ACOUSTICS2008/809
Improving source localisation in multi-source, reverberant conditions: exploiting local spectro-temporal location cues

Heidi Christensen[a], Ning Ma[b], Stuart Wrigley[a] and Jon Barker[b]

[a]University of Sheffield, Regent Court, Dept. of Computer Science, 211 Portobello Road, s1 4dp Sheffield, UK

[b]University of Sheffield, Dept. of Computer Science, Regent Court, 211 Portobello Street, S1 4DP Sheffield, UK

This paper examines the use of binaural cues for estimating the location of individual sources in multi-source acoustic mixtures. In most environments, location cues such as interaural time difference (ITD) and interaural intensity difference (IID) are rendered unreliable by noise due to competing sources and reverberation.

We propose a method that examines spectro-temporal regions where such effects are minimised. Source location is estimated by combining two processing stages. First, following [Christensen et al, Interspeech (2007)], as pitch cues are generally less affected by noise, a multi-pitch estimation and tracking algorithm is used to identify local spectro-temporal 'fragments' in which the SNR is high. Second, the confidence of ITD and IID estimates arising from each time-frequency 'pixel' is estimated based on interaural coherence [Faller and Merimaa, J. Acoust. Soc. Am. 116, 3075–3089 (2004)] and reverberation estimation [Heckmann et al, Int. Conf. Int. Robots and Systems (2006)]. A location estimate is then constructed by integrating suitably weighted cues from each pixel within the fragment.

Experiments are carried out on a large corpus of multi-speaker data, mixed from binaural recordings in a real, reverberant environment. The fragment based processing is shown to provide significant improvements when compared to competitive baseline approaches.