



Discussion

Pitch variation is unnecessary (and sometimes insufficient) for the formation of auditory objects

John G. Neuhoff*

Department of Psychology, The College of Wooster, Wooster, OH 44691, USA

Received 12 December 2001; accepted 21 March 2002

1. Introduction

Kubovy and Van Valkenburg (2001) have posited a theory of auditory “objecthood” that draws heavily on a proposed analogy between the auditory and visual systems. Specifically, they have suggested that the auditory dimension of pitch is analogous to the visual dimension of spatial extent and that pitch is an “indispensable attribute” in the formation of auditory objects. In their development of an auditory “Theory of Indispensable Attributes” they propose that variation in pitch is a necessary and sufficient condition for the perception of auditory figure-ground relationships and thus for the perception of “auditory objects”. In this paper evidence is presented that suggests that differences in pitch are unnecessary for the formation of auditory objects and that under some circumstances, pitch differences are insufficient for the formation of auditory objects.

2. Auditory objects and figure-ground relationships

Kubovy and Van Valkenburg (2001) define a perceptual object as “that which is susceptible to figure-ground segregation” (p. 102), and also state that “any object, be it visual or auditory, must have an edge or boundary” (p. 104). They set out principles for defining “auditory edges” that are reasonable and well grounded. In providing evidence for auditory edges created by pitch, they cite some relevant work on auditory stream segregation. Specifically, they refer to the work of Bregman and Campbell (1971) showing that fundamental frequency can be used to create figure-ground relationships in pitch and thus, by definition, create auditory objects.

However, in defining the existence of figure-ground relationships as the criterion for auditory objecthood, Kubovy and Van Valkenburg (2001) have actually created a much

* Tel.: +1-330-263-2475; fax: +1-928-244-5577.

E-mail address: jneuhoff@wooster.edu (J.G. Neuhoff).

larger category of auditory objects than those specified by differences in pitch. Several other important auditory dimensions can also lead to figure-ground relationships (Bregman, 1990) and thus, according to their definition, lead to the formation of auditory objects. Yet, in the explication of their theory they have underestimated the importance of these additional auditory dimensions. In essence, they propose that pitch variation is both necessary and sufficient for the formation of auditory objects. Moreover, they have implied that spatial separation cannot lead to the creation of meaningful auditory objects.

However, an examination of the relevant work on auditory stream segregation shows that, by their own definition, auditory pitch is not indispensable. It is true that pitch variation *can* lead to auditory figure-ground segregation and thus, to the formation of auditory objects as they define them. In fact, pitch may even be the *strongest* cue to auditory “objecthood”. However, a considerable amount of research has shown that (1) frequency variation is not a necessary condition for auditory figure-ground relationships, and (2) simple frequency separation does not ensure the formation of auditory objects. These points will now be addressed in turn.

3. Frequency variation is not necessary for auditory “objecthood”

Kubovy and Van Valkenburg (2001) suggest that two acoustic sources separated in space that emit the same fundamental frequency will be perceived as a single source (p. 110, Fig. 11b). Under some circumstances, this is true. However, under real-world listening conditions there are many examples where such sounds *are* perceived as separate sources. Thus, to conclude that pitch variation is indispensable for forming auditory objects overstates the importance of pitch. In fact, to reject this hypothesis we need only formulate one example of spatially disparate sources that emit the same fundamental frequency and are perceived as separate sources. There are many such examples that are based on differences in timbre and spatial location.

Consider first a listener presented with two sound sources. To the listener’s left is a flute and to the right is an oboe. If both instruments play the same fundamental frequency at the same time, the listener will likely hear two instruments. Of course, there may be certain acoustic environments in which a listener might experience some ambiguity or even perceive the two sources as one. However, these ambiguities could almost certainly be resolved by interaction with the environment through head movements and navigation between the sources, behaviors that are quite common in natural environments. Nonetheless, Kubovy and Van Valkenburg (2001) propose that separation in fundamental frequency is an indispensable attribute for perceptual numerosity. Thus, according to this view, listeners should not only be unable to tell the difference between these two very different sounds, they should not even be able to tell that there are two sources present. In fact, it is well established that such differences in timbre can cause the very figure-ground relationships that Kubovy and Van Valkenburg have defined as creating auditory objects (McAdams & Bregman, 1979; McNally & Handel, 1977; Singh & Bregman, 1997; Smith, Hausfeld, Power, & Gorta, 1982; Wessel, 1979).

For example, Iverson (1995) played sequences in which tones had the same fundamental frequency but in which alternate tones differed in timbre. Under carefully controlled

conditions, a bassoon (for example) at middle C (262 Hz) was alternated in a sequence with a trumpet at middle C. The results showed that listeners experienced auditory stream segregation and thus, figure-ground relationships based on timbre differences alone. Cusack and Roberts (2000) also used alterations in timbre in a sequence of tones and found clear evidence that such differences influence auditory stream segregation. It should be noted that these experiments involved *sequential* stream segregation and thus, differ slightly from the simultaneous segregation example of the flute and oboe above. However, Kubovy and Van Valkenburg (2001) cite the work of Bregman and Campbell (1971), a similar sequential stream segregation study in the pitch dimension, as evidence for auditory figure-ground segregation. Thus, if sequential stream segregation is the standard by which auditory objects are defined, it is clear that fundamental frequency is not an indispensable attribute in their formation.

In addition to the auditory figure-ground relationships created by timbre, auditory space and motion can create figure-ground relationships between two sources with the same fundamental frequency. Contrary to what Kubovy and Van Valkenburg (2001) have proposed, these spatial figure-ground relationships are real and accomplish more than just cueing attention to the pitch dimension. 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. To be fair, Kubovy and Van Valkenburg do not dismiss the importance of auditory space entirely. They state that "...although spatial cueing may be sufficient to draw attention to a pitch, attention is allocated to the pitch, not to its location." (p. 108). However, arguing that the spatial dimension in this example is simply cueing attention to pitch seems equally untenable. In this example, the spatial location of the bees is much more important than the fundamental frequency of their buzz. Moreover, there is abundant experimental evidence showing that spatial location can lead to auditory stream segregation and thus, by the current definition, to the formation of auditory objects (Axelrod & Guzy, 1968; Axelrod, Guzy, & Diamond, 1968; Axelrod & Powazek, 1972; Ciocca, Bregman, & Capreol, 1992; ten Hoopen & Akerboom, 1982; ten Hoopen, Van Meurs, & Akerboom, 1982; ten Hoopen, Vos, & Dispa, 1982).

4. Frequency variation does not ensure auditory "objecthood"

There are many instances in which listeners presented with two different fundamental frequencies will report two auditory objects. However, there are numerous studies that show that such differences in frequency do not *ensure* that two sources will be heard as distinct from each other (for a review see Darwin & Carlyon, 1995). The relationship between auditory objecthood and fundamental frequency also requires a careful examination of the harmonicity of the two sources.

Most naturally occurring sounds (that have pitch) are composed of a fundamental frequency and an ordered series of harmonics (integer multiples of the fundamental frequency). Whether two sources can be separated in a figure-ground relationship depends as much on their harmonic relationship as on their fundamental frequency. For example,

imagine one source with a fundamental frequency of 100 Hz and harmonics at 200, 300, and 400 Hz. If we played this source from the same location together with a second source that had a fundamental frequency of 500 Hz and harmonics at 1000, 1500, and 2000 Hz, listeners would likely report only one source with a fundamental frequency of 100 Hz. Thus, under certain circumstances, two sources that differ in fundamental frequency can be perceived as a single source. In fact, Moore and colleagues have devoted considerable efforts to try to determine what changes need to be made to this type of ordered harmonic structure in order for listeners to perceive more than one source (Moore, Glasberg, & Peters, 1985, 1986; Moore & Ohgushi, 1993; Moore, Peters, & Glasberg, 1985; Peters, Moore, & Glasberg, 1983). Similarly, Roberts and Bregman (1991) have shown that perceptual grouping based on frequency information is dependent on the harmonic relationship between the grouping elements.

One might argue that these types of laboratory experiment are too contrived. Rarely do two sources line up harmonically with such precision. Thus, studies in which listeners are asked to determine the number of sources present in a static presentation of harmonics leaves much to be desired in terms of ecological validity. However, in complex naturally occurring sounds such as speech, space may be even more important than fundamental frequency. In a series of experiments, Darwin and Hukin (Darwin & Hukin, 1999; Hukin & Darwin, 1995) used speech sounds with interaural time differences (ITD) that corresponded to spatial locations and found that “in following a particular auditory sound source over time, listeners attend to perceived auditory objects at particular azimuthal positions rather than attend explicitly to those frequency components that share a common ITD” (Darwin & Hukin, 1999, p. 617).

5. “What” and “where”

Kubovy and Van Valkenburg (2001) present convincing evidence that suggests that the auditory subsystem that processes localization is in the service of the visual system. They cite behavioral and physiological work that shows the malleability of auditory localization in the presence of conflicting visual information. They point out that the relationship is essentially a one-way proposition, with visual judgments of location dominating auditory judgments. They also cite auditory cueing studies that show that reaction time to correctly localize a visual target is faster with congruent auditory cues than with incongruent auditory cues. However, these studies allow us to conclude only what the results show – that auditory localization facilitates visual localization and that when location cues are discordant, vision dominates. Kubovy and Van Valkenburg have drawn a conclusion that goes beyond these data. They imply that because auditory localization serves vision it cannot lead to the formation of auditory objects. Clearly, another possible alternative is that the auditory “where” system can do both. The finding that visual localization dominates auditory localization when cues are discordant or incongruent does not logically lead to the conclusion that auditory localization itself is incapable of forming figure-ground relationships when visual cues are either consistent or absent.

6. Summary

Kubovy and Van Valkenburg (2001) have proposed a definition of perceptual objects that is useful. Things that are “susceptible to figure-ground segregation” should rightly be called objects, be they auditory or visual. However, when we apply this definition to audition, we find that the dimension of pitch is not indispensable. Auditory figure-ground relationships can also be formed by timbre, spatial location, and even dynamic acoustic characteristics such as amplitude modulation (Carrell & Opie, 1992; Grose, Hall, & Mendoza, 1995). Furthermore, differences in fundamental frequency do not ensure that two sources will be perceived as separate auditory objects. The harmonic relationship between the two sources is a critical factor in this determination (Roberts & Bregman, 1991). Moreover, it appears that whether auditory figure-ground relationships are formed or not depends on an interaction of stimulus information and the context in which that information is perceived. A study by McNally and Handel (1977) showed that two tones separated in frequency will form separate streams (or auditory objects) if they are presented only in a sequence of tones. However, if the same two tones are presented in sequence with atonal clicks and hisses, the tones form one auditory object and the atonal stimuli form another. O’Connor and Sutter (2000) have shown that auditory grouping by frequency can be modulated by spatial location. Like McNally and Handel (1977), they argue for an interaction of cues that lead to stream segregation and conclude that “auditory-perceptual grouping involves global neural processing, i.e., the participation of neurons with very broad frequency input that are also sensitive to spatial location.”

It is tempting to draw analogies between vision and audition. Both systems inform organisms about the environment. Both are instrumental in complex perception and action relationships such as those used in communication and navigation. However, researchers who seek one set of principles to explain both systems should proceed with great caution. The two systems have evolved to meet very different environmental challenges. The deficiencies of one system tend to be compensated for by the strengths of the other. Thus, drawing analogies between the two can be problematic, and attempts to do so might be correctly characterized as “seductive but misleading” (Handel, 1988, p. 315).

Acknowledgements

I am grateful to Al Bregman, Bruno Repp, and Michael McBeath for comments on an earlier draft of this paper. Preparation of this manuscript was supported by a grant from the National Science Foundation.

References

- Axelrod, S., & Guzy, L. T. (1968). Underestimation of dichotic click rates: results using methods of absolute estimation and constant stimuli. *Psychonomic Science*, *12* (4), 133–134.
- Axelrod, S., Guzy, L. T., & Diamond, I. T. (1968). Perceived rate of monotic and dichotically alternating clicks. *Journal of the Acoustical Society of America*, *43* (1), 51–55.
- Axelrod, S., & Powazek, M. (1972). Dependence of apparent rate of alternating clicks on azimuthal separation between sources. *Psychonomic Science*, *26* (4), 217–218.

- Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: MIT Press.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89 (2), 244–249.
- Carrell, T. D., & Opie, J. M. (1992). The effect of amplitude comodulation on auditory object formation in sentence perception. *Perception & Psychophysics*, 52 (4), 437–445.
- Ciocca, V., Bregman, A. S., & Capreol, K. L. (1992). The phonetic integration of speech and non-speech sounds: effects of perceived location. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 44 (3), 577–593.
- Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, 62 (5), 1112–1120.
- Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Eds.), *Hearing. Handbook of perception and cognition* (2nd ed.). (pp. 387–424). San Diego, CA: Academic Press.
- 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 (3), 617–629.
- Grose, J. H., Hall, J. W., & Mendoza, L. (1995). Perceptual organization in a comodulation masking release interference paradigm: exploring the role of amplitude modulation, frequency modulation, and harmonicity. *Journal of the Acoustical Society of America*, 97 (5, Pt 1), 3064–3071.
- Handel, S. (1988). Space is to time as vision is to audition: seductive but misleading. *Journal of Experimental Psychology: Human Perception & Performance*, 14 (2), 315–317.
- Hukin, R. W., & Darwin, C. J. (1995). Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. *Journal of the Acoustical Society of America*, 98 (3), 1380–1387.
- 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.
- Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80, 97–126.
- McAdams, S., & Bregman, A. S. (1979). Hearing musical streams. *Computer Music Journal*, 3, 23–43.
- McNally, K. A., & Handel, S. (1977). Effect of element composition on streaming and the ordering of repeating sequences. *Journal of Experimental Psychology: Human Perception & Performance*, 3 (3), 451–460.
- Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America*, 77 (5), 1853–1860.
- Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, 80 (2), 479–483.
- Moore, B. C., & Ohgushi, K. (1993). Audibility of partials in inharmonic complex tones. *Journal of the Acoustical Society of America*, 93 (1), 452–461.
- Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *Journal of the Acoustical Society of America*, 77 (5), 1861–1867.
- O'Connor, K. N., & Sutter, M. L. (2000). Global spectral and location effects in auditory perceptual grouping. *Journal of Cognitive Neuroscience*, 12 (2), 342–354.
- Peters, R. W., Moore, B. C. J., & Glasberg, B. R. (1983). Pitch of components of complex tones. *Journal of the Acoustical Society of America*, 73 (3), 924–929.
- Roberts, B., & Bregman, A. S. (1991). Effects of the pattern of spectral spacing on the perceptual fusion of harmonics. *Journal of the Acoustical Society of America*, 90 (6), 3050–3060.
- Singh, P. G., & Bregman, A. S. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *Journal of the Acoustical Society of America*, 102 (4), 1943–1952.
- Smith, J., Hausfeld, S., Power, R. P., & Gorta, A. (1982). Ambiguous musical figures and auditory streaming. *Perception & Psychophysics*, 32 (5), 454–464.
- ten Hoopen, G., & Akerboom, S. (1982). The perceived tempi of coherent and streaming tone sequences: II. *Perception & Psychophysics*, 32 (5), 481–485.
- ten Hoopen, G., Van Meurs, G., & Akerboom, S. (1982). The perceived tempi of coherent and streaming tone sequences. *Perception & Psychophysics*, 31 (3), 256–260.
- ten Hoopen, G., Vos, J., & Dispa, J. (1982). Interaural and monaural clicks and clocks: tempo difference versus attention switching. *Journal of Experimental Psychology: Human Perception & Performance*, 8 (3), 422–434.
- Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3 (2), 45–52.