Hearing Research xxx (2017) 1–9

Contents lists available at ScienceDirect

Hearing Research

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

Research Paper

Auditory sequential accumulation of spectral information

Yi Shen

Department of Speech and Hearing Sciences, Indiana University Bloomington, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:
Received 29 May 2017
Received in revised form
3 October 2017
Accepted 4 October 2017
Available online xxx

Keywords:
Temporal integration
Informational masking
Auditory memory
Sequential processing

ABSTRACT

In many listening situations, information about the spectral content of a target sound may be distributed over time, and estimating the target spectrum requires efficient sequential processing. Listeners' ability to estimate the spectrum of a random-frequency, six-tone complex was investigated and the spectral content of the complex was revealed using a sequence of bursts. Whether each of the six tones was presented within each burst was determined at random according to a presentation probability. In separate conditions, the presentation probabilities (p) ranged from 0.2 to 1, the total number of bursts varied from 1 to 16, and the inter-burst interval was either 0 or 200 ms. To evaluate the information acquired by the listener, the burst sequence was followed, after a 500-ms silent interval, by the six-tone complex acting as an informational masker and the listener was required to detect a pure-tone target presented simultaneously with the masker. Greater performance in this task indicates more accurate estimation of the spectrum of the complex by the listener. Evidence for integration of information across bursts was observed, and the integration process did not significantly depend on inter-burst interval.

1. Introduction

The auditory system is highly efficient in processing sequentially presented information. For example, speech recognition in noise is known to depend on contextual cues, and these cues are often distributed across time (e.g., Moore, 2003). Moreover, the processing of this sequential acoustic information is robust against temporal interruptions and masking (e.g., Miller and Licklider, 1950). It appears that the auditory system can keep acoustic information in short-term memory and the stored information is available for retrieval and computation at a later time. In other words, the auditory system is capable of combining “multiple looks” across time (Viemeister and Wakefield, 1991). The current study investigated the efficiency of the auditory system in combining temporally distributed information to determine the spectral content of a target sound.

How the auditory system processes acoustic information over time has been a long-standing issue in psychoacoustics. It is known that absolute threshold decreases and loudness increases with increasing stimulus duration (e.g., Garner and Miller, 1947; Plomp and Bouman, 1959; Zwislocki, 1960; Florentine et al., 1988). Moreover, listeners cannot follow envelope fluctuations at very high rates (e.g., Viemeister, 1979) or detect very brief temporal gaps (e.g., Buunen and Van Valkenburg, 1979; Fitzgibbons and Wightman, 1982; Shailer and Moore, 1983). These phenomena have been explained using the concept of a temporal integration window. According to the temporal integration hypothesis, the inputs into the auditory system, after peripheral processing, are smoothed by a sliding temporal window. The duration of the integration window represents the sluggishness of the auditory system, with longer window duration corresponding to poorer temporal acuity (e.g., Viemeister, 1979; Forrest and Green, 1987; Moore et al., 1988). However, in many situations, the auditory system seems to be able to carry out more sophisticated sequential operations than just summing acoustic energy within a contiguous temporal neighborhood (Viemeister and Wakefield, 1991). This involves keeping prior stimuli in short-term memory and combining information from the memory with that from the on-going stimulus. These computations could involve both bottom-up and top-down processes (e.g., Bregman, 1990; Näätänen and Winkler, 1999; Alain et al., 2001; Sussman et al., 2002; Cusack et al., 2004; Micheyl and Oxenham, 2010).

Many studies have investigated high-level auditory sequential processing through the auditory stream-segregation paradigm. For example, two alternating tones of different frequencies were perceived as a single perceptual stream or two segregated streams, depending on factors such as frequency separation, alternation rate, and sequence duration (Van Noorden, 1975; Bregman and Campbell, 1971; Bregman, 1978; Moore and Gockel, 2002; Carlyon

E-mail address: shen2@indiana.edu.

https://doi.org/10.1016/j.heares.2017.10.001
0378-5955/

and Gockel, 2008; Micheyl and Oxenham, 2010; Richards et al., 2012). Sequential grouping or segregation has also been studied using stimuli other than pure tones, and it has been demonstrated that perceptual grouping across time can be based on fundamental frequency, spatial location, spectral cues, component phase, and level (e.g., Darwin, 1997; Cusack and Roberts, 1999; Vliegen and Oxenham, 1999; Grimault et al., 2002; Roberts et al., 2002; Stainsby et al., 2004; Gaudrain et al., 2007). These stream segregation studies have demonstrated that the auditory system is capable of assessing the consistency of acoustic information over time and grouping the consistent components into the same perceptual stream.

In a realistic auditory scene, the acoustic information associated with individual sound sources may vary over time and exhibit uncertainty. In such situations, the auditory system may have to progressively estimate the properties of each sound source based on acoustic information that is distributed across time, to limit the adverse influences of uncertainty. This process has been revealed using informational-masking paradigms. Informational masking refers to the masking effects that are not related to peripheral interactions between the target and masker, but rather to the similarity between the target and masker or the uncertainty associated with them (e.g., Watson and Yost, 1987; Leek et al., 1991; Durlach et al., 2003).

Kidd et al. (1994) investigated whether listeners can take advantage of temporal consistency in the target sound to reduce the informational masking effect from a random-frequency masker. In the Multiple-Burst-Same (MBS) condition, the masker consisted of a sequence of 50-ms bursts and each burst contained multiple simultaneous pure tones. The frequencies of these tones were randomly drawn for each trial, and were fixed across bursts on the same trial. The target to be detected was a sequence of 1-kHz tones, gated on and off with the masker bursts. Due to the uncertainty in the frequencies of masker components, this task was very challenging even when the masker frequencies were chosen to minimize interactions between the target and masker at the auditory periphery. On the other hand, in the Multiple-Burst-Different (MBD) condition, the frequency components in the masker were redrawn for each burst. This led to a significant improvement in the detection threshold for the target compared to the MBS condition.

Several processes could contribute to this MBD advantage. First, similar to auditory stream segregation, the temporal consistency of the target may contribute to the formation of an auditory stream that is distinct from the masker bursts (Kidd et al., 1994; Huang and Richards, 2006; Micheyl et al., 2007). Second, the listeners may be able to accumulate statistical evidence with regard to the long-term spectrum of the masker (Kidd et al., 2003). These different processes are relatively difficult to dissociate in informational masking (Kidd et al., 1994, 2003; Micheyl et al., 2007) or stream segregation (Elhilali et al., 2009; Akram et al., 2014) experiments using the MBS/MBD stimuli, because the two mechanisms (i.e. stream segregation and information accumulation) would lead to similar effects for simultaneously presented target and masker stimuli.

The current study focused on the ability to accumulate statistical evidence over time. Information with regard to the spectrum of a random-frequency, multi-tone complex was conveyed using multiple brief bursts. Each burst in the sequence contained a subset of the frequency components from the complex, determined at random and separately for each burst. To obtain an accurate estimate of the spectrum of the complex, accumulation of information across multiple bursts is required. Efficiency in estimating the spectral content of the complex was estimated indirectly with an experimental paradigm inspired by previous informational masking studies.

It is known that random-frequency, multi-tone complexes give rise to informational masking (e.g., Neff and Green, 1987; Neff and Callaghan, 1988; Oh and Lutfi, 1998). When a preview of the masker is presented as a precursor to the target and masker, the detection threshold often improves significantly. This phenomenon is often referred to as the auditory enhancement effect (Viemeister, 1980). Sensory mechanisms have been invoked as explanations of the enhancement effect. For example, according to the adaptation of suppression/inhibition hypothesis (Viemeister and Bacon, 1982; Nelson and Young, 2010; Byrne et al., 2011; Shen and Richards, 2012), the presence of the precursor gives rise to suppression or inhibition in the spectral region of the target, and the suppression or inhibition effect adapts over time. When the masker and target are later presented, the response to the target is enhanced since it is subjected to less suppression/inhibition. Recent findings have suggested that sensory mechanisms may be insufficient to fully account for the enhancement effect (e.g., Byrne et al., 2013). Rather, the benefit of the precursor may partially occur because it provides knowledge about the spectral content of the masker (Neff and Green, 1987; Richards and Neff, 2004; Richards et al., 2004; Kidd et al., 2011). Frequency cuing effects have been observed not only in detection tasks, but also in tasks that involve judgements based on pitch (Demany and Ramos, 2005; Erviti et al., 2011; Carcagno et al., 2012) or amplitude modulation (Shen, 2016).

A core assumption of the current study is that if a listener can estimate the spectrum of a multi-tone complex well, then the complex will provide little masking. Thus, higher target detectability indicates higher fidelity of the estimated masker spectrum. Using this approach, the ability to accumulate statistical evidence across precursor bursts was measured as a function of the number of bursts and inter-burst interval (IBI). These stimulus manipulations were intended to alter the way the statistical information was sequentially distributed. When the number of bursts increases, the precursor sequence contains more information regarding the spectrum of the upcoming masker. Therefore, the effectiveness of the masker is expected to decrease with increasing number of bursts.

Besides sequential information processing, manipulations of the total number of bursts and the IBI may change the amount of frequency cuing via sensory mechanisms. For example, increasing the number of bursts increases the total energy of the precursor. The increased total energy of the precursor may then affect target detection through mechanisms related to neural adaptation (Viemeister, 1980). To investigate the effects of stimulus manipulations unrelated to information processing, Exp. I used precursor bursts that were copies of the complex masker in the detection task. Because the precursors were always informative about the masker as the number of bursts and IBI were varied, the effects of these stimulus manipulations reflected mainly sensory processes. In Exp. II, the precursors with reduced informativeness were used, the effects of number of bursts and IBI observed in Exp. II reflected both sensory processes and non-sensory, information-processing mechanisms.

2. Experiment I: cueing effect with deterministic precursor sequences

2.1. Method

2.1.1. Listeners

Nine listeners (six females) were recruited from the student population at Indiana University. The listeners were between 19 and 29 years of age and had audiometric thresholds equal to or better than 15 dB HL between 250 and 8000 Hz in both ears. For each listener, the ear with the lower pure tone average (PTA) threshold (mean of the hearing level at 0.5, 1, and 2 kHz) was tested.

Please cite this article in press as: Shen, Y., Auditory sequential accumulation of spectral information, Hearing Research (2017), https://doi.org/10.1016/j.heares.2017.10.001