Verbal and Non-Verbal Elicitation Techniques in the Subjective Assessment of Spatial Sound Reproduction

Russell Mason, Natanya Ford, Francis Rumsey, and Bart de Bruyn
Institute of Sound Recording and Department of Psychology
University of Surrey
Guildford, Surrey GU2 7XH, UK

Presented at
the 109th Convention
2000 September 22-25
Los Angeles, California, USA

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

VERBAL AND NON-VERBAL ELICITATION TECHNIQUES IN THE SUBJECTIVE ASSESSMENT OF SPATIAL SOUND REPRODUCTION

Russell Mason, Natanya Ford, Francis Rumsey and Bart de Bruyn,

Institute of Sound Recording and Department of Psychology,
University of Surrey,
Guildford,
Surrey
GU2 7XH
UK
e-mail: r.mason@surrey.ac.uk, natford@genie.co.uk, f.rumsey@surrey.ac.uk
and b.de-bruyn@surrey.ac.uk

Abstract - Current research into spatial audio has shown an increasing interest in the way subjective attributes of reproduced sound are elicited from listeners. The emphasis at present is on verbal semantics, however studies suggest that non-verbal methods of elicitation could be beneficial. Research into the relative merits of these methods has found that non-verbal responses may result in different elicited attributes compared to verbal techniques. Non-verbal responses may be closer to the perception of the stimuli than the verbal interpretation of this perception. There is evidence that drawing is not as accurate as other non-verbal methods of elicitation when it comes to reporting the localisation of auditory images. However, the advantage of drawing is its ability to describe the whole auditory space rather than a single dimension.

I. INTRODUCTION

Aspects of sound quality assessment have been researched using subjective listening tests for a number of years. However, to date these have mostly concerned such features as timbre quality or distortion artefacts [1]. With the increasing number of sound reproduction systems that can deliver an enhanced spatial auditory experience to the consumer, a great deal of research is being conducted into describing and quantifying the auditory spatial attributes and how they relate to audio quality. These auditory spatial attributes are qualities or features of the auditory event that relate to space, in other words the first three dimensions of height, depth and width.

One of the many tasks involved in this research is the elicitation of the subjective spatial effect of various physical parameters of sound reproduction systems from participating subjects. This should enable examination of otherwise hidden information about the way in which people perceive and interpret the sound fields they hear. The elicitation task requires the subject to communicate their perception of the stimulus as accurately and completely as possible.

Elicitation experiments are needed for a number of reasons, including the following. Firstly, one might use an elicitation experiment to create meaningful and reliable scales and categories for grading in further subjective experiments. Secondly, they can be used to extract the detailed parameters of expert knowledge, and examine the correlation between the terms and concepts used by two or more experts. Finally, the elicitation could be used as a means of accessing a subject's perception of a stimulus.

For the purpose of this paper, methods of elicitation are examined in order to obtain from the subject an accurate description of their perception of an auditory event. The only direct way of accessing the perception is if the experimenter is the subject. Whilst these self-testing experiments are useful as a preliminary enquiry, it is generally accepted that for greater reliability with potentially less biased results, a listening panel comprising a number of subjects is preferable. This raises the problem of eliciting the perception of the auditory event from a number of subjects as accurately as possible.

One project which has used elicitation experiments to examine auditory spatial attributes is the Eureka 1653 MEDUSA (Multichannel Enhancement of Domestic User Stereo Applications) project. This has focused on the spatial aspects of reproduced sound including the recording, processing and reproduction of sound. Using methods based on quantitative descriptive analysis (QDA) [2], repertory grid technique (RGT) [3] and verbal protocol analysis [4], a range of verbal descriptors of spatial attributes have been elicited.

However, it may be that relying solely on verbal descriptors is limiting for the subject, and that non-verbal techniques may enable additional information to be communicated. It may also be that non-verbal techniques allow some spatial attributes to be elicited more accurately and reliably.

For the purposes of this paper it is useful to define what the authors mean by non-verbal elicitation. Non-verbal representations naturally complement the limitations of verbal language. For example, it is common that during a conversation a person gestures with their hands to convey information such as size or direction. In addition, written language is frequently accompanied by diagrams or pictures that help to explain a certain point more clearly. These are all types of non-verbal representation and can be argued to be in the perceptual domains of vision and motor action. For a person to draw or gesture involves the use of motor action to move the arm, as well as visual perception to provide a 'feedback loop' that monitors the accuracy of the given motion.

Non-verbal elicitation techniques have been used in auditory experiments. The most common kind is the simple localisation experiment conducted by having the subjects point towards the position at which they perceive the source to be [5], [6], [7]. Localisation experiments have also been carried out using alternative non-verbal techniques such as indicating the perceived position on a plan view [8], [9], [10], and [11]. Further experiments have also examined such spatial attributes as perceived source size using non-verbal elicitation methods [12] and [13].

Evans examined the different types of response that could be used for a localisation experiment, but concluded that more work was needed to evaluate the benefits of each and create a standard [14]. Data is available illustrating the use of pointing methods, as this appears to be the most common technique for eliciting spatial information, however sketch map techniques have been less frequently evaluated and employed. Because of the lack of research on this type of method, an experimenter wishing to explore the options needs to cover a large range of interrelated subjects to examine the relative merits of the techniques, and to assess any bias that the techniques may impose on the data obtained.

This paper is an in-depth literature review, and is intended to provide a starting point for experimenters wishing to expand their range of useful tools. The paper attempts to justify the use of non-verbal techniques, to assess what advantages and disadvantages these techniques may have, to cover certain salient factors that need to be considered when conducting non-verbal experiments, and to indicate where more details can be found. The focus is primarily on using a sketch map elicitation technique for evaluating sound reproduction systems. Even so, some of the arguments and results may be applicable to other types of non-verbal elicitation, and for evaluating some other forms of auditory stimulus where spatial attributes are of interest. This paper is not intended to be a definitive work, but a guide based on the experience and research in carrying out such experiments.

II. COMMUNICATION AND ITS ROLE IN SUBJECTIVE EXPERIMENTS

In a subjective experiment where the participant is used to assess a given auditory event, some form of communication must necessarily take place between this participant and the experimenter. The method of communication could be either written, verbal, or diagrammatic, and occurs at many steps along the experimental procedure from the issuing of instructions to the collection of responses. At each stage, the transfer of information between both parties is subject to numerous errors which can reduce the validity of the results. However, communication cannot be avoided as the listener and experimenter must make known to each other their wishes and thoughts, for otherwise there is no benefit to the study. It is the aim of this section of the paper to discuss in more detail some of the difficulties associated with subjective assessment, primarily those of communication, interpretation and the notion of subjective 'correctness'.

The purpose and ambiguity of language

As human beings we try to make sense of our experiences, we impose structure on the objects around us and by doing so, "impose meaning upon the world" [15]. According to Levy [16] it is for the purpose of understanding and communication that we have developed our highly elaborate systems of event coding and individual hypotheses regarding this coding system. Be it speech, writing, drawing, semaphore, Morse code, or a chain of binary numbers, communication requires a 'language'.

The language most often used in subjective assessment is that of text or speech. According to Ogden and Richards [17], this verbal, word based language is used in two different ways. Firstly, there is a symbolic use, whereby items are identified, catalogued or related to one another and secondly, the language can be used emotively, being selected by the communicator to elicit certain responses in the reader or listener.

Ogden and Richards believe symbolic language to be truly empirical and arbitrary, stating that any word could be invented and used to describe an object, providing that everyone who uses the language agrees. Take, for example, the object known as a 'pen' in the English language[1]. Depending upon cultural and linguistic boundaries the object has many different names, yet the object remains constant in its appearance and purpose. Not only this, but the object could just as easily be known by some other name if history had seen fit to adopt for it an alternative symbol. As John Locke wrote in his 'Essay Concerning Human Understanding' in 1689 [19] (cited in [20]), "When we begin to fix by means of words… abstract ideas… there is a danger of error. Words should not be treated as adequate pictures of things; they are merely arbitrary signs for certain ideas – chosen by historical accident and liable to change."

Cherry [20] is of a similar opinion. In his review of human communication, he states that words are simply signs that have achieved significance by convention. Those who are unaware of, or fail to adopt, the convention simply fail to communicate effectively. Even when the convention is adopted, there is an inherent vagueness to these signs, for example as Cherry muses, when does a shrub become a tree?

Emotive language is believed to be an even less precise communication device than symbolic language. Cherry states that if a word acts solely as an 'emotive stimulant' and does not name objects with precision, then language truly must be a source of communicative ambiguity. As an example he suggests that words like 'democracy', 'happiness' and 'civilisation' are interpreted differently by each individual depending upon their history. This is also true for less emotive words, as Spinelli [15] inquires, if the terms we know for objects were to be removed, what would we perceive a simple object to be? In response he proposes that the object would be something, but the definition or meaning given to that something would have as much to do with the individual and the meaning system employed by the individual as it would have to do with the object itself. Not only is language interpreted uniquely according to the history of those who use it, but the language itself is context dependent. Cherry believes that the full meaning of a word does not appear until it is placed in context, the word's meaning altering according to how it is used, the relationship and communication experience of the communicators and the situation in which the communication is presented.

So it can be seen that anomalies in verbal communication occur as a result of the symbolism of language, the knowledge and personal histories of the communicators, and the context of the communication. These problems are apparent in subjective audio assessment. As Letowski [21] surmises, the sheer number of terms used when describing a sound, "is a blessing for artistic freedom, but a problem when it comes to meaningful communication between people". Furthermore, as each individual uses language according to their history we cannot be certain that participants are using the language in the same way as intended by the experimenter. Guski [22], in his paper assessing the psychological methods for evaluating sound quality and acoustic information wrote, "It should be noted that individual human subjects in psychoacoustic laboratories still have their individual history and may use even common language in a slightly different way than the experimenter intends." This ambiguity is highlighted by the findings of Bannister [23] in a 1962 experiment in which participants were asked to rate 20 photographs of people in terms of seven adjectives. The results showed that although individuals may assign common meanings to a set of adjectives, there is still a chance that these adjectives will be applied differently. For Bannister it was apparent

[1] It is worth noting that although the object described by the word pen remains constant, there are different meanings of this word including: an implement for writing or drawing with ink; a small enclosure in which animals are kept; or a female swan [18].

that there was little agreement between the way the individual participants rated the photographs, even though the adjectives chosen to describe those photographs were universal.

Individuality in language comprehension has thus been shown to be an important variable in communication. One which, in experimental procedures, can lead to indeterminate, inconsistent results. It is therefore important that an individual's 'personal constructs'[2] are considered carefully when devising response scales and the questions to be asked of experimental participants.

From experience to communication

We cannot easily remove language from subjective experiments for, regardless of the associated problems, its purpose is to enable sense to be made of our experiences and the communication of these thoughts to other individuals. With language we make sense of auditory events, translating these events into a, hopefully, meaningful set of terms in order that we may communicate effectively what we have heard. Cherry [20] writes: "The only way to pin down a thought before it can slip away and fly out of the window is to jump on it with both verbal feet, to pin it down with language, by diagrams, or by mathematical symbols though", he continues, "such language may be inadequate". It is further suggested by Cherry that this linguistic inadequacy is a result of the inability of language to represent the subtleties of thought. Furthermore he suggests that not enough words are available to express all experiences. This view is not a unique one; Levy [16] states that something is always lost when words are used to describe events, and Cytowic [25] is of the opinion that not everything we do or know can be expressed in language. The findings of Kelly [26], suggest that when the process is reversed and language is used to promote, rather than to describe an experience, the language used makes us sensitive to certain stimuli and not to others, language therefore moulding our ways of thinking and dealing with events. Along similar lines, Novitz [27] states that, "One's ability to describe an object, and more particularly the way in which one describes it, often affects one's ability to recall it".

If thoughts are difficult to express in language, our interpretation of an event which has been described using language is equally marred, as Cherry [20] declares, "the writer or speaker does not communicate his thoughts to us; he communicates a representation for carrying out this function … Speech is like painting, a representation made out of given materials, sound or paint". Not only this but according to Cherry, the language used by the 'communicator' will only have meaning for us if it represents a continuity of our own experience. This is well exemplified when translating experiences between cultures. Although a grammar book may help us decipher a text in a foreign language, Cherry believes, "we may never fully understand if we are not bred in the culture and society that has moulded and shaped the language". He further explains this opinion stating, "The translator of poetry really has an impossible task", for words in this form are more than just symbols.

Interpreting events and the question of truth

According to Olson and Bialystok [28], the way we cognise spatial events is by assigning to them a structural description. When learning about an object or event, a unique description is constructed. To recall the same event, the appropriate description is retrieved from memory. Olson and Bialystok believe that the structural description assigned to a particular event is dependent upon its particular characteristics, the context in which the event occurs and the prior knowledge of the perceiver. This results in a representation that is not an objective copy of reality but a personal context-dependent interpretation. So it appears that whilst a communication is taking place, the event being discussed (in whatever language) is subject to interpretation not only by the communicator but also the individual receiving the communication. Thus Levy [16] believes that any event, be it verbal, visual or auditory, should be viewed with respect to this wider context. He declares that events themselves do not carry any meaning, the meaning attributed to them being imposed by the interpreter according to context, interest and personal history. According to Spinelli, phenomenologists[3] argue that the process of interpretation must be acknowledged in statements about reality. The possibility of a correct interpretation of an event is denied as it would pre-suppose that there was, in fact, an ultimate reality in any given situation. Instead it is believed that interpretations remain open to alternatives in

[2] "Personal constructs are the system of dichotomous contrasts used by individuals as they try to make sense of their experiences" [24]. Further references to personal constructs can be found in Kelly [26].

[3] Husserl adopted the term 'phenomenology' when developing a science of phenomena, to clarify how objects are experienced and presented to our consciousness.

meaning. Spinelli states that what most of us term a 'correct interpretation' is not based upon objective laws or universally accepted truths, rather it is influenced by the viewpoint of a cultural consensus.

For Levy, the interpretation of an event, whatever the method of communication used, is not a search for the true meaning of the event as this is subject to a vast range of interpretations. Instead he concludes that as everything is open to interpretation: "the only empirical question involved in the evaluation of any analytical system is whether it serves the purpose for which it is being used."

In search of an auditory language

So it may be that another analytical system could communicate some aspects of auditory events better than those already in use. Whether this is indeed the case or not, by using a variety of communication tools we will be able to extract a larger amount of information about the event. As Levy surmises, "Events may be viewed from many perspectives and each may suggest a different interpretation or class membership for the event." What we are in essence trying to do is take the individual event along various different paths until all interpretations of it have been exhausted.

Levy asserts that, "while some events seem to resist description in more than one language system, others are not so recalcitrant. In each case however, we select the language which we expect will maximise our ability to deal with the problem at hand" [16].

As a non-verbal communication method, drawing, or graphically representing auditory events, could be considered. For Novitz, "Pictures play a special and very important role in communication." He states that in some cases, the use of a picture "can reveal in a matter of seconds what it would take minutes to describe… And this, of course suggests that pictorial describing, reporting or explaining differs radically from its verbal counterpart, since in some cases at least, it is much more effective than using words."

III. COMPARISON OF VERBAL AGAINST NON-VERBAL RESPONSES

Most elicitation experiments regarding the attributes of auditory stimuli make use of verbal semantics. Usually this elicitation is in terms of listeners describing the perception of a sound in their own verbal language. The advantage of these verbal descriptors is that they are easy to relate to, and they can be frequently used later as a common scale for many subjects [29]. As verbal language is the principal means of human communication, it is therefore logical that it is the most common form of elicitation.

However, verbal semantics may not be the optimum form for communicating some perceived objects or experiences. There is evidence that if the source is more abstract and difficult to describe, then verbal descriptors will be less forthcoming. Olson and Bialystok surmise that verbal descriptions that are based on a few relatively invariant features are constructed more readily than those based on more variable and complex features [28].

Matthews examined the similarity of sketch maps and verbal descriptions of routes made by children aged 6 to 11 years old. In this case, for all age groups, there was at least five times more information included in the sketch maps compared to the verbal descriptions [30]. Whilst most auditory elicitation exercises use adult subjects, whose vocabulary will be more developed, it highlights the possible difficulty of relying solely on verbal descriptors. A similar study was made of adults, comparing verbal elicitation with sketch map drawing of the layout of a number of American cities. It appeared that, in each case, the subject emphasised the parts that were most easily expressed in that particular form and, in some cases, excluded those that were most difficult [31]. Gärling et al. argue that some aspects of spatial cognition are better represented in a sketch map as it is a more simple task. However one has to bear in mind that drawing ability is a confounding factor [32]. Of course, this is similar to a verbal elicitation experiment where verbal ability is a confounding factor.

Mental imagery

The use of verbal or non-verbal response methods in elicitation experiments is also related to how objects and experiences are conceived in the human mind. As mentioned in the previous section, a great deal of thought is conceived as verbal language. However, also important is the recollection and manipulation of mental images. A

detailed theory of the relative roles of verbal semantics and imagery in human cognition was set out by Kosslyn in 1981, based on a computational theory of imagery [33]. He stated that:

"Although no serious researcher today maintains that images are actual pictures in the head, some still find it reasonable to posit quasi-pictorial representations that are supported by a medium that mimics a coordinate space."

In evaluating the relative use of verbal semantic descriptors or mental images, Kosslyn's theory states that if an event or property has not been considered frequently in the past, then it is likely that the fact retrieval will be in the form of a mental image. Another way to look at this is to assume that both semantic and image retrieval occurs simultaneously. Therefore if a required recollection has been regularly described as opposed to pictured, it will be 'closer to hand' and thus will be conceived initially in terms of verbal semantics. This hypothesis is supported by the following examples. If asked to state how many fingers are on a typical human hand, the verbal answer is readily recounted without the need for a mental image. However, when asked to describe how many windows there are in one's house, it is most usual to imagine moving round the house, counting the windows [34]. How this relates to the perception of auditory cues has not been researched so well. Certainly, if asked to mentally conceive where a sound has come from, the chances are that it will be imagined in the visuo-motor domains arranged around the head in an egocentric manner. In other words, it is most likely that a mental gesture will be made towards the perceived sound location. This will commonly be accompanied by a physical gesture, either by eye or body movement [35]. This is a compelling argument for the use of a visuo-motor response.

As has been mentioned in the previous section, verbal language is sometimes inadequate for communicating what is perceived. More specifically, so much of spatial cognition is subconscious, and not readily explicable in words [28]. This means that a verbal elicitation exercise on spatial attributes could be overly difficult for a subject. Evidence from elicitation experiments regarding the spatial attributes of reproduced sound have shown that in fact some subjects do find it very difficult to verbalise certain attributes [36].

Accuracy of quantitative verbal descriptors

A problem associated with verbal descriptors is the inaccuracy of quantitative perceptual estimates, as examined by Leibowitz et al. [37]. They hypothesised that the large cognitive distance from visually processed information to a verbal descriptor is the cause of large error, and a non-verbal technique may provide greater accuracy. An example of this is when asking a subject to give a size or distance estimation, they will frequently indicate using fingers or arms whilst attempting to verbalise the dimension. Indeed, in the experiment carried out by Leibowitz et al., the use of a non-verbal technique reduced the variance in the results.

The inaccuracy of the verbal response may also be due to the inaccuracy of linguistic descriptors. The physical space can be separated into an infinite number of positions or directions, however the language of space tends to represent only binary alternatives such as up / down, left / right, forward / back, etc. [38]. A number of these terms can of course be combined such as 'up, to the right, forward', and this can be related to given reference points. However, beyond a certain point, accurate linguistic descriptions may be impossible, certainly without external references given by the experimenter. On the other hand, a graphical representation such as drawing allows for different aspects of the space to be elicited compared to ordinary language [28].

The inaccuracy of verbal descriptors has also been found in auditory localisation experiments. Haber et al. found that a verbal clockface, whilst being the easiest to use, resulted in the worst results in terms of accuracy and variability. They also assumed that the verbal response was adding a further cognitive load [39]. Evans added that having subjects name angles of azimuth and elevation was non-intuitive and therefore less accurate due to the additional cognitive processing required [14].

Cognition involved in describing a perception

The cognitive processes required to perceive a spatial attribute of a sound and then provide a related description are worth further consideration regarding the possible effect that the descriptive technique chosen may have on the resulting data.

Blauert described the subject of an auditory experiment as a black box with a number of internal segments, whose contents and processes are as yet undefined [40]. This is shown in Figure 1 below. The perceptual processes are the

processes that convert the changes in pressure at the ears into a perception of an auditory scene comprising a number of scene components[4] and the auditory environment. The analysis includes the cognitive processes of breaking down the scene and further analysing a specific component or attribute as required by the experiment. The descriptive processes are the cognitive steps necessary to convert this perception into a description of some form that can be communicated.

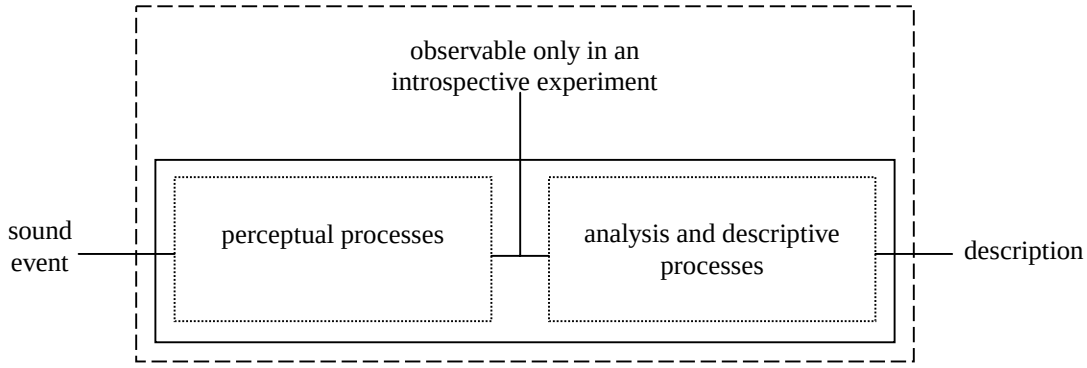

Figure 1: Block diagram of the processes involved in perceiving a sound event and communicating the perception (adapted from [40]).

This model is a very simple overview, and it could be further subdivided into groups of cognitive processes. It may be that the model can be subdivided to consider different descriptive methods used in subjective experiments. Different methods of description may need different cognitive processes and different types of analysis to convert the perceived attribute into the descriptive response. This is included in the modified model as shown in Figure 2 below.

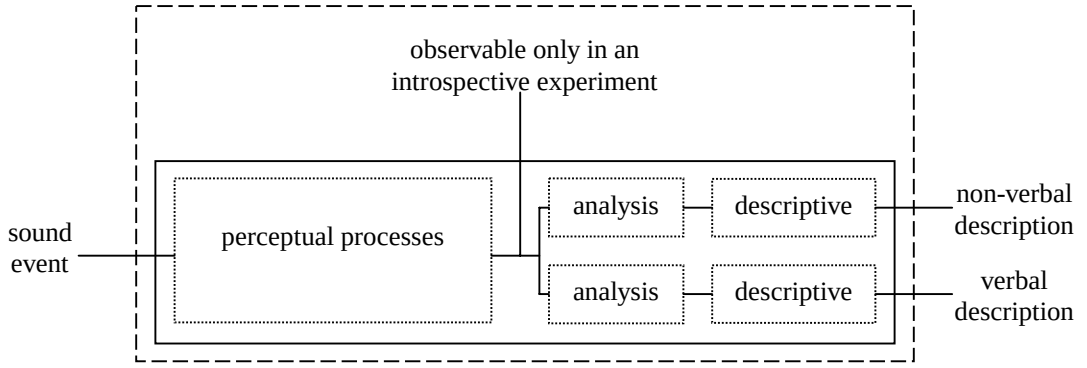

Figure 2: Block diagram of the processes involved in perceiving a sound event and communicating the perception in two different forms (adapted from [40]).

These different processes may then require different amounts of interpretation and computation to be carried by the subject. If this is the case, then an increased amount of required internal computation and interpretation may cause additional variance in the result [41]. This in turn may cause the final description to be further from the perceived event in terms of accuracy and reliability. Therefore, elicitation techniques that require minimum cognitive

[4] For this paper the term 'scene component' has been used instead of the more common terms of 'sound source' or 'sound object'. This is to differentiate that in reproduced sound the source of the sound is in fact usually loudspeakers or headphones, and that for more abstract signals such as noise, separate components may be perceivable with different attributes, though they are part of the same 'object'.

processing and interpretation to convert the perception to a description would be preferable, as long as the required information can still be expressed in that form.

Based on this, it may be that non-verbal elicitation involves fewer layers of interpretation and mental computation than verbal elicitation. Research into human cognitive processing has indicated that the transposition of the perceived auditory space to the visual / motor space of non-verbal representation (as discussed above) contains few steps of interpretation.

Similarities and differences between the senses

Primarily, spatial awareness is independent of any particular mode of sensing [42]. For example, a sound-producing object may be perceived by vision, hearing, or even touch, and these are all interrelated such that the spatial perception of one sense is generally similar and supported by the spatial perception of another sense.

However, the perception and physics of sound and light are obviously very different. Most important for this paper is that the sensing of light by the eye is inherently spatial whereas the transduction of sound in the ear is not. In other words, the spatial properties of vision are mapped spatially on the retina of the eye. In hearing the mapping of the stimuli onto the basilar membrane is not based on location. The response of the basilar membrane that results in the triggering of specific hair cells and therefore the related nerve fibres is based on audio frequency [43]. Of course there is still timing and relative amplitude information in the nerve signals that enable location to be perceived, but this is not mapped to specific peripheral nerve fibres as in vision and touch.

In addition, sound perception is focused more on the source than the sound reflections of the source, whilst vision perception is more focused on the reflections and not usually the light source. Other major differences are that one's visual perception is frontal whilst one's hearing can detect sounds from all directions [44]. Also, human eyes can move independently of the ears, which means that if the stimuli were to be compared between the two senses, variable processing would need to be carried out to compensate for this [45].

Combining the cues from the senses

In order to combine these cues from the different senses, the brain must contain processing that can integrate the vastly different sensory information into a form that can be related and understood. The research of Auerbach and Sperling indicated that auditory and visual spatial perception occupied a 'common space', such that they were perceived on the same spatial map [41]. This means that no translation would be needed to convert auditory spatial perception to visual spatial perception and vice versa. Most importantly, no additional error would be induced by this conversion.

Knudsen et al. provide an overview of the computation that takes place in the brain [46]. The brain uses neural computational maps to combine the cues from the different senses. This computation is laid out in the neurons of the brain in a systematic way across one or more dimensions of the neural structure, much like a map. These maps contain an array of neurons that act as a number of differently tuned filters that respond in parallel to a given stimulus. This provides a place-coded probability distribution that represents the attribute value as peaks of activity on the map. This map-like arrangement means that the information is organised in a form that can be readily accessed by higher-level processors using simple parallel connections.

Computational maps of auditory and visual space

Knudsen et al. also explain that these computational maps are used in auditory localisation [46]. The ear separates the sound spectrum by frequency, and then the frequency bands are processed to elicit the localisation cues. Interaural time difference and interaural intensity difference cues are mapped in different sections of the brain, but in a similar way. The maps contain both a location dimension and a frequency dimension.

These are then combined to create a computational map of auditory space. This is a layer of neurons that are tuned to respond to specific source locations based on the auditory cues. These neurons respond best to broadband signals, as their tuning is a convergence of cues across frequency. There are neural computation maps of both auditory space

and visual space, and they are laid out similarly, enabling a relationship between them to be established relatively simply [43].

The cues from the computational maps of auditory and visual space are combined together, where the auditory space cues are related to the visual space cues. A single cue from either sense results in a triggering of the neurons connected with that perceived position. Two matching cues from the two senses result in a triggering of the neurons with greater intensity, depending on the relative timing [45].

This means that whilst auditory and visual stimuli are very different in their nature, there is physiological evidence that the spatial perception is very closely related between the senses. This suggests that the description of an auditory spatial attribute using a non-verbal representation that involves the visual / motor domains may be a relatively simple task in terms of the levels of interpretation from the perception to the description.

However, to split the model shown in Figure 2 above into further detail, it may be that different methods used for communicating the perception in a non-verbal elicitation experiment may also involve different levels of internal interpretation. This needs to be considered in more detail.

IV. EGOCENTRICITY AND ABSTRACTION

Hart and Moore [47] assert that when an individual endeavours to represent externally an internalised mental reflection of a physical environment, they employ a reference system which allows for their systematic spatial orientation within this environment[5]. For geographical orientation, in other words how an individual relates to objects and spaces around them, Hart and Moore believe this reference system to be initially egocentric.

Howard and Templeton [49] define the term egocentric as the positioning of an object with respect to the body, or some part of the body, of the individual, with no reference being made to any external point or place. The term ‘egocentre’ was introduced by Roelofs [50] to describe the centre, fixed with reference to the body, from which absolute directions are judged, these directions being straight ahead, to the left, to the right, upwards and downwards[6]. Blauert remarks that positional references connected with spatial hearing are usually made in terms of a head-related system of spherical co-ordinates, which shift “in conjunction with movements of the subject’s head” [40]. During the process of this research, it has become clear that the term ‘egocentric’ has more than one interpretation. Roelofs absolute egocentric direction, Blauert’s spatial references and Howard and Templeton’s explication, are joined by a further interpretation whereby egocentricity is defined in terms of the level of abstraction of the elicitation method from the individual. Evans uses this meaning of egocentric in his paper investigating directional listening test responses [14], he asserts,

“If we are giving names to particular directions [as perceived in a listening test], do not assign names based on abstracted angles of azimuth and elevation. Instead, name the directions according to some pattern that listeners will be familiar and comfortable with. In general, this system of response should be inherently egocentric”.

Concluding the paper he pursues this argument stating,

“We have seen that the apparently direct method of having listeners call out the angles of azimuth and elevation of sound sources potentially incurs an excessive amount of inaccuracy due to the non-intuitive and non-egocentric nature of this form of response”.

It is clear from this extract that Evans interpretation of the term egocentric is very different from Roelofs, who uses egocentricity as a purely physical measure. For the purpose of clarity, egocentricity will be used according to the definition of Howard and Templeton and Roelofs in this paper[7], with Evans’ interpretation being referred to as abstraction, intuitiveness or listener familiarity.

---

[5] Wickens and Prevett [48] explain in their paper ‘exploring the dimensions of egocentricity in aircraft navigation displays’ that “The navigator through any space, whether real or virtual must perform two generic types of tasks: Local guidance involves staying on the desired navigational path … and global awareness requires knowing where things are in that space.”

[6] Although Roelofs makes no mention to the direction behind the individual when defining egocentricity, it is suggested that this may be a result of his defining the term with respect to a visual stimulus.

[7] In doing so, the authors do not assume this interpretation of egocentricity to be more ‘correct’ than any other.

A comparison of non-verbal elicitation methods for estimating direction

As geographical orientation is egocentric it can be presumed that the most accurate means of indicating a response to a spatial event would be to use a similarly egocentric elicitation method. Freeman (cited in [47]) states that direction is "represented in the mind" in terms of moving the body either through turning the head or pointing, both methods aligning the individual with the required direction. According to Montello et al. [51], there are a wide range of egocentric response techniques to choose from when studying directional knowledge. It is suggested that participants can point, either with their hands or some other object, turn their heads or eyes towards the direction, rotate their bodies, or in some cases walk along a particular course. An experiment by Montello et al. compared two different 'directional – estimation' methods, one of which was typically egocentric and another which was more abstract. A group of 24 participants were used and were either blindfolded or had their sight partially restricted for the experiment. The participants were instructed to indicate the location of a particular visual object, the position of which they had previously memorised, within the space using either an external 'manual dial' device[8] or by rotating their body to face the item. The results unusually indicated that the 'manual dial' technique was as good as the more intuitive 'body rotating' method for measuring directional knowledge over 360°. Montello et al. suggest that this result was adversely affected by the performance of the blindfolded participants when rotating to face the estimated object location. It is believed by Montello that participants can estimate directions quite well when they can see their feet and the surrounding floor, but once blindfolded have to rely on their short-term vestibular memory to orientate themselves. If a participant falters whilst rotating, they have no access to their initial heading and thus their position with respect to external stimuli is confused. The equivalent results for the 'manual dial' technique showed no such errors, indicating that participants can maintain orientation at all times during pointing.

The results of Montello et al. contradict popular belief about egocentricity and the findings of Haber et al. [39] who conducted a similar study investigating nine methods of direction estimations. Here the greatest accuracy was found for egocentric pointing methods that directed either the nose, chest, or finger at the object, with a manual dial technique producing more variable, and thus less accurate, results.

An important difference between the two studies was in the choice of participants and experiment methodology. Whereas Montello et al. used sighted individuals who, when blindfolded in the experimental situation, recalled the location of an object they had previously been shown, Haber et al. used 20 blind participants to assess the direction of a simple pure tone audio signal reproduced over one of five loudspeakers. Furthermore, the experiment conducted by Montello et al. relied as much on the memory of the individual as to their ability to localise an event

The nine estimation methods employed by Haber et al. included three egocentric 'body pointing' techniques, two egocentric methods where the participant pointed either with a short stick or long cane. Three further methods required the participant to use an entirely external instrument, two of these involving similar manual dials[9] to Montello et al., mounted either on a table or a board attached to the individual, and one where the participant was asked to draw a line on a piece of paper to indicate direction. Finally a single verbal localising method was used, whereby the participant was asked to indicate the direction of the object as if it were a position on a recumbent clock face with 12:00 occupying the location straight ahead.

As previously stated, the results of the experiment showed that the five pointing methods were the most accurate, producing estimations that were substantially better than the two manual dial methods. However these dial techniques were, in turn, more accurate than the similarly variable drawing and verbal techniques.

Sources of error in egocentric and non–egocentric directional methods

According to Howard and Templeton [49], in a pointing experiment the accuracy is limited by the ability of the participant to position their hand in addition to the accuracy with which they can localise the sound. This finding is in agreement with the Haber study where a source of constant error was in the positioning of the 'pointer', in

[8] The device consisted of a single radius line and a rotating radius wire, positioned on the top face of a piece of smooth circular cardboard. The wire was rotated to indicate the estimated direction of the object.

[9] A 15cm long pointer, pivoted at one end, was mounted on a glass board to form the manual dial. The same board, with the pointer removed, was used for the drawing estimation method. It should also be noted that Haber et al. presented target stimuli to the participants in a 180° field of reference to the front of the egocentric space rather than the 360°of Montello et al.

particular the participants found maintaining a straight arm a problem. This error is accentuated when an external pointing device is used, as the participant has to control both their own movement and that of the device.

In the previously introduced paper by Evans [14] he presents a means of eliciting directional responses from listeners characterised by using azimuth and elevation notation to define the position of a sound. Much in the same way as the 'clock face' method mentioned briefly earlier, listeners respond verbally to the stimuli calling out an appropriate direction, for example "30 degrees right, 15 degrees down". In his research of previous studies, Evans found this directional method, although egocentric, produced inaccurate results. It was suggested that as the concept of azimuth and elevation could be unfamiliar to the listener, their use of the method would be unreliable. However the results did not improve when the listener was made familiar with the concept, leading Evans to believe there may have been difficulty translating directional perception into angles due to the lack of intuition involved. Evans suggests a verbal 'clock face' method as used by Haber et al. should be more intuitive to listeners. He proposes that the 12 hours are positioned at 30° intervals, on the clock face as in a conventional clock, but admits that although preferred to the unfamiliar azimuth / elevation method, this limits the directional accuracy of the respondents to 30 degrees. In the method suggested by Haber et al. where a continuous scale was employed on the clock face, the precision problems associated with the Evans technique should have been eliminated. However it was found that although the participants were encouraged to use units as small as one minute, the units of measurement most frequently used were those of 15 or 30 minutes, this inaccuracy being shown in the unfavourable results. The findings would suggest that the participants were unable to use the continuous scale or were unable to translate the direction satisfactorily into the verbal position. Alternatively it could be that a predisposition for time telling in a certain manner was an additional factor in the result. It can be argued that people when estimating time in certain situation approximate to the nearest main unit (i.e. 15, 30, 45 or 60 minutes). Research by Nielson has shown that participants like to quantise their responses to "nice" values, the results of his experiment leading Nielson to believe that "When subjects hear a sound they tend to place the marker on whole numbers or simple fractions thereof" [9]. Evans believes that giving names to particular directions is the origin of many of the problems associated with the clock face technique. An interesting study would be to see if the same results occur for both a verbal and non-verbal methods of elicitation to see if eliminating 'verbal habit' can improve the accuracy of the measurement.

A source of error for the graphical elicitation technique investigated by Haber et al. was not explicitly indicated though it is suggested that once again unfamiliarity with the task could have produced the unfavourable results. Haber et al. propose, "some devices used in the research literature on the blind such as dial pointers or clock face referents are rarely used by the blind in their everyday lives." As a blind person is more likely to be familiar with orientating themselves within an environment than drawing the location of an object on a piece of paper this could be a factor in the results. Whatever the method, Haber et al. conclude by stating that "in selecting response measures for study, researchers would do well to begin with behaviours in which their subjects normally and naturally engage."

Non-verbal methods of eliciting non-directional spatial perceptions

Although the results are by no means conclusive, it appears for the most part that participants can respond best to the location of an unseen object by using egocentric and intuitive pointing responses. However pointing has its limitations, as asserted by Haber et al. "Pointing alone typically defines only direction" [39], and there are cases where pointing does not provide an adequate description of an auditory space, it is obviously difficult to point to an auditory object's size or distance for instance. In such cases, an alternative method of elicitation must be employed. Although geographical orientation is seen as egocentric, for the purpose of these spatial auditory experiments a reliable non-egocentric method would be of benefit. However there needs to be a translation from our egocentric orientation reference to this external description of our perceptions. This translation, as already discussed for the less intuitive directional techniques, is not a simple task and is open to interpretation by the experimental participant.

Translating egocentric spatial perceptions into external representations

According to Evans [14], "an extremely elegant mechanism for listeners to give their responses with complete directional freedom is [to] use a graphical representation of apparent direction". Furthermore a plan sketch enables the whole auditory scene to be displayed at once. However in order to do this the participant has to translate their perception from sound to vision (the similarities and differences of these two senses having been discussed

previously) and from inside their head to a graphical plan on a piece of paper, or computer screen. Not only this but a further complication occurs as a result of translating three-dimensional space into two dimensions. As Arnheim asserts "The pictorial difficulty with which the child has to come to grips is the fact that only two of the three spatial dimensions can be represented directly in the picture plane" [52]. It is for this reason that Arnheim believes that however useful plan drawings are for inferring information about internal representations, they should not be seen as a simple translation from perception, as "misinterpretations are inevitable if the picture is considered a more or less correct replica or derivative rather than a structural equivalent of the object in terms of the medium." Arnheim suggests that when three-dimensional space is represented in two dimensions, the simplest and most characteristic aspect is depicted for each object. In studying the scribblings of children, he also found that space existed only in the two dimensional plane with objects being large, small, close, or distant, to the left or right with nothing in the drawings that distinguished between a flat or a voluminous object. Arnheim found that "The spatial qualities of a dinner plate are not treated differently from those of a football, and all things lie at the same distance from the observer." Following on from Arnheim's research it could be inferred that when depicting graphically in plan view an auditory space, the concept of object depth will be difficult to represent, although left-right width, object size[10] and proximity of the object to the listener will be less so.

The problem of visually depicting internal spatial perceptions has also been studied by Shemyakin (cited in [47]) who found that when children aged between 6 and 8 years draw a plan of some locality familiar to them they usually do this by means of a route map. These routes typically began at the edge of the paper closest to their body and were drawn away from the child so that left and right turns were as they would experience it with their real position in space. When Shemyakin asked the children to draw the routes towards themselves, essentially inverting the plan, it was noticed that the number of errors increased. Although defining the spatial attributes of a sound event does not expressly suggest the drawing of route maps, there is an element of correlation as a listener in a subjective auditory assessment will, on many occasions, be judging the relative locations of auditory events within a space. As this event placement can be considered a form of 'mapping', it is prudent to contemplate the implication of the plan on which these mapped representations will be made. Further to the direction of the plan, a point of note is that participants may find it easier to construct a plan of what they have heard when they have reference to fixed 'landmarks', for instance the position of the walls within the room or the position of their outstretched arms. Hart and Moore [47] suggest that this occurs as a result of a fixed system of reference whereby orientation within space is partially co-ordinated by the use of landmarks.

Minimising problems of abstraction and egocentricity – concluding remarks

So it can be seen that non-verbal methods of eliciting spatial perceptions from listeners are subject to complications on at least two levels. Initially there needs to be a translation from an egocentric point of reference to a plan representation in all but the simplest direction based experiments. Secondly there is the question of how intuitive the elicitation method is for the participant, more errors occurring as this method becomes less intuitive. Ways of minimising errors arising from these difficulties involve easing the translation from egocentric to plan by orienting the view of the response sheet to be that of the listener's perspective and by adding scaling information. Furthermore an acceptance that a graphical response depicting a plan view is limited to the representation of two-dimensional information should keep the technique from being used inappropriately. It is the belief of the authors that perspective too may play a misleading part in the conversion from an egocentric point of reference to a plan view. How do we know if an auditory object, represented smaller on a plan, where it is at a greater distance from the listener than an object in the foreground of larger size, is really smaller? More questions must be asked in order that an intuitive and unambiguous graphical technique for conveying auditory depth, distance and height information can be created. For now, although non-verbal elicitation techniques are proposed as an alternative to the highly problematic verbal methods, it is clear that these methods are selective in the information they can provide, and the results are open to interpretation.

V. ANALYSIS AND INTERPRETATION OF THE RESULTS

The methods used for interpreting and analysing the results of a non-verbal elicitation experiment are very different to those used for a verbal elicitation experiment. A verbal elicitation experiment would allow for analysis of the

[10] Where all objects are perceived to be at the same distance from the participant, (see concluding remarks)

elicited words, and the use of these words as scales or categories in further experiments. The work of Berg on the Repertory Grid Technique shows a good overview of the analysis and interpretation possible [3], [53], [4] and [54]. The numerical results of subjects rating auditory stimuli using such scales can then be analysed using familiar statistical analysis tools such as t-tests or analysis of variance (ANOVA) as described in the standards documents [55] and [56].

Any mathematical analysis relies on the data being represented in a numerical form. Unfortunately, most non-verbal experiment procedures result in data that is not numerical and therefore the data needs to be converted to a numerical form to enable mathematical analysis. This conversion is possible for most experiment types, but must be carried out with care, and may not be able to represent the full information elicited from the subject. This section of the paper will consider some of the methods available for this conversion, along with a consideration of some of the potential problems.

Conversion of elicited results to a numerical form

The method used for interpreting and analysing the results depends on the experimental procedure used. If the experiment is a simple localisation experiment carried out using an egocentric pointing method, then the resulting data can be transformed into a set of error angles away from the true direction, as shown in [39]. This data set can then be entered into conventional statistical analysis. A slightly different method was employed by Damaske, who plotted charts of actual location against perceived location, and denoted the percentage of choices at a particular position on the plot as points of various sizes [57] and [58]. In addition, a special branch of statistics termed ‘circular statistics’ exists that allows analysis of angular data. More information can be found in [59].

The data elicited in sketch map experiments has to be transformed to a numerical representation if statistical analysis methods are required to examine the data. It may be argued that the data does not need complicated statistical analysis as it is already in a form that can be examined intuitively. In other words, it is easier to comprehend the data contained in a graphical representation than the data contained in a table of numbers [60]. However, conventional data analysis techniques may be required in order to simplify description of the data and to enable inferences to be drawn.

Analysis by scene component

One might assume that the sketch map of an auditory scene represents an actual map of perceived positions [61]. If this is so, the dimensions of the attributes can be measured to create numerical data. It is worth considering the type of attributes that can be measured from an elicitation exercise that has used a sketch map method. Assuming that information in three dimensions has been represented, then the following data for each component in the scene may be displayed and therefore may be measurable.

Attribute	Dimensions
Position of centre of scene component	Azimuth Elevation Distance
Size of scene component	Width Height Depth

The position of the centre of a scene component is commonly termed localisation. This is a single value for each of the three dimensions. The measurements are made with reference to the egocentre and are therefore azimuth, elevation and distance. Measuring the position of a single point of a scene component allows the positions of the different scene components to be compared, even though they may not be all of the same perceived size. The centre position is a logical standard to use for each of these measurements.

The size of the scene components can also be measured in terms of width, height and depth. The method used for measuring the height and width of the scene component needs to be chosen with care. This is because the distance of

the scene component from the egocentre determines the ratio between the subtended angle and the height or width represented on the sketch map. This is shown in Figure 3 below.

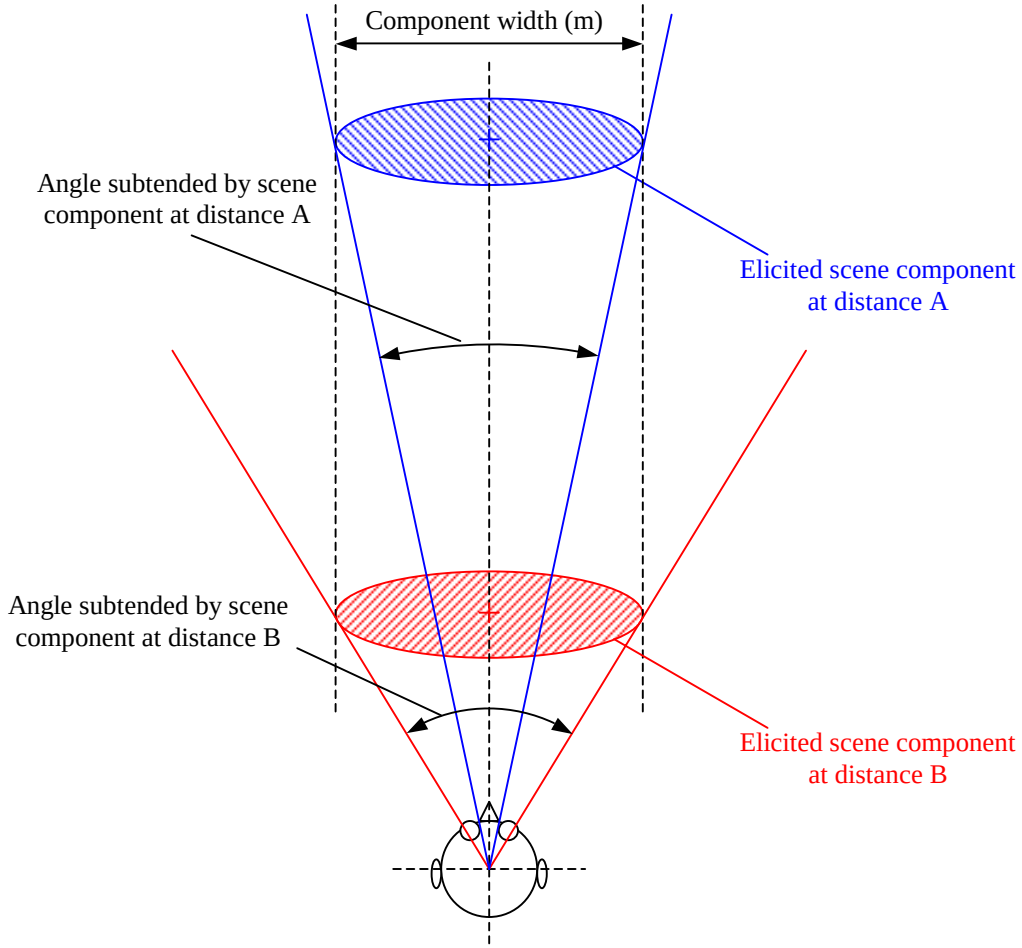

Figure 3: Diagram depicting the difference in subtended angle of the same scene component at two distances.

If it is assumed that the perceived height or width is depicted most accurately as a dimension, then it is advisable that these attributes are measured as a length rather than as a subtended angle. This eliminates the confounding factor of source distance, which is judged very differently between subjects, especially if there is no reference stimulus [9].

The width of a scene component should logically be measured along the dimension parallel to the left / right plane with respect to the head. In a similar manner, height and depth should be measured up / down and front / back respectively. However, when measured on a sketch map, the actual measurement depends on whether the head is considered to be fixed with respect to the response sheet, or free to rotate.

This can be examined using width as an example. If the head is fixed, then the width can always be measured in the dimension parallel to the left / right axis of the head as may be depicted on the response sheet. However, if the head is free to rotate to point towards scene components that are to the side, then the width dimension that is parallel to the left / right plane of the head will also rotate. For the purposes of this paper the two measurement strategies will be termed ‘fixed’ and ‘free’, and are shown for a width measurement in Figure 4.

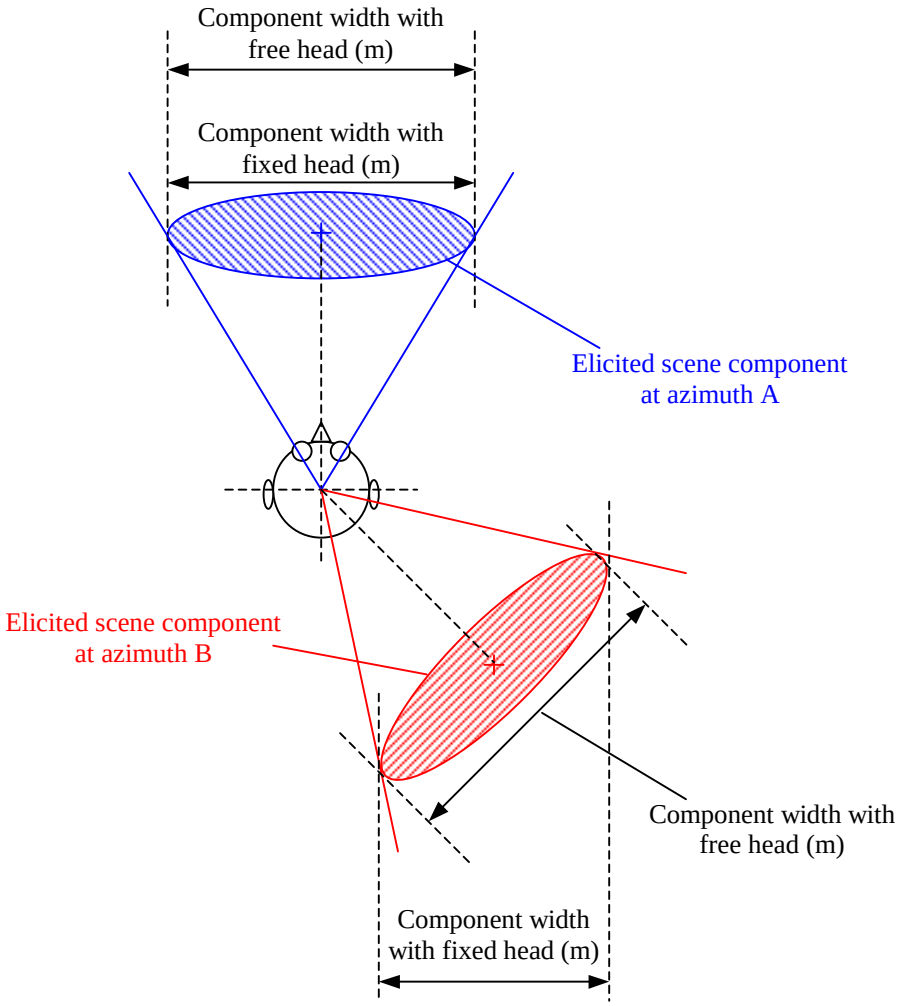

Figure 4: Diagram depicting the measurement of width of two scene components at two positions around the head using two measurement strategies.

The choice of measurement strategy depends on the experiment. If it is expected that one or more scene components will be perceived to be within the head and overlapping the egocentre then it may be that the fixed measurement strategy is most appropriate. This is because it is unlikely that the head will be turned to consider the dimensions of this component. If, however, it is expected that the scene components will be externalised and arranged all around the subject, then it may be most appropriate that the free measurement strategy is used. This is because the subject may turn their head to face each scene component as they consider it.

There is of course a problem of comparing measurements made using the different strategies. Therefore the experimenter must consider the results from all the stimuli when making the decision of how to measure the dimensions of height, width and depth.

Figure 5 and Figure 6 below show the two measurement strategies for a two-dimensional plan view.

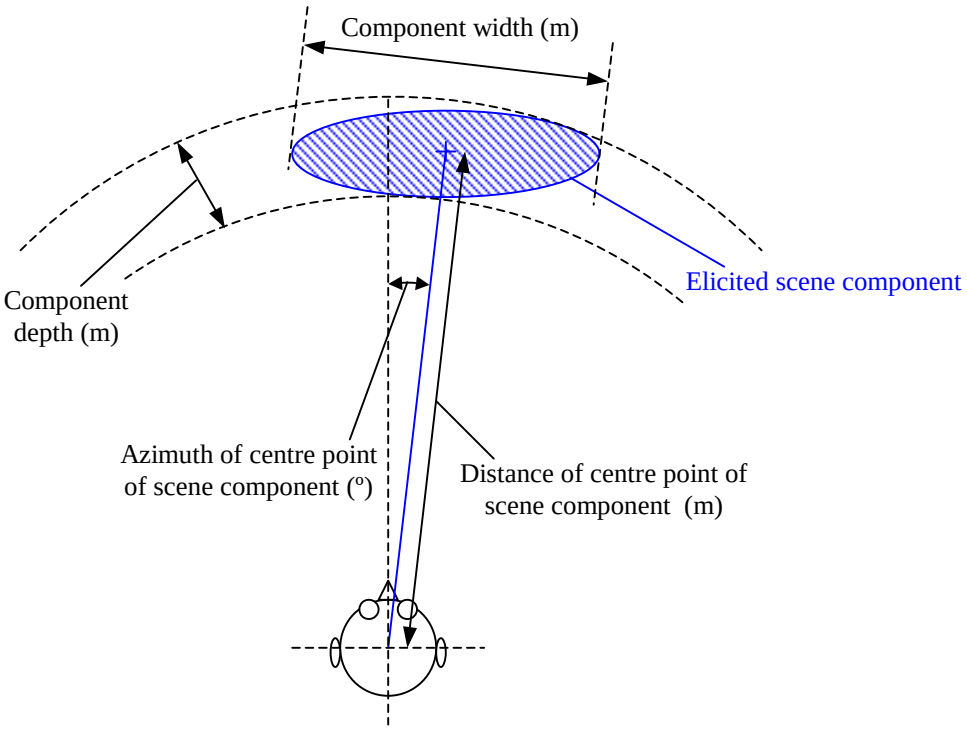

Figure 5: Diagram depicting the measurements made using the free measurement strategy of azimuth, distance, width and depth of a scene component represented in a sketch map plan view

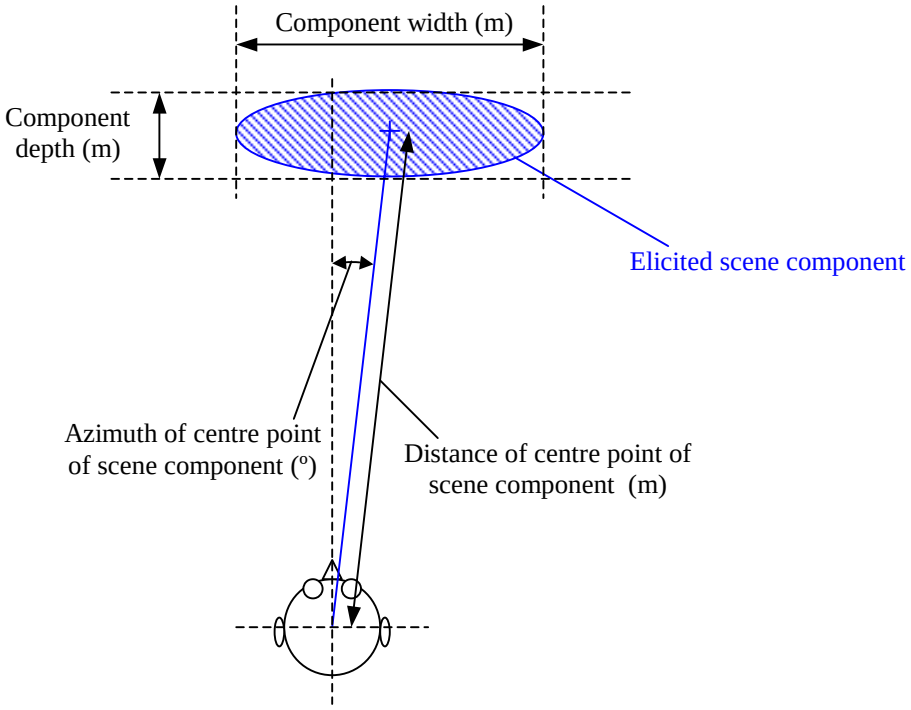

Figure 6: Diagram depicting the measurements made using the fixed measurement strategy of azimuth, distance, width and depth of a scene component represented in a sketch map plan view

Depending on the shape of the representation of the scene component, simple height, width and depth attributes may not describe the component accurately. If this is the case, then more complex measurements could be taken, though this makes the task of comparing the attributes of the scene components more difficult.

Assuming that the primary scene components will be perceived to be located in front of the subject, then for ease of mathematical analysis the azimuth should be measured across a range of –180º to +180º, with 0º directly in front of the subject. This eliminates the numerical break that would occur by measuring clockwise from directly in front of the subject.

Analysis of the whole scene

The attributes for each scene component can be analysed separately and give useful information as described above. In addition, they can be combined to provide information related to the whole scene. Such attributes are listed below.

Attribute	Dimensions
Overall dimensions of scene	Width Height Depth
Number of scene components	Count

The dimensions of the complete scene can be measured in a number of ways depending on the information that is required by the experimenter. Examples include the width of the frontal image of a sound reproduction, the angle subtended by a frontal image that extends to the sides of the subject, the depth of a specific part of the scene or the whole scene from front to back, or other measurements either of the whole scene or specific parts. Each of these dimensions could be measured either from the centre points of the scene components at the outside edges of the measured dimension, or from the outside edges of these components.

The number of scene components represented can also be a measured factor. For example, a stimulus presented using one reproduction system may result in some scene components being masked whereas presentation using a different reproduction system will reveal them. This is a factor that can be measured and analysed.

Once any of these attributes has been measured and represented in some numerical form, they can be analysed using conventional statistical methods. The choice of method depends on the experiment requirements including whether measuring differences between subjects or differences between stimuli.

Density plots

It is often necessary to combine the results from a number of subjects or test runs in a visual form for further examination. The most popular technique for representing this data from sketch map elicitation experiments is as a density plot. Density plots are a summation of the data from a number of subjects or test runs. They can be used when the subject has drawn points or areas to represent a scene component. For each test run, a response in a particular area of the response sheet is counted as a 1. A number of these response sheets are summed to give a density plot. If two response sheets include a response at the same point they sum to give a value of two. If more or less response sheets have responses at a certain point, then they sum to give the respective value. A simple example is given in Figure 7 and Figure 8 below.

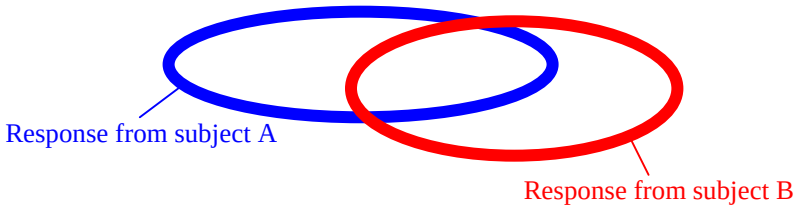

Figure 7: Example sketch map responses from two subjects depicting the perceived size and position of a scene component

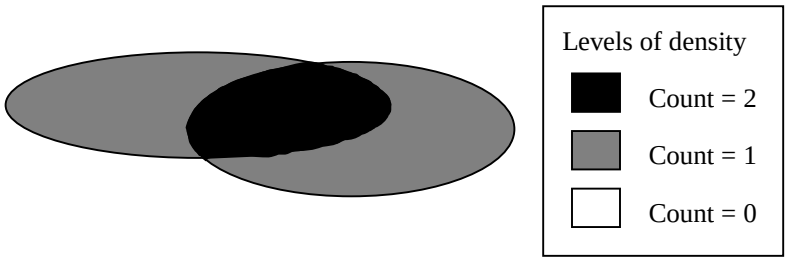

Figure 8: Example density plot calculated from the two responses shown in Figure 7 above

Blauert and Lindemann used a similar technique. They required the subjects to draw the edge of sound sources on paper. The resulting circles on each subject's sheet were filled with black, and then placed one by one in front of a camera and exposed onto the same frame of film. The result was a density plot showing the darker regions where more subjects had judged the sounds to be placed, and lighter regions where less subjects had judged the sound to be positioned [12]. This can be achieved more simply by computer analysis of the results, although there does not appear to be any standard software package for this.

The representation of the data in a density plot has a number of uses, even though it is difficult to analyse the results mathematically. Firstly, it is useful to get an overall impression of the data. Any trends in the data should be apparent with an intuitive overview of the data in a density plot. Secondly, the density plot allows the experimenter to examine the repeatability of a single subject over a number of identical stimulus presentations, or the amount of agreement between a number of subjects.

This repeatability can be calculated mathematically from the density plot. The area that each density level occupies on the plot needs to be calculated. This can then be entered into the following equation.

$$\frac{\sum_{n=1}^{N}(n-1)A_n}{(N-1)\sum_{n=1}^{N}A_n}$$

A_n = area of response with a density level of n
N = number of summed response sheets

Using the density plot shown in Figure 8 as an example, the area covered by a density of 1 (A_1) is approximately 800 mm^2 and the area covered by a density of 2 (A_2) is approximately 200 mm^2. As there are two response sheets,

N equals 2. Entered into the equation above, this gives a result of 0.2. If the separate responses that make up the density plot are very similar, then the resulting value will be close to 1. However, if the separate responses that make up the density plot are very different, then the resulting value will be closer to 0.

Finally, the density plot allows the experimenter to examine aspects of the data that cannot be easily represented mathematically on one response sheet. These can again be compared either between subjects or between stimuli. Examples of attributes that can be examined include the distribution of the scene components and whether there are any effects such as angular distortion at any point of the scene, or extremes of this such as a 'hole in the middle' of the image.

Normalisation

When visually comparing a number of density plots, it would be useful if the effect of the individual subject could be removed in order to focus on the differences between the stimuli. Some form of normalisation of the data could enable this to be carried out. There are arguments against normalisation of data in any type of statistical analysis, and this is no exception.

The simplest type of normalisation is in the dimension of depth or distance. The comparison made between the responses of the subjects to establish the normalisation factor should be for a single stimulus. This is to ensure that the measurements are comparable. From this a specific key scene component should be chosen as a basis for normalisation. The key scene component needs to be selected based on whether all the subjects have represented it in their response, and whether the responses are in some way comparable (i.e. whether the scene component has been represented at the same azimuth and elevation within a reasonable tolerance). The distance from the egocentre to the centre of the key scene component can then be measured to give the normalisation measurement. For scene components that are perceived to be located within the head and overlapping the egocentre, the maximum size of the component can be taken as the normalisation measurement. Once the normalisation measurements have been determined, they can be converted into a set of scale factors. The responses then can be re-drawn with the distances from the egocentre scaled by the normalisation scale factors.

This will then enable trends to be examined across a number of separate stimuli without the confounding variable of different size and distance perceptions and representations of the separate subjects. It must be borne in mind that this normalisation results in data that is no longer a representation of absolute position. It can however be considered as a set of relative positions to be compared between stimuli. It is similar to the z-transformation of numerical scaling data as recommended in [55]. The result of z-transformation is to convert the sets of numerical results from the separate subjects into a set of data with a mean of 0 and a standard deviation of 1. This removes any absolute value of the scores and reference to the outside world, but is a relative measure within the data set of which the z-transform has been calculated.

Using non-verbal results as a prompt for verbal elicitation exercises

It is also possible to use the information elicited in a non-verbal experiment as a prompt for a verbal elicitation experiment. For instance, the first part of the experiment can use a sketch map method to elicit non-verbal data. These sketch map results can then be returned to the subject for them to verbally describe the differences between the representations they have drawn, and the reasons for the differences. This process may help the subject to analyse the auditory scene in a slightly different way, and therefore enable different verbal descriptors to be elicited. It may also help the subject to concentrate on location-based spatial attributes if these are required by the experimenter.

VI. DISCUSSION

To summarise, this paper has attempted to justify the use of non-verbal elicitation methods, and has examined numerous details that need to be considered when conducting elicitation experiments using non-verbal techniques.

Communication and its role in subjective experiments

The communicative nature of subjective assessment cannot be avoided, however, as outlined in the opening section of this paper, communication, and in particular the use of a verbal language, is open to interpretation. From the listener's individual understanding of what is required of them in the assessment, through their perception and subsequent communication (via whatever medium) regarding what they perceived, to the analysis of what has been communicated, and the recording of this representation using the experimenter's own personal constructs, it is clear that what is later digested by the reader, who may or may not fully understand the language of the communication, cannot be thought of as a correct interpretation of the initial auditory event. Rather, the event has been subjected to a series of interpretations with effects similar to those achieved by a game of 'Chinese whispers'. By using a variety of methods there will be an increased redundancy in the information elicited, and by not relying on one method of communication the risk of misinterpretation is reduced.

Comparison of verbal against non-verbal responses

It may be that non-verbal elicitation methods may be preferable to verbal elicitation for communicating some attributes of auditory events. Evidence was discussed that suggested that certain perceptual attributes are difficult to describe using some methods of communication, and it was shown that responses are more forthcoming using the medium in which they are easiest to describe. The limitations of language were discussed, especially with reference to describing locations and positions. It was debated that the representation of auditory space using non-verbal methods may be closer to the perception than using verbal methods. This is due to the similarity of the internal neural processes between the auditory and visual senses. Therefore the representation using non-verbal methods may need less internal interpretation and conversion than verbal methods.

Egocentricity and abstraction

Two basic problems associated with non-verbal auditory assessment were discussed, namely those of communicating an essentially egocentric experience by using an externalised plan view and questions surrounding the intuitiveness of the experimental elicitation method. It was found that the most egocentric and intuitive methods of locating an object within a space were those which produced the least variance in results. However these methods are limited to only being able to describe the direction of an object from the egocentre. Therefore, when other auditory spatial attributes need to be investigated, less egocentric methods of elicitation must be used. The necessary translation from egocentric to plan view was discussed, including ways of increasing the accuracy of the results, for instance orienting the view of the plan response sheet to be the same as that of the listener's, and the inclusion of a meaningful scale on the plan.

Analysis and interpretation of the results

As the responses elicited from a non-verbal experiment will be very different to those elicited from a verbal experiment, the methods of analysis and interpretation will be different. This was discussed and the types of mathematical data that can be obtained from non-verbal responses were considered. Methods of measuring the results from sketch map experiments were examined, and forms of analysis were outlined. In addition, the possibility of using the sketch maps as a prompt for verbal elicitation was discussed.

Limitations of non-verbal elicitation methods

For the advantages that have been outlined above, non-verbal techniques are still limited in their scope. The use of either pointing or sketch map methods can only express quantitative physical dimensions of perceived locations and sizes of scene components. Even these are open to interpretation with the representation of object size raising the question of how individuals deal with perspective.

It may be that if the localisation performance of a reproduction system is accurate, then all spatial attributes will be recreated accurately [1]. If this is the case, then a non-verbal localisation experiment will be sufficient to test the

spatial quality of a sound reproduction. Even so, the separate scene components and their attributes are sometimes combined into single descriptors that are not purely location-based, such as envelopment and spaciousness. These cannot be easily expressed in a non-verbal elicitation experiment, even though they may be interpreted from the results obtained. In addition, such factors will be valuable to assess in what manner a reproduction with imperfect localisation will be compromised in terms of the perception of the reproduced sound.

It is also difficult to represent the reverberation or ambience of a scene using a non-verbal elicitation experiment. The perceived position of the reverberation could be drawn, though often this is difficult to determine. Other spatial attributes that have been elicited in verbal elicitation experiments include qualitative and emotive sentiments such as confined / open, natural / unnatural and prefer / don't prefer [54]. These cannot be easily represented using non-verbal methods, if at all.

The limitations of non-verbal elicitation methods may be an advantage in some experiments where the experimenter wishes to limit the possible responses to purely location-based attributes. However, if the experimenter needs to examine the whole auditory spatial perception then it is advisable that both verbal and non-verbal techniques are used.

Conclusion

It appears from the work shown above that carefully conducted non-verbal elicitation experiments are as valid as verbal elicitation experiments, and are useful in eliciting certain attributes and perceptions that are difficult for the subject to describe verbally. However, the type of experiment and the information required should ultimately decide the choice of verbal or non-verbal elicitation methods.

Further Work

The authors plan to use the information included in this paper to conduct two separate elicitation exercises. The first experiment is to elicit the spatial attributes of programme material played over a number of pairs of loudspeakers, in order to examine the differences between loudspeaker types, loudspeaker positions, and listener positions. This will use a range of both verbal and non-verbal elicitation methods. The second experiment is a non-verbal elicitation exercise examining the perception of mono noise signals with the addition of various interaural time difference modulations. These will be presented to the subjects over headphones.

There are two major differences between the stimuli used in these experiments. Firstly, the method of presentation is different, with one using loudspeakers and the other using headphones. Secondly, the programme material is different, with one experiment using musical programme material in which there are recognisable scene components that can be referred by the subject to a known reference, and the other experiment using abstract noise samples that will be unfamiliar to the subject. Because of this, the methods used for the elicitation exercise will differ.

Further details of the experiments and the complete results will be reported in due course.

ACKNOWLEDGEMENTS

The authors would like to thank the MEDUSA team, especially Søren Bech for comments on the paper and Andi Hodgson for his editing.

Russell Mason would like to thank David Meares and BBC R&D for sponsorship and provision of equipment, the MEDUSA team for their thought-provoking discussions and guidance, the AES Educational Foundation for financial assistance and Lin Oskam for her patient proof reading.

Natanya Ford would like to thank New Transducers Ltd. for sponsorship.

REFERENCES

(1) Rumsey, F. 1998: 'Subjective Assessment of the Spatial Attributes of Reproduced Sound', **Proceeding of the 15th International Audio Engineering Society Conference**, Copenhagen, Denmark, pp. 122-135.

(2) Bech, S. 1999: 'Methods for Subjective Evaluation of Spatial Characteristics of Sound', **Proceedings of the 16th International Audio Engineering Society Conference**, Rovaniemi, Finland, pp. 487-504.

(3) Berg, J. and Rumsey, F. 1999: 'Spatial Attribute Identification and Scaling by Repertory Grid Technique and other methods', **Proceedings of the 16th International Audio Engineering Society Conference**, Rovaniemi, Finland, pp. 51-66.

(4) Berg, J. and Rumsey F. 2000: 'In search of the spatial dimensions of reproduced sound: Verbal Protocol Analysis and Cluster Analysis of scaled verbal descriptors', **Audio Engineering Society Preprint**, 108th Convention, preprint no. 5139.

(5) Thurlow, W. R. and Runge, P. S. 1967: 'Effects of induced head movements on localization of direct sound', **Journal of the Acoustical Society of America**, vol. 42, pp. 480-487.

(6) Oldfield, S. R. and Parker, S. P. A. 1984: 'Acuity of Sound Localisation: a topography of auditory space', **Perception**, vol. 13, pp. 581-617.

(7) Middlebrooks, J. C. 1992: 'Narrow-band sound localization related to external ear acoustics', **Journal of the Acoustical Society of America**, vol. 92, pp. 2607-2624.

(8) Feree, C. E. and Collins, R. 1911: 'An experimental demonstration of the binaural ratio as a factor in auditory localization', **American Journal of Psychology**, vol. 22, pp. 250-297.

(9) Nielsen, S. H. 1991: 'Depth perception – finding a design goal for sound reproduction systems', **Audio Engineering Society Preprint**, 90th Convention, preprint no. 3069.

(10) Møller, H., Sørensen, M. F., Jensen, C. B. and Hammershøi, D.1996: 'Binaural technique: do we need individual recordings?', **Journal of the Audio Engineering Society**, vol. 44, pp.451-469.

(11) Begault, D. R. and Wenzel E. M. 2000: 'Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source', **Audio Engineering Society Preprint**, 108th Convention, preprint no. 5134.

(12) Blauert, J. and Lindemann, W. 1986: 'Spatial mapping of intracranial auditory events for various degrees of interaural coherence', **Journal of the Acoustical Society of America**, vol. 86, no. 3 (March), pp. 806-813.

(13) Martens, W. L. 1999: 'The Impact of Decorrelated Low-Frequency Reproduction on Auditory Spatial Imagery: Are Two Subwoofers Better Than One?', **Proceedings of the 16th International Audio Engineering Society Conference**, Rovaniemi, Finland, pp. 67-77.

(14) Evans, M. J. 1998: 'Obtaining Accurate Responses in Directional Listening Tests', **Audio Engineering Society Preprint**, 104th Convention, preprint no. 4730.

(15) Spinelli, E. 1989: **The Interpreted World: An introduction to Phenomenological Psychology**, (London: Sage Publications).

(16) Levy, L.H. 1963: **Psychological Interpretation**, (New York: Holt, Rinehard and Winston).

(17) Ogden, C.K., and Richards, I. A. 1966: **The Meaning of Meaning**, (London: Routledge & Kegan Paul).

(18) Oxford University Press. 1998: **The New Oxford Dictionary of English**, (Oxford: Oxford University Press).

(19) Locke, J. 1689: **An Essay Concerning Human Understanding**, (London: Ward, Locke and Bowden).

(20) Cherry, C. 1966: **On Human Communication**, 2nd edition, (Cambridge, MA: The MIT Press).

(21) Letowski, T. 1989: 'Sound Quality Assessment: Concepts and Criteria', **Audio Engineering Society Preprint**, 87th Convention, preprint no. 2825.

(22) Guski, R. 1997: 'Psychological Methods for Evaluating Sound Quality and Assessing Acoustic Information', **Acustica**, vol. 83, pp. 765-774.

(23) Bannister, D. 1962: 'Personal Construct Theory: A Summary and Experimental Paradigm', **Acta psychologica**, vol. 20, pp. 104-120.

(24) Landfield, A.W. 1968: 'The Extremity Rating Revisited Within the Context of Personal Construct Theory', **British Journal of Social and Clinical Psychology**, vol. 7, pp. 135-139.

(25) Cytowic, R. E. 1993: **The Man Who Tasted Shapes**, (London: Abacus).

(26) Kelly, G. A. 1963: **A Theory of Personality: The Psychology of Personal Constructs**, (New York: Norton).

(27) Novitz, D. 1977: **Pictures and Their Use in Communication**, (The Hague: Martinus Nijhoff).

(28) Olson, D. R. and Bialystok, E. 1983: **Spatial Cognition: The Structure and Development of Mental Representations of Spatial Relations**, (Hillsdale, NJ: Lawrence Erlbaum Associates).

(29) Deese, J. 1965: **Structure of Associations in Language and Thought**, (Baltimore, USA: Johns Hopkins Press).

(30) Matthews, M. H. 1985: 'Young Children's Representations of the Environment: a comparison of techniques', **Journal of Environmental Psychology**, vol. 5, pp. 261-278.

(31) Lynch, Kevin. 1960: **The Image of the City**, (Cambridge, MA: MIT Press).

(32) Gärling, T., Selart, M. and Böök, A. 1997: 'Investigating Spatial Choice and Navigation in Large-scale Environments', in Foreman, N. and Gillett, R., eds.: **A Handbook of Spatial Research Paradigms and Methodologies – Volume 1: Spatial Cognition in the Child and Adult**, (East Sussex: Psychology Press).

(33) Kosslyn, Stephen Michael. 1981: 'The Medium and the Message in Mental Imagery: A Theory', **Psychological Review**, vol. 88, no. 1, pp. 46-66.

(34) Shepard, Roger N. 1966: 'Learning and Recall as Organization and Search', **Journal of Verbal Learning and Verbal Behavior**, vol. 5, pp. 201-204.

(35) Warren, D. H. 1970: 'Intermodality Interactions in Spatial Localization', **Cognitive Psychology**, vol. 1, pp. 114-133.

(36) Berg, J. 2000: Personal communication with the authors.

(37) Leibowitz, H., Guzy, L. T., Peterson, E. and Blake, P. T. 1993: 'Quantitative perceptual estimates: verbal versus nonverbal retrieval techniques', **Perception**, vol. 22, pp. 1051-1060.

(38) Ogden C. K. 1932: **Opposition: A linguistic and psychological analysis**, (Bloomington: Indiana University Press).

(39) Haber, L., Haber, R. N., Penningroth, S., Novak, K. and Radgowski, H. 1993: 'Comparison of nine methods of indicating the direction to objects: data from blind adults', **Perception**, vol. 22, pp. 35-47.

(40) Blauert, J. 1997: **Spatial Hearing: The Psychophysics of Human Sound Localization**, 3rd edition, (Cambridge, MA: MIT Press).

(41) Auerbach, C. and Sperling, P. 1974: 'A common auditory-visual space: Evidence for its reality', **Perception and Psychophysics**, vol. 16, no. 1, pp. 129-135.

(42) Kritchevsky, M. 1988: 'The elementary spatial functions of the brain', in Stiles-Davis, J., Kritchevsky, M. and Bellugi, U., eds., **Spatial Cognition: Brain Bases and Development**, (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 111-140.

(43) Stein, B. E. and Meredith, M. A. 1993: **The Merging of the Senses**, (Cambridge, MA: MIT Press).

(44) Bregman, A. S. 1990: **Auditory Scene Analysis: The Perceptual Organization of Sound**, (Cambridge, MA: MIT Press).

(45) Knudsen E. I. and Brainard, M. S. 1995: 'Creating a unified representation of visual and auditory space in the brain'. **Annual Review of Neuroscience**, vol. 18, pp. 19-43.

(46) Knudsen, E. I., du Lac, S. and Esterly, S. D. 1987: 'Computational maps in the brain', **Annual Review of Neuroscience**, vol. 10, pp. 41-65.

(47) Hart, R. A. and Moore, G. T. 1973: 'The Development of Spatial Cognition: A Review', in Downs, R. M. and Stea, D., eds.: **Image and Environment: Cognitive Mapping and Spatial Behavior**, (Chicago: Aldine Publishing).

(48) Wickens, C.D. and Prevett, T.T. 1995: 'Exploring the Dimensions of Egocentricity in Aircraft Navigation Displays', **Journal of Experimental Psychology**, vol. 1, no. 2, pp. 110-135.

(49) Howard, I. P. and Templeton, W. B. 1966: **Human Spatial Orientation**, (London: John Wiley & Sons).

(50) Roelofs, C.O. 1959: 'Considerations on the Visual Egocentre', **Acta Psychologica**, vol. 16, pp. 226-234.

(51) Montello, D.R, Richardson, A.E, Hegarty, M. and Provenza, M 1999: 'A Comparison of Methods for estimating directions in egocentric space', **Perception**, vol. 28, pp. 981-1000.

(52) Arnheim, R. 1956: **Art and Visual Perception**, (London: Faber and Faber).

(53) Berg, J. and Rumsey, F. 1999: 'Identification of perceived spatial attributes of recordings by repertory grid technique and other methods', **Audio Engineering Society Preprint**, 106th Convention, preprint no. 4924.

(54) Berg. J. and Rumsey F. 2000: 'Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction', **Audio Engineering Society Preprint**, 109th Convention.

(55) ITU-R BS 1116, 1994: 'Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems', **International Telecommunications Union**, Recommendations ITU-R BS 1116, pp. 276-297.

(56) IEC 268-13, 1987: 'Sound system equipment - Part 13. Guide for listening tests on loudspeakers'.

(57) Damaske, P. 1970: 'Directional dependence of spectrum and correlation functions of the signals received at the ears', **Acustica**, vol. 22, no. 4, pp. 191-204.

(58) Damaske P. 1971: 'Head-related two-channel stereophony with loudspeaker reproduction', **Journal of the Acoustical Society of America**, vol. 50, no. 4 (October), pp.1109-1115.

(59) Batschelet, E. 1981: **Circular Statistics in Biology**, (London: Academic Press).

(60) Weiss, Neil A. 1989: **Elementary Statistics**, (Reading, MA: Addison-Wesley).

(61) MacEachren, A. M. 1995: **How Maps Work: Representation, Visualization and Design**, (New York: Guilford Press).