Hearing Research 353 (2017) 17–25

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Detecting temporal changes in acoustic scenes: The variable benefit of selective attention

Laurent Demany [a, *], Yann Bayle [a], Emilie Puginier [a], Catherine Semal [a, b]

[a] Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, CNRS, Université de Bordeaux, 146 rue Leo-Saignat, F-33076, Bordeaux, France
[b] Institut Polytechnique de Bordeaux, Ecole Nationale Supérieure de Cognitique, 109 avenue Roul, F-33400, Talence, France

ARTICLE INFO

Article history:
Received 19 April 2017
Received in revised form
21 July 2017
Accepted 25 July 2017
Available online 27 July 2017

Keywords:
Change deafness
Selective attention
Auditory salience
Amplitude modulation
Sound onset
Sound offset

ABSTRACT

Four experiments investigated change detection in acoustic scenes consisting of a sum of five amplitude-modulated pure tones. As the tones were about 0.7 octave apart and were amplitude-modulated with different frequencies (in the range 2–32 Hz), they were perceived as separate streams. Listeners had to detect a change in the frequency (experiments 1 and 2) or the shape (experiments 3 and 4) of the modulation of one of the five tones, in the presence of an informative cue orienting selective attention either before the scene (pre-cue) or after it (post-cue). The changes left intensity unchanged and were not detectable in the spectral (tonotopic) domain. Performance was much better with pre-cues than with post-cues. Thus, change deafness was manifest in the absence of an appropriate focusing of attention when the change occurred, even though the streams and the changes to be detected were acoustically very simple (in contrast to the conditions used in previous demonstrations of change deafness). In one case, the results were consistent with a model based on the assumption that change detection was possible if and only if attention was endogenously focused on a single tone. However, it was also found that changes resulting in a steepening of amplitude rises were to some extent able to draw attention exogenously. Change detection was not markedly facilitated when the change produced a discontinuity in the modulation domain, contrary to what could be expected from the perspective of predictive coding.

1. Introduction

A perfectly detectable change in a complex visual image may pass unnoticed if the observer's attention is not focused on the detail that changes. Numerous studies have been devoted to this phenomenon, called “change blindness” (see, e.g., Rensink et al., 1997). They have inspired other studies in which auditory counterparts of change blindness were looked for, using in particular acoustic “scenes” spontaneously perceived by the listeners as combinations of separate auditory objects. The objects in question were, in most of the experiments, naturalistic stimuli such as animal calls, productions of musical instruments, or human speech (Eramudugolla et al., 2005; Backer and Alain, 2012; McAnally et al., 2010; Gregg and Samuel, 2008, 2009; Gregg and Snyder, 2012; Gregg et al., 2014; Pavani and Turatto, 2008). A typical change to be detected was the disappearance of one of the stimuli composing the scene. The task was performed in the absence or the presence of a cue drawing the listener's attention to the critical component of the scene before the change. In the presence of this cue, detection performance was essentially perfect. In the absence of a cue, by contrast, disappearance detection was poor when the scene initially contained more than three or four stimuli. The authors thus concluded that humans are prone to change deafness, as well as change blindness, and that attention is very important for auditory change detection.

From the purely acoustic point of view, however, the changes used in these experiments were subtle and complex, because each of the stimuli mixed in a scene was already complex, both spectrally and temporally. It can therefore be doubted that change detection entirely occurred in the sensory domain, as the term “deafness” would suggest. Instead, change detection might have followed a categorization of the stimuli (which were typically easily categorizable) and then be limited by a nonsensory form of memory, with a smaller capacity than auditory sensory memory (Demany et al., 2008; Pavani and Turatto, 2008; McAnally et al., 2010). Support

Abbreviation: SAR, shortening of amplitude rise

* Corresponding author.
E-mail addresses: laurent.demany@u-bordeaux.fr (L. Demany), bayle.yann@live.fr (Y. Bayle), emilie.puginier@free.fr (E. Puginier), catherine.semal@bordeaux-inp.fr (C. Semal).

http://dx.doi.org/10.1016/j.heares.2017.07.013
0378-5955/

for this hypothesis was provided by Gregg and Samuel (2009) and Gregg and Snyder (2012), who found in their experiments that change detection was more affected by the semantic relationship between the components of the two successive scenes than by their acoustic distance.

Chait and coworkers (Cervantes Constantino et al., 2012; Sohoglu and Chait, 2016a) investigated change detection in acoustic scenes consisting of at least 4 and up to 14 amplitude-modulated pure tones, sufficiently spaced in frequency to be resolvable by the cochlea. As the tones were modulated at different frequencies (ranging from 3 to 35 Hz), the scenes were perceived as sets of separate auditory objects; but of course these objects were both acoustically simpler and less easily categorizable than real-world sounds. Listeners had to detect the appearance or the disappearance of a single amplitude-modulated tone in a scene. The authors mainly found that, in comparison with disappearance detection, appearance detection was markedly better, quicker, and also less affected by the number of concurrently presented tones (see also, in this respect, Schnuerch et al., 2014). This suggests that appearance detection was less attention-demanding than disappearance detection. As pointed out by the authors, the perceptual advantage of appearance over disappearance might have originated from neural adaptation: in a tonotopically organized neural array such as the auditory nerve, an appearing tone stimulated units that had been only weakly stimulated previously, as the tone frequency was novel; thus, due to adaptation, the initial response to an appearing tone was expected to be stronger than the response to a tone which was already part of the scene, and this was liable to make the appearing tone “pop-out" perceptually (Palmer et al., 1995).

In the study reported here, as in those of Chait and coworkers, listeners had to detect changes in scenes composed of concurrently amplitude-modulated pure tones. However, whereas the changes produced by Chait and coworkers were mostly spectral, ours were purely temporal: each change in a scene affected only the frequency or the shape of the modulation imposed on a single tone, randomly selected among the five components of the scene; the carrier frequency and the overall level of this tone did not change. In order to assess the effect of selective attention on temporal change detection, detection performance was measured when the scene was preceded by a cue (“pre-cue") pointing to the critical component tone, and compared to the performance obtained when the cue followed the scene (“post-cue"). A post-cue provided no *perceptual* advantage, in contrast to a pre-cue, but it provided the same advantage as a pre-cue with regard to the *decision* processes involved in the task, from the point of view of signal detection theory (see, e.g., Kinchla et al., 1995).

To our knowledge, there exists only one previous study in which listeners had to detect purely temporal changes within single auditory streams in scenes perceived as sets of concurrent streams. In that study (Botte et al., 1997), the scenes consisted of three streams of tone bursts. On each trial, the listener's attention was drawn to one of these three streams by a pre-cue. The change to be detected (a local violation of isochrony) took place in the cued stream on 70% of trials, and in one of the other two streams on the remaining trials. Change detection was found to be much worse in the latter case. However, in the framework of signal detection theory, this deficit could be ascribed at least in part to a problem of decision: the authors state that “the subjects were told to respond to any temporal irregularities in the multistream sequences but were not told explicitly that the irregularity could occur in noncued streams” (Botte et al., 1997, p. 422).

We wished to determine if the influence of attention on temporal change detection depends on the nature of the change. In this regard, we initially compared the detection of increases and decreases in modulation frequency. In addition, as the modulation of each tone before or after a change was periodic and very simple, it was interesting to determine if an *instantaneous* change in its frequency would be treated as a transient by the auditory system and therefore would draw attention automatically, like a sudden luminance transition in the visual domain (Phillips, 1974). To this end, two scene continuity conditions were compared. In the “Continuous” condition, the modulation frequency of the changing tone was shifted instantaneously to a new value, but this always occurred at a modulation phase corresponding to zero amplitude, so that for this tone there was a discontinuity in the modulation domain but not in the waveform fine structure; the rest of the scene was completely continuous. In the “Break” condition, on the other hand, the scene as a whole was briefly interrupted between the pre-change and post-change parts of the changing tone, and this discontinuity was independent of the change to be detected.

Another significant aspect of our study is that we investigated change detection using two different tasks, in separate experiments. In experiments 1 and 3, the scene presented on a given trial either changed (because one of its components changed) or did not change; the task was simply to detect the changes; this could be done, hypothetically, on the basis of global perceptual properties of the scene as well as by attending selectively to the changing tone (McAnally et al., 2010; Cervantes Constantino et al., 2012). In experiments 2 and 4, on the other hand, each scene included a component that changed, and the task was to indicate if this component was, or was not, the one to which attention was drawn by the pre- or post-cue.

2. Experiment 1

In this experiment, listeners had to detect changes in scenes consisting of five sinusoidally amplitude-modulated pure tones. On the trials in which a change occurred, this change was, in separate blocks of trials, an increase or a decrease in the modulation frequency of a single tone. Detection sensitivity (d') was measured in the Break and Continuous conditions briefly described above.

2.1. Method

The experiments reported here were carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.1.1. Stimuli

The carrier frequencies of the five tones composing a scene varied randomly from trial to trial (on a logarithmic frequency scale), but were always separated by intervals of 850 cents (1 cent = 1/100 semitone = 1/1200 octave). The lowest possible carrier frequency was 250 Hz and the highest one 4000 Hz. Each tone had a nominal sound pressure level (SPL) of 53 dB at full amplitude. The amplitude modulations, sinusoidal in shape, had a depth of 100%. At the onset of a scene, the modulation frequency of each tone was selected randomly between 2 and 32 Hz (using again a logarithmic scaling of frequency), with two constraints: (1) the modulation frequencies of the five tones had to be separated by intervals of 850 cents; (2) the modulation frequency of any tone had to be at least 20 times smaller than its carrier frequency; this second constraint ruled out the possibility of detecting a change in the spectral (tonotopic) domain rather than the temporal domain (we come back to that issue below). When the scene changed, the modulation frequency of a single tone, selected at random, was increased or decreased by 600 cents, i.e., half an octave; this change in modulation frequency was not expected to have a substantial effect on “long-term" loudness (Moore, 2014).