University of Huddersfield Repository

Moore, David J. and Wakefield, Jonathan P.

Surround Sound for Large Audiences: What are the Problems?

Original Citation

Moore, David J. and Wakefield, Jonathan P. (2007) Surround Sound for Large Audiences: What are the Problems? In: Proceedings of Computing and Engineering Annual Researchers' Conference 2007: CEARC'07. University of Huddersfield, Huddersfield, pp. 1-7.

This version is available at http://eprints.hud.ac.uk/id/eprint/3704/

http://eprints.hud.ac.uk/

SURROUND SOUND FOR LARGE AUDIENCES: WHAT ARE THE PROBLEMS?

J. D. Moore, Dr J. P. Wakefield

[1] University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK

ABSTRACT

When replaying surround sound to large audiences it is inevitable that some people will not be in the ideal central listening position relative to the loudspeakers. As a result of this, their surround sound experience is degraded. The purpose of this paper is to outline how and why the listening experience is affected. This work is part of an ongoing study which aims to improve surround playback in large spaces both for listeners in central positions and off-centre positions.

Keywords surround sound psychoacoustics

1   INTRODUCTION

Surround sound is now commonplace in the home. This is largely due to the demand for more realism in movies, music reproduction and computer games (Davis 2003). In this environment a single listener will have a good surround sound experience when positioned in or around the ideal central listening position known as the "sweet spot". In large-scale listening situations (e.g. cinema, art installations), however, the audience will be widely distributed with many of the audience members positioned at a significant distance from the "sweet spot". This can result in their surround sound experience being degraded. As a general rule of thumb, the further away the listener is from the "sweet spot" the worse the surround sound experience will be affected (see figure 1). This paper looks at the some of the issues that can occur as a result of off-centre listening.

The following section reviews the psychoacoustic mechanisms we use for localising sound sources. Section 3 will then discuss methods of surround sound reproduction over the standard surround loudspeaker layout known as ITU 5.1 (Theile 1993). Section 4 will focuses on the problems with off-centre listening. Specifically, sound source localisation issues introduced by time differences between sound arriving at the listener from different loudspeakers. Finally, a summary will be given with future work outlined.

2   SOUND SOURCE LOCALISATION

When sound arrives at the listener from a sound source a number of different auditory cues combine allowing the listener to localise the source. The main auditory cues are; the interaural time difference (ITD), the interaural level difference (ILD), and spectral filtering caused by sound reflecting off the outer ear (pinnae) and the upper body (Blauert 2001). The efficiency of each cue is dependent on the nature of the sound. A general outline of each cue follows.

2.1 Interaural Time Difference

The ITD is used throughout the entire human hearing range for localising sounds. It occurs because of the different path lengths traveled by sound to the two ears. ITD localisation is known to be most effective for impulsive sounds (e.g. speech). In this case, it is onset differences between sound waves arriving at each ear which are important. The ITD varies from 0µs to approximately 700µs depending on the position of the sound source around the listener and also the distance separating the listener's ears (see figure 2a). A minimum ITD is experienced when a sound is directly in front of the listener. A maximum ITD is experienced when a sound is directly to the side of the listener. The ITD can be approximated using equation 1 which assumes the listener's head is spherical:

$$ITD = \frac{D}{2c}(1 + \cos\theta)$$

(1)

where θ is the horizontal angle of the sound source, D is the diameter of the head and c is the speed of sound.

For single frequency signals (e.g. sinusoids) the auditory system can use phase differences between sound waves arriving at each ear to locate a sound source. A time difference can be converted into a corresponding phase difference by using equation 2. This is often referred to as the Interaural Phase Difference (IPD).

$$IPD = ITD(Frequency * 2\pi) \tag{2}$$

However, the IPD is only useable up to about 1400Hz. Above this, the wave length is shorter than the diameter of the head rendering the phase information ambiguous (Rayleigh 1907). This is because the auditory system cannot determine absolute phase differences. For example, a phase difference of 500° could also be interpreted as a phase difference of 140° (i.e. 500° – 360° = 140°).

2.2 Interaural Level Difference

For frequencies above about 1400Hz level differences between sound arriving at the ears are used to locate a sound source. Above this frequency a sound's wavelength is shorter than the diameter of the listener's head causing sound waves to be attenuated on their route to the furthest ear (see figure 2b). Unlike the ITD, the ILD is frequency dependent. Feddersen et al (1957) showed that ILDs may be as large as 20dB for high frequency sounds that are directly to the side of the listener.

2.3 Spectral cues

The ITD and ILD are not enough on their own for localising sounds. For sound sources in the horizontal plane there are always two points around the listener with identical ITDs and ILDs. For example, sound arriving from a source at 45° will have an identical ITD and ILD as sound arriving from a source at 135°. If the vertical plane is considered as well then there will be a whole series of points on the surface of a cone which have the same ITD and ILD. This is known as the cone of confusion (see figure 3). To resolve this ambiguity spectral cues are used which occur as a result of the directional-dependent filtering caused by sound reflecting off the ear's pinnae, the listener's torso and head shape. Spectral cues (and the ITD and ILD) can be described by a head related transfer function (HRTF). Every individual has their own unique HRTF for each sound source direction (Cheng and Wakefield 2001).

2.4 Head movements

Localisation of sound sources is assisted further by head movements (Thurlow and Runge 1967). Small head movements result in slight changes in ITD, ILD and spectral filtering helping us focus in on the sound source (Moore 2003). Head movements play a very important role when there is limited cue information available. For example, they can help us resolve front-back localisation confusions which can occur when limited localisation information is available from a sound.

2.5 Localisation accuracy and resolution

The accuracy with which we can determine a sound source's position varies by angle. Greater accuracy is possible at the front than at the rear and the sides (Blauert 2001). Not only does it vary by angle, it is also dependent on the frequency content of the sound. Generally, the more cues that are in agreement with each other, the more accurately we can locate a sound.

Work by Mills (1958) investigated the resolution of human localisation in the horizontal plane by trying to determine what the smallest noticeable change in a sound source's position is. Results from his experiments (and others since) have shown that localisation resolution is at it lowest when in front of the listener, and at its highest when at the sides of the listener (Hartmann 1989). Under ideal conditions a resolution of 1° in the front is possible and at the side about 10°. The lower limit on localisation resolution is called the Minimum Audible Angle (MAA).

3   SURROUND SOUND TECHNIQUES

A sound can be made to appear to come from between a pair of loudspeakers by outputting the sound from both of the loudspeakers. This is an auditory illusion which is often referred to as a phantom image. Phantom images form as a result of sound combining at a point giving rise to a single perceived location (Rumsey 2007). The position of a phantom image can be controlled by changing the ratio of level differences or time differences between the loudspeaker outputs (referred to as panning). One of the most common methods for panning is amplitude panning (as used in stereo). One common technique uses a cosine-sine law for generating sound level weightings for a pair of loudspeakers (see equation 3).

$$Left = S * \cos\theta$$
$$Right = S * \sin\theta$$
$$0 >= \theta <= \pi / 2 \tag{3}$$

where S is the audio signal and θ is the angle of the phantom source.
An extension of the above panning law is the most commonly used method for the five-speaker ITU layout. This algorithm is used almost exclusively in mixing desks and in software audio sequencers. Whilst it works reasonably well for positioning sound sources between closely spaced speakers, Theile (1977) has shown that problems can occur with generating stable phantom images between loudspeakers angled further apart than about 60°. This is significant for the ITU layout as spacing between loudspeakers at the side is 90° and rear loudspeakers is 140° (see figure 1 for the angular positions of the loudspeakers). More recent work by Martin et al (1999) found similar localisation issues at the sides and the rear of listeners during an experiment using the ITU configuration. One of the localisation issues that can occur is because of conflicting auditory cues (i.e. ITD and ILD). Pulki (2001) showed that ITDs and ILDs generated by amplitude panning can indicate sources are in different positions. Benjamin (2007) has since shown this problem is significant in the mid-frequency range of human hearing.

It has long been known in specialist circles that the Ambisonic surround sound technique is superior to pair-wise panning in terms of sound source localisation. This is because it takes into account the different localisation mechanisms used by the auditory system. Unfortunately though, the design of Ambisonic systems for any irregular loudspeaker layout (e.g. ITU 5.1) is complicated. A non-linear system of equations needs to be solved in order to produce a decoder that will output suitable loudspeaker feeds. Recently, however, work by the authors has helped develop improved Ambisonic 5.1 decoders (Moore and Wakefield 2007). This work used a heuristic search algorithm to find good decoder coefficients. The quality of a set of coefficients derived by the search is measured by a fitness function. The fitness function evaluates the performance of the decoders at the central listening position (Moore and Wakefield 2007) using respected models of auditory localisation (Gerzon 1992). The next stage of the authors' work is to improve spatial audio quality for listeners in non-central positions. The following section discusses some of the problems which need to be considered before meeting this goal.

4   OFF-CENTRE LISTENING

A listener in an off-centre position will be nearer or further away from some loudspeakers resulting in time differences between sound waves arriving from each loudspeaker. This leads to the loss of temporal synchronization from the contributing sound waves causing phantom images to be distorted. The main factor which influences a change in a phantom image's quality is the precedence effect. This effect says that listeners will tend to localise a sound source in the direction of the earliest arriving wave front causing the listener to perceive sound as coming from the nearest loudspeaker.

In his classic study of the precedence effect, Wallach (1957) explained that correlated sound waves arriving in close succession will be fused together and heard as a single sound with a single location. The window of time that fusion takes place over is dependent on the nature of the sound, but the lower limit has been estimated at 5ms for short transient sounds, and the upper limit at 50ms for wide-band like speech (Litovsky 1999). For sound arriving before this window of time a single phantom source image will be perceived at a location determined by the contributing sounds. If sounds arrive after this window of time they will be heard as separate sources (i.e. echos).

The influence of the precedence effect in off-centre listening was seen in a recent comparative study of several surround sound panning techniques (see Kearney et al 2007). The authors found that for all techniques sound source localisation was compromised as sound was drawn towards the nearest loudspeaker for listeners in off-centre positions. It should be noted, however, none of the techniques evaluated in this work were specifically designed for off-centre listening.

In order to gain a better understanding of the significance of the precedence effect in surround sound reproduction, a number of different listener positions within the ITU 5.1 layout were evaluated. At each listening position a time difference was calculated between 10 different loudspeaker pairs (see table 1). By doing this every possible loudspeaker pair combination is considered when the loudspeaker layout is left-right symmetrical.

Loudspeaker 1	Loudspeaker 2
Centre	Right
Centre	Right Surround
Centre	Left Surround
Centre	Left
Right	Right Surround
Right	Left Surround
Right	Left
Right Surround	Left Surround
Right Surround	Left
Left Surround	Left

Table 1: Loudspeaker pairs tested for time differences

Two different time difference scenarios where chosen for the evaluation. In scenario 1 if the calculated time difference between a pair of loudspeakers was less than 5ms (i.e. before localisation fusion for transients occurs) then a score of 1 was awarded. If it was above 5ms then it was given zero. In scenario 2 the time difference evaluated was 50ms (i.e. before localisation fusion for longer sounds occurs). If it was less that 50ms a score of 1 was awarded, if it was greater than 50ms a score of zeros was awarded. The time difference information was taken from Litovsky (1999) with 5ms and 50ms by estimated as the upper and lower limits of sound fusion respectively.

By checking the time difference between each loudspeaker pair, a mark out of ten can be given for each position in the listening area allowing performance to be quantified at different positions in the reproduction area. The resulting data has enabled quality maps of the surround sound reproduction area to be compiled (see figure 4 and 5).

The reproduction area in Figure 4 is 10m^2. When evaluating a time difference of 5ms the optimal listening area is approximately 2m^2 which is big enough to accommodate a small number of listeners (see figure 4a). It is envisaged that listeners in this area will perceive phantom sound source images as coming from the approximately the same location. However, when moving away from this area, it is likely that phantom images will become biased towards the nearest loudspeaker to the listener because sound will be arriving earlier. The most problematic listening positions are at the front and to the sides of the reproduction area close to the loudspeakers. However, when evaluating the reproduction area using a 50ms time difference the optimal listening area is extended throughout the entire reproduction area (see figure 4b).

Figure 5 displays the performance in a reproduction area of 20m^2. As might be expected, the same quality pattern is obtained as the 10m^2 reproduction area. However, for time differences of 5ms there is now a much larger problem area because sound waves from opposite speakers now have to travel a larger distance (see figure 5a). This highlights that larger arrays can be more problematic for listeners situated off-centre. However, when evaluating for time differences of 50ms the problem area is significantly reduced and the optimal listening area is big enough for a large number of listeners (see figure 5b).

5   SUMMARY AND FUTURE WORK

Auditory localisation can be measured at the sweet spot using Gerzon's models of auditory localisation. This has enabled surround sound decoders to be optimised for playback in this position. However, localisation quality needs to be evaluated at different positions throughout the reproduction area in order to enable surround sound to be improved for a distributed audience. Steps towards this goal have been made.

A method for evaluating the influence of the precedence effect has been developed. This has enabled quality maps of the reproduction area to be compiled. Quality maps were created for two reproduction areas of different sizes. In both cases 5ms time differences and a 50ms time differences were evaluated. When comparing the 5ms and 50ms evaluations there is a significant difference. For 50ms the listening area is much larger than for 5ms. This suggests that any surround sound playback method would benefit from some form of transient suppression applied to the output signals.

The above model is able to predict the influence of the precedence effect. However, it does not predict sound reproduction quality. Recent work by Poletti (2007) has introduced a method for evaluating sound field reconstruction at off-centre positions. It is envisaged that by including this method in future optimisation strategies as well as the precedence effect model discussed above surround sound decoders will be able to be improved for off-centre listeners. This is the subject of future work.

REFERENCES

DAVIS, M.A. (2003), *History of Spatial Audio Coding,* Journal of the Audio Engineering Society, Vol 51, No. 6, pp 554-569.

THEILE G. (1993), *The New Sound Format "3/2 Stereo".* Presented at the 94th Audio Engineering Society Convention, Berlin, Germany.

BLAUERT, J. (2001), *The Psychophysics of Human Sound Localization.* 3rd Edition. The MIT Press, Cambridge.

RUMSEY F. (2007), *Basic Psychoacoustics for Surround Recording.* Presented at "Illusions in Sound" the 22nd UK Audio Engineering Society Conference, Cambridge, UK.

RAYLEIGH L, (1907), *On Our Perception of Sound Direction.* Philosophical Magazine.

FEDDERSEN, W. E., SANDEL, T. T., TEAS, D. C. and JEFFRESS, L. A. (1957), *The Localization of High-Frequency Tones,* The Journal of the Acoustic Society of America, Vol 29, No. 3, pp 988- 991.

CHENG, C.I., WAKEFIELD, G.H. (2001), *Introduction to Head-Related Transfer Functions (HRTFs): Representation of HRTFs in Time, Frequency and Space.* Journal of the Audio Engineering Society, Vol 49, No. 4, pp 231-249.

MOORE, B.C.J. (2003), *An Introduction to the Psychology of Hearing.* Academic Press

THURLOW, W.R., RUNGE, P.S. (1967), *Effect of Induced Head Movements on Localization of Direction of Sounds.* The Journal of the Acoustic Society of America, Vol 42, No. 2, pp 480- 488.

THEILE, G. (1977), *Localization of Lateral Phantom Sources.* Journal of the Audio Engineering Society, Vol. 25, No. 4, pp 196-200.

MARTIN, G., WOSZCZYK, W., COREY, J., QUESNEL, R (1999), *Sound source Localisation in Five-Channel Surround Sound Reproduction,* Presented at the 107th Audio Engineering Society Convention, New York, USA.

MOORE, J.D., WAKEFIELD, J.P (2007), *The Design and Detailed Analysis of First-Order Ambisonic Decoders for the ITU Layout.* Presented at the 122nd Audio Engineering Convention, Vienna, Austria.

MOORE, J.D., WAKEFIELD, J.P (2007), *The Design of Improved First-Order Ambisonic Decoders by the Application of Range-Removal and Importance in a Heuristic Search Algorithmt.* Presented at the 31st Audio Engineering Convention, London, UK.

GERZON, M.A. (1992), *General Metatheory of Auditory Localisation,* Presented at the 92nd Audio Engineering Convention, Vienna, Austria.

KEARNEY, G., BATES, E., BOLAND, F and FURLONG, D (2007), *A Comparative Study of the Performance of Spatialization Techniques for a Distributed Audience in a Concert Hall Environment.* Presented at the 31st International Audio Engineering Convention, London, England.

LITOVSKY, R.Y., COLBURN, S.H., (1999) *The Precedence Effect,* The Journal of the Acoustic Society of America, Vol 106, No. 4, pp 1633-1654.

POLETTI, M. (2007), *Robust Two-Dimentional Surround Sound Reproduction for Non-Uniform Loudspeaker Layouts.* Journal of the Audio Engineering Society, Vol. 55, No. 7/8, pp 598-610.
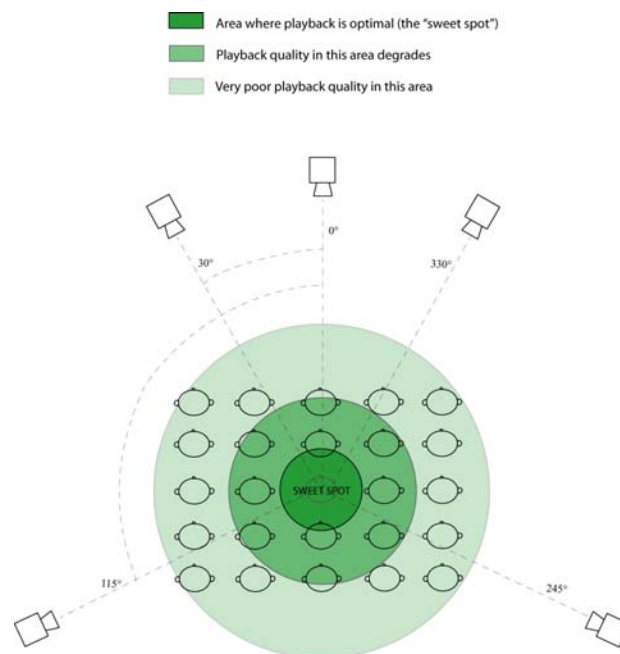
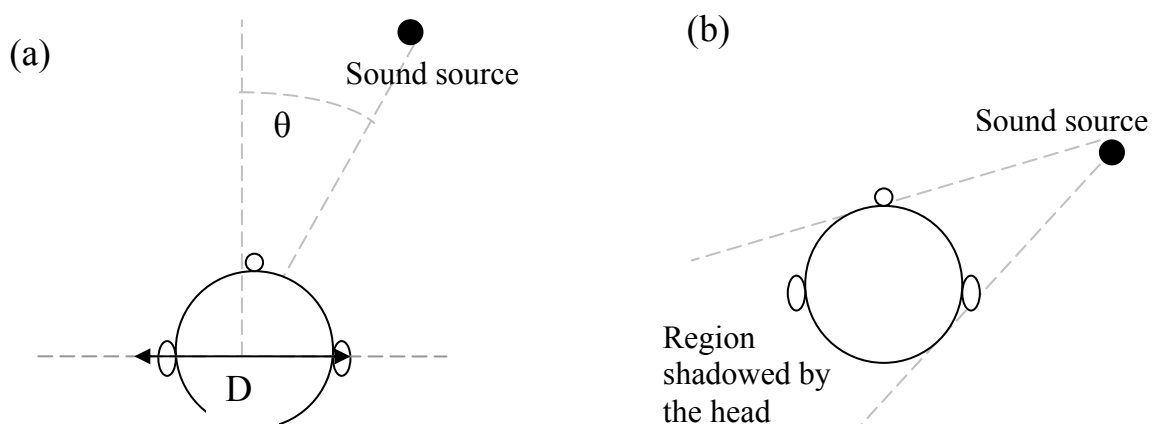Figure 1: Multiple listeners in the ITU loudspeaker configuration

Figure 2: The ITD is dependent on the angle of the sound source (φ) and distance separating the listener ears (D). ILDs are cause by the head casting an acoustic shadow
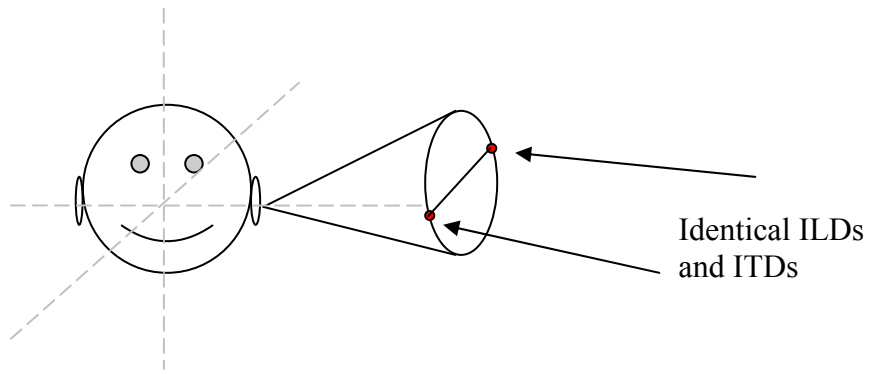
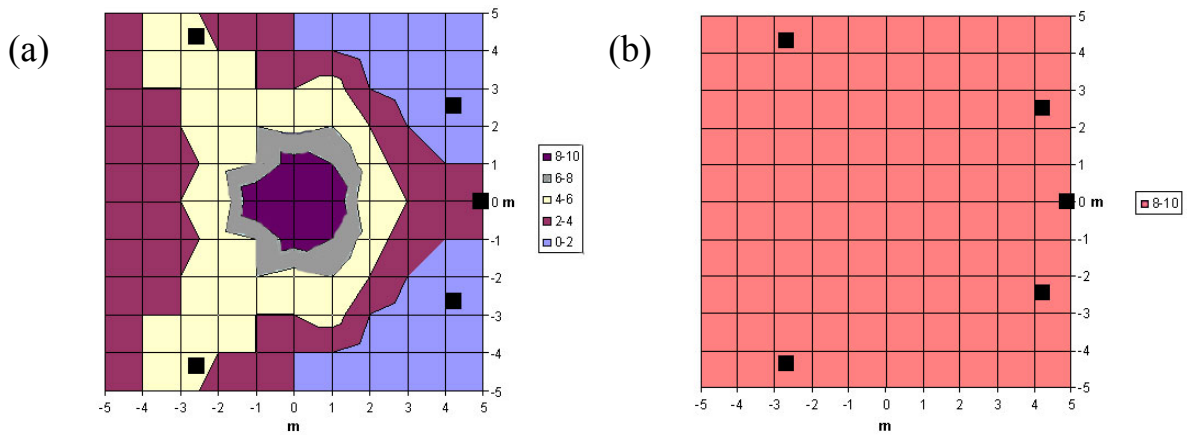**Figure 3: The Cone of Confusion – identical ITDs and ILDs can be perceived.**

Figure 4: Quality maps of 10m^2 reproduction array. The left plot is for a 5ms time difference. The right plot is for a 50ms time difference.
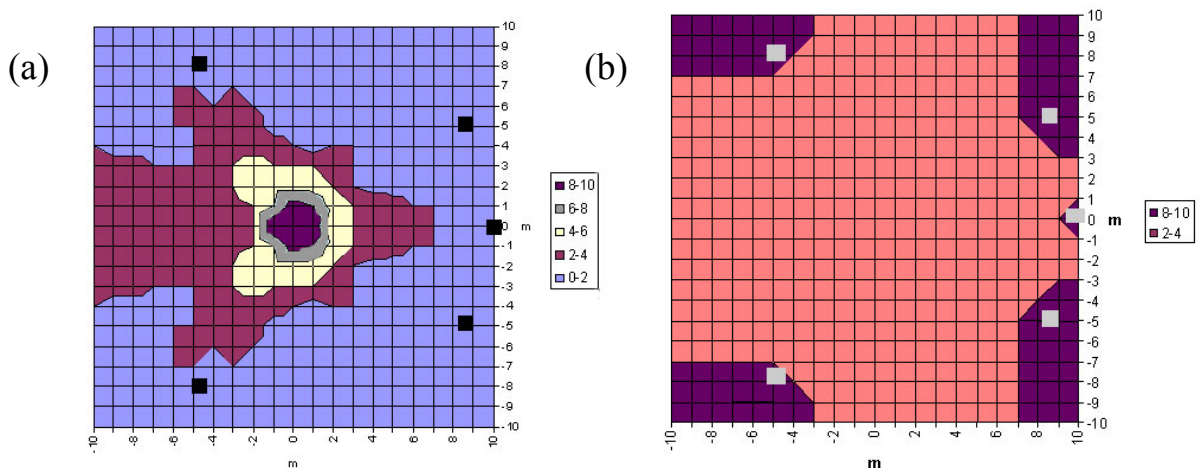

Figure 5: Quality maps of 20m^2 reproduction array. The left plot is for a 5ms time difference. The right plot is for a 50ms time difference.