Localisation of Sound— A New Zealand Revelation

George Dodd
Acoustics Research Centre, School of Architecture, University of Auckland
(g.dodd@auckland.ac.nz)

Yoshimasa Sakurai
Experimental House, 112 Gibbons Rd., RD 2, Kaiwaka

Introduction

Psycho-acousticians now feel they have a good understanding of the main mechanisms used by our hearing system for localising the position of a sound source. The need to generate convincing sound for virtual reality systems has produced a substantial body of research which has identified and confirmed these mechanisms.

Much of this research on localisation has been carried out in controlled conditions (e.g. an anechoic chamber, [1], [2]) where efforts are also made to remove visual cues because visibility of the sound source would be a confounding factor when trying to identify subjects' ability to localise by sound alone. However, whilst there is probably an evolutionary advantage conferred by an ability to localise prey and predators (or other sources of danger) purely by sound, it is unrealistic – at least for sighted humans — to treat sight and sound as being separated elements in the process of localisation.

In the vast majority of cases of attended-to sounds we turn to look at the source and thus we receive a strong visual feedback. This feedback, we may argue, constantly reconfirms our auditory tracking ability and may even be necessary for maintaining a good and accurate functioning of it.

If this latter conjecture is correct we should be able to demonstrate some loss or change in localising ability if we remove or disrupt visual feedback. This would then be consistent with a concept of the cortex as a plastic organ which – as the expansion of the auditory cortex in non-sighted individuals demonstrates – is able to change and adapt as the need or experience of its owner demands.

The purpose of this paper is to report an experience which, we suggest, illustrates such a change happening and therefore supports the idea that our ability to localise may require regular training and updating. It also suggests it may be possible to effect major changes during a surprisingly short period of time.

After a review of localisation mechanisms the particular experience which gave rise to these thoughts is described and analysed.

Processes used in localisation

The hearing system uses a combination of three main methods to localise a source-

a comparison of the intensities arriving at the left and right ears – i.e. it uses the inter-aural level differences (ALD)

a comparison of the times of arrival of the onset of a sound at the left and right ears - i.e. it uses the inter-aural time differences (ITD),

recognition of angle-of-arrival-dependent and movement-dependent spectral details – i.e. it uses a memory of spectral features associated with particular directions of arrival resulting from the interaction of sound waves with the body's geometry producing the so-called anatomical transfer functions (ATF) - also known as head-related-transfer functions (HRTF)

The availability of these depends on the frequency range of the sound radiated by the target source.

Inter-aural level difference

A difference in level between the ears will result if the source is in a position such that the distance to

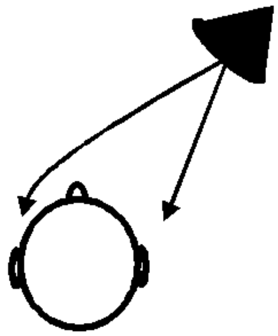

Figure 1. A source position that will result in an ILD

the 2 ears is different (see fig 1). The size of the ILD will depend on (a) the ratio of the path lengths to the 2 ears, and (b) the amount of shading of the contra-lateral ear by the head.

The minimal ILD we can detect appears to be around 0.5 dB independent of frequency [1]. If this were to be the result of path length difference alone then – for a typical size of head – a source would need to be within 10 mm of one of the listener's ears! Thus in the majority of situations path length difference can be discounted, and useful ILD's must be the result of head shadow.

Diffraction around the head determines the depth of shadow for the further ear and hence the ILD depends on the frequency content of the sound. Because diffraction is more efficient at low frequencies the ILD is least at these frequencies. Where the wavelength of incident sound is more than 4 – 5 times head size (e.g 0.6-0.75m) diffraction will be essentially 100% complete and hence we cannot expect ILD's to play any part in the localisation of sounds which have no frequency content above about 500 Hz.

The information on which the direction decision is based when ILD's are not available is derived from the phase difference between the signals at the 2 ears. In effect the hearing system detects inter-aural time differences (ITD).

Inter-aural time difference

Modelling the head as a smooth sphere of radius, a, allows us to make a simple prediction for the ITD. For sources subtending an angle q to the mid-plane (see Fig 2) straightforward geometry shows that the ITD depends only on the radius of the sphere (i.e. head) and θ:

$$ITD = \frac{2\pi a\theta}{180c} \qquad (1)$$

where c= speed of sound.

As we might expect, ITD does not depend on how far the source is from the head[1].

Experiments [Kuhn] have shown that we are able to discriminate the angular position of sources to an accuracy of 1°–2°, at least up until frequencies where the separation between the ears approaches half a wavelength.

When the source is nearly straight ahead (e.g. q = 1°) equation (1) predicts an ITD of 9 ms which is much smaller than the processing time that we associate with hearing (e.g. an integration time of the order of 50 ms for speech). This ability to resolve such small changes in ITD suggests our

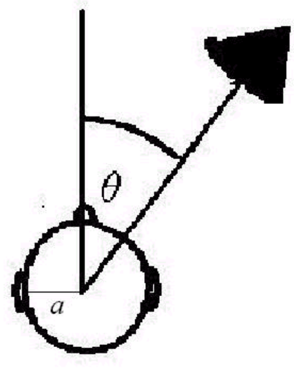

Figure 2. ITD depends on angle of incidence and head size

1. This differs from the expression obtained by Kuhn [4] but is more satisfactory in the limiting case of a source at 90° and is more simply derived.

hearing depends on a peripheral mechanism (i.e. not cortical processing) to do this, and is taken as support for the idea that a cross-correlation is performed between the signals from the 2 cochleas. Correlation of the 2 signals is thought to occur in the superior olive in the mid-brain (suggested by the work of Yin and Chan cited in [3]).

At frequencies beyond the point where half a wavelength matches the width of the head (i.e. frequencies above about 1.5 kHz) there is potential for phase ambiguity to mislead the ear and it is interesting to note that we become insensitive to ITD at these frequencies. So the ear must find other sources of localisation cues for frequencies in this range.

ILD's seem to be the likely source as they increase in a stable manner and become strongly dependent on angle of incidence – certainly for frequencies above about 3 kHz. But a comparison of difference thresholds for ILD's with the change of ILD with angle of incidence indicates that ILD cannot be the basis for our fine angular resolution unless we have frequency content above 3 kHz

The explanation for why we can localise signals which are band limited at intermediate frequencies (i.e. 1 – 3 kHz) is provided by anatomical (or head related) transfer functions (ATF or HRTF) – decribed later in this paper. HRTF's are also the reason why we can localise in the vertical plane, distinguish front and back source positions (i.e. $\theta = 0^0$ and $\theta = 180^0$), and discriminate positions on the so-called 'cone of confusion' which have the same ITD or ILD.

Lateralisation vs Localisation

This is an appropriate point at which to mention lateralization and how it differs from localisation. When hearing stereo recordings through headphones listeners will perceive the position of the sound changing (in response to changes in ILD's and ITD's) but the source of sound is experienced as being "inside the head" and simply swinging between the left ear and the right ear. This is referred to as lateralization.

By contrast, when we localise a source (which necessarily must be some distance from the head) our perception is that it is situated outside the head (and we can estimate its distance away). Our sensitivity to the information in HRTF's is what makes the difference between lateralising and localising a sound.

Head related transfer functions

The waves entering the ear canal of a listener consist of the direct sound plus a host of components scattered from the outer ear, the head, the shoulders and upper torso of the listener. These combine to give a resultant spectrum which has a fine detail determined by the relative amplitudes and phases of these contributions. These amplitudes and phases depend on the direction from which the original sound arrives at the listener, hence there is a fine filtering linked to the direction from which the sound arrives. This fine filtering is measured as an HRTF for each particular direction of arrival.

During the early years of our development we relate the positions of sources (by vision if we are sighted) with what we hear and so unconsciously come to recognise the direction-dependent nuances in the HRTF's. These then become

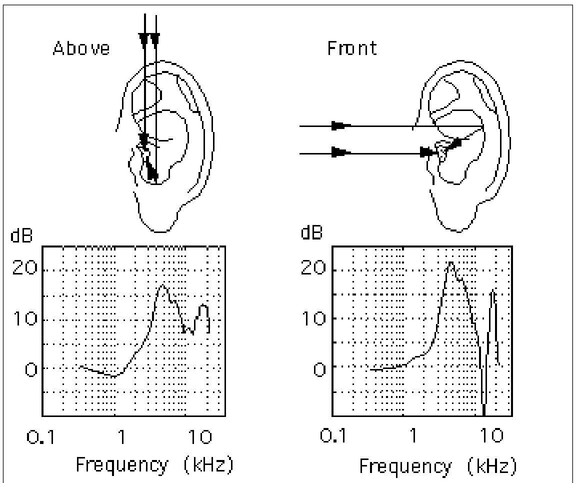

Figure 3. Head Related Transfer Functions measured on a manikin. These show how the response changes as the source moves from above the head to a position in front. (After Duda [5])

available to us to use instinctively as localisation cues.

As a source moves with respect to the head big differences are seen in the HRTF between the frequencies of 5 kHz and 12 kHz. Figure 3 gives 2 examples of HRTF's measured on a manikin and the effects of the changing constructive and destructive interference can be clearly seen.

HRTF's are quite specific to each individual because every listener is physically different. Thus reproduced sound which has been processed with manikin HRTF's in an attempt to recreate directional fidelity can be confounded if the listener's own HRTF's are sufficiently different from those of the manikin. Discrimination of front versus rear positions is particularly affected but there is often difficulty in correctly identifying positions on the conical planes either side of the head which produce identical ILD and ITD (i.e. on the aptly named cones of confusion).

Since HRTF's vary from listener to listener their use must be a learned process for each individual. This process must also be capable of revision and adaptation so as to respond to the changes produced in HRTF's as the body grows from that of a tiny baby into adulthood. If localising is equally necessary at all stages of life then – in principle – this process must be one which continues throughout our lives since it is part of the human condition that we change physically as we age.

The Kaiwaka experience

The events

Early in 2003 Prof Yoshi Sakurai (a member of our Acoustical Society now living in Kaiwaka [6]) observed some surprising changes in localisation when watching screenings of films for the local community in Kaiwaka. These screenings take place not in a purpose-designed cinema but in a large gallery/workshop belonging to a local artist. For convenience the loudspeakers for the sound track are placed at the opposite end of the room from the wall on which the picture is projected (see fig. 4) and high off the floor. However, it was Yoshi's impression that during a film the position of the source of the sound did not always locate at the rear of the room. In particular when there was activity on the screen which related to the sounds there seemed a tendency for the sounds to be localised on the screen.

At the invitation of Professor Sakurai the first author and his wife attended a screening to experience the effect. In reporting the experience it is important to relate the sequence of events. Immediately the film began we felt in no doubt that all the sounds were localised to the rear at the position of the loudspeakers. We watched attentively for the first 5 minutes or so to see if there was any shifting of the apparent source position as the images on the screen changed between landscape scenes and close-ups of the characters. But all sounds seemed to arrive resolutely from behind.

At this point – having, we felt, made our decision on the issue – we began to attend more to the film. We soon found ourselves engrossed (perhaps it is necessary to explain that the film had a French sound track and we were delighted with an opportunity to immerse ourselves in that delightful language once again and brush up our comprehension, and also the content of the film was rather arresting – an apt description would be that it was somewhat 'blue') so we forgot about the purpose of our visit and simply enjoyed the film.

Figure 4. Views of the gallery/workshop (length 11.1m, width 5.8m, average height 5m) at Kaiwaka where the shift of localisation was experienced, (a) View to the front i.e. the screen; (b) View to the rear showing the alcove at the top where the loudspeakers are placed.

At the conclusion – approximately 1½ hours later – the projector was turned off whilst the credits were still running (the film was played from a video cassette through a data projector for the vision and through a separate amplifier for the sound to the loudspeakers) and the video player was left feeding the sound track to the loudspeakers.

It was at this point we had the startling realisation that we now localised this sound quite definitely on the bare, dark screen!

However, on turning towards where we knew the loudspeakers were located we quickly readjusted and the localisation on the blank wall disappeared.

We have not repeated the experience to confirm it and make further investigations but we remain convinced of the validity of the phenomenon. Over the duration of the film we had changed and at its conclusion were responding to our localisation cues in a different way.

Discussion

An explanation for this experience would be that, given the plasticity of our brain and the requirement (noted above) that the process of interpreting localisation cues needs to be modifiable, our instinctive processes had recalibrated our responses. This would have occurred in order to harmonise the sounds with their obvious sources – i.e. the images on the screen. What is surprising, however, is the short time scale over which the readjustment happened.

It must be admitted, however, that in this case since the source at the rear was being relocalised to the front the localisation relied predominantly on HRTF's. These may be relatively easy to recalibrate. If the loudspeakers had been to the side and therefore creating ILD and ITD the process might have proved more resistant. Clearly further experimentation is required.

On reflection, the experience of our hearing responses being mouldable is not unfamiliar. Loudspeakers provide a good example. During so-called A-B comparisons where we switch quickly from one loudspeaker to another all but the subtlest of differences become very apparent. However, we readily adapt to the sound of a particular loudspeaker, e.g. in our home TV or hi-fi, and the result of familiarity with it is that we lose sensitivity to all but its major defects and accept what we hear as valid.

Conclusion

This experience of listening to sound in an unusual situation has prompted the realisation that localisation of sounds is a 'plastic' process. It also suggests that it may be relatively easy to recalibrate our hearing to localisation cues in quite short periods of time. Thus we may conjecture that regular reinforcement by visual (or other) feedback is necessary for maintaining our precision of localisation.

References

1. J.Hebrank & D.Wright 'Are 2 ears necessary for localisation of sound sources on the median plane?' J.Acoust.Soc.Am. (56) 935-938, 1974

2. E. Wenzel et al 'Localisation using nonindividual HRTF's' J.Acoust.Soc.Am. (94) 111-123, 1993

3. W. Hartmann 'How we localize sound' Physics Today Vol 52 (11), 1999

4. G. F. Kuhn 'Model for the interaural time differences in the azimuthal plane' J. Acoust. Soc. Am. (62) 157 (1977).

5. R.O.Duda 'Head Related Transfer Functions' Dept of Elect Eng., San Jose State University (www-engr.sjsu.edu/~knapp/ HCIROD3D/3D_home.htm)

6. Yoshimasa Sakurai 'Welcome to the Nature-friendly Experimental House', (www.ecohouse.co.nz/) ❑

(Continued from page 6)

- Noise is one of the biggest subjective neighbourhood environmental problems
- Survey annoyance responses are consistent over all surveys

The third was the brief paper by Guski, entitled "How to Predict Future Annoyance in Planning?", where he argued that the assumptions that annoyance does not change over time may not be true. This is at least in relation to transport noise where the noise created by individual sources has reduced. He showed, for example, that despite a decrease in the L_{dn} of aircraft, the annoyance of residents exposed to aircraft noise has increased over the years. If this relationship holds true for other sources as well, planning for future noise sources may require calculation of the annoyance trend for the last twenty years to enable a prediction for the future based on extrapolating the statistical trend.

The ideas these three presentations created for me were well worth the almost five days travel time and the two days confined in a metal tube travelling at 900+ km/hr. The next ICBEN World Congress will be held in 2008 in Connecticut, United States of America. Should be well worthwhile if this one is anything to go by.

Terence Moody (abbreviated)