This Journal’s editorial board comprises largely students of visual perception, so as a humble hearing specialist in such august company I thought I’d offer a few thoughts about ways that ideas about visual perception have been used as ‘visual metaphors’ to try to understand hearing. The explanatory failures of such metaphors have been quite spectacular.

One vivid example of such a failure is known in the trade as ‘the lateralisation model of binaural unmasking’. The visual metaphor here is a straightforward observation, which is that when one visual object obscures another from view the objects will lie in the same direction at different distances, whereas when the same objects lie in different directions the nearer object will no longer obscure the other. The auditory observations are first that the presence of one sound can obscure another, an effect called masking, and second that the direction heard for a sound varies with the relative times of arrival and intensities of the sound at each of the ears, i.e. the interaural relationships of the sound. The metaphor fails because the masking is not always reduced when the sounds are heard to come from different directions. Thus, sounds from different directions can obscure one another, while it is also the case that two or more sounds from the same direction are often distinguishable. In the crucial experiments it can be shown that altering interaural relationships in ways that do not change apparent directions of sounds can sometimes reduce masking. Also, there are certain other alterations of interaural relationships that do change apparent directions of sounds but which do not result in masking reductions. Such experiments indicate that the way that sound sources are heard in auditory space does not really resemble the properties of objects in visual space. Indeed, many writers in the field eschew the term ‘auditory space’ in favour of a dichotomy between ‘localisation’, which is the extraction of information about the directions of things, and ‘selection’, which is the separation of information about one thing from information about others. This dichotomy is reflected in the differing processes that are used to model the underlying mechanisms.

A related failure of the auditory-space metaphor is again based on a straightforward visual scenario. When, say, a house is seen from across a lake on a still day, the house’s reflection in the lake is seen as separate from the house itself, and in a different direction. Also, if the reflection is obscured while the house remains visible, the house looks the same. If a breeze should ruffle the surface of the lake, the reflection is distorted, but perception of the house itself is unchanged by this. The auditory perception of reflected sound does not exhibit any of these properties. Although very long delays between a sound and its echo are heard as separate things, in most rooms there are numerous reflections that arrive sooner, and these are not heard separately as they all ‘fuse’ perceptually together with the unreflected sound. Accompanying this fusion of the reflections there are alterations in the character of the single sound that is heard, so that when the nature of the reflecting surfaces is altered there are consequent changes in perception of the sound.

A slightly more sophisticated visual metaphor is the ‘frequency versus time’ graphical display. Here, time is plotted along the abscissa while audio frequency is plotted up the ordinate. This allows you to ‘see’ groupings and separations of sounds and parts of sounds, and you may then wonder whether the Gestalt principles that describe the
visual experience—such as grouping by similarity, proximity, continuation, and so on—can similarly describe what we hear when the sounds are played. Such displays have been considered helpful in showing that principles of perceptual grouping do apply to the organisation of auditory events. This is because the auditory and visual groupings sometimes correspond, as has long been known. Nevertheless, a more literal translation at this phenomenological level is inappropriate in many ways. As an example, consider a diagonal line. Visually there is likely to be 'good continuation' of this line, for various reasons, but a sound that follows such a line on a frequency-versus-time graph does not possess this property. An upward rise of this kind becomes more likely to change direction downward the longer it continues. This is because the frequencies of vibrating sound sources, or of resonators, are typically limited in their range. Such ecological considerations account for there being groupings in frequency-versus-time displays that are seen, but never heard, and vice versa.

The extent to which frequency-versus-time graphs 'work' as a metaphor depends largely on the careful selection of relatively simple stimuli. The problems start to emerge when more complex, but ecologically typical, stimuli are represented in this way. An example is the auditory 'continuity' of a tone that is heard when it is periodically interrupted, and when the gaps are filled by noise. The visual metaphor of a 'picket fence' is often used to describe what is heard here. However, this demonstration only works when the sounds are presented over headphones, and not when they are heard in the reverberant conditions that are typical of everyday listening.

Frequency-versus-time graphs can be helpful in some respects. For example, when speech sounds are viewed with their aid, the paradoxes and puzzles that speech raises for perceptual theories are made stark. One can 'see' that there are illusions in the perception of speech events, in that its perception takes account of the surrounding context. Nevertheless, the explanation for such 'taking account' is not to be had from any visual metaphor.

The rationale behind the ordinate of frequency-versus-time graphs is of course 'Ohm's acoustical law', the idea that the ear performs a Fourier analysis on incoming sounds so that the different frequencies in the sound are filtered into different auditory nerve fibres, which therefore act like 'frequency channels'. The limitations of this particular metaphor are becoming increasingly clear. At one time explanations of diverse auditory phenomena tended to be based on the putative activity in a single frequency channel, a prominent example being the several 'single-channel' theories of the way that the pitches of complex sounds are effected. Nowadays it is clear that even in the simplest of psychoacoustic tasks, such as in the detection of a change in the dB level of a sound with a single frequency component, there is a perceptual synthesis of information from across a range of frequency channels. Indeed, early attempts to measure the characteristics of auditory filters with psychoacoustic methods were plagued by difficulties in finding tasks that depend primarily on information from only a single auditory channel. Having eventually overcome these difficulties this research has demonstrated the fundamentally nonlinear nature of the ear's filters, reflecting as they do the frequency- and level-dependent amplification within the cochlea. Such nonlinearity is a further blow to the Fourier-analysis metaphor, and points up the oversimplifications that inhere in metaphors from frequency-versus-time displays.

In some ways it is hardly surprising that visual metaphors fail to account for auditory phenomena. Hearing gives us information about the environment that we cannot obtain in any other way, including through seeing. From an evolutionary perspective this is necessarily the case, as there would be no selection pressure for a perceptual mechanism whose information simply duplicates that from another sense. Auditory information is therefore unique, arising as it does from small movements that often go undetected by our other senses. Even when a sound source is in full view, its movements
can be invisibly small, such as a cello string on a concert-hall stage. But such sounds loudly fill the whole room. As auditory information is unique, the mechanisms of hearing and auditory perception that are needed to obtain this information are in many ways quite different from the mechanisms of other senses. Indeed, recent work on auditory cognition even suggests differences in mechanism at post-perceptual levels.

From a scientific point of view it is, of course, no bad thing to have considered a failed metaphor. Optimists would say that the processes of science should allow us to acknowledge the error, learn from it, and move on.

Viewed in this way a failed metaphor may be seen to have had heuristic value in that it led to the experiments that were its demise. Moreover, it may be that metaphors represent only the less formal, embryonic stages of scientific explanation, so that formal models should eventually replace them as the science matures. A more pessimistic viewer might point to the decidedly nonmonotonic progress of the discipline, as old ideas tend to be revisited periodically, and this is certainly true of visual metaphors for hearing phenomena. Even so, some progress should still be possible. A trend that seems to be gaining strength recently is towards a sort of ‘learning to live’ with poor metaphors. For example, Huckvale (1996) describes how engineers who developed devices for automatic speech recognition were always aware that the workings of their artifacts were fundamentally different in nature from those of human speech perception. They nevertheless pursued the development along the lines that they did as their aim was purely and simply to make a device that worked. Despite this, Huckvale argues, much new knowledge was obtained about the nature of speech and its perception along the way, without there being any direct pursuit of such knowledge, or of a coherent theoretical structure for it. It would seem then that although a metaphor may not pass muster as a basis for theory, it might serve as the basis for a measuring instrument, and thereby lead to the acquisition of new knowledge. A more prosaic example of this would be the sound pressure level meter. Although at one time the working principles of these meters had some status in loudness theory, it is now well known that the loudness of sounds is at best only loosely related to the dB readings from such an instrument. Nevertheless, these meters continue to be useful instrumentation in auditory research.

Visual metaphors can be attractively neat and simple, and so can be extremely useful when one is starting to get to grips with complex material for the first time. However, if these or other metaphors ultimately fail to capture important aspects of new knowledge, then there is the danger of ‘negative transfer’. That is to say, one might be misguided by inappropriate metaphors to the extent that learning is ultimately more difficult than it would have been without them, because of the need to ‘unlearn’ initial attempts at an understanding. Perhaps that is why there are, to my knowledge, rather few perceptual scientists around these days who make important contributions in both hearing research and vision research.

Anthony J Watkins
Department of Psychology, University of Reading, Reading RG6 6AL, UK

Acknowledgements. I am grateful to Mary Dyson and John Harris for their comments on an earlier draft of this piece.

Reference
Huckvale M, 1996 “Learning from the experience of building automatic speech recognition systems”, University College London, Department of Phonetics and Linguistics, Speech Hearing and Language: Work in Progress, volume 9, pp 133–147