Learning & Perception 5 (2013) Suppl. 2, pp. 33–35
DOI: 10.1556/LP.5.2013.Suppl2.3

PREFACE

In the winter of 2010–2011 Ilona Kovács *and* István Winkler – *renowned vision and audition scientists – conceived a special issue on bistable perception. This special issue followed a conference, named "Cognition at Christmas", at the Budapest University of Technology and Economics where audition and vision scientists discussed the similarities and differences in bistable perception across the two sensory domains.*

All papers in this special issue share examination of the process of resolving sensory ambiguities, i.e. the process of forming Gestalts, in perception. This is currently a cardinal issue in perception research as ambiguities in the sensory signals, including noise, play an intrinsic role in any perceptual interpretation. One prominent ambiguity in vision concerns what is called 'figure-ground segregation'. In audition the counterpart of this ambiguity concerns what is called 'auditory stream segregation'. The first three papers in this special issue examine this latter ambiguity: Auditory scene analysis, involving the decomposition of a mixture of sounds into meaningful perceptual units.

The auditory system is often exposed to discrete sound events that need to be bound together in order to form meaningful Gestalts. At the same time, binding events together that are actually being emitted by two different sources must be avoided. This process has been termed auditory stream segregation.

The first study by Bendixen et al. compares the effects of feature similarity (based on linking sounds with similar characteristics) and feature predictability (based on linking sounds that follow each other in a predictable manner) in auditory stream segregation. The data demonstrate that similarity and predictability are both relevant in stream segregation, but that they exert their influence at different stages of auditory processing. Only feature similarity seems to contribute to the first stage of auditory scene analysis, in which alternative sound organizations are first formed. Both feature similarity and predictability appear to contribute to the second stage of auditory scene analysis, where competition between alternative sound organizations takes place. It may be the case that predictability over time depends on attention entailing a top-down component while similarity detection can take place in a bottom-up fashion.

Bőhm et al. proceed by asking whether sound source motion is used as a primary cue in auditory stream segregation. They hypothesize that sounds emitted by two sources moving on a common spatial trajectory would be more likely to be grouped together, and sounds emitted by two sources moving on separate (independent) trajectories would be more likely

to be perceived as segregated. However, although the results demonstrate a clear effect of spatial separation, they find no clear effect of sound source motion on the perceptual organization of the test sequences. Thus, their data do not support the hypothesis that auditory motion can serve as a primary cue in auditory stream segregation. In other words, the results provide no evidence that the effects of auditory motion can be described as a case of the Gestalt principle of common fate.

Szalárdy et al. examined, on the one hand, whether differential amplitude modulation assists in separating concurrent sound sequences and, on the other hand, whether this cue would interact with previously studied static cues (carrier frequency and location difference) in segregating concurrent streams of sound. They conclude that separation in the frequency of amplitude modulation is a primary aspect in auditory stream segregation, thereby increasing the stability of auditory perception.

Denham et al. then proceed capitalizing on a slightly different approach. While they use the same *auditory streaming paradigm* as used above, their paper does not focus on segregation as such, but more on the underlying perceptual competition process. To be more precise, Denham et al. are particularly interested in the dynamics of resolving perceptual ambiguity. Using tone sequences of as long as 4 minutes they demonstrate that listeners hold their first percept of the sound sequence for a relatively long period, after which perception switches between two or more alternative sound organizations, each held on average for a shorter duration. They report that the first percept differs from subsequent ones in that stimulus parameters influence its quality and duration to a greater extent than the subsequent ones. To explain these data they propose two processes that fit well with what has been proposed in the visual ambiguity literature: The first process is a choice process depending on priming and intrinsic choice bias (for example employing known alternative interpretations); The subsequent process is a stochastic switch process (depending on adaptation of the prevailing percept).

Changing gears from audition to vision, the next paper in line by Pastukhov and Braun resembles the previous paper by Denham in that it also studies the competition process. They focused on priming in the choice process thereby distinguishing positive priming (reflecting some kind of neural facilitation) and negative priming (likely caused by neural adaptation). They were able to isolate negative from positive priming effects using sequences of intermittent presentations, consisting of a prime stimulus, a blank period, and a test stimulus. They report that negative priming builds up and decays in seconds, whereas positive priming builds up in seconds and decays in minutes.

In the final paper of this special issue, Reiz et al., also using vision as the sensory modality of study, suggest that the long-range connectivity pattern of the primary visual cortex provides an architecture in which spreading neural activity may lead to pertinent figure-ground segregation. They present a computational model containing three basic features of the long-range connectivity network: local orientation, distance selectivity, and spreading neural activity. They found their model capable of detecting contours in a noisy background, with sensitivity to background noise and a preference for closed contours that reportedly resemble human performance.

This figure-ground model of the final paper brings us back to the theme of the first papers in this special issue that study auditory scene segmentation, in fact another example of

figure-ground segmentation, but then in audition. On a more general note, it is inspiring to see this variety of insights emerging from a conference where both audition and vision scientists learned from the similarities and differences in each other's research field. It is amazing to see how many similarities there seem to be for auditory and visual bistability, pointing to similar underlying neural processing. Such progression in knowledge invites studies that examine disambiguation in one sensory modality through true multi-modal interactions. This is a relatively new area of research. The first insights on multi-modal interactions with unambiguous stimuli are promising. They suggest that disambiguation of unreliable sensory information is a primary purpose of multi-modal interactions, underlining the importance of combined auditory and visual research for which this special issue will constitute a facilitating effect.

Raymond van Ee,
Department of Experimental Psychology, University of Leuven
Department of BioPhysics, Radboud University Nijmegen