# SOUND SOURCE SEPARATION IN REVERBERANT ENVIRONMENTS USING INTERAURAL COHERENCE IN A PROBABILISTIC MODEL OF LOCALISATION

JOACHIM FAINBERG
IoSR, 2014

A sound source separation and localisation model for reverberant environments is formulated following a review of the literature. The main components are a localisation algorithm based upon lookup tables, a precedence effect model based upon interaural coherence and an extraction procedure using time-frequency masking techniques.

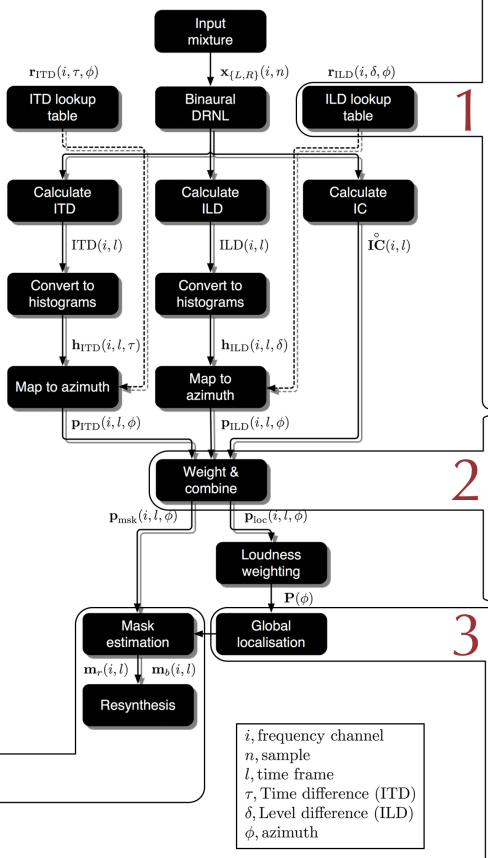
The localisation algorithm calculates the probability of a source arising from a given angle by comparing the input signal to a lookup table that relates localisation strength to lateral angle for each binaural cue. The probabilities arising from either cue are then weighted by a cross-weighting formula, which broadly mimics the duplex theory, and subsequently combined by multiplication.

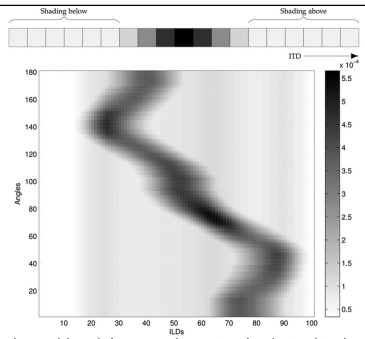
The precedence effect model provides a measure of whether the binaural cues represent the true direction of a source by calculating the interaural coherence. An interaural coherence threshold determines the input to the localisation algorithm.

Finally, a binary mask is calculated to form the separated output. The probabilities used to determine the pattern of the mask are linearly weighted by the interaural coherence, to ensure that the mask does not become unnecessarily sparse.

An experiment is undertaken to assess the performance of the model, and in particular to find the thresholds of interaural coherence for which the model performs optimally. A threshold of approximately 0.8 is found to provide a reasonable compromise in performance across all conditions. Thresholds larger than 0.9 distort localisation performance severely.

The model is compared to a range of established models in a variety of acoustic conditions and it is found to perform comparably for source separations of 10° and 20°, and favourably for source separations of 40° and 60°. The estimates of source location at small source separations are found to be imprecise.



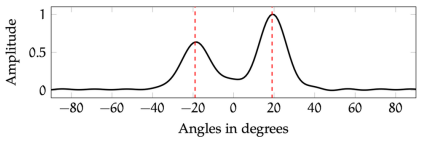

Lookup tables define a 3-dimensional relationship between azimuth, frequency channels and a range of ITDs and ILDs. The tables build upon the research by Supper [2005] and Dewhirst [2008]. By combining the lookup tables with calculated binaural cues in a sound mixture, the probability that a source has arisen from a given azimuth can be calculated.

The probabilities of a source arising from a given azimuth is frequency weighted by a formula that broadly mimics the duplex theory, and subsequently linearly multiplied with the interaural coherence for the mask estimation, and with a a thresholding function for the localisation algorithm:

$$\mathbf{p}_{\mathrm{loc}}(i,l,\phi) = \psi \times w_{\mathrm{ITD}}(i)\mathbf{p}_{\mathrm{ITD}}(i,l,\phi) \times w_{\mathrm{ILD}}(i)\mathbf{p}_{\mathrm{ILD}}(i,l,\phi),$$

$$\mathbf{p}_{\mathrm{msk}}(i,l,\phi) = \mathring{\mathbf{IC}} \times w_{\mathrm{ITD}}(i)\mathbf{p}_{\mathrm{ITD}}(i,l,\phi) \times w_{\mathrm{ILD}}(i)\mathbf{p}_{\mathrm{ILD}}(i,l,\phi),$$

$$\text{where } \psi = \begin{cases} 1 & \mathring{\mathbf{IC}} \geq \mathrm{IC}_0 \\ 0 & \text{otherwise} \end{cases}$$




The local probabilities are sharpened, loudness weighted and summed across frequency channels and time frames to yield a global estimate of the location of the sources. The algorithm explicitly assumes two stationary sources at separate azimuthal angles. The estimate of location of each source is derived from the two greatest global maxima in the global estimate.

$i$, frequency channel
$n$, sample
$l$, time frame
$\tau$, Time difference (ITD)
$\delta$, Level difference (ILD)
$\phi$, azimuth

Finally, the resulting target and interferer azimuthal locations from the global estimate are used to create a binary mask:

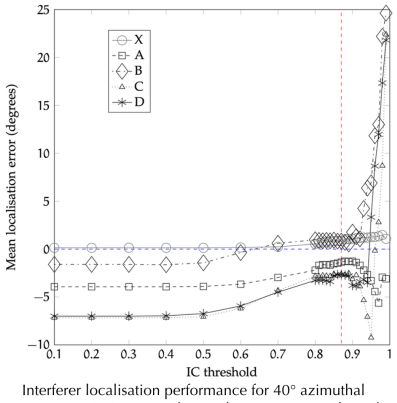
$$\mathbf{m}_{\mathrm{B}}(i,l) = \begin{cases} 1 & \mathbf{p}_{\mathrm{msk}}(i,l,\phi_{\mathrm{T}}) > \mathbf{p}_{\mathrm{msk}}(i,l,\phi_{\mathrm{I}}) \\ 0 & \text{otherwise} \end{cases}$$

The masked mixture is subsequently resynthesised by passing it through an array of band-pass sinc filters, and summing the output of the filterbank across frequency.

## Values of interaural coherence

Faller [2004] showed that the interaural coherence in a binaural signal can be used to simulate the precedence effect. This is achieved by only selecting binaural cues with an interaural coherence than a certain threshold. A range of research has found different levels of thresholds to be appropriate (Faller [2004], Hummersone [2011b]).

The present source separation model was tested for a variety of speech signals with four azimuthal separations (10, 20, 40 and 60 degrees) in an anechoic room ('X') and four increasingly reverberant rooms ('A' through 'D'). It was shown that localisation varied greatly with varying IC thresholds. In general, IC thresholds lower than 0.5 had little, to no effect on the localisation algorithm. The IC threshold in most situations converged towards a minimum mean square error (MMSE) at about IC0 = 0.8. For IC thresholds higher than 0.8, particularly higher than 0.9, the localisation precision generally deteriorated across all conditions with a notable exception of the pseudoanechoic room.

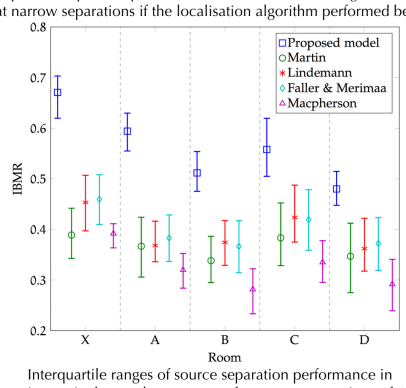



Interferer localisation performance for 40° azimuthal separation in increasingly reverberant rooms (X through D). The horizontal, blue line indicates correct localisation, and the vertical red line shows the MMSE across all conditions.

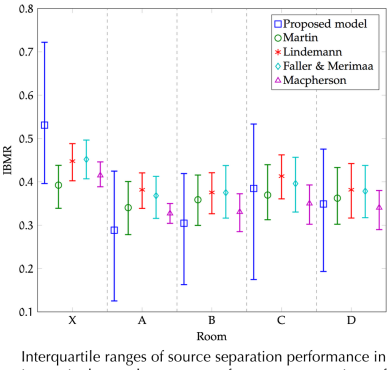
## Comparison to other models in literature

The model was tested against a range of established models in the literature, and evaluated using the Ideal Binary Mask Ratio metric proposed by Hummersone [2011a].The proposed source separation algorithm performs significantly better than a range of established models for azimuthal separations of 40° and 60°, across a range of acoustic conditions. As there were a number of untested assertions made in the computational implementation of the model, the model has potential for optimisation across several avenues.

The proposed localisation algorithm performed poorly for sources spatialised at 10° and 20°. Because the source separation is directly dependant upon the performance of the localisation algorithm, it can be hypothesised that the source separation performance would improve at narrow separations if the localisation algorithm performed better.




Interquartile ranges of source separation performance in increasingly reverberant rooms for source separations of 40°.




Interquartile ranges of source separation performance in increasingly reverberant rooms for source separations of 10°.

## Conclusions

Optimal values of interaural coherence thresholds were determined to approximately 0.8. When the thresholds were set to their optimal values, the results indicate that the model outperforms a range of established models in the literature for source separations of 40° and 60°. The model is relatively robust to reverberation and its performance in the most reverberant room aligns with the performance of some of the established models in the anechoic room. However, the model does not perform particularly well for source separations at 10° and 20°. This was attributed to the wide peaks in the global estimate and the related peak selection heuristic. Further, a number of assertions were made in the computational implementation. The performance of the model with respect to the established models may be further improved if these assertions are tested more thoroughly.

## References

Dewhirst, Martin. *Modelling Perceived Spatial Attributes of Reproduced Sound*. Doctoral thesis, University of Surrey, 2008.

Faller, Christof and Merimaa, Juha. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *The Journal of the Acoustical Society of America*, 116(5): 3075, 2004.

Hummersone, C.; Mason, R., and Brookes, T. Ideal binary mask ratio: A novel metric for assessing binary-mask-based sound source sepa- ration algorithms. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2039–2045, 2011a.

Hummersone, Christopher. *A psychoacoustic engineering approach to machine sound source separation in reverberant environments*. PhD thesis, University of Surrey, 2011b.

Supper, Ben. *An Onset-Guided Spatial Analyser for Binaural Audio*. Phd thesis, University of Surrey, Institute of Sound Recording, 2005.