Exp Brain Res (2014) 232:273–282
DOI 10.1007/s00221-013-3738-3

RESEARCH ARTICLE

Received: 1 July 2013 / Accepted: 3 October 2013 / Published online: 20 October 2013

as in Experiment 1. Listeners were required to make two alternative forced choices (2AFC) to report the perception of a given Ternus display: element motion (auditory apparent motion from sound A to B to C) or group motion (auditory apparent motion from sound ‘AB’ to ‘BC’). The results indicate that the perceptual grouping of short auditory sequences (materialized by the perceptual decisions of the auditory Ternus display) was modulated by temporal and spectral cues, with the latter contributing more to segregating auditory events. Spatial layout plays a less role in perceptual organization. These results could be accounted for by the ‘peripheral channeling’ theory.

auditory object and more sophisticated auditory streams. This is accomplished with minimum interference from any background distracter auditory inputs. The process of separating a target auditory event or auditory stream from these distracters was first understood as the ‘cocktail party problem’ (Cherry 1953). Afterward, this phenomenon spawned extensive studies (Cooper and Roberts 2007; Denham and Winkler 2006; Takegata et al. 2005; Yabe et al. 2001 and perceptually uncertain auditory stimuli (Cusack 2005; Pressnitzer and Hupé 2006; Szalárdy et al. 2013). Similar to the processes related to vision, there are common principles of perceptual organization within the auditory domain. The different perceptions of auditory bi-stabil-

graduate students (four females, aged between 21 and 25; average age 23.1 years) participated in Experiment 1b. All participants reported having normal hearing and were naïve to the purposes of the study. The experiment was performed in compliance with all institutional guidelines set by the Academic Affairs Committee, Department of Psychology at Peking University.

Apparatus and Stimuli

Three hamburger mini-speakers (DK-601, diameter 3.6 cm) were placed horizontally on a desk (see Fig. 2a). The center-to-center distances between the speakers were set at 45 cm in Experiment 1a and 25 cm in Experiment 1b. A monitor was placed behind the speakers. A normal PC—interfaced with a sound card (RME Fireface UFX)—was used for all stimuli presentation, instruction presentation (with 17-inch CRT monitor) and data collection (by key-press). The computer program used to control the experiment was developed with Matlab (Mathworks Inc.) and the Psychophysics Toolbox (Brainard 1997; Pelli 1997). The test cabin was semi-anechoic. No light was present, except that which was emitted by the monitor. The viewing distance was set at 70 cm.

The auditory stimuli consisted of four sequentially presented, identical 50-ms burst of white noise (65 dB) to generate auditory apparent motion. The initial noise was provided by the first (flanker) speaker. The second and third noises were generated by the middle speaker. The fourth noise was emitted from the third (flanker) speaker. The first two and final two sounds were treated as two frames. The IFI was the interval between the offset of the second tone in the first frame and the onset of the first tone in the second frame (see Fig. 2b). The IFI was chosen between 50, 80, 110, 140, 170, 200 and 230 ms on a trial-by-trial basis. This was similar to the settings in a visual or tactile display (Chen et al. 2010l. rre

The ANOVA of the estimated JNDs, with WFIs of 5, 10 and 20 ms, also revealed the main effect to be insignificant—$F(2,26) = 0.324$, $p = 0.726$ (Experiment 1a), $F(2,22) = 0.321$, $p = 0.729$ (Experiment 1b).

A repeated-measures ANOVA toward the percentages of GM perception (using WFI and IFI as the two within-participant independent factors) revealed a significant main effect of IFI in both Experiment 1a—$F(6,78) = 230.875$, $p < 0.001$ and Experiment 1b—$F(6,66) = 500.326$, $p < 0.001$. Nevertheless, no significant main effect for the WFI was observed—$F(2,26) = 0.257$, $p = 0.775$ (Experiment 1a), $F(2,22) = 1.148$, $p = 0.336$ (Experiment 1b). Furthermore, no significant effect on the interaction between the WFI and IFI was observed—$F(12,156) = 0.387$, $p = 0.967$ (Experiment 1a), $F(2,26) = 1.362$, $p = 0.192$ (Experiment 1b).

We then performed a cross-experiment analysis to discover the effects of the spatial layout, if any. A Univariate ANOVA was carried out for PSE with WFI and spatial layout (45 cm in Experiment 1a and 25 cm in Experiment 1b) as dependent factors. Importantly, the analysis results revealed no significant effect of spatial layout, $F(1,72) = 1.557$, $p = 0.216$. The effect of WFI was insignificant, $F(2,72) = 0.227$, $p = 0.759$, and no significant interaction between spatial layout and WFI was found, $F(2,78) = 0.091$, $p = 0.913$. The cross-experiment analysis for JND likewise yielded no statistical differences.

The results showed that, similar to visual and tactile Ternus, the perception of auditory Ternus motion was mainly modulated by the IFIs. The perception of 'GM' was dominant under longer IFIs conditions. The distinction between 'EM' and 'GM' in auditory Ternus was generally based on the principles of temporal grouping. Here, the longer IFIs made the temporal boundary of two auditory frames ('AB' and 'BC') distinctive. A longer IFI also enhanced the perceived separation of the two grouped auditory events. This led to a dominant perception of GM. In a precedence effect (Hartung and Trahiotis 2001), the lagging sound might fuse to the leading sound when both are in short temporal separations (less than 10 ms). Here, we adopted three within-frame delays (WFIs of 5, 10 and 20 ms as brief gaps) and found that the WFI imposed no discernible influence on the participants' ability to discriminate and perceive apparent relative easy to separate the two frames) and (3) distinctly separated (800 vs. 1,000 Hz, easy to separate two frames).[1]

The amplitude of each tone was set according to the equal-loudness level. This was done because the WFI has little influence on the perception of apparent motion (as concluded from Experiment 1). Only two WFIs (5 and 20 ms) were employed in Experiment 2, in order to reduce the number of trials. In addition, due to the fact that the vast majority of participants had made virtually 100 % GM judgments for long IFIs in the previous experiments, the range of IFIs was adjusted to from 30 to 210 ms, with increased step sizes of 30 ms.

A 2 (WFI) × 7 (IFI) × 3 (frequency separation: low, medium and high) block design was adopted. There were still 840 trials throughout the experiment, which were divided into 5 blocks. The presentation order of the standard frame (auditory pair of 800 Hz) and the comparative frame (auditory pair of 820, 860 and 1,000 Hz), and the directions of apparent motion (left or right) were fully randomized and balanced. The participants received the same amount of practice as in Experiment 1, in order to assure a clear distinction between EM and GM. In the following formal experiment, participants were asked to concentrate on discriminating their perceptions of apparent motion, rather than the pitch differences between the two frames. The data collection method was the same as in Experiment 1.

[illegible]

A repeated-measures ANOVA toward the percentages of GM perception (using WFI and IFI as the two within-participant independent factors) revealed a significant main effect of IFI, $F(6,72) = 330.67$, $p < 0.001$.

A 2 × 3 ANOVA was conducted with WFI (5 vs. 20 ms) and frequency separation (small vs. medium vs. large) as within-subject independent factors and PSE as dependent factor revealed nonsignificance of the main effect of WFI, $F(1,12) = 0.249$, $p = 0.627$. The effect of frequency separation was significant, $F(2,24) = 13.378$, $p < 0.001$. For both WFI conditions, Bonferroni-corrected comparisons showed that the PSE at a higher frequency (1,000 Hz) was lower than the PSE at a lower frequency (820 Hz) ($p < 0.05$) and lower than the PSE at the medial frequency (860 Hz) ($p < 0.01$). However, the interaction between the WFI and frequency was insignificant, $F(2,24) = 1.341$, $p = 0.280$. Therefore, an obvious decrease in the PSE was observed

[1] In order to confirm that the frequencies selected justified our research purposes, we asked 10 participants to do a pitch discrimination task. Two frames of pure tones were presented, and the frequencies of the reference frame were kept at 800 Hz, while the comparative frame had a frequency selected from 660, 700, 740, 780, 820 Hz, 860, 900 and 940 Hz. ANOVA with the frequency as the single independent factor showed a significant frequency effect, $F(2,16) = 20.201$, $p < 0.001$. Bonferroni-corrected pairwise comparisons for 820, 860 and 940 Hz conditions confirmed that performance was better at 940 Hz (90.4 %) than at 860 Hz (83.8 %) ($p < 0.05$), better at 860 Hz (83.8 %) than at 820 Hz (74.3 %) ($p < 0.01$) and better at 940 Hz (90.4 %) than at 820 Hz (74.3 %) ($p < 0.01$).

continuity (Bregman and Campbell 1971; Bregman 1990). Segregation is also aided when auditory objects differ in their spectral content or temporal structure, such as occurs with repetition rate (Perrott 1984; Stellmack 1994). Theories have been proposed to account for how temporal cues and spectral cues contribute to auditory streaming. An influential theory, known as the 'Peripheral Channeling stream segregation, which takes from 5 to10 s. The quick perceptual decisions for short tone sequences were seemingly at odds with Bregman's proposal that auditory scene analyses start at the same coherent position and are subsequently segregated into separate streams after a sufficient number of cues are collected. The current findings therefore suggest that bi-stable perception could be both an active exploration of the sensory environment and a fundamental aspect of sensory cognition, which supports flexible decision making (Kim et al. 2006). Considerable studies have been conducted to explore neural mechanisms' mediating of auditory stream segregation (Gutschalk et al. 2005; Micheyl et al. 2007; Rauschecker 2005). Recent research has indicated an important role for both primary (A1) and non-primary auditory cortexes, and one study has suggested a role for the intra-parietal sulcus (Cusack 2005). Using an ABA-stimulus paradigm, Cusack (2005) found that regions in the intra-parietal sulcus (IPS) showed greater activity when two streams were perceived rather than one stream. Indeed, the auditory system contains several subcortical nuclei, which are generally believed to establish basic feature encoding even before perceptual organization starts at the cortical level (Griffiths and Warren 2002; Nelken 2004). Using an ABA-stimulus paradigm, Pressnitzer et al. (2008) found that ASA starts much earlier in the auditory pathways, by recording single units from one peripheral structure of the mammalian auditory brainstem, the cochlear nucleus. Peripheral responses were similar to cortical responses and displayed all of the functional properties required for streaming. During the presentation of long auditory sequences, adaptation in peripheral auditory neurons may also be influenced by the descending feedback from upper processing stages, including the auditory cortex. However, at present, the explorations of neural substrates that correspond to the roles of temporal and spectral cues (differential frequencies) and the temporal courses for perceptual grouping in short auditory sequences are lacking and await future investigations. (Getzmann and Lewald 2012; Getzmann 2011; Hall et al. 2002).

This study was supported by grants from the Natural Science Foundation of China (11174316, 31200760), National High Technology Research and Development Program of China (863 Program) (2012AA011602) and Strategic Priority Research Program of the Chinese Academy of Sciences (XDA06020201).

Anstis S, Saida S (1985) Adaptation to auditory streaming of frequency modulated tones. J Exp Psychol Hum Percept Perform 11(3):257–271
Aydın M, Herzog MH, Öğmen H (2011) Attention modulates spatio-temporal grouping. Vis Res 51:435–446
Beauvois M (1998) The effect of tone duration on auditory stream formation. Percept Psychophys 60:852–861
Beauvois MW, Meddis R (1991) A computer model of auditory stream segregation. Q J Exp Psychol A 43:517–541
Beauvois MW, Meddis R (1996) Computer simulation of auditory stream segregation in alternating-tone sequences. J Acoust Soc Am 99:2270–2280
Bee MA, Klump GM (2005) Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. Brain Behav Evol 66:197–214
Blauert J (1997) Spatial hearing: the psychophysics of human sound localization. MIT Press, Cambridge
Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436
Bregman AS (1978) Auditory streaming is cumulative. J Exp Psychol Hum Percept Perform 4:380–387
Bregman AS (1990) Auditory scene analysis. MIT Press, Cambridge
Bregman AS, Campbell J (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. J Exp Psychol 89:244–249
Carlyon RP (2004) How the brain separates sounds. Trends Cogn Sci 8:465–471
Chen L, Shi Z, Müller HJ (2010) Influences of intra- and crossmodal grouping on visual and tactile Ternus apparent motion. Brain Res 1:152–162
Chen L, Shi Z, Müller HJ (2011) Interaction of perceptual grouping and crossmodal temporal capture in tactile apparent-motion. PLoS ONE 6:e17130
Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975–979
Cooper HR, Roberts B (2007) Auditory stream segregation of tone sequences in cochlear implant listeners. Hear Res 225:11–24
Cusack R (2005) The intraparietal sulcus and perceptual organization. J Cogn Neurosci 17:641–651
Cusack R, Carlyon RP (2004) Auditory perceptual organization inside and outside the laboratory. In: Neuhoff JG (ed) Ecological psychoacoustics. Emerald Group Publishing Limited, Bingley, pp 15–48
Darwin CJ, Carlyon RP (1995) Auditory grouping. In: Moore BCJ (ed) Hearing, vol 6. Academic, Orlando, pp 387–424
Denham SL, Winkler I (2006) The role of predictive models in the formation of auditory streams. J Physiol Paris 100:154–170
Eggermont JJ (2001) Between sound and perception: reviewing the search for a neural code. Hear Res 157:1–42
Fastl H, Zwicker E (1999) Psychoacoustics: facts and models. Springer, New York, Incorporated
Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151:167–187
Forrest TG, Formby C (1996) Temporal gap detection thresholds in sinusoidal markers simulated with a single-channel envelope detector. Aud Neurosci 3:21–33
Füllgrabe C, Moore BCJ (2012) Objective and subjective measures of pure-tone stream segregation based on interaural time differences. Hear Res 291:24–33
Getzmann S (2011) Auditory motion perception: onset position and motion direction are encoded in discrete processing stage. Eur J Neurosci 33:1339–1350
Getzmann S, Lewald J (2012) Cortical processing of changes in sound location: smooth motion versus discontinuous displacement. Brain Res 1466:119–127
Griffiths TD, Warren JD (2002) The planum temporale as a computational hub. Trends Neurosci 25:348–353
Grimault N, Bacon SP, Micheyl C (2002) Auditory stream segregation on the basis of amplitude-modulation rate. J Acoust Soc Am 111:1340–1348

Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ (2005) Neuromagnetic correlates of streaming in human auditory cortex. J Neurosci 25:5382–5388

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. Cereb Cortex 12:140–149

Harrar V, Harris LR (2007) Multimodal Ternus: visual, tactile, and visuo-tactile grouping in apparent motion. Perception 36:1455–1464

Hartmann WM, Johnson D (1991) Stream segregation and peripheral channeling. Music Percep 9:155–183

Hartung K, Trahiotis C (2001) Peripheral auditory processing and investigations of the "precedence effect" which utilize successive