Speech communication in outdoor soundscapes

Davies, WJ, Mahnken, PZ and Plack, CJ

Speech Communication in Outdoor Soundscapes

W. J. Davies[1], P. Z. Mahnken[1], C. J. Plack[2]

[1] *University of Salford, Acoustics Research Centre, UK, Email: w.davies@salford.ac.uk*
[2] *University of Manchester, Human Communication & Deafness Division, UK*

Abstract

The Positive Soundscape Project is a large multi-disciplinary project investigating the perception of soundscapes. Recent findings indicate that speech communication is a principal factor in users' perceptions of urban soundscapes. The project has therefore explored how this factor might be quantified. This paper reports on an attempt to use speech intelligibility index (SII) to characterise time-varying speech intelligibility in real outdoor soundscapes. The relationships between SII, subjective intelligibility and subjective quality were explored, as functions of signal-to-noise ratio. Possible applications in sound quality mapping of soundscapes and potential implications for rational planning of soundscapes are discussed.

Introduction

This paper reports on an investigation carried out as part of the Positive Soundscape Project (PSP). PSP is a highly inter-disciplinary project, which aims to characterise the significant factors in the perceptual and emotional response to soundscapes. Researchers on the project come from many disciplines: acoustics, art, sound quality, psychoacoustics, physiology, neuroscience and social science. PSP has investigated the human response to soundscapes with several techniques, including qualitative (sound walks, interviews and focus groups), quantitative (rating scales, principal component analysis, physiological measures and fMRI brain scanning) and artistic (soundtoy, recording/composition and mapping). The fieldwork of the project has focussed on urban soundscapes in two UK cities: London and Manchester. Each specific location (such as a city square) has been studied with multiple methods. PSP has developed a conceptual framework to organise the significant variables in the subjective response to soundscapes [1] and is currently drawing together the results from the many different methods to find overlaps and differences. The working hypothesis is that if a perceptual factor emerges as significant from more than one method, then this increases confidence in that factor as a real characteristic of subjective response. Where appropriate, quantitative indicators will be sought for the emergent factors.

One such factor which is emerging from PSP's work is speech communication. Participants on sound walks evaluate some soundscapes in terms of the ability to hold a conversation, while focus group participants say that background speech hubbub is a component of one kind of positive soundscape. One previous factor analysis of soundscape perception [2] 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.

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 [3] proposed by Rherbergen and Versfeld [4] 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.

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.

SII is a method of estimating speech intelligibility that is based on AI (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.

Based on the code and methods of [4] and [5], a model for calculating the SII of speech files was used. Analysis was performed over the 200 low context speech samples from the SPIN test set to determine five target SII values with relative gains calculated for each noise sample (based on a unity gain for all speech samples). Analysis using these gain values acros the sample set calculated SII values within a standard deviation of 0.05 and SNR values within a standard deviation of 1.5dB.

The SII 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. [6]. 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, Lopez-Poveda, and colleagues [7]. 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

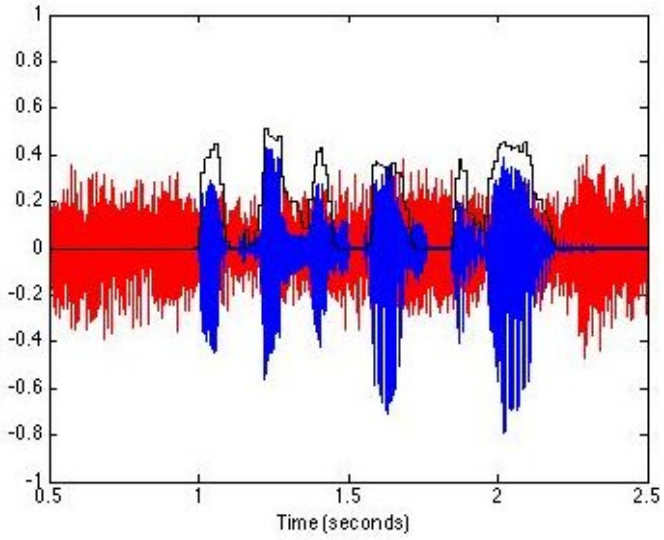

Figure 1: Example SII calculation with low SNR. Red: noise, black: speech, blue: SII output.

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 SII 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.

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 testing, 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

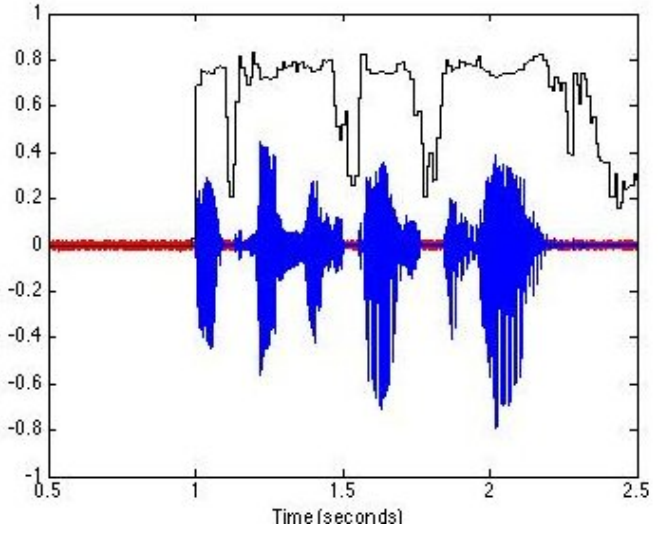

Figure 2: Example SII calculation with high SNR. Red: noise, black: speech, blue: SII output.

were from the Revised Speech Perception in Noise (SPIN) [8]. 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. For this testing, 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 hhave 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.

Results and Discussion

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 SII. Two different methods were tried: a simple mean SII and the 90th centile of the SII. It was found that the 90th centile SII 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

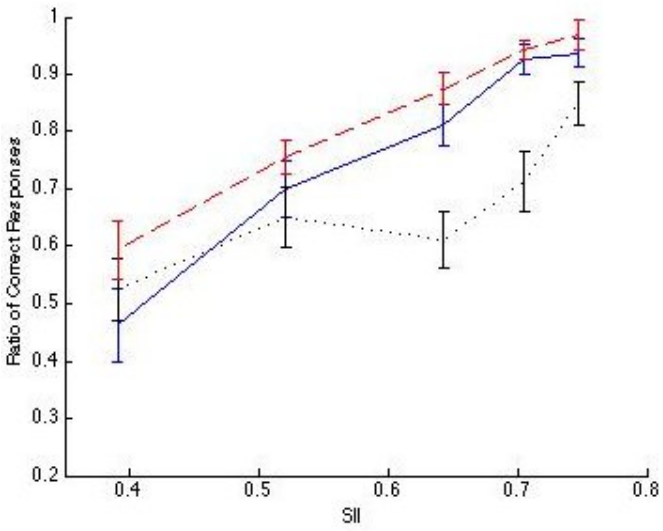

Figure 3: Subjective intelligibility (proportion of correct responses) as a function of 90th centile of SII. Red dashed: white noise, blue solid: soundscape sample 1, black dotted: soundscape sample 2.

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. A two-way analysis of variance with factors noise type (df=2) and SII (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 SII into a single figure.

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

Table 1: ANOVA P-value for final word responses

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.

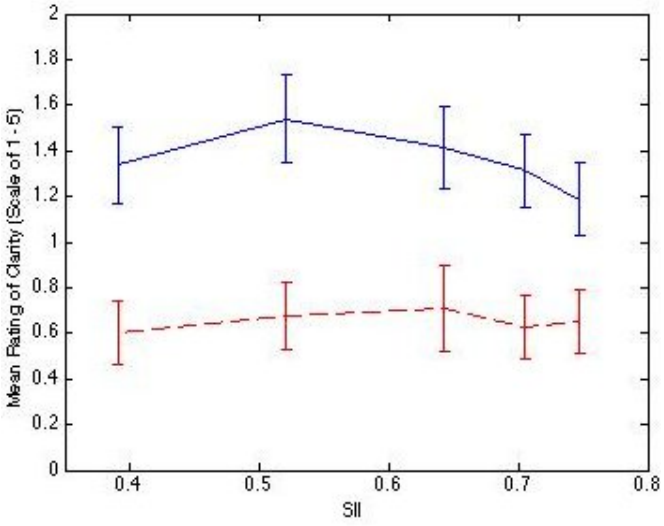

Figure 4: Mean subjective rating of clarity as a function of 90th centile of SII. Blue solid: soundscape sample 1, red dashed: soundscape sample 2

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

Table 2: ANOVA P-values for quality and clarity

SII with impaired hearing

When the SII 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

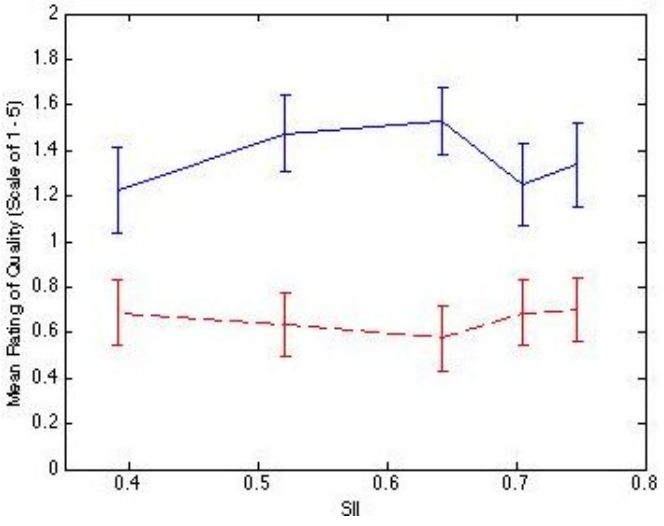

Figure 5: Mean subjective rating of quality as a function of 90th centile of SII. Blue solid: soundscape sample 1, red dashed: soundscape sample 2

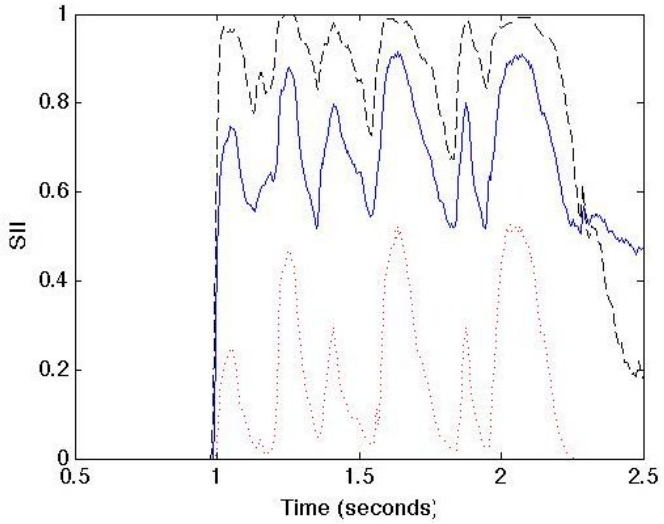

Figure 6: Example SII calculation with hearing impairment at 40 dB SNR. Solid blue: normal hearing, black dashed: moderate high frequency loss, red dotted: severe flat loss.

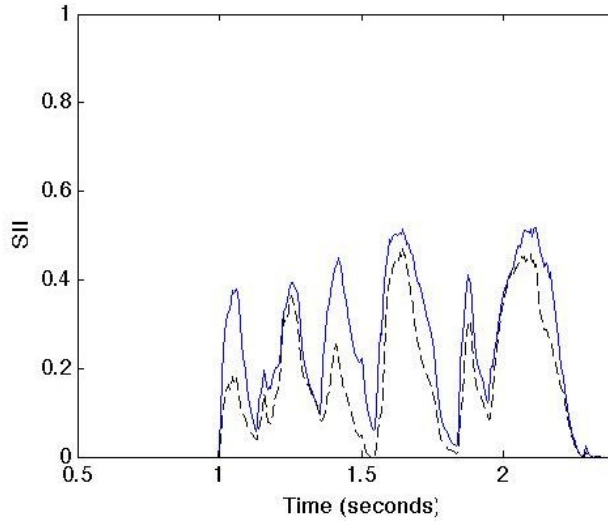

Figure 7: Example SII calculation with hearing impairment at 0 dB SNR. Solid blue: normal hearing, black dashed: moderate high frequency loss. The severe flat loss lies below SII=0.

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.

Conclusion

SII 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 SII 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 SII 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.

References

[1] Cain, R., et al., SOUND-SCAPE: A framework for characterising positive urban soundscapes, Acoustics 08, Paris, 2008

[2] Kang, J., Urban Sound Environment. Taylor and Francis, London, 2007

[3] ANSI S3.5-1997, American national standard methods for calculation of the speech intelligibility index. American National Standards Institute: New York, 1997

[4] Rhebergen, K.S. and Versfeld, N.J., 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** (2005), 2181-2192

[5] Acoustical Society of America Working Group S3-79, Speech Intelligibility Index. URL: http://www.sii.to/index.html

[6] Plack, C.J., Oxenham, A.J., and V., D., Linear and nonlinear processes in temporal masking. Acustica **88** (2002), 348-358

[7] Lopez-Poveda, E.A. and Meddis, R., A human nonlinear cochlear filterbank. J. Acoust. Soc. Am. **110** (2001), 3107-3118

[8] Bilger, R.C., et al., Standardization of a test of speech-perception in noise. J. Speech & Hearing Res. **27** (1984), 32-48