Hearing Research 350 (2017) 244–250

Contents lists available at ScienceDirect

Hearing Research

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

Research Paper

Comparison of perceptual properties of auditory streaming between spectral and amplitude modulation domains

Shimpei Yamagishi [a, b, *], Sho Otsuka [c], Shigeto Furukawa [c], Makio Kashino [c, d]

[a] Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama, Kanagawa, 226-8503, Japan
[b] Reseach Fellow of Japan Society for the Promotion of Science, Japan
[c] NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa, 243-0198, Japan
[d] School of Engineering, Tokyo Institute of Technology, Yokohama, Kanagawa, 226-8503, Japan

A R T I C L E I N F O

Article history:
Received 4 October 2016
Received in revised form
20 February 2017
Accepted 15 March 2017
Available online 18 March 2017

Keywords:
Auditory scene analysis
Bistable perception
Amplitude modulation
Sequential organization of sounds

A B S T R A C T

The two-tone sequence (ABA_), which comprises two different sounds (A and B) and a silent gap, has been used to investigate how the auditory system organizes sequential sounds depending on various stimulus conditions or brain states. Auditory streaming can be evoked by differences not only in the tone frequency ("spectral cue": ΔF_{TONE}, TONE condition) but also in the amplitude modulation rate ("AM cue": ΔF_{AM}, AM condition). The aim of the present study was to explore the relationship between the perceptual properties of auditory streaming for the TONE and AM conditions. A sequence with a long duration (400 repetitions of ABA_) was used to examine the property of the bistability of streaming. The ratio of feature differences that evoked an equivalent probability of the segregated percept was close to the ratio of the Q-values of the auditory and modulation filters, consistent with a "channeling theory" of auditory streaming. On the other hand, for values of ΔF_{AM} and ΔF_{TONE} evoking equal probabilities of the segregated percept, the number of perceptual switches was larger for the TONE condition than for the AM condition, indicating that the mechanism(s) that determine the bistability of auditory streaming are different between or sensitive to the two domains. Nevertheless, the number of switches for individual listeners was positively correlated between the spectral and AM domains. The results suggest a possibility that the neural substrates for spectral and AM processes share a common switching mechanism but differ in location and/or in the properties of neural activity or the strength of internal noise at each level.

1. Introduction

In natural acoustic environments, the auditory system analyzes incoming complex sounds and organizes their components into auditory objects or streams to make sense of the current auditory scene. The auditory streaming paradigm has been widely used to study the sequential organization of incoming components in the auditory system (van Noorden, 1975; Bregman, 1990; Moore and Gockel, 2012). With the repeated ABA_ sequence, where A and B represent two tones with different frequencies and "_" represents a silent gap, van Noorden (1975) showed that listeners tended to perceive a single coherent stream ("ABA_ABA_ …") when the frequency difference between the A and B tones was small (or the presentation rate was slow). On the other hand, listeners tended to perceive two distinct streams (a fast-tempo "A_A_A_A_ …" and a slow-tempo "_B____B___ …") when the frequency difference was large (or the presentation rate was fast). In the present study, the former percept is referred to as a one-stream percept (S1) and the latter is referred to as a two-stream percept (S2). A large number of studies have examined how auditory streaming depends on the difference in frequency between the A and B (referred to as the "spectral cue"), but auditory streaming can also be evoked by other cues, such as a difference in the fundamental frequency of bandpass filtered harmonic complexes with unresolved harmonics (Vliegen and Oxenham, 1999) and a difference in amplitude modulation (AM) rate of two sounds that produce almost identical long-term excitation patterns (Grimault et al., 2002; Dolležal et al., 2012; Szalárdy et al., 2013). The cue based on AM rate difference is referred to here as the "AM cue". Several studies reported that neural activity is correlated with auditory streaming based on such

* Corresponding author. NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa, 243-0198, Japan.
E-mail addresses: shimpei.yamagishi@gmail.com (S. Yamagishi), otsuka.s@lab.ntt.co.jp (S. Otsuka), furukawa.shigeto@lab.ntt.co.jp (S. Furukawa), kashino.makio@lab.ntt.co.jp (M. Kashino).

http://dx.doi.org/10.1016/j.heares.2017.03.006
0378-5955/

temporal cues (Gutschalk et al., 2007; Itatani and Klump, 2009, 2011; Dolležal et al., 2014).

Spectral cues have a strong influence on auditory streaming and this perceptual predominance can be accounted for by peripheral channeling theory (Hartmann and Johnson, 1991). According to this theory, a one-stream percept occurs if the two sounds have highly overlapping excitation patterns (excite the same peripheral channels), while a two-stream percept occurs if the sounds produce excitation patterns with little or no overlap. Several perceptual (Rose and Moore, 2000), physiological (Fishman et al., 2001, 2004; Bee and Klump, 2004, 2005) and modeling studies (Beauvois and Meddis, 1996; McCabe and Denham, 1997) support this idea, not only for peripheral excitation pattern but also at multiple levels of the auditory pathway.

Recently, several studies have explored the neural basis of auditory streaming by taking advantage of the bistable characteristic of stream perception. Prolonged exposure to a repeated ABA_ sequence often produces stochastic perceptual switching between S1 and S2 without any change in the stimulus (Denham and Winkler, 2006; Pressnitzer and Hupe, 2006; Kashino and Kondo, 2012; Kondo et al., 2012). This phenomenon provides key information about how the brain organizes sequential sounds, because changes in neural activity in response to the unchanged stimulus can reflect the change in stream perception. Several studies using this approach have reported that auditory streaming involves widely distributed neural sites, including both subcortical and cortical areas (Gutschalk et al., 2005; Winkler et al., 2005; Pressnitzer et al., 2008; Kondo and Kashino, 2009; Schadwinkel and Gutschalk, 2011; Kashino and Kondo, 2012; Yamagishi et al., 2016).

Studies concerning channeling theory or the bistability described above have typically examined a single cue, mostly the spectral one. Therefore, the extent to which the findings of the studies are specific to the cue examined or can be generalized to a streaming percept evoked by various cues remain unclear. The effects of cue difference on auditory streaming have been compared for spectral and fundamental frequency cues (Vliegen et al., 1999) and spectral, location, and AM cues (Szalárdy et al., 2013). However, the former study did not focus on the channeling theory or bistability properties, and the latter study did not compare the effects of cue difference on bistability properties at compatible perceptual ambiguity.

The present study examined the relationship between the perceptual properties of auditory streaming based on spectral and AM cues. The principal aim was to explore differences and similarities in cue-dependent effects on auditory stream perception. First, we examined how well the difference in internal representation of a feature could explain the ratio of S1 and S2 percepts, as hypothesized in the channeling theory. We assumed that the "perceptual distance" for each domain is associated with the difference in the internal representation of a feature and is determined by the bandwidths of the auditory and modulation filters. The modulation filter bank is a set of channels in the modulation domain. It is considered to be located at a higher level relative to the peripheral level at which the auditory filter bank is assumed to be located (Dau et al., 1997; Ewert and Dau, 2000; Sek and Moore, 2003; Moore et al., 2009). These studies indicated that the modulation filters are about three to five times broader than auditory filters, when the bandwidth is expressed relative to the center frequencies. Assuming that perceptual distance is related to bandwidth (a change of one bandwidth representing a fixed perceptual step), a given perceptual difference would require a larger separation on a logarithmic scale in the AM domain than in the spectral domain. Furthermore, assuming that the S1/S2 ratio is monotonically related to perceptual difference, the perceptual difference should be the same for the AM and spectral domains if the S1/S2 ratio is the same for the two domains. Second, we compared the number of perceptual switches between the spectral and AM domains at compatible perceptual ambiguity to assess the relationship between the properties of auditory bistability in the two feature domains.

2. Materials and methods

2.1. Listeners

Fourteen adults with normal hearing participated in the experiment. They ranged in age from 21 to 39 years (mean = 33, twelve females and two males). The experimental protocols were approved by the Research Ethics Committee of Nippon Telegraph and Telephone (NTT) Communication Science Laboratories. All listeners gave written informed consent prior to the experiment.

2.2. Stimuli

We used two types of stimuli (Fig. 1). One was a sequence of pure-tone bursts (*TONE* condition). The other was a sequence of sinusoidally amplitude-modulated tones (*AM* condition). For both conditions, the stimuli consisted of four hundred repetitions of the ABA_ sequence in one trial. The duration of one cycle of the ABA_ sequence was 440 ms. The duration of each signal was 50 ms, including 10-ms rising and falling raised-cosine ramps. The silent interval between A and B tones within a triplet was 60 ms. The duration of the silent gap (_) between ABA triplets was 170 ms.

For the TONE condition, the A-tone frequency was fixed at 600 Hz. The frequency difference (ΔF_{TONE}) between A and B was

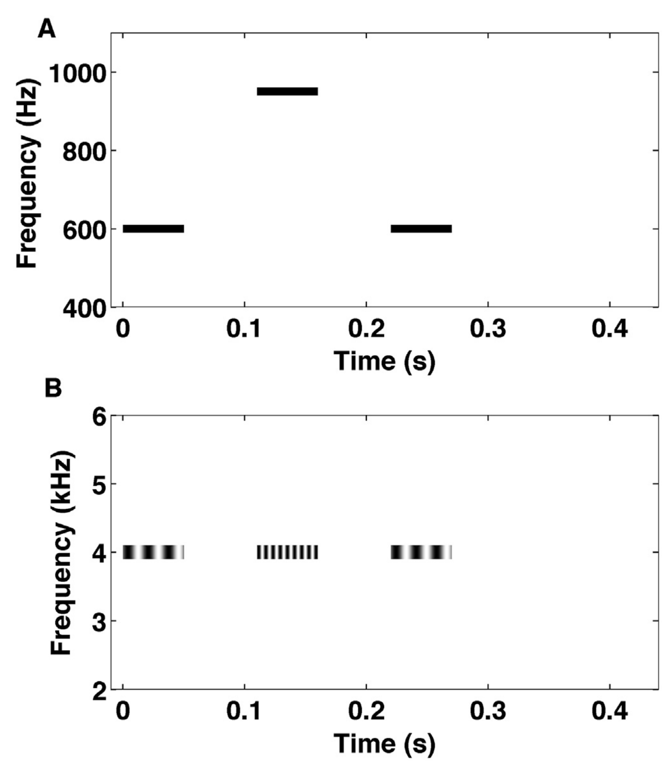

Fig. 1. Schematic spectrograms of the stimuli. **(A)** An example of one cycle of the ABA_ sequence used in the eight-semitone ΔF_{TONE} condition. The A and B tones differed in frequency but not in envelope fluctuation. **(B)** An example of one cycle of the ABA_ sequence used in the two-octave ΔF_{AM} condition. The A and B signals differed in envelope fluctuation but not in carrier frequency.