A Perceptually–Inspired Approach to Machine Sound Source Separation in Real Rooms

Chris Hummersone, Russell Mason and Tim Brookes
Institute of Sound Recording
University of Surrey, Guildford, GU2 7XH, United Kingdom

Automated separation of the constituent signals of complex mixtures of sound has made significant progress over the last two decades. Unfortunately, completing this task in real rooms, where echoes and reverberation are prevalent, continues to present a significant challenge. Conversely, humans demonstrate a remarkable robustness to reverberation. An overview is given of a project that set out to model some of the aspects of human auditory perception in order to improve the efficacy of machine sound source separation in real rooms. Using this approach, the models that were developed achieved a significant improvement in separation performance. The project also showed that existing models of human auditory perception are markedly incomplete and work is currently being undertaken to model additional aspects that had previously been neglected. Work completed so far has shown that an even greater improvement in separation performance will be possible. The work could have many applications, including intelligent hearing aids and intelligent security cameras, and could be incorporated in to many other products that perform automated listening tasks, such as speech recognition, speech enhancement, noise reduction and medical transcription.

Separating Sounds in Real Rooms

Research into machine sound source separation has been undertaken for many years, due to the numerous applications for such systems, including front-end processing for missing data speech recognition, and enhancement of hearing prostheses and communication devices such as mobile phones. Many systems have been proposed that utilise a variety of techniques that work well in ideal acoustic conditions. Unfortunately, the reverberation that arises in rooms continues to present a major obstacle to this task because it breaks down many of the acoustical cues that these systems rely on for separation. Consequently, reducing the detrimental effects of reverberation continues to be an important research goal not only for researchers in sound source separation, but also for researchers working in other areas of signal processing.

The separation algorithm in the proposed system attempts to separate the constituent signals arising from two spatially–separate sound sources. The system estimates the azimuth of each source in every processing frame. These data are used to calculate a binary mask that is used to separate the signals. The performance of the system is assessed by comparing the calculated mask to the ideal mask in terms of how closely they match. The metric is referred to as the Ideal Binary Mask Ratio (IBMR) [1]. The metric removes the effect of the room in order to fairly determine how effective the separation has been.

Modelling the Precedence Effect

Numerous human psychophysical and perceptual mechanisms for suppressing the effects of reverberation are well documented, which have occasionally provided a source of inspiration for researchers in signal processing. One such mechanism is the precedence effect. The precedence effect is a perceptual phenomenon that appears to weight the first few wavefronts of a sound over later wave fronts arriving as reflections from surrounding surfaces. It is an essential mechanism that allows humans to localise sounds in rooms, a task that would otherwise be very difficult. The perceptual literature has described the precedence effect in sufficient detail for numerous computational models of precedence to be proposed. Despite this, relatively little work has been carried out on incorporating the precedence effect into source separation systems that use spatial cues (i.e. features relating to a sound source's location in space).

This project initially investigated using models of the precedence effect to improve source separation in real rooms. The investigation [2] used a previously proposed precedence and separation system [3] and investigated the effect on performance of changing the precedence model for others previously proposed in the literature. The architecture of the model is given in Fig. 1.

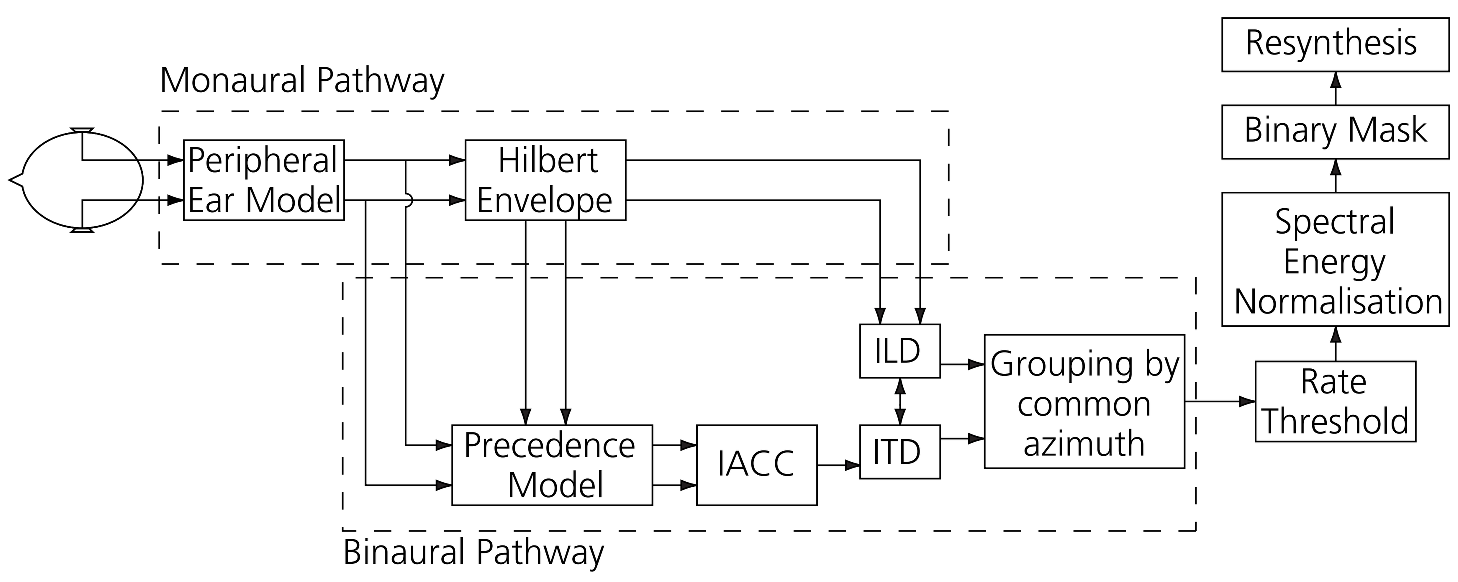


Figure 1: Model architecture

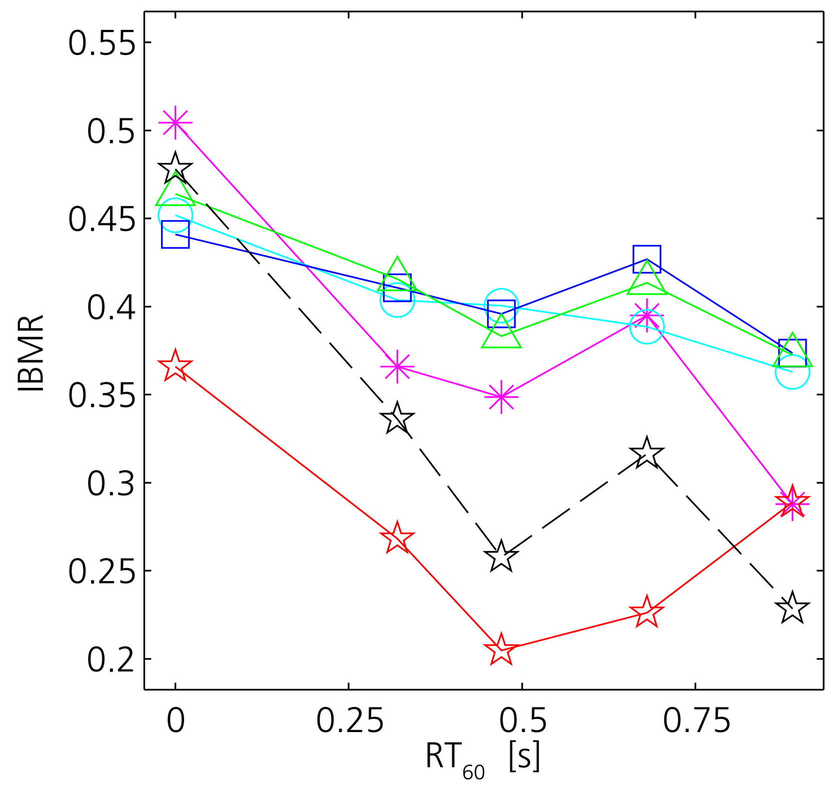


Figure 2: Performance of each of the precedence models compared with the baseline model.

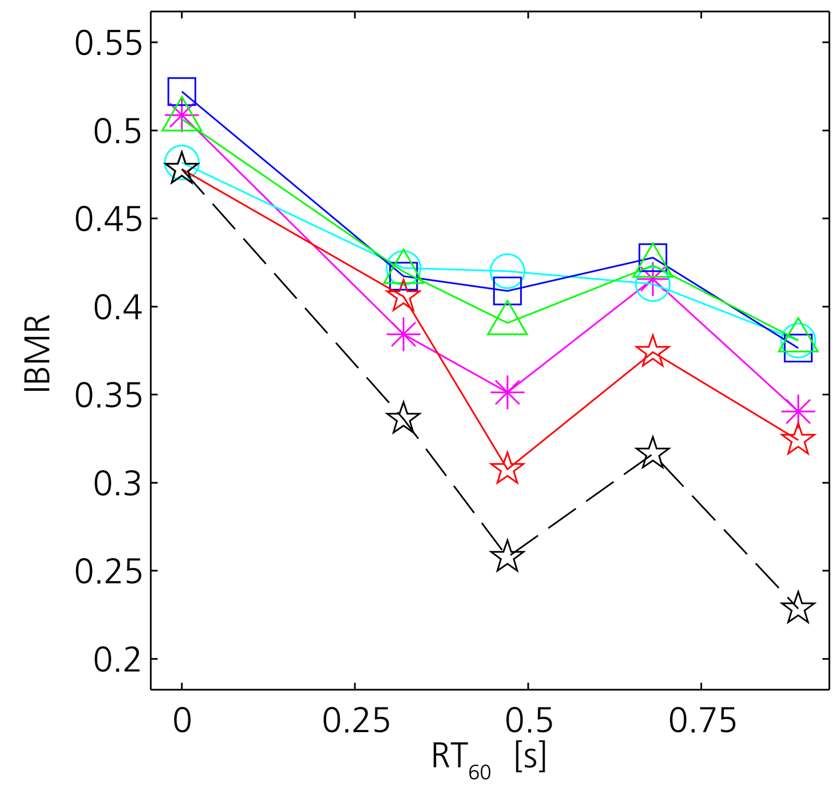
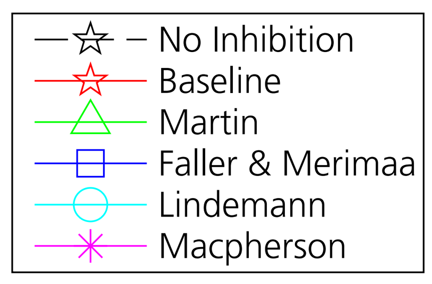


Figure 3: Performance after optimising the precedence parameters for each room.



The results of this investigation are given in Fig. 2. The results indicated that Faller & Merimaa's model [4] based on Interaural Coherence (IC) provided significant performance improvement over the baseline model. The results also indicated that it might be necessary to adapt the precedence model to room under test

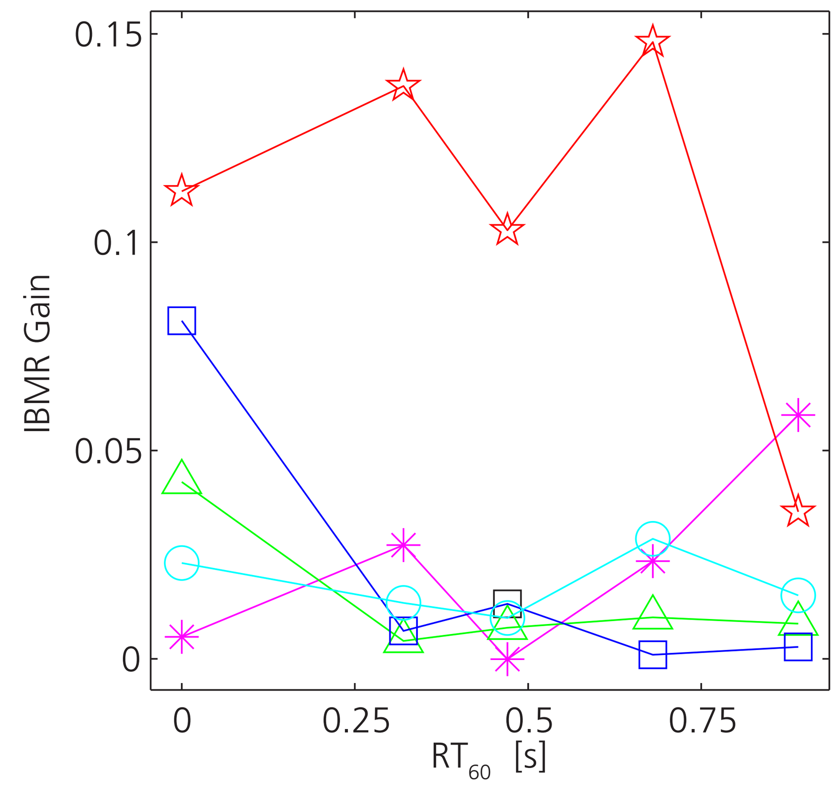


Figure 4: Performance gain achieved by optimising the precedence models in a manner analogous to the perceptual Clifton effect.

Towards a Model of the Clifton Effect

Dynamic processes in the precedence effect are well documented. Research agrees that this adaptation, known as the Clifton effect, occurs when an implausible reflection pattern is heard. This may include, for example, an unrealistically large interaural time delay, implying that the time of arrival difference is due to a reflection. This implausible reflection pattern causes the precedence effect to breakdown whilst the listener rescans the room. The precedence effect then builds up again in response to the new reflection pattern. The research conducted by the authors [5,6] confirms that such a mechanism is essential for all of the previously tested computational precedence models, at least when the model is tasked with separating sounds. The results show that the inhibition parameters need to be tuned to the acoustic characteristics of the room under test. As shown in Fig. 3 and Fig. 4, this can lead to a large gain in separation performance.

Conclusions

This project has shown that modelling the precedence effect can be beneficial to sound source separation systems that rely on spatial cues. However, the research also shows that further work needs to be done to model the Clifton effect so that the precedence model can adapt its processing to the acoustics of any room. Such a mechanism has lead to a large performance gain for some models and more is likely with further refinement of the research. The project has also highlighted the importance of choosing other model parameters, such as the window length and window shape; careful choice can lead to large performance gains. The research could be incorporated into intelligent listening devices (such as hearing aids and security cameras) where acoustic conditions are likely to change. It could also be incorporated into existing models of human auditory perception, especially those that perform localisation and separation.

References

[1] C. Hummersone, R. Mason, and T. Brookes, "Ideal Binary Mask Ratio: A Novel Metric for Assessing Binary–Mask–Based Sound Source Separation Algorithms," *IEEE Trans. Audio, Speech, Lang. Process.* (submitted).

[2] C. Hummersone, R. Mason, and T. Brookes, "A comparison of computational precedence models for source separation in reverberant environments," in *Proc. 128th Audio Eng. Soc. Conv.*, London, U.K., May 2010.

[3] K. J. Palomäki, G. J. Brown, and D. Wang, "A binaural processor for missing data speech recognition in the presence of noise and small-room reverberation," *Speech Commun.*, vol. 43, pp. 361–378, Sep. 2004.

[4] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Amer.*, vol. 116, pp. 3075–3089, Nov. 2004.

[5] C. Hummersone, R. Mason, and T. Brookes, "Dynamic precedence effect modeling for source separation in reverberant environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, pp. 1867–1871, Sep. 2010.

[6] C. Hummersone, "A Psychoacoustic Engineering Approach to Machine Sound Source Separation in Reverberant Environments", PhD Thesis, University of Surrey, 2010 (in preparation).