

# Machine Listening for Sound Quality Evaluation

Tim Brookes & Chris Hummersone  
Institute of Sound Recording  
University of Surrey, Guildford GU2 7XH. UK.  
www.surrey.ac.uk/soundrec

The overall perceived quality of audio depends, at least partly, on the perceived quality of lower-level timbral and spatial attributes (brightness, warmth, locatedness, envelopment, etc). These attributes depend, in turn, on acoustic parameters (e.g. frequency spectrum, inter-aural cross-correlation coefficient). Using acoustic measurement and human listening tests, the connections between acoustic parameters and perceived timbral and spatial attributes, and also between these perceptual attributes and overall quality, can be established. Quantification of these connections can then inform the development of machine listening systems to assess the quality of audio as it would be perceived by human subjects.

## Audio Quality

Audio quality can be taken to mean an emotion-free characterisation of sound in terms of its perceived timbral and spatial attributes, or a hedonic judgement based on these attributes [Letowski 1989], with spatial attributes contributing approximately 30% and timbral attributes approximately 70% to overall quality [Rumsey et al 2005]. The spatial side of audio quality involves attributes such as source width, source distance, locatedness, room size, and envelopment [Rumsey 2002]; the timbral side relates to attributes such as warmth, brightness, roughness, softness and fullness [Stepanek 2006].

## Perceptual vs Objective

Each timbral and spatial attribute relates to one or more acoustic cues, such as spectral slope, relative harmonic amplitudes, inter-aural signal relationships, and reverberation level [Zwicker & Fastl 1999; Blauert 1997]. Determination of the exact relationships between acoustic cues and perceptual attributes is an ongoing research task since these relationships can be more intricate than they may initially appear and a relationship that holds for one particular category of sound sources may not hold for another. Inter-aural cross-correlation coefficient, for example, is an important cue for source width perception but this percept is also influenced by several other factors [Mason et al 2005]; perceived softness is often correlated to harmonicity, but this is not always the case [Williams & Brookes 2009]. The relationships between objective acoustic parameters, perceptual attributes and overall audio quality is summarised in figure 1.



Figure 1. Audio quality depends on timbral and spatial attributes of sound which, in turn, depend on objective acoustic parameters.

## Listening Tests

In order to establish the connections between acoustic parameters and perceived timbral and spatial attributes, and also between these perceptual attributes and overall sound quality, extensive tests with human listeners and carefully-controlled or accurately-quantified stimuli are required. These tests must be carefully designed in order to avoid corruption of the results by the many forms of bias which can occur [Zielinski et al 2008] and will often require the development of novel or hybrid analysis techniques [Neher et al 2006a].

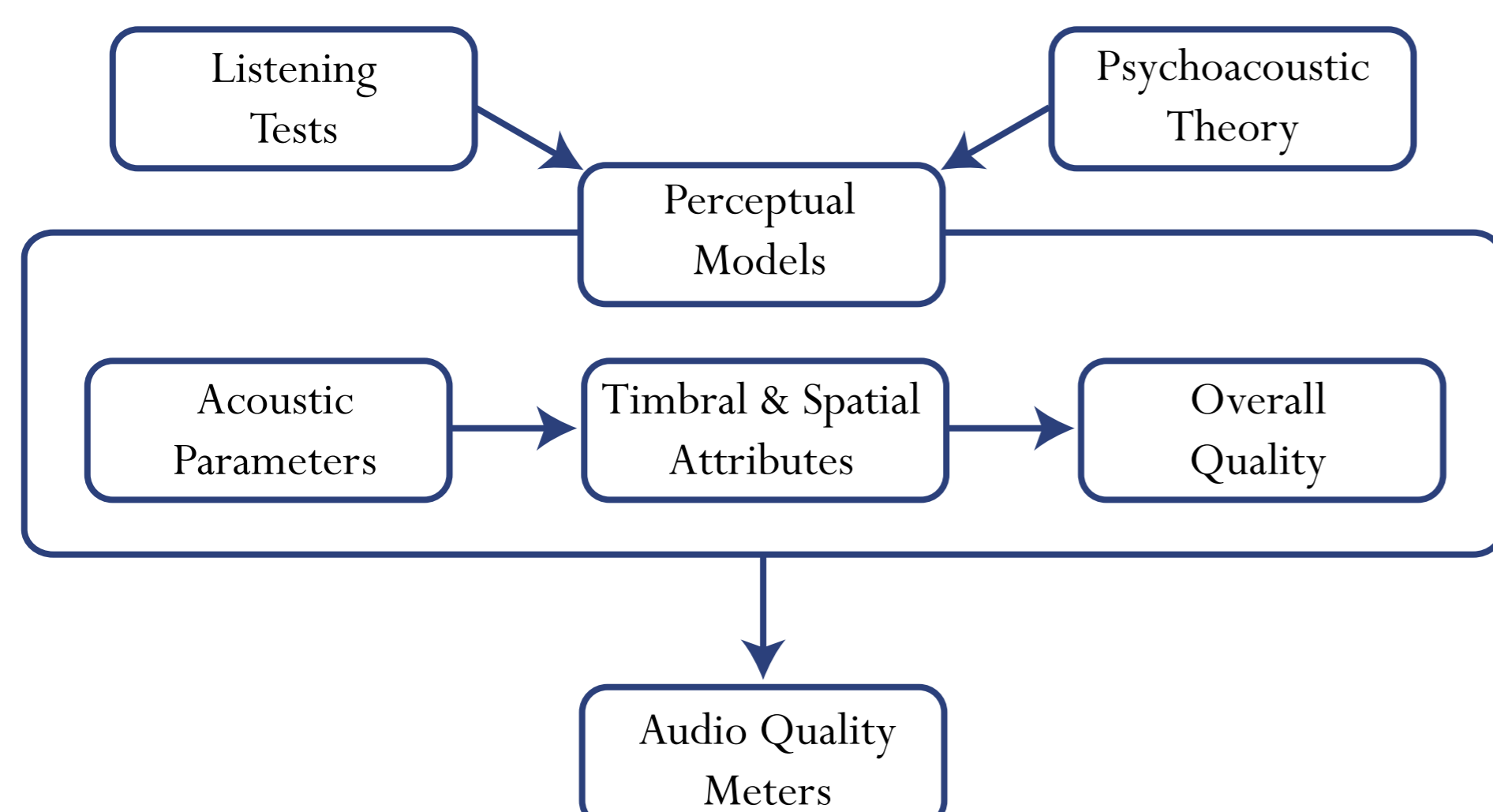


Figure 2. Extensive listening tests, together with established psychoacoustic theory, can inform the development of perceptual models able to predict how a particular combination of acoustic parameters will be perceived in terms of both timbral and spatial attributes, and how this combination of perceptual attributes will relate to overall quality perception. Such perceptual models can then form the core of audio quality metering systems.

## Modelling Quality Perception

Once the acoustic-perceptual and descriptive-hedonic relationships are established they can inform the development of mathematical and computer models of perception. Such models may be additionally informed by established psychoacoustic theory and may resemble physiological or neurological processes; however they may also be, or may include elements which are, purely functional, employing an engineering or statistical approach to signal processing. Representations of auditory processes such as signal onset detection [Supper et al 2006] and reverberant source separation [Hummersone et al 2010] can be important additional components. Realising a practical audio quality meter from a perceptual model is sometimes straightforward but will sometimes require significant additional effort to design appropriate test signals [Neher et al 2006b] or audio capture devices [Kim et al 2010]. The full development process is illustrated in figure 2, while figure 3 illustrates the essence of the intended end product.

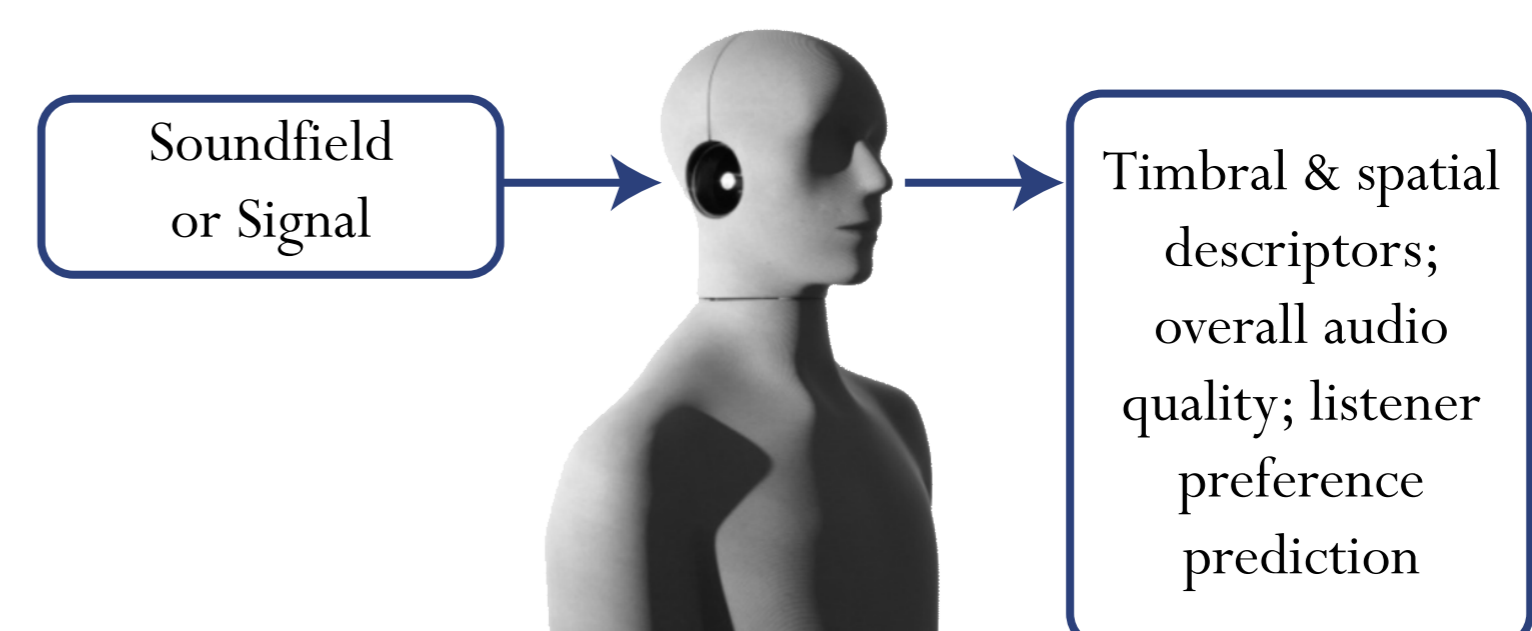


Figure 3. The ultimate goal of a full machine listening system for sound quality evaluation.

## References

- J.Blauert (1997): "Spatial Hearing: The Psychophysics of Human Sound Localization", MIT Press
- C.Hummersone, R.Mason & T.Brookes (2010): "Dynamic Precedence Effect Modelling for Source Separation in Reverberant Environments", IEEE Transactions on Audio, Speech & Language Processing, vol. 18 (7), p1867-1871
- C.Kim, R.Mason & T.Brookes (2010): "Development of a head-movement-aware signal capture system for the prediction of acoustical spatial impression", Proceedings of the 20th International Congress on Acoustics
- T.Letowski (1989): "Sound Quality Assessment: Cardinal Concepts", Journal of the Audio Engineering Society, vol. 37 p.1062
- R.Mason, T.Brookes & F.Rumsey (2005): "Frequency dependency of the relationship between the perceived width and the interaural cross-correlation coefficient of time-invariant stimuli", Journal of the Acoustical Society of America, vol. 117 (3), p.1337-1350
- T.Neher, T.Brookes & F.Rumsey (2006a): "A Hybrid Technique for Validating Unidimensionality of Perceptual Variation in a Spatial Auditory Stimulus Set", Journal of the Audio Engineering Society, vol. 54 (4), p259-275
- T.Neher, T.Brookes & R.Mason (2006b): "Musically Representative Test Signals for Use in Interaural Cross-Correlation Coefficient Measurement", Acta Acustica united with Acustica, vol. 92 (5), p787-796
- F.Rumsey (2002): "Spatial quality evaluation for reproduced sound: Terminology, meaning, and a Scene-Based Paradigm", Journal of the Audio Engineering Society, vol. 50 (9), p.651-666
- F.Rumsey, S.Zielinski, S.Bech & R.Kassier (2005): "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality", Journal of the Acoustical Society of America, vol. 118 (2), p.968-976
- J.Stepanek (2006): "Musical Sound Timbre: Verbal description and dimensions", Proceedings of the 9th International Conference on Digital Audio Effects (DAFx06)
- B.Supper, T.Brookes & F.Rumsey (2006): "An auditory onset detection algorithm for improved automatic source localization", IEEE Transactions on Audio, Speech and Language Processing, vol. 14 (3)
- D.Williams & T.Brookes (2009): "Perceptually-motivated audio morphing: softness", 126th Convention of the Audio Engineering Society
- S.Zielinski, F.Rumsey & S.Bech (2008): "On Some Biases Encountered in Modern Audio Quality Listening Tests - A Review", Journal of the Audio Engineering Society, vol. 56 (6), p.427-451
- E.Zwicker & H.Fastl (1999): "Psychoacoustics: Facts and Models", Springer