved in listening to a soundscape.[5, 6] It is clear that modelling listening can tell us much about emotional response to a soundscape and how to make places sound better.

However, the strong focus on listening in current research does rather overlook the important effect of soundscapes on speech communication. Instead of passively (even if intently) listening to a soundscape, we can also contribute to it ourselves with our speech. Urban soundscapes are, most of the time, the background setting for conversations. While we want our soundscapes to sound positive, to be relaxing, vibrant and convey meaning, we also need them to act as a supportive context for speech. The Positive Soundscape Project has recently found some evidence that when people think and talk about soundscapes in their own words, that the concept of speech communication is important. Participants on sound walks and focus groups evaluated some soundscapes in terms of the ability to hold a conversation[7], and background speech hubbub was found to be a component of one kind of positive soundscape.[5]

[a] Email address: w.davies@salford.ac.uk

One previous factor analysis of soundscape perception[6] has found that 'communication' is one of four significant factors. There is, of course, also a large literature demonstrating that people with a hearing impairment often experience extreme difficulty in understanding speech with other sound sources present.

This raises the question of how we might assess and model the effect of a soundscape on speech communication. Most existing work on speech with background sound uses simple constant noise signals, quite unlike a real soundscape with its complex masking and sounds which distract attention. The work reported here is a small trial which explored how speech intelligibility in an urban soundscape could be assessed or predicted. It uses an adaption of the standard metric speech intelligibility index[8] proposed by Rherbergen and Versfeld[9] which extends SII for non-stationary noise. This is applied in this paper to the situation where the 'noise' is a binaural recording of a real urban soundscape.

2. METHOD

There were two elements to this investigation, both using the same soundscape recordings. One was predicting speech intelligibility using SII and the second was making subjective measurements of intelligibility, clarity and quality using twenty listeners.

A. Prediction of speech intelligibility

SII is a method of estimating speech intelligibility that is based on Articulation Index and estimates the average overall understanding of speech information by a listener. It uses a scale of 0.0 (unintelligible) to 1.0 (perfect intelligibility). Rherbergen and Versfeld proposed breaking the SII calculations into smaller time windows based on frequency and showed that this increased the predictive accuracy of SII estimates for various types of noise signal. This mmethod is now commonly referred to as extended SII (ESII).

Based on the code and methods of [9] and [10], a MATLAB code for calculating ESII of speech files was used. Analysis was performed over the 200 low context speech samples from the SPIN test set to determine five target ESII values with relative gains calculated for each noise sample (based on a unity gain for all speech samples). Analysis using these gain values across the sample set calculated ESII values within a standard deviation of 0.05 and SNR values within a standard deviation of 1.5dB.

Figures 1 and 2 show examples of the ESII calculated for a speech sample set in a stationary white noise signal. Generally, the ESII predicts good speech intelligibility at instants when the speech signal has more energy than the noise signal. The ESII sometimes decays more slowly than the speech signal due to the variable window length. At instants where there is no speech output, the ESII falls to zero.

The ESII model described thus far predicts intelligibility for normal hearing. Predictions were also made for hearing impaired listeners, using the temporal window model of the ear developed by Plack et al.[11] The temporal window model is a model of auditory temporal resolution and temporal aspects of masking. The model includes a simulation of cochlear frequency selectivity based on the dual-resonance nonlinear filterbank of Meddis et al.[12] The parameters of the filterbank are derived from fits to recent forward masking data. For each frequency channel, the output of the filterbank is squared, and then convolved with a linear intensity-weighting function (the temporal window), with a time constant of approximately 10 ms. The temporal window acts as a leaky integrator, and simulates temporal sluggishness in the auditory pathway. This relatively simple model can account for a wide range of temporal masking phenomena. After quasi-instantaneous cochlear compression, the auditory system seems to behave as a linear energy integrator with respect to many aspects of temporal masking. The temporal window model allows predictions of time-varying ESII for many different

types of hearing impairment. Results are shown here for two types only: moderate hearing loss and 'severe flat,' a 'dead' region above 2 kHz.

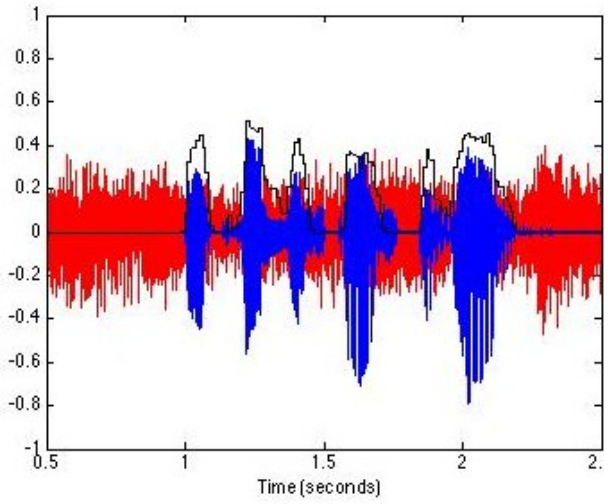

Figure 1: Example ESII calculation with low SNR. Red: noise, blue: speech, black: SII output.

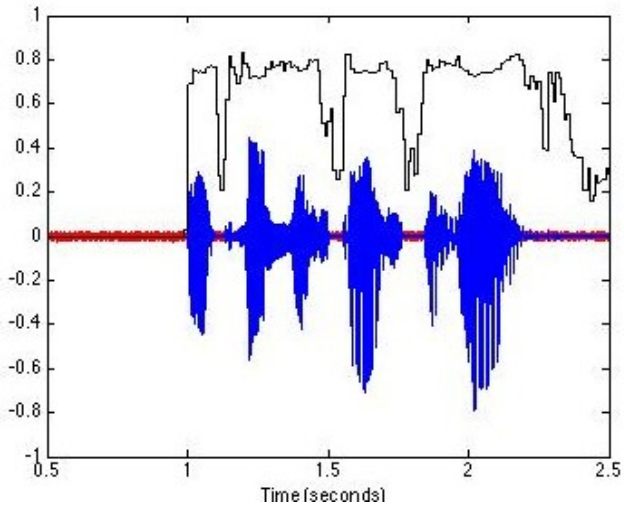

Figure 2: Example ESII calculation with high SNR.

B. Subjective testing

There were two main components of the subjective testing. The first test methodology was to require the listener to identify the final word of a speech sentence presented in a noise background. The second was to have the listener rate preference of the clarity and quality of speech in two different soundscape noise samples, using a five-point rating scale.

For this project, the noise sources were white noise and two different samples of binaural soundscape noise recorded in St. Ann's Square in Manchester. The speech samples used were from the Revised Speech Perception in Noise (SPIN).[13] In the SPIN sample set, there are two types of sentences, with either low or high contextual probabilities for identification of the final word of the sentence. Only the samples of low contextual probability were used.

St Ann's Square is largely pedestrianised, with road traffic on only one of its four sides. It has shops, a fountain, a church and is used both as a thoroughfare and a meeting place. Two recordings of the soundscape were used for this investigation. Both have similar ambient sounds (mainly distant road traffic and indistinct voices). Each recording also had more prominent sources. On soundscape sample 1, the fountain and some conversation could be clearly heard. On sample 2, the fountain, footsteps and close traffic on a cobbled street were prominent. Two different samples were used to explore whether different identifiable sources have different effects on intelligibility. It was also thought possible that different sources might have differential effects on intelligibility, quality and clarity. That is, two recordings might have the same intelligibility but different perceived clarity.

Subjects listened to speech mixed with either soundscape sample 1, soundscape sample 2, or white noise. Playback of the binaural recordings was performed using a pair of circumaural headphones and a high quality audio soundcard. Twenty native speakers of English with an average age of 31 with no known hearing impairment were used as test subjects. Randomization for each subject was performed using balanced Latin squares for both the sample order and SNR level. This resulted in two main blocks of stimuli. The first block consisted of 80 speech in noise samples for final word identification. Of these 80, the first 40 had a 'noise' of either white noise or soundscape sample 1 in a randomized order, and the second 40 had white noise or soundscape sample 2 in a randomized order. The second main

block of stimuli consisted of 20 pairs of speech in noise samples to directly compare intelligibility, quality and clarity, using a 'noise' of either soundscape sample 1 or sample 2 in a randomized order. Both main stimuli blocks featured random and equal distribution of each of the five SNR values in pairs throughout.

3. RESULTS AND DISCUSSION

A. Predicted vs. measured intelligibility

The following represents the results for the final word response testing for the 20 test subjects. Over the course of the testing, 80 responses (160 for white noise) for each noise/SNR combination were given. To compare the subjective scores with the predicted SII, one needs to derive a single figure from the time-varying ESII. Two different methods were tried: a simple mean ESII (as used by Rherbergen and Versfeld) and the 90th centile of the ESII. It was found that the 90th centile ESII agreed fairly well with the subjective data, as shown in Fig. 3. It is tentatively hypothesised that the 90th centile is the better predictor because it effectively picks out the portions of the signal when significant speech energy is present. This may not work so well for soundscape sample 2 because this recording features more identifiable distracting sources than sample 1. In particular, car tyres on cobbles make an unusual sound which many listeners could not identify and which may therefore have imposed an extra cognitive load. A two-way analysis of variance with factors noise type (df=2) and ESII (df=4) was conducted and the results appear in Table 1. The ANOVA shows that both factors are highly significant, with the different shape of soundscape sample 2 also producing an interaction significant at the 4% level. More work is clearly needed to explore prediction ability with both more soundscape recordings and more schemes for averaging ESII into a single figure.

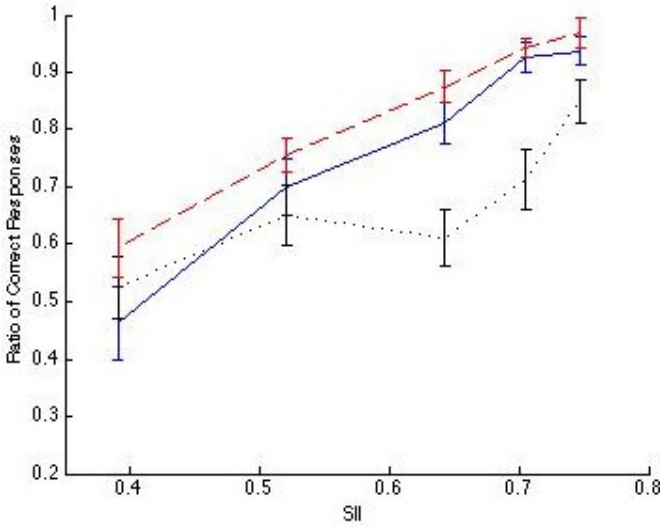

Figure 3: Subjective intelligibility (proportion of correct responses) as a function of 90th centile of ESII. Red dashed: white noise, blue solid: soundscape sample 1, black dotted: soundscape sample 2. Error bars are +/- one standard error.

Table 1: *ANOVA P-value for final word responses*

Factor	Noise Type	SII	Noise Type x SII
P	<0.0001	<0.0001	0.0395

B. Clarity and quality ratings

The results for the clarity and quality preference testing are shown in Figures 4 and 5. Each subject was played 20 samples of each noise sample (same soundscape samples as in the previous section), providing 400 decisions and ratings for both the quality and clarity scales. It can be seen that SII does not predict either rated clarity or quality well. Other metrics would be needed for these. This is confirmed by a two-way ANOVA with factors noise type (df=2) and SII (df=4). The p-values in Table 2 show that the soundscape sample is highly significant but SII is not significant at the 50% level for either clarity or quality.

It is interesting that soundscape sample 2 is rated significantly poorer than sample 1. Sample 2 includes more traffic and footsteps. It may be that cognitive features of identifiable sources, such as their meaning, have influenced this evaluation, besides the purely physical features of the recorded sound.

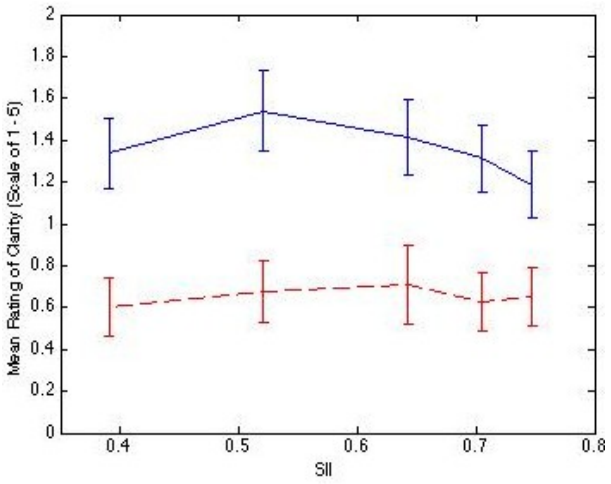

Figure 4: Mean subjective rating of clarity as a function of 90th centile of SII. Blue solid: soundscape sample 1, red dashed: soundscape sample 2

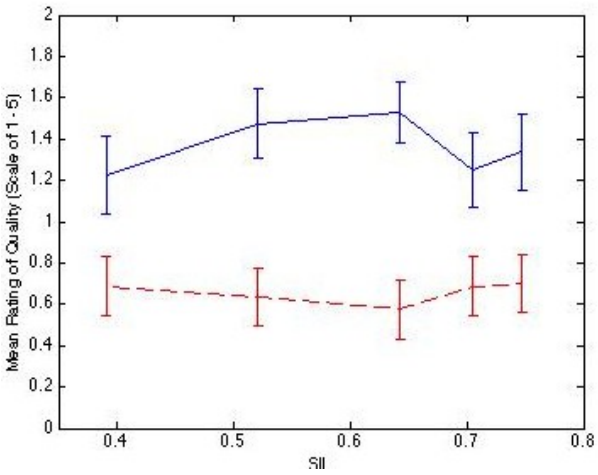

Figure 5: Mean subjective rating of quality as a function of 90th centile of SII. Blue solid: soundscape sample 1, red dashed: soundscape sample 2

Table 2: *ANOVA P-values for quality and clarity*

Factor	Noise Type	SII	Noise Type x SII
p (clarity)	<0.0001	0.9556	0.6329
p (quality)	<0.0001	0.7728	0.9045

C. SII with impaired hearing

When the ESII evaluation is coupled with the temporal window model, predictions can be made on the effects of different kinds of hearing loss on speech intelligibility. These have not yet been compared to subjective measures of intelligibility in people with a hearing loss, but the model indicates the potential for making soundscape measurements or predictions specific to users with a hearing loss. Figures 6 and 7 present predictions for high and low signal-to-noise ratios,

for three different conditions: normal hearing, mild loss and severe loss. It is immediately noticeable that, with 40 dB SNR, the SII is predicted to be better for the mild hearing loss than for normal hearing! One possible explanation for this better-than-normal performance for the impaired simulation is that loss of compression in the impaired model (due to outer hair cell loss) increases the effective SNR when the speech is more intense than the noise.

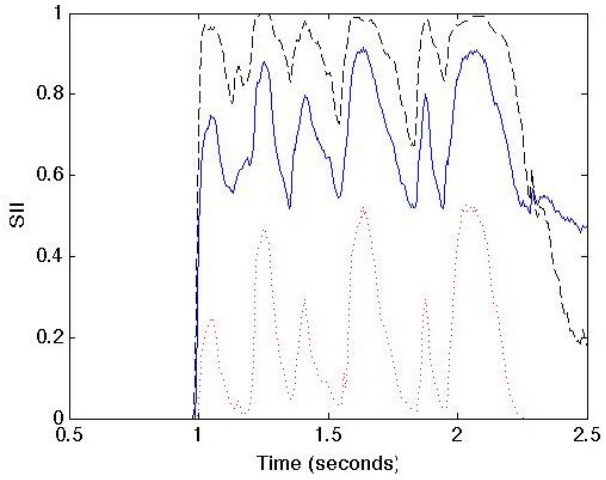

Figure 6: Example ESII calculation with hearing impairment at 40 dB SNR. Solid blue: normal hearing, black dashed: moderate high frequency loss, red dotted: severe flat loss

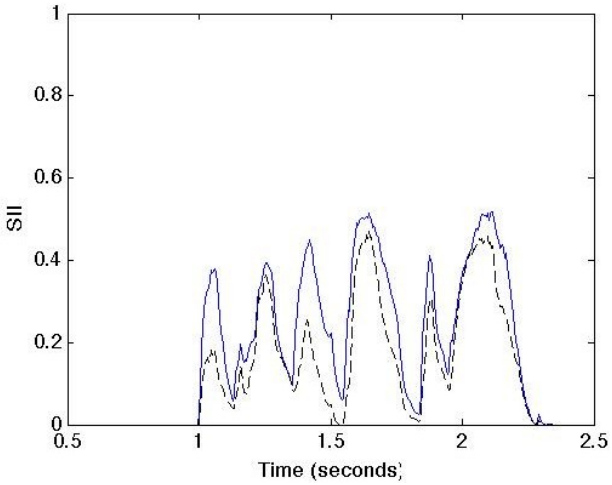

Figure 7: Example ESII calculation with hearing impairment at 0 dB SNR. Solid blue: normal hearing, black dashed: moderate high frequency loss. The severe flat loss lies below SII=0.

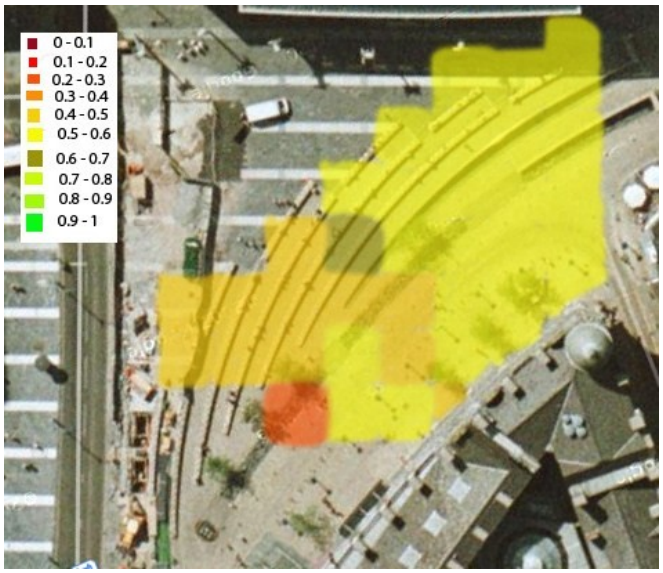

Figure 8: Predicted speech intelligibility mapped at Exchange Square, Manchester. The large BBC TV screen can be seen bottom centre.

D. Mapping speech intelligibility

To show how predicted speech intelligibility could be used, an intelligibility map has been made for another urban square in Manchester, called Exchange Square. This was based on 5-minute soundscape recordings made on a grid across the square, from which single-figure ESII values were calculated as described above. These values have been superimposed onto an aerial

photograph. This square has a similar mix of audible sound sources as St Ann's square, with the addition of more traffic and a large television screen operated by the BBC. The map predicts that directly in front of this screen would be a poor spot to hold a conversation.

4. CONCLUSION

ESII seems to offer a promising way of developing a metric to predict an important component of a perceived urban soundscape. Much more work is needed to refine the metric and to determine the circumstances under which it works best. The ability to combine ESII with a hearing loss model to predict intelligibility for people with impaired hearing could bring a useful benefit to soundscape designers and planners. One can envisage supplementing the existing predicted noise level maps with intelligibility maps. Certainly there would seem to be an application to urban squares, where speech communication is an important component of soundscape perception.

Of course, speech intelligibility is just one component of soundscape perception and there are many other perceptual aspects where metrics are lacking. The present work has shown that percepts which would seem closely related – intelligibility, clarity and quality – are not predicted by the same metric. Many other indicators will need to be developed and tested to explore these other aspects of soundscape perception.

ACKNOWLEDGMENTS

The Positive Soundscape Project is wholly funded by the UK Engineering and Physical Sciences Research Council, under grant number EP/E011624/1.

REFERENCES

1. R.M. Schafer, The tuning of the world. New York: Knopf. 1977.
2. R.J. Hawkes and H. Douglas, Subjective acoustic experience in concert auditoria, *Acustica* **24**, pp. 235-250, (1971).
3. B. Truax, Acoustic Communication. 2nd ed., Santa Barbara, CA, USA: Greenwood Publishing Group. 2000.
4. M. Adams, et al., "Soundwalking as a methodology for understanding soundscapes", in *Proc. I. o. A.*, Reading, U.K., 2008.
5. W.J. Davies, et al., "The positive soundscape project: A synthesis of results from many disciplines", in *Internoise 2009*, Ottawa, Canada, 2009.
6. J. Kang, Urban Sound Environment. London: Taylor and Francis. 2007.
7. W.J. Davies, et al., "A positive soundscape evaluation system", in *Euronoise 2009*, Edinburgh, U.K., 2009.
8. ANSI, *American national standard methods for calculation of the speech intelligibility index*, S3.5-1997, Editor. 1997, American National Standards Institute: New York.
9. K.S. Rhebergen and N.J. Versfeld, A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners, *J. Acoust. Soc. Am.* **117**(4), pp. 2181-2192, (2005).
10. *Speech Intelligibility Index Matlab Implementation*. Acoustical Society of America Working Group S3-79, 2005 Last accessed 8 July 2009; Available from: http://www.sii.to/index.html.
11. C.J. Plack, A.J. Oxenham and D. V., Linear and nonlinear processes in temporal masking, *Acustica* **88**, pp. 348-358, (2002).
12. E.A. Lopez-Poveda and R. Meddis, A human nonlinear cochlear filterbank, *J. Acoust. Soc. Am.* **110**, pp. 3107-3118, (2001).
13. R.C. Bilger, et al., Standardization of a test of speech-perception in noise, *J. Speech & Hearing Res.* **27**(1), pp. 32-48, (1984).