Available online at www.sciencedirect.com
SCIENCE DIRECT

Cognition 87 (2003) 225–233

COGNITION

www.elsevier.com/locate/cognit

Discussion

In defense of the theory of indispensable attributes

David Van Valkenburg*, Michael Kubovy

Department of Psychology, The University of Virginia, Charlottesville, VA, USA

Received 12 June 2001; received in revised form 17 December 2001; accepted 23 December 2002

1. Introduction

Neuhoff (in press) has criticized Kubovy's (1981) Theory of Indispensable Attributes (TIA), which is the one of the building blocks of Kubovy and Van Valkenburg's (2001) theory of auditory objecthood. Specifically, he claims that (1) "simple frequency separation does not ensure the formation of auditory objects" and that (2) "frequency variation is not a necessary condition for auditory figure-ground relationships". Neuhoff is not the first to criticize TIA. Handel (1988) offered similar criticisms (to which Kubovy, 1988, replied), and we have received an abundance of personal communication, some of which brings up similar issues. For this reason, we are grateful to have the opportunity to clarify aspects of the theory which have been misunderstood and to correct some of our errors.

We divide our response into two sections. In the first, we clarify our theory. Although Neuhoff's critique is focused on a particular aspect of our theory (i.e. the role of frequency in auditory objecthood), in order to reject this portion of our theory he must also reject our ideas about the nature of grouping and objecthood. That is why we (1) summarize our theory of auditory objecthood, (2) specify the role of the TIA in our theory of auditory objecthood and state exactly what the TIA claims, and (3) discuss other implications of our theory. In the second section of our response, we consider Neuhoff's claims and his counter-examples.

2. Objects, grouping, and the TIA

2.1. Auditory objecthood

Object perception is generally not the result of one modality (Gibson, 1966, 1979, 1982;

* Corresponding author. Department of Psychology, P.O. Box 400400, The University of Virginia, Charlottesville, VA 22904-4400, USA.
E-mail address: dlv6b@virginia.edu (D. Van Valkenburg).

0010-0277/03/$ - see front matter
doi:10.1016/S0010-0277(03)00005-2

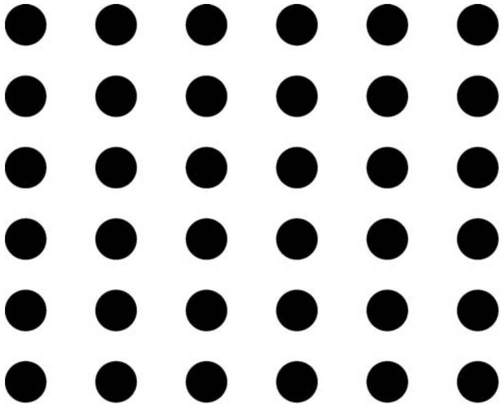

Fig. 1. An example of a set $\mathscr{E}$: a lattice of 36 dots distributed over a medium $\mathscr{M}$, the plane. Perceptual grouping (PG) by spatial proximity produces a phenomenal partition (PP): six columns (*blocks*) of dots.

Stein & Meredith, 1993), yet most researchers focus on one modality. As a result, we have much modality-specific information, but few frameworks that enable us to discuss and compare results across modalities (Jones, 1993, is an exception). The purpose of Kubovy and Van Valkenburg (2001) was to provide a modality-neutral definition of objecthood.

There are two parts to the theory. Firstly, in a manner analogous to Milner and Goodale's (1995) dual pathway conception of the visual system, we believe that the auditory system is both physiologically and functionally decomposable into two systems, or information streams: an anterior-ventral stream which deals with determining 'what' is in the environment, i.e. with grouping and object formation, and a posterior-dorsal stream for determining 'where' things are in the environment, i.e. with spatial localization (Rauschecker, 1997, 1998a,b; Rauschecker & Tian, 2000; Romanski, 2003).[1] *Our theory of objecthood* (Kubovy & Van Valkenburg, 2001) *is applicable only to the 'what' part of the auditory and visual systems.*

The second part of the theory consists of a modality-neutral definition of objecthood. To this end, Kubovy and Van Valkenburg (2001) defined a perceptual object, be it auditory or visual, in terms of its susceptibility to figure-ground segregation. A number of researchers (e.g. Bregman, 1990; Brochard, Drake, Botte, & McAdams, 1999; Izumi, 2002) have used the concept of figure-ground segregation to describe and study auditory objecthood. When they do, they do not discuss the logic and assumptions involved. In the following section, we examine this further.

2.2. Grouping and claims of the TIA

Figure-ground segregation implies that some parts of the environment are perceived to

[1] Evidence for this assertion is still accumulating, and it is not universally accepted (Middlebrooks, 2002; Zatorre, Bouffard, Ahad, & Pascal, 2002, offer balanced reviews).

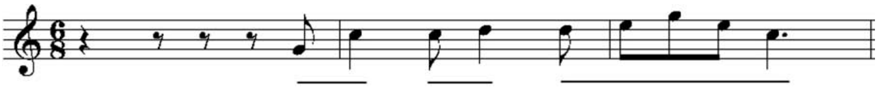

Fig. 2. Another example of a set $\mathscr{E}$: the first phrase of *Pop Goes the Weasel* distributed over two media, frequency and time. PG by temporal proximity produces a PP of three sub-phrases (underlined).

go together whereas others do not. When elements in the environment *are* perceived to go together, we say that perceptual grouping (PG) has occurred.

The purpose of the TIA is simple: to specify necessary conditions for PG to occur. It assumes that PG is a transformation, which has an *input* and an *output*.

The input to PG is a set of discrete elements, $\mathscr{E}$, distributed over a medium, $\mathscr{M}$. One example (Fig. 1) is a case of visual grouping: the set $\mathscr{E}$ consists of dots. These dots are distributed over a plane in space, which is the medium $\mathscr{M}$. A second example (Fig. 2) is a case of auditory grouping: the set $\mathscr{E}$ consists of tones. These tones are distributed over frequency and time, i.e. two media. We call these media *indispensable attributes* (IAs). The TIA tells us which stimulus dimensions can serve as IAs.

The output of PG is a *phenomenal partition* (PP) of $\mathscr{E}$ into subsets (called *blocks*), E_1, E_2, …, E_m (Figs. 1 and 2).[2]

We usually talk of grouping *by* a feature: proximity, similarity, good continuation, common fate, etc. In general the elements in block i share at least one feature, F_i (it is possible that $F_i = F_j$ for some pairs of blocks $\{E_i, E_j\}$).

The description of any PP, even the simplest, requires three concepts: a set of discrete elements ($\mathscr{E}$), a set of one or more media over which these elements are spread ($\mathscr{M}$), and a set of features ($\mathscr{F}$). A standard way to describe a PP is to say that the elements of $\mathscr{E}$ are grouped by $\mathscr{F}$. Such descriptions elide the media $\mathscr{M}$. We therefore recommend that this formula be amended to read: the elements of $\mathscr{E}$, spread over $\mathscr{M}$, are grouped by $\mathscr{F}$.

Such a careful formulation is beneficial because it eliminates any doubt about whether a feature is playing the role of medium or of grouping feature. For example, when we talk of grouping by visual proximity (which is a spatial property) we may be unaware of the fact that space plays two roles: it is that over which the elements are distributed (the medium), as well as that in which distances are not uniform (the feature).

The criterion that an attribute must satisfy to be an IA is that in its absence (and the absence of any other IA for that modality) perceptual numerosity is impossible.

> *If* you distribute elements over a medium *and* perceptual numerosity is perceived, *then* the attribute is indispensable.

Kubovy and his colleagues have – by and large – based their claims on which attributes are IAs and which are not on thought experiments. Their conclusions are summarized in Table 1. The TIA makes no claims aside from those summarized in the table.

We believe that Neuhoff (in press) has misunderstood the central point of the TIA. As we will show in the second part of this response, Neuhoff has failed to appreciate the

[2] A partition of a set X is a subdivision of X into subsets which are disjointed and whose union is X.

Table 1
IAs for vision and audition

	Vision	Audition
The proper description of PG takes the form: grouping *of* **perceptually numerous**	**spatio-temporal entities** *by* ⟨one or more $\mathscr{F}$s⟩[a], *distributed over* ⟨space, time⟩.	**frequency-temporal entities** *by* ⟨one or more $\mathscr{F}$s⟩, *distributed over* ⟨frequency, time⟩.
The set $\mathscr{M}$ of IAs *over which* a set $\mathscr{E}$ of	visual entities can be distributed has two members: **space** and **time**.	auditory entities can be distributed has two members: **frequency** and **time**.
The set $\mathscr{F}$ of features *by which* a set of	visual entities can be partitioned is large: spatial proximity, similarity (color, luminance, shape, …), good continuation, common fate, …	auditory entities can be partitioned is large: timbre (i.e. synchrony, intensity, harmonicity, frequency, proximity, attack/decay, …), space (interaural time differences, interaural level differences, interaural phase differences, …), …
There can be no grouping *in*	color or shape, only *by* color or shape.	space or timbre, only *by* space or timbre.
Object boundaries cannot be formed *in*	color or shape, they are formed *by* color or shape *in* space and/or time (Shih & Sperling, 1996).	space or timbre, only *by* space or timbre.

[a] The notation, '⟨a, b, c, …⟩' means 'at least one element in the set {a, b, c, …}'.

important distinction between the features $\mathscr{F}$ *by* which grouping occurs and the IAs, or media $\mathscr{M}$ *in* which or *over* which the elements are distributed. In addition, we will argue that Neuhoff has ignored the fact that we hold time, in addition to frequency, to be an auditory IA.

2.2.1. A note on terminology

In Kubovy and Van Valkenburg (2001) we incorrectly used the term 'pitch' instead of 'frequency' in our discussion of the TIA. It is an elementary fact that pitch refers to a perceptual quality, whereas frequency represents a physical quantity. When components of a sound (the set of elements $\mathscr{E}$) are grouped by harmonicity (a feature: $\mathscr{F}$) over a medium ($\mathscr{M}$ = frequency and/or time), the resulting percept can have a pitch. Many combinations of $\mathscr{E}$ can lead to the same pitch; our point is that these $\mathscr{E}$s are grouped *over* spectral-temporal $\mathscr{M}$s.

2.2.2. The implied mapping

We (Kubovy & Van Valkenburg, 2001) used the TIA to argue for an implied theoretical mapping between the auditory and visual modalities: specifically that experiments and theories of objecthood in the respective modalities should be compared with respect to the media ($\mathscr{M}$) in which the objects exist. In other words, the IAs in vision (space and time) correspond to the IAs in audition (frequency and time). We call this the *TIA mapping*. Our position on this matter is not unique. Belin and Zatorre (2000) argued for the same sort of mapping based on the fact that that auditory spectral motion and visual spatial motion are “both related to changes of energy across the sensory epithelium” (p. 965). In other words, instead of mapping based on functional characteristics of the two systems (where auditory space is analogous to visual space), they prefer to map based on physiological-sensory characteristics. Woods, Alain, Diaz, Rhodes, and Ogawa (2001) conducted four experiments designed to assess the role of space and frequency cueing in auditory selective attention – they concluded that auditory frequency plays a role analogous to visual space. In fact they propose an auditory version of Treisman’s (1993) feature integration theory which they call the frequency-based feature integration theory. This is just the kind of mapping we have been proposing.

2.2.3. On the concept of edges

TIA mapping, in combination with the idea that figure-ground segregation defines perceptual objecthood, suggests that edges are important. In our view, an edge is where a PP, or an object boundary, occurs within a given medium ($\mathscr{M}$). In visual figure-ground segregation therefore, edges can occur in space and/or in time, whereas in auditory figure-ground segregation edges occur in frequency/time (see Kubovy & Van Valkenburg, 2001). At this point we can only speculate about the exact nature of auditory edges. If we think of auditory objects as harmonic complexes (such as a voice), then we might think of them as having only lower edges (e.g. the fundamental frequency of the voice). In this respect, it is suggestive that a mistuned low harmonic in an auditory complex is more easily detected than a high one (Hartmann, McAdams, & Smith, 1990; Lee & Green, 1994). But if we think of auditory objects as more complex combinations of sounds, then perhaps they have lower and upper edges. Brochard et al. (1999) found that when observers try to attend to one of a number of simultaneous subsequences, they find it easier to attend to subsequences at the lowest and highest frequencies than to those sandwiched between them.

2.2.4. The role of IAs in concurrent vs. sequential figure-ground segregation

Kubovy and Van Valkenburg’s (2001) theory of auditory objecthood applies to both concurrent and sequential figure-ground segregation. In concurrent segregation, spectral components ($\mathscr{E}$) are grouped by common features ($\mathscr{F}$; harmonicity, attack/decay, onset/offset, etc.) *within* time and across frequency (the $\mathscr{M}$s) to produce PPs. The subjective perception of these PPs is pitch and timbre. In sequential segregation, spectral components ($\mathscr{E}$) are grouped by common features ($\mathscr{F}$; harmonicity, attack/decay, onset/offset, etc.) *across* time and across frequency (the $\mathscr{M}$s) to produce PPs. The subjective perception of these PPs is streaming.

Our theory makes no claims about either time or frequency being more important than the other – we consider them to be equally important.

3. Countering the counter-examples

We now take up our discussion of Neuhoff's critique. He claims that (1) "simple frequency separation does not ensure the formation of auditory objects" and that (2) "frequency variation is not a necessary condition for auditory figure-ground relationships".

3.1. On the sufficiency of frequency and time

Neuhoff argues that "differences in frequency do not *ensure* that two sources will be heard as distinct from each other". We have *never* said or implied that a separation in frequency ensures perceptual numerosity; we have only maintained that frequency and/or time are indispensable (i.e. necessary) for perceptual numerosity. The TIA does not claim that frequency separation is a sufficient condition for perceptual numerosity, but that frequency separation makes perceptual numerosity *possible*.

3.2. On the necessity of frequency and time

Neuhoff claims that we have overstated the importance of frequency in our theory of auditory objecthood, and that in fact frequency is not an IA. He offers a number of examples which purportedly show that sounds can segregate without being distributed over frequency. Here they are.

3.2.1. Sequential timbre segregation

Neuhoff describes two studies which show that two sounds with different timbres but the same fundamental frequency will segregate if they are presented sequentially (Cusak & Roberts, 2000; Iverson, 1995). This is not a problem for our theory of auditory objecthood or the TIA.[3] Although these are demonstrations of segregation without frequency variation, they are also demonstrations of the segregation of sounds *by* timbre *in* time. Neuhoff has failed to appreciate that the TIA offers two IAs for hearing: frequency and time. He only mentions frequency.

3.2.2. Segregation by space, timbre, and motion

3.2.2.1. The flute and the oboe Neuhoff claims that a listener presented with two sound sources – a flute to the listener's left and an oboe to the listener's right, played simultaneously and at the same fundamental frequency – will hear them as two instruments. We disagree. We believe that the listener would hear one 'floboe' – a hybrid sound resembling both source instruments (Krumhansl, 1989). Unfortunately, we can think of no empirical way to test our hypothesis using behavioral methods,[4] and we do not know of any non-behavioral data on this topic. There are some empirical questions worth pursuing here, but

[3] In fact, Iverson (1995) concludes with ideas similar to our own.

[4] Briefly, there is a recognition-and-inference problem – faced with a 'floboe', and given the choice between 'one' and 'two' instruments, listeners might be able to *infer* that there are two.

we have no reason to believe that careful empirical studies would support Neuhoff's claims.

Neuhoff claims that even if the two instruments *were* heard as one, this percept would be undone "by interaction with the environment through head movements and navigation between the sources". We agree. This is an example of segregation of the elements (in this case the frequency components emitted by the instruments) *by* a feature (in this case space: interaural time differences and interaural level differences cues) which changes *across* time. The result would be two spectral-temporal entities grouped *by* space, but *not in* space. The thrust of our argument is not to undermine the importance of space in PG, but to differentiate between features and media.

3.2.2.2. The flight of the bumblebees Neuhoff's second example involves two sounds that are distributed over space (but not frequency or time or timbre): "Imagine that two bumblebees emit the same fundamental frequency and buzz around the left and right ears of a listener respectively. Arguing that these two sources will be perceived as one auditory object simply because they have no separation in fundamental frequency seems somewhat untenable."

According to Neuhoff, you can tell that two bumblebees are coming at you from two directions because spatial disparity and relative motion produce perceptual numerosity. From these conjectures he concludes that frequency is not an IA. Unfortunately Neuhoff has created a couple of straw bees: Doppler shifts would ensure that the bees would not buzz at the same frequency. If, however, the bees could be induced to buzz at the same frequency, *and* to fly in formation (to ensure that our two ears would receive the same information), we would probably hear one bumblebee.

3.3. Loose ends

Two citations used by Neuhoff in his critique do not support his views. We have already pointed out that Iverson's (1995) experiments and conclusions are in accordance with our theory. He also cites Darwin and Hukin (1999) to support the assertion that "in complex naturally occurring sounds such as speech, space may be even more important than fundamental frequency". Actually, the thrust of Darwin and Hukin's thesis is in perfect agreement with our theory (Kubovy & Van Valkenburg, 2001), and is inconsistent with Neuhoff's. According to them, auditory objects are the product of non-spatial grouping processes (e.g. harmonicity and onset time) and once an object is formed, listeners can attend to it and to its features.

4. Conclusion

For the most part, Neuhoff agrees with us on the nature of auditory objecthood: "Things that are 'susceptible to figure-ground segregation' should rightly be called objects, be they auditory or visual", and he agrees that our principles for defining 'auditory edges' are "reasonable and well grounded". Where we seem to differ is in our conception of the nature of figure-ground segregation itself and the grouping process which leads to it.

We have claimed that grouping involves three components: elements ($\mathscr{E}$), distributed *over* a medium ($\mathscr{M}$), grouped *by* features ($\mathscr{F}$). In disagreement with us, Neuhoff gives all auditory dimensions an equal opportunity to play a crucial role in the formation of auditory objects, thus implying that $\mathscr{M}$ and $\mathscr{F}$ are interchangeable. We do not understand the basis of this claim. Object formation, in both vision and audition, is constrained by the very nature of the sensory epithelium over which perception unfolds. Neuhoff's position unnecessarily relaxes these real environmental/sensory constraints, and the result of this is (1) an unsound logical platform on which to rest theories of objecthood, and (2) necessarily incompatible theories of visual and auditory objecthood, with no way to compare experimental results across the two domains. In this response we hope that we have clarified our position, and we hope that we have shown that Neuhoff's counter-examples miss their target.

Acknowledgements

This study was supported by NEI grant No. R01 EY12926-06.

References

Belin, P., & Zatorre, R. J. (2000). 'What', 'where', and 'how' in auditory cortex. *Nature Neuroscience*, *3* (10), 965–966.

Bregman, A. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: MIT Press.

Brochard, R., Drake, C., Botte, M., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: effects of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology: Human Perception & Performance*, *25* (6), 1742–1759.

Cusak, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, *62* (5), 1112–1120.

Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: the role of interaural time differences. *Journal of Experimental Psychology: Human Perception & Performance*, *25*, 617–629.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum.

Gibson, J. J. (1982). What is involved in surface perception? In J. Beck (Ed.), *Organization and representation in perception* (pp. 151–157). Hillsdale, NJ: Lawrence Erlbaum.

Handel, S. (1988). Space is to time as vision is to audition: seductive but misleading. *Journal of Experimental Psychology: Human Perception & Performance*, *14*, 315–317.

Hartmann, W. M., McAdams, S., & Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America*, *88*, 1712–1724.

Iverson, P. (1995). Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception & Performance*, *21* (4), 751–763.

Izumi, A. (2002). Auditory stream segregation in Japanese monkeys. *Cognition*, *82*, B113–B122.

Jones, D. (1993). Objects, streams, and threads of auditory attention. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: selection, awareness and control: a tribute to Donald Broadbent* (pp. 87–104). Oxford: Clarendon Press.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzen & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43–53). Amsterdam: Elsevier.

Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 55–99). Hillsdale, NJ: Lawrence Erlbaum.

Kubovy, M. (1988). Should we resist the seductiveness of the space:time:vision:audition analogy? *Journal of Experimental Psychology: Human Perception & Performance*, *14*, 318–320.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*, 97–126.

Lee, J., & Green, D. M. (1994). Detection of a mistuned component in a harmonic complex. *Journal of the Acoustical Society of America*, *96*, 716–725.

Middlebrooks, J. C. (2002). Auditory space processing: here, there, or everywhere? *Nature Neuroscience*, *5* (9), 824–826.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Neuhoff, J. G. (in press). Pitch variation is unnecessary (and sometimes insufficient) for the formation of auditory objects. *Cognition*.

Rauschecker, J. P. (1997). Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Oto-Laryngologica Supplement*, *532*, 34–38.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinions in Neurobiology*, *288*, 516–521.

Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiology and Neurootology*, *3*, 86–103.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. *Proceedings of the National Academy of Sciences USA*, *97*, 11800–11806.

Romanski, L. M. (2003). Anatomy and physiology of auditory-prefrontal interactions in non-human primates. In A. A. Ghazanfar (Ed.), *Primate audition: ethology and neurobiology* (pp. 259–278). Washington, DC: CRC Press.

Shih, S., & Sperling, G. (1996). Is there feature-based attentional selection in visual search? *Journal of Experimental Psychology: Human Perception & Performance*, *22* (3), 758–779.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Treisman, A. M. (1993). The perception of features and objects. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: selection, awareness, and control: a tribute to Donald Broadbent* (pp. 5–32). Oxford: Clarendon Press.

Woods, D. L., Alain, C., Diaz, R., Rhodes, D., & Ogawa, K. H. (2001). Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception & Performance*, *27* (1), 65–74.

Zatorre, R. J., Bouffard, M., Ahad, P., & Pascal, B. (2002). Where is 'where' in the human auditory cortex? *Nature Neuroscience*, *5* (9), 905–910.