Hearing Research xxx (2013) 1–12

Contents lists available at ScienceDirect

Hearing Research

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

Review

Representation of frequency-modulated sounds in the human brain

Christian F. Altmann[a,b,*], Bernhard H. Gaese[c]

[a] Human Brain Research Center, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan
[b] Career-Path Promotion Unit for Young Life Scientists, Kyoto University, Kyoto 606-8501, Japan
[c] Institute for Cell Biology and Neuroscience, Goethe University, Max-von-Laue-Str. 13, 60438 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 10 May 2013
Received in revised form
26 July 2013
Accepted 27 July 2013
Available online xxx

ABSTRACT

Frequency-modulation is a ubiquitous sound feature present in communicative sounds of various animal species and humans. Functional imaging of the human auditory system has seen remarkable advances in the last two decades and studies pertaining to frequency-modulation have centered around two major questions: a) are there dedicated feature-detectors encoding frequency-modulation in the brain and b) is there concurrent representation with amplitude-modulation, another temporal sound feature? In this review, we first describe how these two questions are motivated by psychophysical studies and neurophysiology in animal models. We then review how human non-invasive neuroimaging studies have furthered our understanding of the representation of frequency-modulated sounds in the brain. Finally, we conclude with some suggestions on how human neuroimaging could be used in future studies to address currently still open questions on this fundamental sound feature.

This article is part of a Special Issue entitled <Human Auditory Neuroimaging>.

1. Frequency modulation as a ubiquitous sound feature

While the visual system is remarkably well adapted to resolve fine-scale spatial differences, our auditory system shows a similar aptitude in the temporal domain (see for example Recanzone, 2003). In particular, even small continuous changes in sound frequency – frequency modulations (FM) – are detected with high reliability (Fig. 1a shows examples of FM stimuli commonly used in research).

For us as humans, precise FM detection and recognition appears to be necessary, as frequency modulations are ubiquitous in human language. Early perceptual studies concluded that formant transitions are primary cues for the locus of articulation in stop consonant–vowel syllables (Liberman et al., 1967; Liberman and Studdert-Kennedy, 1978). For example, the stop consonant–vowel syllables /ba/, /da/, and /ga/ contain formant transitions (see also Fig. 1b) after the plosive and measurements showed that the rates of these transitions are in the range of about 16–30 octaves/s for the first and about 8 octaves/s for the second formants (based on table A1 in Kewley-Port, 1982). However, the information provided by the formant transition may not be sufficient to distinguish the place of articulation in different vowel contexts (Kewley-Port, 1982), so other cues such as spectrum at the time of the plosive might play an important role, too (Stevens and Blumstein, 1978; Blumstein and Stevens, 1979, 1980). Interestingly, a highly degraded form of speech – sine-wave speech – that consists of time-varying sine waves following the lower formants of the speech signal can be perceived as speech with prior knowledge of the content (Remez et al., 1981; Möttönen et al., 2006). Deficits in rapid auditory processing have been proposed to underlie language and literacy learning deficits (Tallal and Piercy, 1973, 1974; for a general review see Habib, 2000) and treatment strategies have been developed to improve these by training (Tallal et al., 1996; Gaab et al., 2007). In a recent review, Hämäläinen et al. (2012) summarized studies that investigated individuals with dyslexia either using behavioral or EEG/MEG measures. Amongst other measures (frequency, rise time, and duration discrimination and amplitude modulation/AM detection), FM detection was persistently impaired in dyslexic individuals compared to normal controls when modulation rates were below about 60 Hz.

In tonal languages (e.g., Mandarin Chinese, Thai, Vietnamese, see Fig. 1c for an example), suprasegmental modulations in frequency of formants signify word meaning. For example, in a series

---

Abbreviations: AM, amplitude modulation; aSSR, auditory steady-state responses; BOLD, blood oxygenation level-dependent; EEG, electroencephalography; FM, frequency modulation; IC, inferior colliculus; MEG, magnetoencephalography; MMF, mismatch field; MMN, mismatch negativity; MVPA, multivariate pattern analysis; PET, positron emission tomography; PT, planum temporale; STG, superior temporal gyrus

[*] Corresponding author. Career-Path Promotion Unit for Young Life Scientists, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. Tel.: +81 75 753 9295; fax: +81 75 753 9281.

E-mail addresses: altmann@cp.kyoto-u.ac.jp, chfaltmann@gmail.com (C.F. Altmann).

0378-5955/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.heares.2013.07.018

Please cite this article in press as: Altmann, C.F., Gaese, B.H., Representation of frequency-modulated sounds in the human brain, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.07.018

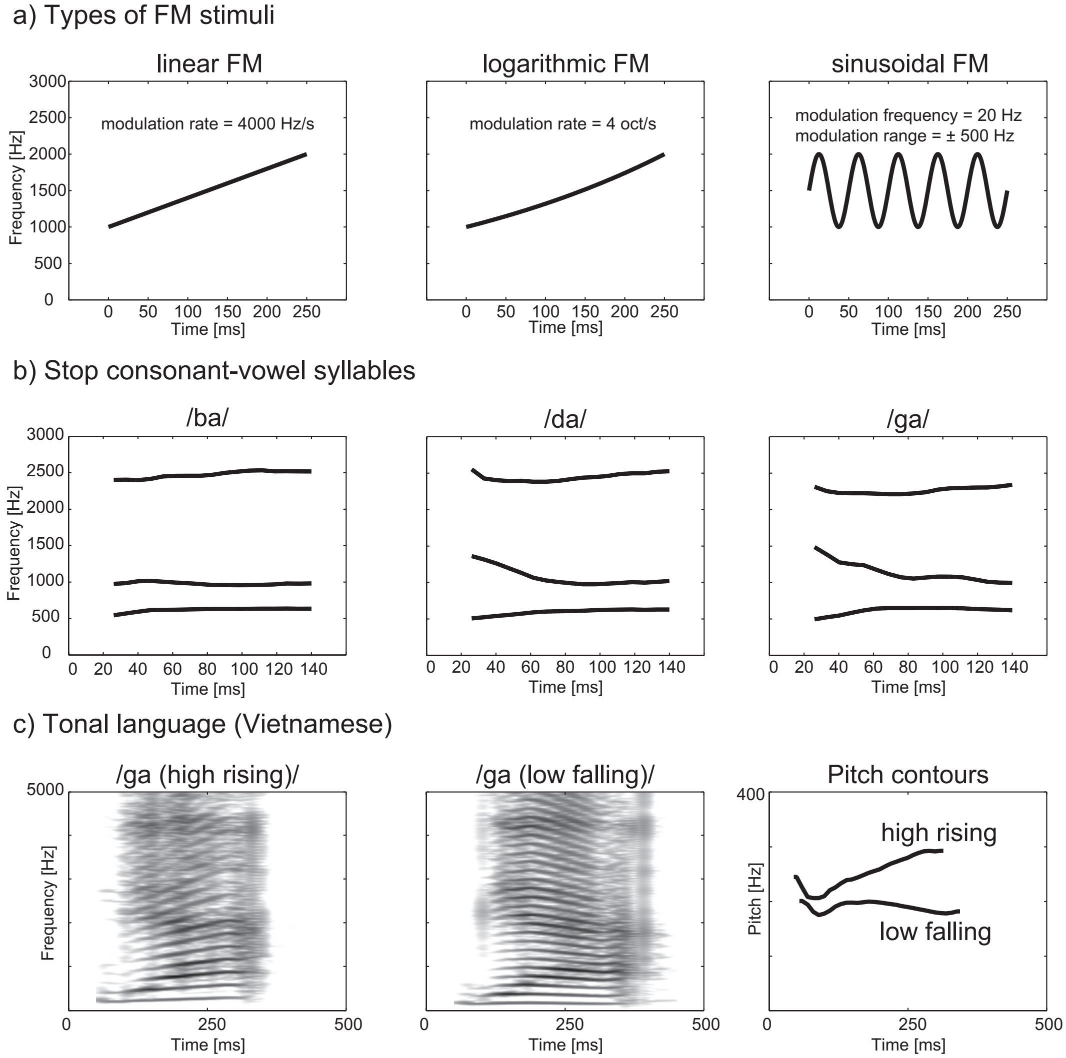

Fig. 1. Examples of FM stimuli and human speech vocalizations. a) Three different FM types: linear frequency modulated sweeps, logarithmic frequency modulated sweeps, and sinusoidal frequency modulated sweeps. Logarithmic FM sweeps derive their name from the fact that the logarithm of their instantaneous frequency changes linearly with time. b) First to third formants of three different stop consonant–vowel syllables: /ba/, /da/, and/ga/. The syllables were spoken by a male native German speaker and the formants were extracted with the "to Formant" function of the Praat software (http://www.fon.hum.uva.nl/praat/). c) Examples of tonal language, spoken by a female native Southern Vietnamese speaker. The left and center graph depicts spectrograms of the syllable /ga/ pronounced with high rising and low falling tone. The right graph shows the pitch contours of these syllables, extracted with the Praat software.

of recent studies (Xi et al., 2010; Zhang et al., 2012), the Chinese phoneme /pa/ with either high rising or falling lexical tone and intermediate morphed stimuli were used, with modulation rates of about 3.5–7.5 octaves/s (based on Fig. 2 in Xi et al., 2010). The importance of frequency modulations in the processing of tonal language has been shown by Zeng et al. (2005) by testing normal hearing and cochlear implant users with either AM or FM stimuli derived from Mandarin Chinese speech sounds. Their results suggested that FM plays a particularly important role for speech recognition under noisy conditions or in the presence of competing voices. Interestingly, native speakers of a tonal language (Mandarin Chinese) have been shown to identify FM direction better than speakers of a non-tonal language (English) while FM discrimination was at a similar level (Luo et al., 2007a). Thus, long-term experience with a language that requires reliable FM identification has an effect on non-speech sound processing. Slow modulations of the fundamental frequency (F0) of speech sounds can also convey linguistic (e.g., question or statement) and emotional information (Majewski and Blasdell, 1969; Lakshminarayanan et al., 2003).

FM is not exclusively found in human speech sounds, but also occurs in communication calls of other species. Non-human primates exhibit FM in their communicative behavior, for example New World monkey twitter calls contain FM with typical rates of 30–50 octaves/s (Bieser, 1998; Atencio et al., 2007). Japanese macaques use coo sounds for communication that comprise rising and falling FM patterns with the time of inflection as a discrimination cue. The macaques' classification of these cues exhibits properties similar to categorical perception in human speech (May et al., 1989). In rodents, rats have been shown to categorize FM sweep direction (Mercado et al., 2005; Gaese et al., 2006) as well as Mongolian gerbils (Wetzel et al., 1998). Bats constitute a special case: they not only use FM for communication but also for echolocation (Klug and Grothe, 2010), for which all bat species emit downward frequency sweeps. Evolutionary, this specialization has led to a "hypertrophy" of the auditory system (Pollak and Casseday, 1989), which renders them a model species for FM processing. The models describing the neural mechanisms underlying FM processing have in great part been developed for the bat model (e.g.,