Neuroscience Letters 441 (2008) 11–15

Contents lists available at ScienceDirect

Neuroscience Letters

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

Processing of pitch and time sequences in music

Christiane Neuhaus*, Thomas R. Knösche

Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany

ARTICLE INFO

Article history:
Received 16 January 2008
Received in revised form 20 May 2008
Accepted 29 May 2008

Keywords:
Pitch
Time
Randomisation
Encoding
Processing effort
Event-related potentials

ABSTRACT

Pitch and duration – either as written symbols or in auditory form – are the basic structural properties in tones that form a melodic sequence. From the cognitive perspective, it is still a matter of debate whether, and at which processing stage, these two factors are processed independently or interdependently. The present study addresses this issue from the neuroscientist's point of view by measuring event-related potentials (ERPs) in musicians and non-musicians. Either the pitches or the durations of the tones, or both, were permuted randomly over a set of melodies in order to remove all sequential ordering with respect to these factors. Effects of both, pitch and time order, on the peak amplitudes of the P1-N1-P2 complex were observed. ANOVA revealed that sequential processing may depend on the different levels of skill in analytical hearing. For musicians, strong interaction effects for all three ERP components corroborated the *interdependence* of pitch and time processing. Musicians also seem to rely on coherent time structure more than non-musicians and showed enlarged P1 and P2 components whenever tone duration, either with or without preserved pitch, was at random. Non-musicians tend to use ordered pitch relations for perceptual orientation, and main effects without any interactions might indicate some kind of *independent* processing of both dimensions at some processing stages.

When listening to melodies in an everyday situation, people implicitly understand the overall form and do not focus on melodic details. Nonetheless, every melody is a sequence of tones and unfolds step by step over time. In order to comprehend how a melody is processed as an entity, we suggest to zoom-in on the local structural level and consider how *adjacent* tones within a sequence are generally arranged. The approach is in accordance with the 'feature-integration theory of attention' [11] which claims that features within the same focus of attention can be related to each other and integrated to unitary percepts.

In a previous event-related potentials (ERP) study with non-musicians, Neuhaus and Knösche [6] investigated the neural correlates of tone sequence processing by contrasting structured melodic sequences with completely randomised versions. When regular melodies were processed, some early ERP components (P1 and P2) were significantly reduced, as compared to randomised sequences. The less predictable the tone progression was (due to randomisation), the higher was the processing effort to integrate the actual tone input, and early ERP components, presumably reflecting this encoding process, were enlarged. Using a similar experimental approach in conjunction with fMRI, Levitin and Menon [4] found that Brodmann Area 47 and its right-hemisphere homologue were significantly activated when musical syntax/structure was *preserved*, as compared to arbitrarily scrambled counterparts, having the same spectral energy, but lacking any temporal coherence. Thus, while ERP results of Neuhaus and Knösche [6] suggest an increase of activity for the *random* condition, the fMRI results of Levitin and Menon [4] demonstrate some extra activity for the *ordered* condition. This apparent contrast might be a consequence of different functional processes and/or different measuring methods. While event-related potentials focus on the time course of a cognitive process, neuroimaging methods show a much higher spatial resolution so that activation in small brain areas might also be detected.

A melody may be described as a rule-based sequence of tone items characterized by mainly two dimensions, 'pitch' and 'duration'. For the processing of spatio-temporal relationships on the local level, interval and rhythm are therefore the most relevant structural attributes. It is still a matter of debate whether the *structural* independence of pitch and time has a counterpart in *perceptual* independence, or if both factors are processed in an interwoven manner. Results of brain lesion studies seem to corroborate the independence theory, as lesions in the right auditory cortex affect the processing of pitch distance while damage of the left auditory cortex causes impairments in rhythm discrimination [10]. In line with that, Palmer and Krumhansl [8] demonstrated in experienced musicians that either pitch or time was sufficient to make judgments about the completeness of a musical phrase, sug-

* Corresponding author. Tel.: +49 3425 8875 16; fax: +49 3425 8875 11.
E-mail address: neuhaus@cbs.mpg.de (C. Neuhaus).

0304-3940/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2008.05.101

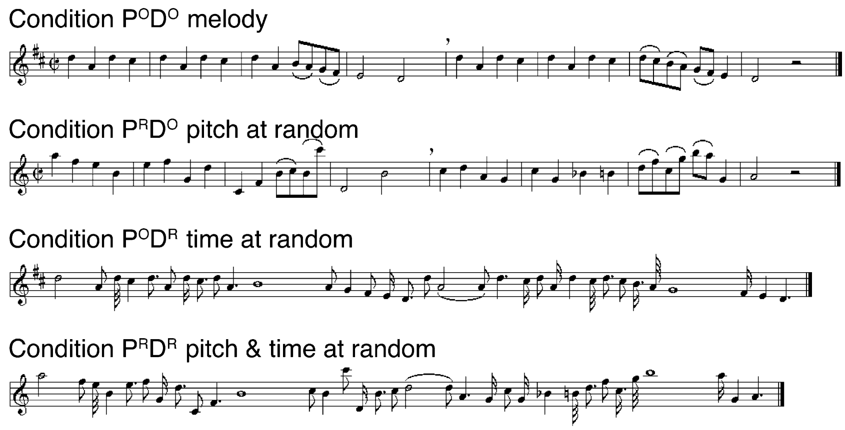

Fig. 1. Note examples, illustrating tone sequence structure. (a) Melody, pitch and time in balance ($P^O D^O$). (b) Same instrumental piece with preserved time structure, but randomised pitch ($P^R D^O$). (c) Pitch structure as defined in the original version. Duration values scrambled over all note examples ($P^O D^R$). (d) Pitch and time in randomised order ($P^R D^R$).

gesting the perceptual independence of both factors. As regards long-term memory for pitch and time, familiar folk songs were still recognizable when either isochronous tone sequences or pure rhythms were presented [2]. The score of correct naming responses of musicians and non-musicians in this study revealed the primacy of pitch over time (49% vs. 6% of correct answers), and that musical experience did not influence the identification result. Boltz [1], by contrast, demonstrated the *interdependence* between pitch and time in non-musicians by varying the type of instruction when folk tunes had to be learned. While focussing either on pitch, on duration, or on both during encoding, participants had lowest reproduction scores for the prior attention to pitch while disregarding rhythm, and highest performance scores for the prior attention to time when pitch was considered as the secondary information.

The present study investigates from the neuroscientist's perspective, (a) whether and to what extent pitches and durations of tones are processed as conjoined or as separate dimensions, and (b) to what extent tone sequence processing is influenced by formal musical training. For these purposes, we extended the previous experimental design [6] and used four types of tone sequences with the following properties, (a) melodies with unchanged order of pitches and durations (condition $P^O D^O$)[1], (b) pitch order preserved and time order at random (condition $P^O D^R$), (c) time order preserved and pitch order at random (condition $P^R D^O$), and (d) both factors permuted over the entire set of tones (condition $P^R D^R$). Our hypotheses were as follows: (1) the perception of melodies basically relies on the processing of pitch. Results supporting this claim have already been obtained with behavioural methods [2]. Sequences with scrambled pitches in contrast to scrambled durations might be processed with higher mental effort as reflected by increased ERP amplitudes. (2) Tone sequence processing is influenced by formal musical training in such a way that ERP amplitude differences between ordered vs. unordered pitch or time structure are larger in musicians than in non-musicians. Musicians might reveal training-induced plasticity effects in the auditory cortex. Pantev et al. [9] have demonstrated that instrumental experience during childhood leads to an increase of the early N1 for octave tones in the timbre of the learned instrument. Musacchia et al. [5] could even show a training-induced enlargement of brainstem responses due to the particular periodicity in individual sounds (syllables and cello tones). (3) *Interdependent* processing of pitch and time complies with the economical principle of mind. That is, sequence processing might be facilitated, as indicated by reduced ERP components, when tone structure is built on regular and rule-based conjunctions of both dimensions. (4) Encoding sequential information of pitch and/or time affects early but not late ERP components.[2] In line with this assumption are the results of Neuhaus and Knösche [6], indicating that some top-down strategies, such as cognitive integration effort, are already taking effect at an early stage of sequence processing.

Fourteen musically trained and 15 untrained normal-hearing subjects participated in the experiment. Musically trained persons (seven females, mean age 23.5 years, range 18–30 years) had started playing an instrument at the average age of 8.93 years, and daily practicing at the point of measurement was between 1 and 3.5 h. All of them were undergraduate students in music or musicology, their main instrument was the piano, oboe, cello, saxophone, or French horn. Musically untrained persons (eight females, mean age 25.17 years, range 22–30 years) had played an instrument, if at all, for altogether less than 1.6 years.

The original form ($P^O D^O$) of the stimulus material comprised tonal melodies with symmetric phrase structure as exemplified in Fig. 1a. Fig. 1b shows the same sequence where time structure is left unchanged, but pitch values are randomised over the entire set of examples ($P^R D^O$). Fig. 1c depicts the overall randomisation of time values (i.e. tone duration plus subsequent offset–onset distance, combined as inter-onset interval, IOI), while pitch structure is preserved ($P^O D^R$), and Fig. 1d shows simultaneous, but independent, permutation of the pitch and time dimensions ($P^R D^R$). For each condition, 80 full-length examples (average length of 10.3 s, S.D. = 3.2) were generated. Original melodies were played on a programmable keyboard (Yamaha PSR 1000; timbre 'piano') and were stored in MIDI format for subsequent randomisation. Original and modified MIDI versions were transformed to Soundblaster™ audio format. As the original melody was played the natural way by connecting adjacent notes with finger movements (legato), duration overlaps

[1] P stands for pitch, *D* for tone duration (time), while ^{O} and ^{R} classify ordered versus random structure. Thus, to specify a sequence type, we used four types of abbreviation: $P^O D^O$ (melody), $P^R D^O$ (random pitch and regular time order), $P^O D^R$ (regular pitch and random time order), and $P^R D^R$ (double permutation of pitch and time).

[2] We define early components as all deflections before 250 ms after stimulus onset. They mainly reflect the bottom-up processing of physical stimulus properties, although top-down influences might also occur (see also [5]). Late components of at least 300 ms onset-latency are considered to exclusively indicate higher, e.g. task-related, processing.