Hearing Research 333 (2016) 247–254

Contents lists available at ScienceDirect

Hearing Research

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

Research paper

Harmonic fusion and pitch affinity: Is there a direct link?

Damien Bonnard, René Dauman, Catherine Semal, Laurent Demany*

Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, CNRS and Université de Bordeaux, 146 Rue Leo-Saignat, F-33076 Bordeaux, France

ARTICLE INFO

Article history:
Received 6 April 2015
Received in revised form
19 August 2015
Accepted 27 August 2015
Available online 1 September 2015

Keywords:
Harmonicity
Fusion
Pitch
Octave
Musical interval
Frequency ratio

ABSTRACT

Simultaneous pure tones approximately one octave apart tend to be fused perceptually and to evoke a single pitch sensation. Besides, sequentially presented pure tones show a subjective "affinity" or similarity in pitch when their frequency ratio is close to one octave. The aim of the study reported here was to determine if these two perceptual phenomena are directly related. Each stimulus was a triplet of simultaneous or successive pure tones forming frequency ratios varying across stimuli between 0.96 and 1.04 octaves. The tones were presented at a low sensation level (15 dB) within broadband threshold-equalizing noise, in order to prevent them from interacting in the cochlea when they were simultaneous. A large set of stimulus comparisons made by 18 listeners indicated that: (1) when the tones were simultaneous, maximal fusion was obtained for a mean frequency ratio deviating by less than 0.2% from one octave, and fusion decreased less rapidly above this frequency ratio than below it; (2) when the tones were presented successively, maximal pitch affinity was obtained for a mean frequency ratio significantly larger than one octave, and pitch affinity decreased more rapidly above this frequency ratio than below it. The differences between the results obtained for simultaneous and successive tones suggest that harmonic fusion and pitch affinity are unrelated phenomena.

1. Introduction

A sum of simultaneous pure tones with harmonic relationships (i.e., simple frequency ratios) is normally heard as a single sound, evoking a single pitch sensation, even though the component tones may be largely "resolved" (i.e., segregated) in the cochlea (see, e.g., Moore et al., 1986; Carlyon et al., 1992; Roberts and Bailey, 1996; Lin and Hartmann, 1998; Kalluri et al., 2008). This perceptual phenomenon, which is thought to play a major role in the auditory analysis of everyday acoustic scenes (Bregman, 1990), has been called "harmonic fusion". In addition, there is ample evidence that pure tones presented successively rather than simultaneously are perceived as more similar in pitch when they have a harmonic frequency ratio, especially 2:1 (one octave), than when their frequency ratio is substantially inharmonic. The perceptual affinity of non-simultaneous pure tones one octave apart – often called "tone chroma" following Bachem (1937) – has been demonstrated in a number of experiments on melody recognition (Dowling and Hollombe, 1977; Idson and Massaro, 1978; Kallman and Massaro, 1979; Massaro et al., 1980), as well as using tasks that did not require musical judgments (Deutsch, 1973; Borra et al., 2013); moreover, this affinity has been observed not only in human adults but also in three-month-old human infants (Demany and Armand, 1984), in rhesus monkeys (Wright et al., 2000), and in rats (Blackwell and Schlosberg, 1943). It is natural to hypothesize that the perceived affinity in pitch of non-simultaneous pure tones one octave apart is directly related to their propensity to fuse when they are presented simultaneously. The present study was intended to test this hypothesis, hereafter termed "hypothesis *H*". According to hypothesis *H*, crucially, simultaneous pure tones one octave apart are perceptually fused after an initial identification of their individual pitches by the central auditory system. Fusion may then occur because the relation between these pitches is recognized as an octave due to the existence of internal templates of the pitch intervals formed by harmonically related simultaneous tones. The existence of an internal template for the simultaneous octave could explain why an affinity is perceived between *non*-simultaneous pure tones one octave apart.

The mechanism of harmonic fusion has been the subject of intense speculation since the time of Helmholtz (1877/1954) and is still unclear (for reviews, see Carlyon and Gockel, 2008; Micheyl and Oxenham, 2010). As shown by Roberts and his coworkers (Roberts and Bregman, 1991; Roberts and Bailey, 1996; Roberts and

---

* Corresponding author.

E-mail addresses: damien.bonnard@chu-bordeaux.fr (D. Bonnard), rene.dauman@chu-bordeaux.fr (R. Dauman), catherine.semal@bordeaux-inp.fr (C. Semal), laurent.demany@u-bordeaux.fr (L. Demany).

http://dx.doi.org/10.1016/j.heares.2015.08.015
0378-5955/© 2015 Elsevier B.V. All rights reserved.

Brunstrom, 1998, 2001; Brunstrom and Roberts, 2000), harmonicity is not a necessary condition for the perceptual fusion of simultaneous and peripherally resolved pure tones: regularity of spectral spacing is sufficient to induce some amount of fusion. For instance, a sum of sinusoids with frequencies equal to 230, 430, 630, … 2430 Hz tends to be heard as a coherent sound. The components of this inharmonic stimulus are difficult to hear out individually. Disrupting the spectral regularity by shifting the frequency of one component by a few percent, in either direction, induces a perceptual segregation of this component (Roberts and Brunstrom, 1998). This is inconsistent with the idea that spectral fusion is entirely controlled by internal templates of harmonic relationships in the central auditory system. Roberts and Brunstrom (2001) argued that spectral fusion, as a whole, can be better understood on the basis of the temporal autocorrelation model of pitch perception proposed by Licklider (1951) and Meddis and Hewitt (1991). The crucial assumption of Roberts and Brunstrom is that, at odds with hypothesis *H*, the neural information used for spectral fusion (or segregation) is not a neural representation of the frequencies or pitches of the stimulus components; it is instead a richer and more "primitive" representation of the stimulus components, in the time domain; spectral fusion and pitch perception are supposed to be based on distinct processes, contrary to hypothesis *H* and contrary to the idea that "Nature gave us pitch to sort out the world" (Hartmann, 1996, p. 3502). Nevertheless, harmonically related tones are perceptually fused to a greater extent than tones with a regular spectral spacing but without harmonic relations (Roberts and Bailey, 1996; Roberts and Brunstrom, 1998; Brunstrom and Roberts, 2000). This gain in fusion caused by harmonicity might be due to the existence of harmonic templates in the central auditory system and might be related to the perception of pitch affinity. Therefore, the findings of Roberts and his coworkers do not rule out hypothesis *H*.

The classical pitch perception model proposed by Terhardt (1974) assumes that harmonic fusion is not innate and that human listeners acquire harmonic templates in early life through repeated exposure to periodic complex tones. Before this learning process, the resolved spectral components of a periodic complex tone are supposed to be perceived individually; each of them then evokes a separate "spectral pitch". The templates acquired during the learning process represent intervals between spectral pitches rather than frequency intervals. That is an important feature of the model because, according to Terhardt (1970, 1971, 1974), the pitch evoked by a partial in a periodic complex such as those typically heard by humans tends to be slightly different from the pitch evoked by the same pure tone presented in isolation. In consequence of these small pitch shifts, supposedly due to the existence of cochlear interactions between the partials of a complex tone, the acquired harmonic pitch intervals correspond to frequency ratios which, for pure tones presented successively rather than simultaneously, are slightly different from small-integer ratios. In this way, Terhardt provided a tentative explanation of a well-established oddity known as the "octave enlargement" phenomenon: For human listeners, successive pure tones must typically have a frequency ratio slightly exceeding 2:1 in order to form a perfectly tuned melodic octave (Ward, 1954; Walliser, 1969; Sundberg and Lindqvist, 1973; Dobbins and Cuddy, 1982; Demany and Semal, 1990; Hartmann, 1993). Terhardt (1970, 1971, 1974) argued that as the cochlear interactions between partials result in repulsions between their pitches, these interactions must be the source of the octave enlargement. In Terhardt's theory, therefore, both harmonic fusion and the perceived affinity of successive pure tones with harmonic or quasi-harmonic frequency ratios stem from one and the same learning process. This is in line with hypothesis *H*. The same basic idea was more recently supported by Schwartz et al. (2003) and Ross et al. (2007).

Terhardt's theory has been challenged on several grounds. First, it has been suggested that the pitch shifts supposedly explaining the octave enlargement phenomenon do not really exist (Peters et al., 1983; Hartmann and Doty, 1996). Second, listeners appear to have difficulty in identifying the sign (positive or negative) of small but detectable deviations from an octave interval for simultaneous pure tones (Bonnard et al., 2013) but this is not the case for successive pure tones (Dobbins and Cuddy, 1982; Bonnard et al., 2013); this difference seems to be at odds with the assumption that the same internal octave templates are used to detect mistunings with simultaneous and successive tones. Third, the detection of inharmonicity in a pair of simultaneous tones becomes extremely poor above about 2 kHz (Demany and Semal, 1988, 1990; Demany et al., 1991) and it is also extremely hard to identify, within complex tones containing many harmonics, a mistuned component when its frequency exceeds 2 kHz (Hartmann et al., 1990); in contrast, melodic octaves can still be recognized very accurately up to 4 kHz (Ward, 1954; Demany and Semal, 1990); how could this be the case if the internal templates of the melodic octave were entirely derived from experience with harmonic complex sounds? Nonetheless, the experimental evidence that we just summarized does not definitely discredit Terhardt's theory, and more generally hypothesis *H*.

Recently, hypothesis *H* has been reactivated by Borchert et al. (2011). In their study, listeners had to detect a small difference in fundamental frequency (F0) between two simultaneous groups of harmonics in separate frequency regions. According to Borchert et al., the perceptual cue used in this task was perceived fusion. Detection performance was poorer when the spectrally higher group of harmonics had the higher F0 than when it had the lower F0. Thus, positive mistunings were less well detected than negative mistunings. Borchert et al. suggested that this asymmetry had the same origin as the octave enlargement effect. A similar asymmetry had been previously reported by Demany et al. (1991), who found that for two simultaneous pure tones, positive deviations from an octave interval (perfectly tuned from the physical point of view) are less well detected than negative deviations. However, these authors did not interpret their findings in the same manner as Borchert et al. (2011). In fact, the studies of Demany et al. and Borchert et al. provide information about listeners' ability to discriminate (physically) harmonic sounds from (physically) inharmonic sounds, but they provide no information about the relationship between frequency ratio and perceived fusion because they did not assess perceived fusion directly. So, the observed asymmetry in the detection of inharmonicity is not necessarily related to the octave enlargement effect and it has no clear implication regarding hypothesis *H*.

The present experiment was based on the following reasoning. If hypothesis *H* is correct, then simultaneous pure tones approximately one octave apart should be maximally fused when the perceived affinity of their pitches is maximal; moreover, fusion and pitch affinity should decay similarly above and below the pitch interval corresponding to a perceptually perfect octave. We tested this prediction by determining precisely how perceived fusion — assessed with simultaneously presented tones — and pitch affinity — assessed using sequential presentation — depend on frequency ratio, for frequency ratios in the vicinity of 2:1. An important point is that, in the two experimental conditions ("simultaneous" versus "sequential"), the tones were presented at a low sensation level (nominally 15 dB) in a background of threshold-equalizing noise (Moore et al., 2000). Given the frequency selectivity of the auditory system (see, e.g., Moore, 2012), the cochlear representation of a tone in the "simultaneous" condition was affected to a much greater extent by the noise than by the tonal context. Thus, the