

Subjectively preferred octave size is resolved at the late stages of cerebral auditory processing

Jussi Jaatinen^{1,2}  | Jani Vääntänen^{2,3} | Viljami Salmela²  | Kimmo Alho² 

¹Musicology, Faculty of Arts, University of Helsinki, Helsinki, Finland

²Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

³Psychiatric Assessment and Consultation Clinic, Wellbeing services county of Pirkanmaa, Tampere, Finland

Correspondence

Jussi Jaatinen, Musicology, Faculty of Arts, University of Helsinki, P.O. Box 24, Helsinki FI-00014, Finland.
Email: jussi.jaatinen@iki.fi

Funding information

Alfred Kordelin Foundation

Edited by: Sophie Molholm

Abstract

Human listeners prefer octave intervals slightly above the exact 2:1 frequency ratio. To study the neural underpinnings of this subjective preference, called the octave enlargement phenomenon, we compared neural responses between exact, slightly enlarged, oversized, and compressed octaves (or their multiples). The first experiment ($n = 20$) focused on the N1 and P2 event-related potentials (ERPs) elicited in EEG 50–250 ms after the second tone onset during passive listening of one-octave intervals. In the second experiment ($n = 20$) applying four-octave intervals, musician participants actively rated the different octave types as ‘low’, ‘good’ and ‘high’. The preferred slightly enlarged octave was individually determined prior to the second experiment. In both experiments, N1-P2 peak-to-peak amplitudes attenuated for the exact and slightly enlarged octave intervals compared with compressed and oversized intervals, suggesting overlapping neural representations of tones an octave (or its multiples) apart. While there were no differences between the N1-P2 amplitudes to the exact and preferred enlarged octaves, ERP amplitudes differed after 500 ms from onset of the second tone of the pair. In the multivariate pattern analysis (MVPA) of the second experiment, the different octave types were distinguishable (spatial classification across electroencephalography [EEG] channels) 200 ms after second tone onset. Temporal classification within channels suggested two separate discrimination processes peaking around 300 and 700 ms. These findings appear to be related to active listening, as no multivariate results were found in the first, passive listening experiment. The present results suggest that the subjectively preferred octave size is resolved at the late stages of auditory processing.

Abbreviations: A4, commonly used tuning reference tone for pianos and orchestras (≈ 440 Hz); ANOVA, analysis of variance; EEG, electroencephalography; ERP, event-related potential; ET, equal-tempered tuning (a tuning system where all half steps in the musical scale have an equal frequency ratio of $\sqrt[12]{2}$ and octaves have a ratio of 2:1); f_0 , fundamental frequency; fMRI, functional magnetic resonance imaging; ICA, independent component analysis; MVPA, multivariate pattern analysis; N1, first negative-going evoked potential; P2, second positive-going evoked potential; SD, standard deviation; SPL, sound pressure level; SVM, support vector machine.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *European Journal of Neuroscience* published by Federation of European Neuroscience Societies and John Wiley & Sons Ltd.

KEYWORDS

EEG, multivariate pattern analysis, octave enlargement, pitch perception, psychoacoustics, tuning

1 | INTRODUCTION

The octave interval is one of the essential concepts in most musical systems worldwide (Honning & Bod, 2011). It is defined mathematically as a 2:1 ratio between the frequencies of two tones. However, psychoacoustic studies have suggested that human listeners tend to prefer enlarged octave intervals that slightly exceed the exact 2:1 ratio (Demany & Semal, 1990; Hartmann, 1993; Jaatinen et al., 2019; Sundberg & Lindqvist, 1973; Terhardt, 1971; Walliser, 1969; Ward, 1954). At the same time, the preferred tuning of the musical scale stretches; for example, the size of half steps increases as one progresses farther away from the reference tone ($A4 \approx 440$ Hz). The main reason for using octaves as measuring intervals is the simple 2:1 frequency ratio, which is relatively easy for humans to judge. The slightly enlarged preferred octave interval has been observed when sinusoidal, complex (a tone with harmonics) or Huggins' (Cramer & Huggins, 1958) tones are used (Hartmann, 1993) or when two tones are presented successively or simultaneously (Demany & Semal, 1990; Ward, 1954). The most common example of stretched tuning can be heard in modern pianos, where it is a de facto tuning standard (Jaatinen & Pätynen, 2022; Schuck & Young, 1943). Some variability has been observed in the preferred enlarged octave size between listeners, but no systematic differences have been reported. For example, musical background or culture does not affect the results significantly (Dowling & Harwood, 1985).

Several theories have been introduced to explain the possible neural origin of the octave enlargement phenomenon. Terhardt (1970, 1974) proposed the central template theory, suggesting that the tuning of a tonal memory matrix in semantic memory, acquired from hearing speech (vowels) since birth or even prenatally, would be stretched. This stretching would be explained by the masking effect caused by simultaneous partials of complex tones; the fundamental frequency of the octave-higher tone masks the second harmonics of the lower tone, resulting in a downward shift in the perceived pitch of the lower tone. Although complex tones are required to explain the masking effect and pitch shift, even when sinusoidal tones are used, listeners adjust their octave intervals sharp to match the learned stretched tonal matrix. Ohgushi (1983), in turn, proposed that pitch sensation is determined by a neural representation based on

an interspike interval histogram. He observed that the interspike intervals of a sinusoidal tone are not precisely locked to the stimulus waveform. Instead, the interspike interval histogram's first and second peaks are delayed due to neural refractory effects. The refractory effect is more important for higher frequencies, and therefore, the shift is greater in higher frequencies. According to this theory, the octave enlargement in frequency compensates for a refractory effect that increases with frequency (see also McKinney & Delgutte, 1999). While Terhardt's (1970, 1974) theory is based on the place theory of pitch perception and suggests the learned origin of the octave enlargement phenomenon, Ohgushi's (1983) theory is based on the timing theory of pitch perception and suggests an inherent origin of the octave enlargement phenomenon in the peripheral auditory system.

Hartmann (1993) compared Terhardt's (1970, 1974) and Ohgushi's (1983) theories in experiments, where he also used Huggins' tones as stimuli. He did not reject either of the theories but made an interesting additional finding; namely, the octave enlargement phenomenon was observed even with Huggins' tones. Huggins' tones are produced by presenting white noise to both ears; although the signals are identical in amplitude, a narrow frequency region is 180° out of phase in one ear. The shifted region can be perceived as a pitch sensation. Hartmann supposed that the pitch of the Huggins' tone is perceived somewhere in the central auditory system, much later than in the cochlear nerve, where Ohgushi's (1983) data were measured.

Bell and Jędrzejczak (2017) recently introduced a theory based on measurements of spontaneous otoacoustic emissions from the inner ear. These emissions suggested that cochlear frequency representations have steps at semitone intervals and that this semitone (frequency ratio between the higher and lower tone $1.063 \pm .005$) would be slightly wider than its equal-tempered counterpart (frequency ratio 1.059). The widened semitone steps would, therefore, cumulatively cause the tuning to widen. According to this theory, it also seems unlikely that the use of widened semitones in many (but not all) cultures would appear by chance. It may be that the ratio reflects an innate musical characteristic of the inner ear. This theory also supports the inherent origin of the octave enlargement phenomenon at the cochlear level.

The most recent theory is presented by de Cheveigné (2023). He suggests that human listeners are more

tolerant of positive than negative mistuning of the higher tone in an octave pair. Due to this asymmetry, a compressed octave is discriminated better than an enlarged octave. This theory is based on the hypothesis that a neural circuit tuned to cancel the lower tone also cancels the higher tone if the octave interval is in tune. If the cancellation is imperfect, it indicates the mistuning of the octave. In the neural circuits of the auditory system, excitatory–inhibitory interaction has been found from the cochlear nucleus to the dendritic fields of the inferior colliculus. Inhibition requires an extra synapse and thus may be delayed. A mismatch between the time constants of the two synapses causes an asymmetry in mismatch sensitivity. Therefore, there is a greater tolerance to positive than to negative mistuning. This theory also supports the inherent origin at the brainstem level.

A major problem with theories proposing that the neural basis of the octave enlargement phenomenon is located in the most peripheral auditory system is that the perceived pitch of a complex tone is not unequivocally represented in the cochlea but must instead be computed at the higher stages of the auditory system (Bendor et al., 2012). Pitch perception can be divided into two dimensions, height and chroma (Shepard, 1964). The pitch height increases as the fundamental frequency (f_0) of a tone increases, whereas the pitch chroma is repeated at octave intervals. Because of this two-dimensional approach, the pitch is often graphically depicted as a helix (Shepard, 1982). This pitch helix is illustrated in Figure 1. The fundamental frequency, f_0 , (height) is increased when ascending the spiral, and vertically aligned tones, one or more octaves apart from each other, share the same chroma. In practice, musical tones with

the same pitch class have identical pitch chroma (e.g., A1, A2, A3 and A4 have the same chroma or pitch class). Several studies have proposed that the neural organization of pitch is consistent with the helical model of distinct chroma and height representations in the auditory cortex (Briley et al., 2013; Moerel et al., 2015; Warren et al., 2003). The existence of pitch height-selective and pitch chroma-selective neurons in the auditory cortex is supported by observations of pitch-selective neuron populations (Gutschalk et al., 2002; Hall & Plack, 2009; Patterson et al., 2002; Puschmann et al., 2010) and neuron populations sensitive to one-octave intervals in the auditory cortex (Moerel et al., 2013, 2015). However, although the pitch height seems to be processed at an early (presumably preattentive) processing stage, there is some evidence that resolving chroma may require higher cognitive processing (Regev et al., 2019).

The present two studies aimed to examine whether event-related potentials (ERPs) elicited by tones in scalp-recorded electroencephalogram (EEG) would reveal a neural correlate of the octave enlargement phenomenon. In both studies, we used an adaptation paradigm in which an adapter tone is presented first, which causes a temporary reduction in the sensitivity of the population of neurons responsive to that tone. The adapter is then followed by a probe tone that shares some of the adapter's features (common harmonics or chroma). The similarity of neural processes evoked by the adapter and the probe can then be estimated by measuring the differences in neural responses to the probe. Stronger responses reflect weaker adaptation and thus different neural populations; conversely, weaker responses reflect stronger adaptation and more similar neural populations.

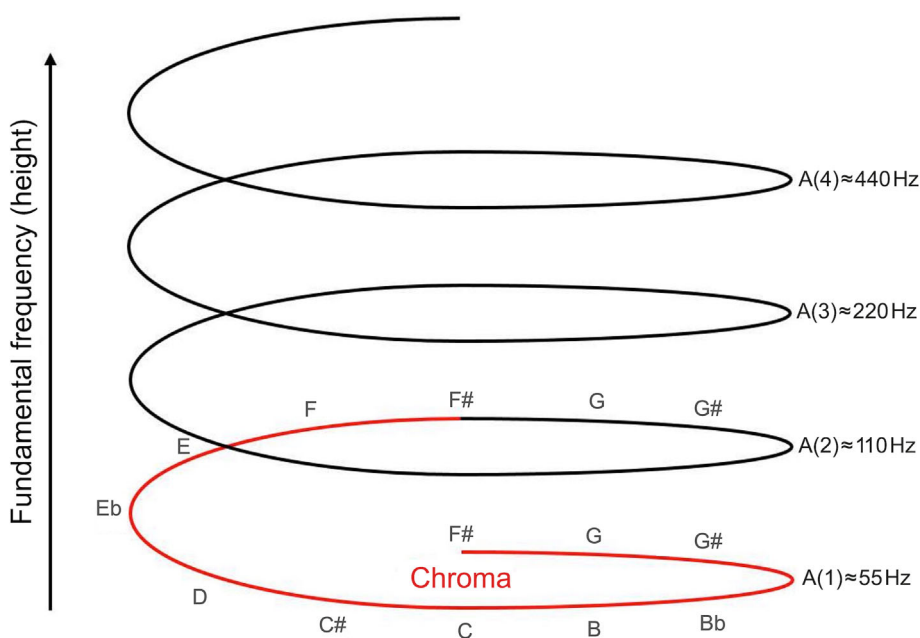


FIGURE 1 The pitch helix. The fundamental frequency, f_0 , (height) of a tone is increased when ascending the spiral, and vertically aligned tones, one or more octaves apart from each other (e.g., the tones A1, A2, A3 and A4 on the musical scale), share the same chroma. The vertically aligned tones with the same chroma are supposed to activate the same neural populations.

In both experiments, we measured adaptation of the negative-polarity N1 and positive-polarity P2 ERP responses elicited ca. 50–250 ms from the probe tone onset at and close to the vertex of the head. These ERP potentials are generated predominantly in the bilateral auditory cortices (e.g., Alho et al., 1998; Hämäläinen et al., 1993; Picton, 2010). Previous studies have shown that the N1 and P2 are smaller in amplitude the closer in frequency the adapting sinusoidal tones and the following probing sinusoidal tones are (Butler, 1968, 1972; Näätänen et al., 1988; Picton et al., 1978) and when adapting and probing tones are an octave apart (Briley et al., 2013). This adaptation could be due to an overlap between frequency- or pitch-specific auditory-cortex neuron populations activated by adapters and those activated by probes (Näätänen et al., 1988; Näätänen & Picton, 1987).

2 | THE FIRST EXPERIMENT

In the first experiment, we presented during EEG recording tone pairs of different sizes of octave-type intervals in a passive listening condition to the participants. To measure the adaptation effects on auditory ERP responses, we used a complex or sinusoidal tone as an adapter, followed by a complex tone probe. We hypothesized that (1) in addition to an increasing N1-P2 amplitude with an increasing frequency difference between the adaptor and probe tones, there would be a significant reduction in the N1-P2 peak-to-peak amplitude to probe tones around an octave (Briley et al., 2013), and (2) this adaptation effect would be the strongest for a slightly enlarged octave. Such results from a passive listening condition would suggest that the octave enlargement phenomenon originates from preattentive processing of tones in the auditory cortex. To our knowledge, the second hypothesis has not been tested before. To address a possible confounding of results by overlapping harmonics between adapters and probes, we used sinusoidal tones as adapters with no higher harmonics overlapping with those of the complex tone probes in half of the participants.

2.1 | Methods

2.1.1 | Participants

Eighteen volunteers, including two of the authors, participated in the experiment. Ten volunteers (mean age 40.3 years, SD 13.3, four females), including the two authors, participated in a condition with a complex tone adapter. Two of the volunteers were professional musicians; the remaining volunteers reported lesser degrees of

musical proficiency ranging from occasional to regular musical practice. The two professional musicians had self-reported absolute pitch.

Another eight volunteers and the two authors (mean age 37.1 years, SD 11.2, two females) participated in a condition with a sinusoidal tone adapter. Five of the participants were professional musicians; the remaining had some musical knowledge. Two of the professional musicians had self-reported absolute pitch.

The participants did not report any psychiatric, neurologic or auditory disorders. However, two participants self-reported mild tinnitus, a common ailment in professional musicians. All participants provided informed consent.

2.1.2 | Stimuli

The complex tone stimuli were gathered from the Vienna Symphony Library (Vienna Symphonic Library GmbH, Vienna, Austria), which encompasses a full range of professionally recorded instrument samples. In our experiments, we used string instrument samples in mezzo-forte dynamics. The applied tones were based on steady-state wavetable synthesis. The procedure comprised the following steps: a tone waveform was upsampled to a sample rate of 384 kHz/32 bits, after which a single period was isolated from the waveform. Single-period waveforms were imported to MATLAB 2021b (Mathworks Inc, Natick, MA, USA) to regulate the fundamental frequency accurately. The single-period waveform was repeated with an integer until a duration of 2 s was reached. The fundamental frequencies of the applied tones were modified to the correct tunings by resampling the repeated waveform by a ratio of integers that was iteratively found based on the number of fundamental periods in the repeated waveform and the ideal duration for the intended tuning frequency. The amplitudes of each resulting signal were equalized with C-weighting to the same value. C-weighted sound pressure level (SPL) was measured at the cup of the headphone with a professional-level decibel metre in a soundproof chamber. The intensity of the tones was 74-dB SPL. All tones were tuned according to A4 = 442-Hz reference pitch. For playback, the stimuli were downsampled to 48 kHz/16 bits. A custom MATLAB script was used to create sinusoidal tones for the sinusoidal tone adapter condition. The sampling and intensity of the sinusoidal tones were similar to those of the complex tones. For different experiments, the tones were truncated to appropriate lengths (400, 800, 1000 or 1500 ms). All tones had a 10-ms linear fade-in and fade-out to eliminate clicks at onset and offset. All tones were unfiltered.

2.1.3 | EEG acquisition

The present psychophysical and EEG data were collected in a shielded room in the Cognitive Brain Research Unit Laboratory, University of Helsinki, Helsinki, Finland.

The EEG data were recorded with 32 active scalp electrodes (BioSemi ActiveTwo System and ActiView605-Lores, BioSemi B.V., Amsterdam, the Netherlands) placed according to the standard 10–20 arrangement. Sampling rate was 2 kHz, and bit depth was 24. All EEG electrodes were online referenced to the Common Mode Sense (CMS) electrode at a standard location (PO1). Additional electrodes were placed on the left and right mastoids and the top of the nose.

In both conditions, the stimuli were presented in a Microsoft Windows (Microsoft Inc., Redmond, WA, USA) environment by Presentation software (Neurobehavioral Systems Inc., Berkeley, CA, USA), a Sound Blaster Audigy (Creative Technology Ltd., Singapore) sound card, a Yamaha AX-390 (Yamaha Corporation, Shizuoka, Japan) stereo amplifier and Sony MDR-7506 (Sony Corporation, Tokyo, Japan) headphones.

2.1.4 | Experimental design

The first EEG experiment had two adapter conditions and six or seven probe types (tone pairs). In the condition with a complex tone adapter, the stimuli were presented in six consecutive 15-min 36-s blocks. Each block contained 288 trials (48 for each probe type). In the condition with a sinusoidal tone adapter, the stimuli were presented in six consecutive 16-min 14-s blocks, where each block contained 336 trials (48 for each probe type). Each adapter-probe tone pair was presented 288 times (6 blocks \times 48 trials) per participant in both conditions.

In both conditions, the auditory trial (tone pair) consisted of a 1500-ms adapter tone (lower f_0) followed immediately by a 400-ms probe tone (higher f_0), consistent with the methods of Briley and colleagues (Briley et al., 2013). In both conditions, the inter-trial interval (ITI) was 1350 ms with a ± 10 -ms jitter.

The adapters (complex tone or sinusoidal tone) were randomly assigned a pitch of either G4 (393.8 Hz), A4 (442 Hz) or B4 (496.1 Hz). The probes consisted of complex tones with an f_0 of 600, 900, 1200, 1210, 1220, 1300 or 1800 cents higher than the adapter f_0 (one semitone = 100 cents). For example, for the A4 adapter, the probes were Eb5, G#5, A5, A5 +10 cents, A5 +20 cents, Bb5 and Eb6. In the following, the lowest and highest probes will be called *m600* and *p600*, respectively, as their frequencies were a mathematical octave minus 600 cents and a mathematical octave plus 600 cents.

Correspondingly, the second lowest and the second highest probes will be called *m100* and *p100*. The probes with an exact mathematical 2:1 octave apart from the adapter will be called *mat*, and the 10 and 20 cents higher probes will be called *p10* and *p20*, respectively. However, *p10* was used only in conditions with a sinusoidal tone adapter for a denser detection of a stretched octave. In a musical sense, the measured intervals were the augmented fourth (*m600*), the major seventh (*m100*), the exact mathematical octave (*mat*), the stretched octaves (*p10* and *p20*), the minor ninth (*p100*) and the augmented eleventh (*p600*). In the analysis phase, the data for all probe types were averaged across the G-, A- and B-based trials. Figure 2 presents the fundamental frequencies of all stimuli used in the first and second experiments (only A-based stimuli of the first experiment are illustrated; G- and B-based stimuli were similar except that they were one whole step lower or higher, respectively). Adapter-probe pairs (trials) were presented in pseudo-random order so that no more than two similar pairs were presented in succession. While presenting the auditory stimuli, participants viewed a silent subtitled movie in a shielded chamber.

EEG preprocessing

EEG data were preprocessed with EEGLAB 2021.1 toolbox (Delorme & Makeig, 2004) in the following steps: individual sets were appended to one file, downsampled to 250 Hz, band-pass filtered between .1 and 30 Hz (second-order Hamming-windowed sinc finite impulse response filter) and re-referenced to an average across all electrodes. Moreover, independent component analysis (ICA) weights were calculated for removing eye blinks and horizontal eye movements. ICA components were classified by the IClab plugin (Pion-Tonachini et al., 2019). Eye components with a 90% probability threshold were marked for removal. After removing the labelled components, the EEG data were epoched to -100 to 2900 ms from the onset of the first tone in the interval pair. The 0- μ V baseline was determined as the mean EEG amplitude over -100 to 0 ms (adapter onset at 0 ms). Epochs with EEG amplitudes exceeding ± 200 μ V were excluded before averaging the epochs to obtain ERPs separately for each participant and probe type.

ERP analysis

Amplitudes and latencies of the N1 responses (local negative ERP peak at 50–150 ms after probe onset) and the P2 response (local positive peak at 100–250 ms after probe onset) were measured at all electrodes separately for each participant and probe type. The Cz electrode was chosen for the analyses due to the largest mean N1-P2 peak-

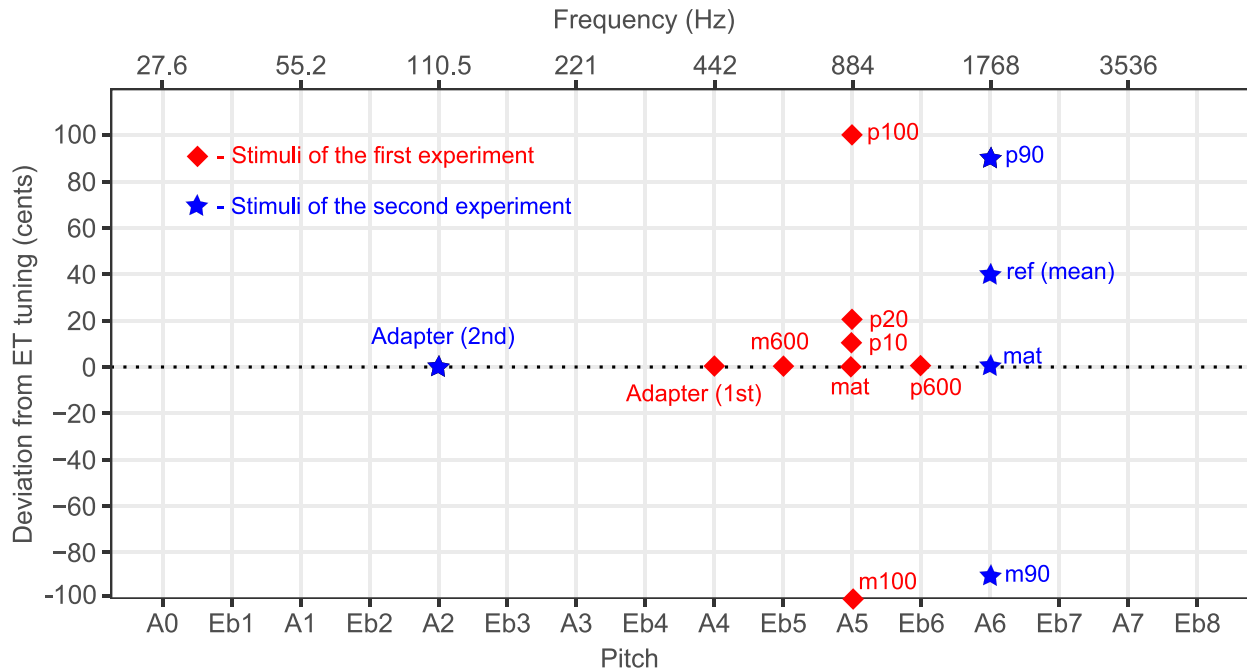


FIGURE 2 Adapter and probe tones used in tone pairs (trials) in the experimental setups. The x-axis corresponds to the equal-tempered tuning (ET). The y-axis represents a deviation from ET tuning. In the first experiment (red diamonds and labels), A4 is the adapter tone against which all probe tones are compared. The octave-type probes (around A5) are *m100*, *mat*, *p10*, *p20* and *p100*, where *m* and *p* refer to ‘minus’ and ‘plus’, respectively, and the following number to cents and *mat* is the mathematically exact one-octave interval (2:1). *m600* and *p600* correspond to Eb4 and Eb5 tones. In the second experiment (blue stars and labels), A2 is the adapter tone. *ref* (personal reference) denotes the mean of the individually preferred size of a subjective octave (18–70 cents higher than the *mat*, mean +39 cents). The other four octave-type probes (around A6) are *m90*, *mat* and *p90*. *mat* is the mathematically exact four-octave interval (16:1).

to-peak amplitude differences between probe types among the electrodes. Mixed analyses of variance (ANOVAs) were conducted using the MATLAB Statistics and Machine Learning Toolbox on both adaptation conditions (complex tone or sinusoidal tone) as between-participant factors and six probe types (*m600*, *m100*, *mat*, *p20*, *p100* and *p600*) common to both conditions as within-participants factors. The degrees of freedom were Greenhouse–Geisser corrected when needed, as indicated by the correction term ϵ reported together with the corrected *p* value. Multiple comparisons for estimated marginal means were calculated for significant factors. For the analyses, the 0- μ V baseline was re-determined as the mean ERP amplitude over -100 to 0 ms from probe onset.

2.2 | Results

In both the complex tone and sinusoidal tone adapter conditions, the adapter and probe elicited N1 and P2 ERP components at several electrodes. They were the largest at and close to the vertex of the head, as seen in the N1 and P2 scalp maps of Figure 3b. Figure 3a shows prominent

N1 and P2 responses for each probe type in ‘grand-average’ ERPs obtained by averaging across participants and conditions ERPs, recorded at the vertex electrode location (Cz). The mean latencies of N1 peaks, determined at Cz separately for each participant and probe type, were 93 ms (SD \pm 6 ms) in the complex tone adapter condition and 87 ms (SD \pm 7 ms) in the sinusoidal tone adapter conditions. The respective mean P2 peak latencies were 178 ms (SD \pm 20 ms) and 173 ms (SD \pm 22 ms). Mixed ANOVAs of N1 ($F_{1,18} = 4.10$, $p = 0.058$) and P2 ($F_{1,18} = .31$, $p = 0.586$) peak latencies did not show significant differences between adaptation conditions.

Figure 3c shows mean N1, P2 and N1-P2 peak-to-peak amplitudes at Cz for each probe type and adapter condition and amplitudes averaged across the adapter conditions. As expected, in both adapter conditions, there was an increase in the N1-P2 peak-to-peak amplitude as the pitch of the probing tone increased from *m600* to *m100* due to an increasing adapter-probe pitch separation. However, for the *mat* probe, at the mathematically exact octave separation, as well as the nearby higher probes *p20* and *p100*, the N1-P2 peak-to-peak amplitude decreased, implying adaptation. A mixed ANOVA with the adapter condition as a between-subjects

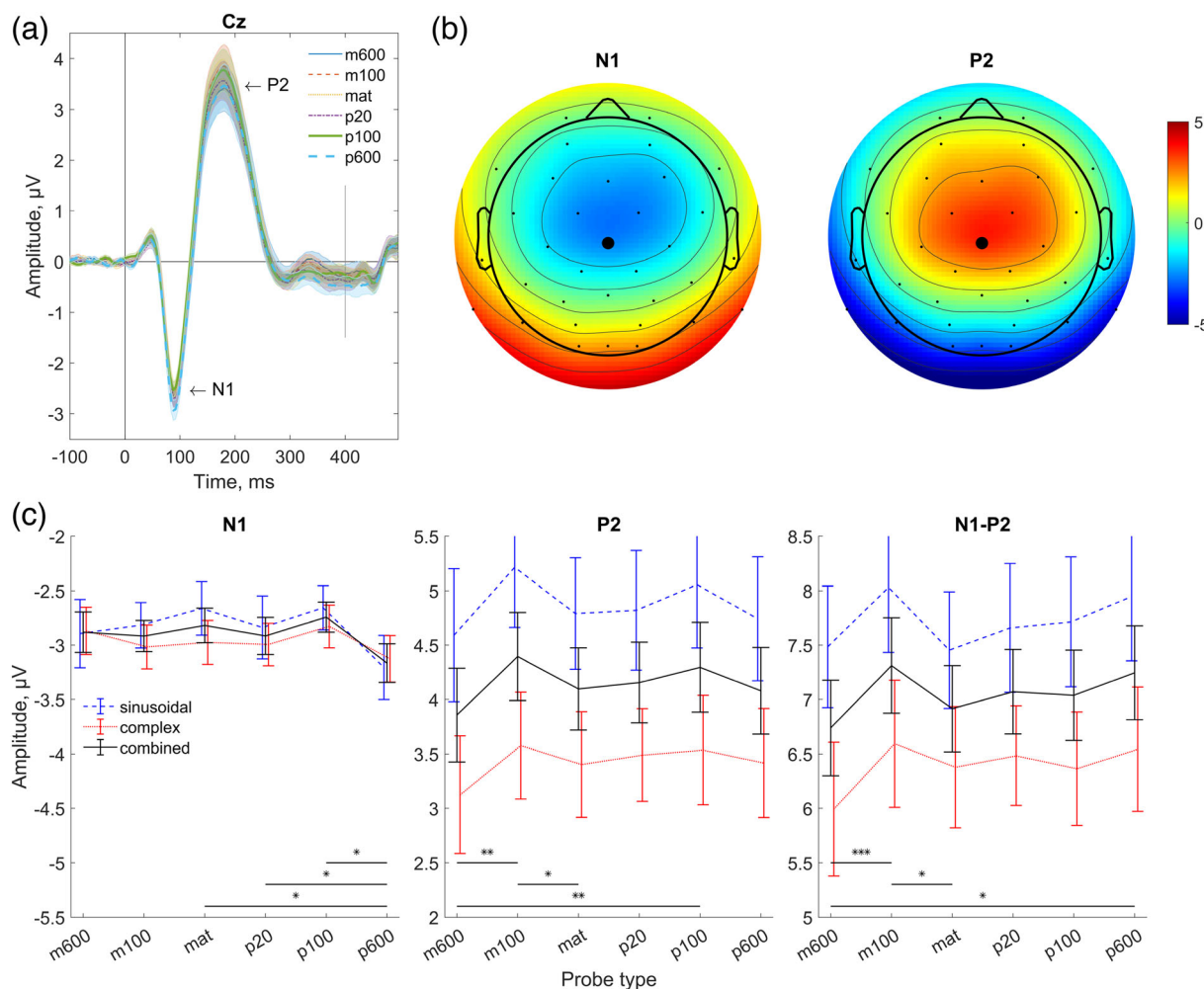


FIGURE 3 (a) Averaged grand-average event-related potentials (ERPs) across participants (combined data from the complex tone and sinusoidal tone adapter conditions) elicited by *m600*, *m100*, *mat*, *p20*, *p100* and *p600* probes at the Cz electrode. The grey vertical line at 400 ms denotes probe offset. (b) Scalp maps of N1 (84–96 ms) and P2 (168–184 ms) from grand-average ERPs (adapter conditions and probe types combined). The Cz electrode at the central vertex is denoted with a black dot. (c) Mean N1, P2 and N1-P2 peak-to-peak differences were measured from Cz using both sinusoidal tones ($n = 10$) and complex tones ($n = 10$) as adapters and a combined average of the two ($n = 20$). Error bars depict the standard errors of the mean. For combined adapter conditions, significant amplitude differences between probe types are marked (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

factor and probe type as a within-subject factor indicated significant differences between the probe types ($F_{5,90} = 5.70$, $p < 0.001$). However, the main effect of the adapter condition (complex tone vs. sinusoidal tone) ($F_{1,18} = 2.84$, $p = 0.109$) and the interaction between probe types and adapter conditions ($F_{5,90} = .85$, $p = 0.521$) were not significant. Post hoc multiple comparisons of the estimated marginal means for different probe types (data from both conditions combined) indicated significant differences between *m600* and *m100* ($p < 0.001$), *m100* and *mat* ($p = 0.010$) and *m600* and *p600* ($p = 0.016$).

Separate mixed ANOVAs for the N1 and P2 peak amplitudes at Cz showed significant differences between probe types both for the N1 ($F_{5,90} = 3.23$, $p = 0.039$,

$\epsilon = 0.498$) and P2 ($F_{5,90} = 5.61$, $p < 0.001$). There was no significant difference between the conditions or significant interaction between probe types and conditions. For post hoc multiple comparisons of N1, estimated marginal means for different probe types (data from both conditions combined) indicated significant differences between *p600* and *p100* ($p = 0.011$), *p600* and *p20* ($p = 0.016$) and *p600* and *mat* ($p = 0.011$). For post hoc multiple comparisons of P2, the estimated marginal means for different probe types (data from both conditions combined) indicated significant differences between *m600* and *m100* ($p = 0.007$), *m100* and *mat* ($p = 0.020$) and *m600* and *p100* ($p = 0.002$).

The sinusoidal tone adapter condition included an additional *p10* probe. Post hoc multiple comparisons did

not show significant N1, P2 or N1-P2 peak differences between *p10* and the other probes. As we analysed data from the second experiment with multivariate pattern analysis (MVPA) methods, we applied the same MVPA methods to the data of the first experiment afterward. However, we found no significant differences between different conditions (adapter types) or octaves (probe types) in spatial or temporal decoding (for details of MVPA, see Section 2.1 for the second experiment).

2.3 | Discussion of the first experiment

We found a significant reduction in the N1-P2 amplitude for *mat* probes, exactly one octave apart from the adapter, compared with one semitone lower *m100* probes. This suggests a greater neural overlap between tones sharing the same chroma than tones one semitone lower. Because the participants were instructed to a silent movie and to ignore the tones, this result may be regarded as suggesting the preattentive processing of chroma. However, some participants may still have paid attention to the tones and perhaps even actively compared the adapter and probe tones, as many of them were professional musicians. Moreover, no significant differences in the N1-P2 amplitudes were found between *mat*, *p10*, *p20* or even *p100* probe types. Thus, there was no evidence for differences between neural processing of mathematically exact and enlarged octaves.

The adaptation effect on the N1-P2 amplitudes to probes one octave higher than the adapters is congruent with previous results (Briley et al., 2013). However, they observed this effect only when both the adapter and probe were complex tones (sharing common harmonics) but not when both were sinusoidal tones (no common harmonics). Therefore, their result may indicate that overlapping harmonics explain the observed adaptation effect (cf. Regev et al., 2019). However, we used both complex tones and sinusoidal tones as adapters and complex tones as probes. Especially in the sinusoidal adapter condition, the strongest adaptation effect on the N1-P2 amplitudes was observed for the mathematically exact octave (see Figure 3c). Thus, sinusoidal tones created an adaptation effect without spectral overlap of the adapter and probe harmonics.

3 | THE SECOND EXPERIMENT

Since the first EEG experiment did not show electrophysiological evidence for the octave enlargement phenomenon, we decided to improve our experimental design in the second experiment in the following ways: (1) changing

from passive listening to an active interval classification task, (2) expanding the evaluated interval from one octave to four octaves (since the larger the interval, the larger the enlargement between the tones) (Jaatinen et al., 2019) and (3) using a pre-measured individual size of a subjectively preferred four-octave interval (personal reference) instead of fixed size. In addition, all participants recruited for the second experiment were professional musicians with extensive experience evaluating musical intervals.

We hypothesized that since there are significant between-participant differences in the size of an individually preferred octave (Jaatinen et al., 2019), the individually preferred enlarged octave rather than a fixed enlarged octave may better reveal the neural basis of the octave enlargement. Moreover, if the N1-P2 amplitudes during active listening, unlike during passive listening of the first experiment, would indicate the octave enlargement phenomenon, this would suggest that the phenomenon is based on attention-dependent brain processes rather than on early preattentive auditory processing. However, it is possible that resolving whether an octave-type interval is the individually preferred one only occurs at the late stages of auditory processing beyond the N1 and P2 latencies. Therefore, in addition to univariate ERP analyses, we used decoding analysis (an MVPA method; Bode et al., 2019) to investigate differences between ERPs to different probe types over 990 ms after probe onset. In the present case, the traditional univariate ERP analysis compares only the mean amplitudes between the probe types. Because MVPA considers both the mean amplitudes and amplitude variability, it is a more sensitive data analysis method than univariate analysis. Therefore, MVPA may reveal additional neural processes underlying the octave enlargement phenomenon not distinguishable with univariate ERP analysis. In the present MVPA, we used support vector machines (SVMs) in the pairwise classification of ERP amplitudes for different probe types. In separate decoding analyses, the amplitude patterns were classified across the EEG electrodes (spatial decoding, electrodes as features) or time (temporal decoding, time points as features).

3.1 | Methods

3.1.1 | Participants

Because the second experiment required outstanding pitch perception accuracy, it was necessary to have only professional musicians as participants. Twenty volunteer professional musicians (19 naïve and one of the authors; three participants had also participated in the first experiment) attended the experiments (mean age 46.7 years, SD

7.3, nine females). Two of these volunteers had self-reported absolute pitch. All participants were of the highest professional level and are employed in symphony or jazz orchestras in the Greater Helsinki area. No participants reported any psychiatric, neurologic or auditory disorders.

3.1.2 | Stimuli

The tones used in the second experiment were similar in spectral structure to the complex tones in the first experiment. The intensity of tones was 74-dB SPL (SPLs equalized and measured as in the first experiment). For the EEG experiment, A6 (1768 Hz) tones were tuned to 91 different variants from -90 cents to $+90$ cents, with steps of two cents. The length of the lower (adapter) and upper (probe) stimulus tones was 800 ms.

According to Jaatinen et al. (2019), the preferred size of a subjective four-octave interval (A2–A6) with similar stimuli as the current ones (string instrument samples in mezzo-forte dynamics) was, on average, about 36 cents wider than the mathematically exact (16:1) four-octave interval. However, the between-participant variation was quite large. Because of this, we determined the individually preferred size for the subjective four-octave interval for each participant. To this end, the participants adjusted the preferred size of the four-octave interval as follows: the lower A2 tone and the upper A6 tone were presented successively by MaxMSP 8 software (Cycling'74, San Francisco, CA, USA) using the headphone amplifier of a Macbook Pro 2019 laptop computer (Apple Inc, Cupertino, CA, USA) and AKG K550 headphones (AKG Acoustics, Los Angeles, CA, USA). The length of each tone was 1 s, as in Jaatinen et al. The fundamental frequency of the lower tone was constant (A2 = 110.5 Hz), and the fundamental frequency of the upper tone (A6 \pm 0 cents = 1768 Hz) was adjustable by the participant to one-cent resolution by the pitch-bend algorithm in MaxMSP 8. This was controlled by a ShuttleXpress (Contour Design A/S, Denmark) rotary controller. The tone pair was repeated until the participant had determined the preferred four-octave interval and pressed the next button. There were no upper or lower limits in adjusting the fundamental frequency of the upper tone. The initial tunings of the upper tone of the presented octave pairs were in order ± 0 , $+90$, $+20$, $+70$, $+30$, $+60$, $+40$ and $+50$ cents. We calculated each participant's average tuning of eight adjusted pairs, determining the individual personal reference (subjective four-octave interval) for the second EEG experiment.

All participants adjusted the subjective four-octave interval (*ref*) larger than the mathematically exact 16:1

four-octave interval (*mat*). The mean was $+39.0$ cents (standard deviation 16.9 cents). There were significant differences in octave preference between participants. The lowest determined *ref* was $+16$ cents, and the highest was $+70$ cents. However, because the *ref* was limited to the range $+30$ to $+60$ cents in the EEG experiment, the mean and standard deviation of the *ref* used were $+40.0$ cents and 14.0 cents, respectively.

3.1.3 | EEG acquisition

EEG was measured with 65 active electrodes (64 electrode locations of the 10–20 system and a ground electrode) using the actiCAP snap electrode cap, actiCHamp Plus amplifier and BrainVision Recorder software (Brain Products GmbH, Gilching, Germany). The acquisition parameters were f_s 1000 Hz and bit depth 24. The FCz electrode served as a reference in the recording phase, and the ground electrode was located at FPz. The stimulus presentation equipment and software were the same as in the first experiment.

3.1.4 | Experimental design

The second EEG experiment had four different probe types (tone pairs; see Figure 1). The participants were instructed to classify the probe types (successively presented alternating tone pairs) into three categories ('low', 'good' and 'high'). The categorization was made by pressing one of the three buttons on a Cedrus RB-740 response pad (Cedrus Corporation, San Pedro, CA, USA).

The tone pairs were A2–A6–90 cents (labelled as *m90*), A2–A6 ± 0 cents (*mat*), A2–A6 personal reference (*ref*; $+20$ to $+70$ cents depending on the participant) and A2–A6 $+90$ cents (*p90*) (see Figure 2). As in the first experiment, the constant lower tone (A2) is called an adapter, and the higher tone (A6) is called a probe. The structure of a trial was as follows: 1000-ms silence, 800-ms A2 tone, 10-ms silence and 800-ms A6 tone variant (total length 2610 ms), followed by a response window. The response time was not limited; the participant pressed a button to start the next trial. A 1-s silence part at the beginning of the trial was intended to eliminate movement-related muscular and brain activity (associated with a button press) from probe-related EEG data.

Data were collected in three blocks of the experiment. Each block included 200 tone pairs, with 50 of each tone pair type (*m90*, *mat*, *ref* and *p90*) presented in a semi-random order. The total number of tone pairs was 600, with 150 for each pair type. The duration of each block was 10–15 min, depending on the participant.

Before the experiment, the participants trained in the classification task (recognition of different octave sizes and appropriate use of response buttons) under the guidance of one of the experimenters.

In the early phase of the experiment (the first five participants), we noticed that if a personal reference was too close to the adjacent alternatives (± 0 or $+90$ cents), the classification task became even more demanding. For compensation, we limited the personal reference for the remaining participants between $+30$ and $+60$ cents even if the average result in the listening experiment was $< +30$ or $> +60$ cents. The first five participants performed the EEG test without this limitation. However, in behavioural data analysis, we did not find any significant difference between their results and the results of the 15 other participants.

3.1.5 | Analysis methods

EEG preprocessing

EEG preprocessing was similar to the first experiment, except the data were epoched to 0–2800 ms from the beginning of the trial (1000-ms silence, 800-ms adapter, 10-ms silence and 800-ms probe). The baseline was corrected as the mean EEG amplitude over 900–1000 ms

(the last 100 ms of the silence period or 100 ms before the adapter).

Behavioural analysis

The percentages of classification into three categories for each participant were calculated for four probe types (tone pairs). From these, means and standard deviations were calculated for the whole set and are illustrated in Figure 4.

ERP analysis

ERP data were processed with MATLAB R2021b using the EEGLAB toolbox (Delorme & Makeig, 2004). N1-P2 peak-to-peak amplitude differences at Cz were calculated, and repeated measures ANOVAs were conducted in correspondence with the first experiment, excluding a between-subjects factor. Additionally, ANOVAs were conducted on all probe types, and F values were obtained for all channels and ERP time points using EEGLAB functions. Post hoc difference t values between all six tone combinations ($m90\text{-mat}$, $m90\text{-ref}$, $m90\text{-p90}$, $mat\text{-ref}$, $mat\text{-p90}$, $p90\text{-ref}$) were calculated for all channels and ERP time points. ERP ANOVA and post hoc t test significance thresholds were corrected with the Fieldtrip-lite plugin (Oostenveld et al., 2011) cluster correction using the triangulation method that calculates electrode neighbours based on a 2D projection of the electrode positions. The alpha level

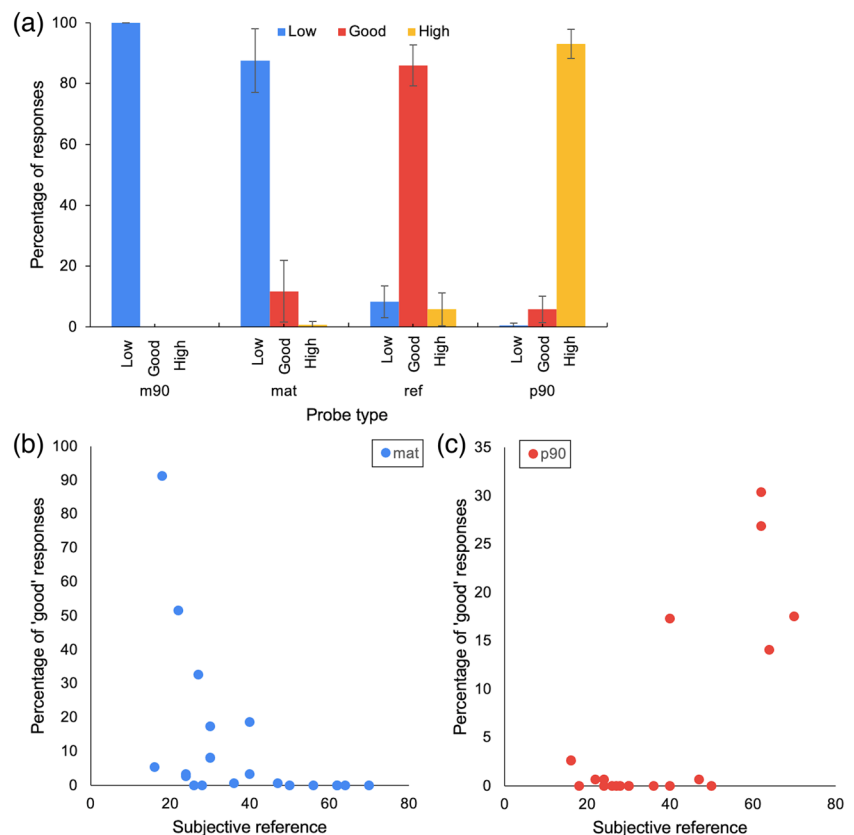


FIGURE 4 Behavioural results of the electroencephalography (EEG) experiment. (a) All participants systematically classified $m90$ as 'low', whereas mat was mainly classified as 'low' and secondarily as 'good', ref mainly as 'good' and secondarily as 'low' or 'high' and $p90$ mainly as 'high' and secondarily as 'good'. The subjective reference affected how likely (b) mat and (c) $p90$ were classified as 'good'. The error bars indicate the 95% confidence intervals.

for correction was set to 0.05. In the analysis phase, the 0- μ V baseline was re-determined as the mean ERP amplitude over -100 to 0 ms from probe onset.

Decoding analysis (MVPA)

Pairwise decoding analysis of all four probe types was conducted spatially and temporally. Each channel was a feature in spatial coding, and the decoding was repeated for an average of five time points (20-ms window) in 10-ms steps. The temporal decoding analysis was conducted separately for each channel, and 64 time points were used as features. Because the data were sampled at 250 Hz, the window for temporal decoding was 256 ms in duration. All decodings were conducted with the Decision Decoding Toolbox (Bode et al., 2019). The decoding accuracy above the chance level was statistically tested with a one-sided t test, separately for each time point. The obtained p values were corrected for multiple comparisons with false detection rate (FDR) correction. To quantify the differences in the shape of the decoding accuracies across time, a cumulative normal distribution (spatial decoding; three free parameters) or sum of two normal distributions (temporal decoding; six free parameters) was next fitted to the decoding accuracy functions with MATLAB's `fmincon` function. The difference in the parameters of the fitted functions was tested with paired t tests (without correction for multiple comparisons).

3.2 | Results

3.2.1 | Behavioural results

Figure 4a indicates that the *ref* was mainly rated as 'good'. The *m90* was unequivocally classified as 'low'. The *mat* was usually classified as 'low' and as 'good'. Interestingly, the *p90* was not only classified as 'high' but also as 'good' to some extent, which is not surprising as some participants had a *ref* quite close to *p90*.

The magnitude of the per-participant *ref* influenced the classification in the EEG experiment. Some participants with a *ref* close to the lower limit were more likely to rate their *mat* as 'good' (Figure 4b; Spearman's $\rho = 0.69$). If the *ref* was close to the upper limit, this was reflected in some participants rating *p90* as 'good' more often than others (Figure 4c; Spearman's $\rho = 0.51$).

3.2.2 | ERP results

Similarly to Figure 3a–c showing the ERP results from the first experiment, Figure 5a presents the Cz ERPs to probes, Figure 5b the N1 and P2 scalp distributions and

Figure 5c the mean N1, P2 and N1-P2 amplitudes from the second experiment.

As seen in Figure 5a, the probes (upper tones) elicited an N1 response followed by a P2 response. However, after the P2 peak, the ERP amplitude returned to the 0- μ V baseline much later (around 400 ms) than in the first experiment (see Figure 3a), presumably due to an overlapping P3 (P300) response, typically associated with the target processing (Alho et al., 1998). However, as seen in Figure 5b, the N1 and P2 scalp distributions were quite similar to those in the first experiment (cf. Figure 3b).

As seen in Figure 5c, in concordance with the results of the first experiment (Figure 3c), N1-P2 amplitudes at Cz were smaller for the *mat* probe and for the higher *ref* and *p90* than for *m90*, implying adaptation for the mathematical octave and for slightly stretched octaves. A repeated-measures ANOVA showed a significant difference between the probe types ($F_{3,57} = 16.21$, $p < 0.001$). Post hoc pairwise comparisons of the estimated marginal means for different probes indicated significant differences between *m90* and *mat* ($p < 0.001$), *m90* and *ref* ($p < 0.001$) and *m90* and *p90* ($p < 0.001$). Separate repeated-measures ANOVA for N1 and P2 amplitudes at Cz showed significant differences between probes for P2 ($F_{3,57} = 7.46$, $p < 0.001$) but not for N1 ($F_{3,57} = 2.67$, $p = 0.056$). For post hoc pairwise comparisons of P2, estimated for marginal means for different probes indicated significant differences between *m90* and *mat* ($p = 0.003$) and *m90* and *p90* ($p = 0.009$).

ANOVAs were performed to determine whether significant differences between the probe types extended to other electrodes. Figure 6 shows a heatmap of the ANOVAs at each ERP time point. Electrodes with non-significant values are blacked out. Significant differences were widely dispersed beginning after 100 ms and continuing after tone offset. The maximum F value was observed at FC3 at 190 ms after probe onset.

According to post hoc t tests comparing more dissonant *m90* and *p90* probes with more consonant *mat* and *ref* probes, significant ($p(\text{cluster-corrected}) < 0.05$) differences between probe types spanned multiple electrodes and time points continuing in some cases even to the end of the analysed period. The differences between *m90* and *mat*, defined as a cluster of at least four significant electrodes, began at 134 ms and peaked at the electrode P1 at 746 ms ($t_{19} = -7.81$). The respective significant differences between *m90* and *ref* also began at 134 ms and peaked at C3 at 194 ms ($t_{19} = 7.76$). The respective onset for the difference between *p90* and *mat* was at 226 ms and peak at FC2 at 414 ms ($t_{19} = 6.43$), whereas for the difference between *p90* and *ref*, the onset was at 182 ms and peak at P8 at 314 ms ($t_{19} = -5.71$). The differences between *m90* and *p90* probes peaked at C3 at 198 ms

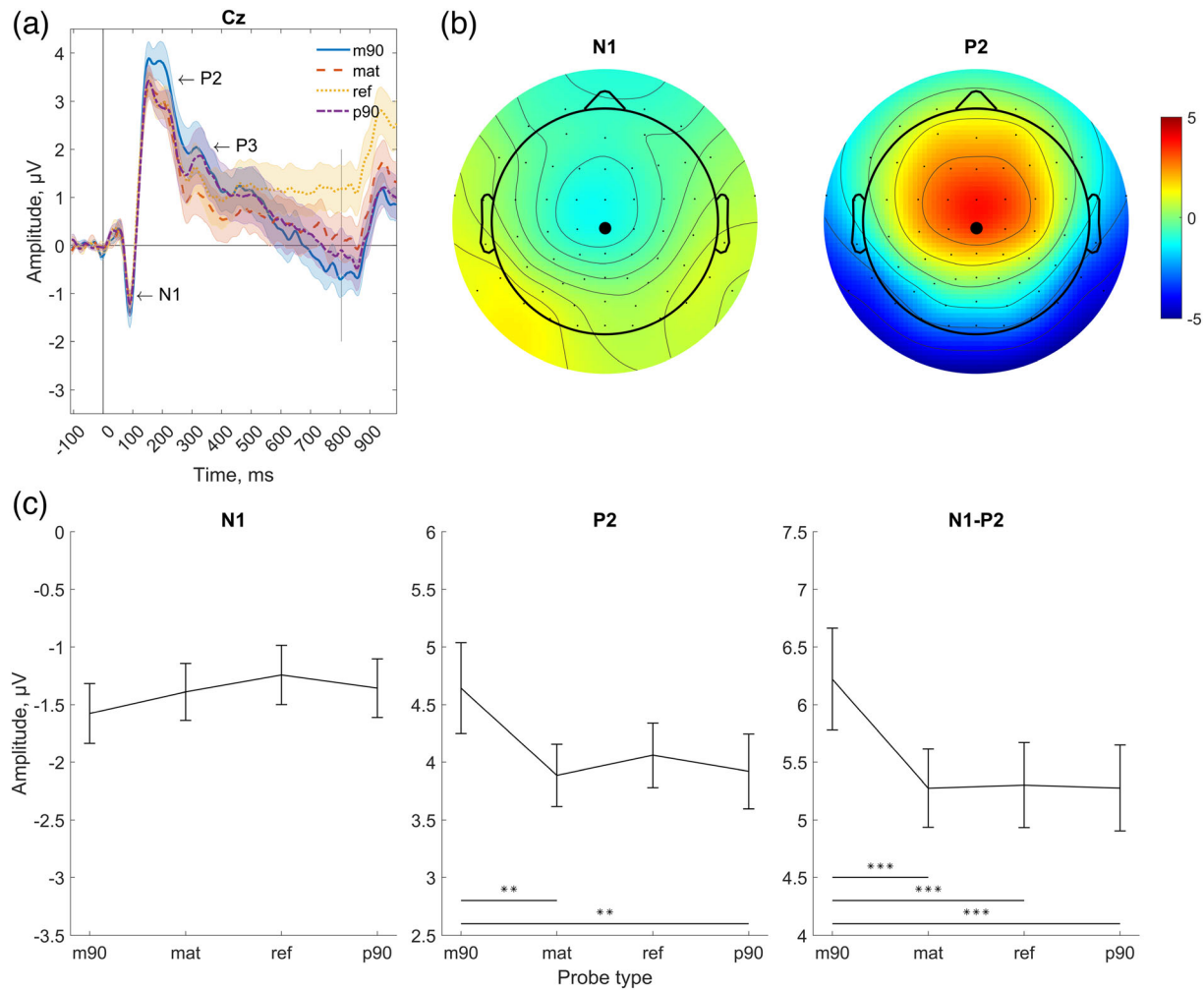


FIGURE 5 (a) Grand-averaged event-related potentials (ERPs) to *m90*, *mat*, *ref* and *p90* probes at Cz electrode. Shaded areas indicate standard errors of the mean. The grey vertical line at 800 ms denotes probe offset. (b) Scalp maps of N1 (82–94 ms) and P2 (142–158 ms) from grand-average ERPs. The Cz electrode in the central vertex is indicated with a black dot. (c) Mean N1, P2 and N1–P2 peak-to-peak differences individually measured from Cz. Error bars indicate standard errors of the mean. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

($t_{19} = 5.91$). These were not significant after cluster correction.

Figure 7 shows that when comparing *ref* and *mat* probes, significant cluster-corrected differences were found at only four fronto-central/right-frontal electrodes (FCz, F2, AF4 and AF8). The cluster of electrodes showing significant differences appeared at 578 ms with its maximum t value at FCz at 862 ms ($t_{19} = -7.36$) and continued to the end of the analysis period. As seen in Figure 5a, this was due to a more positive ERP to *ref* than to *mat*, this difference commencing around 500 ms from probe onset.

3.2.3 | Decoding results

The ERP analyses showed widespread changes across multiple channels and multiple time points. We next

conducted decoding analysis (MVPA), which is more sensitive than ERPs and utilizes either information from all channels (spatial decoding) or across several time points (temporal decoding).

Figure 8 shows the results of spatial decoding analyses. For all pairwise decoding analyses, significantly above-chance classification accuracy was found starting around 200 ms after probe onset. The classification remained significantly above chance until the end of the epoch. However, there was some variation across the probe types on the onset, slope and maximum accuracy of classification. Overall, the fastest classification accuracy was found for *mat* versus *m90* and *ref* versus *m90*; the decoding of *mat* versus *p90* and *ref* versus *m90* reached the highest accuracy (>15% above chance). The poorest and slowest decoding accuracy was found for *ref* versus *mat*.

To quantify all the differences between the pairwise classifications, cumulative Gaussian functions were fitted

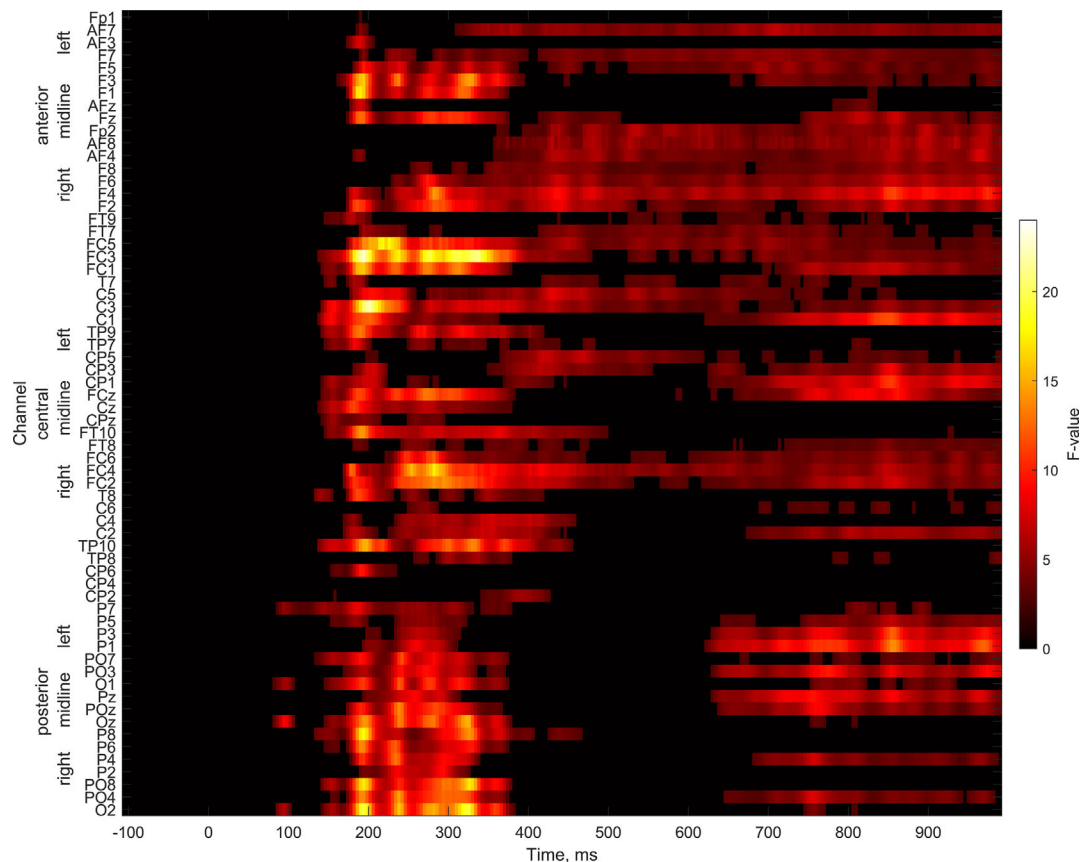


FIGURE 6 Heatmap of analysis of variance (ANOVA) across all event-related potential (ERP) time points and channels. Electrode channels are sorted into anterior, central and posterior groups and into left, midline and right channels according to the 10–10 electrode placement system. Non-significant ($p(\text{cluster-corrected}) > 0.05$) F values are shown in black.

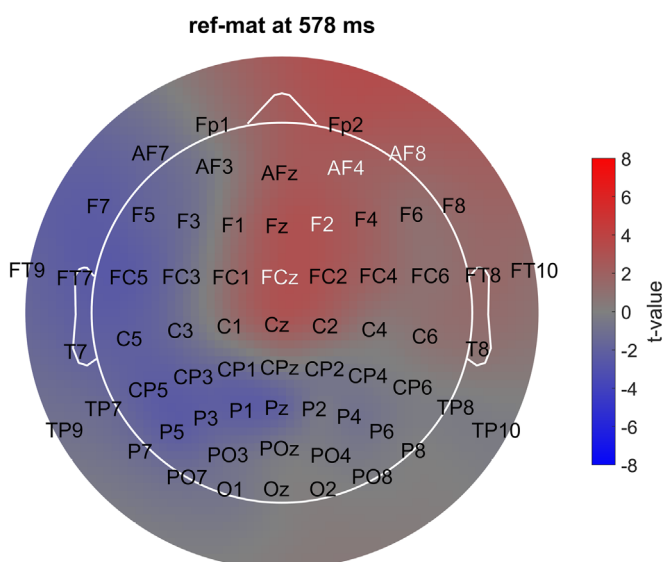


FIGURE 7 Scalp map of t values at 578 ms that is at the onset of a cluster of four electrodes (FCz, F2, AF4 and AF8 written in white) showing a significant difference ($p(\text{cluster-corrected}) < 0.05$) between the probes *ref* and *mat*.

to the individual decoding results, and the parameters of the fitted functions were compared. When classification accuracies were compared to *mat* (Figure 8a), the classification accuracy for *mat* versus *p90* was significantly higher than for *mat* versus *m90* ($t_{19} = -2.52$, $p = 0.021$) or *mat* versus *ref* ($t_{19} = -3.02$, $p = 0.007$). The classification accuracy for *mat* versus *p90* was also higher than for *m90* versus *p90* ($t_{19} = 2.35$, $p = 0.030$). Moreover, *mat* versus *m90* was classified significantly earlier than *mat* versus *ref* ($t_{17} = -2.26$, $p = 0.037$). The tested latency was estimated at the time point when decoding reached (arbitrarily selected) 4% accuracy. The slope of decoding accuracy was also shallower for *mat* versus *ref* than for *mat* versus *m90* ($t_{19} = 2.91$, $p = 0.009$) or *m90* versus *p90* ($t_{19} = -2.51$, $p = 0.021$).

When classification accuracies were compared to *ref* (Figure 8b), the classification accuracy was significantly higher for *ref* versus *m90* than for *ref* versus *mat* ($t_{19} = 2.81$, $p = 0.011$). Furthermore, the classification accuracy for *ref* versus *mat* was slower than for *ref* versus *m90* ($t_{12} = -3.70$, $p = 0.003$), *ref* versus *p90* ($t_{11} = 2.31$,

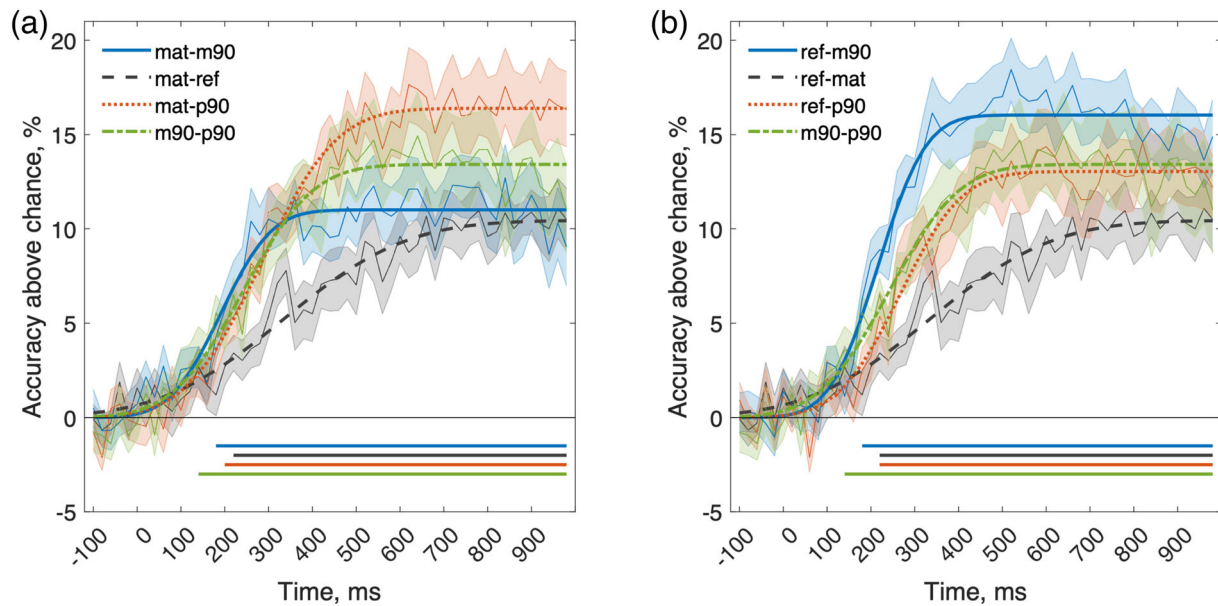


FIGURE 8 Spatial decoding results. Jagged thin lines indicate means of classification accuracy; smooth, thicker lines indicate cumulative normal distributions fitted to the data. Straight solid lines indicate statistical significance (one-tailed t test, $p(\text{FDR}) < 0.05$) above chance. (a) Pairwise decodings when classification accuracies were compared with the *mat*. (b) Pairwise decodings when classification accuracies were compared to *ref*. Note that *m90-p90* and *mat-ref/ref-mat* are identical in both plots.

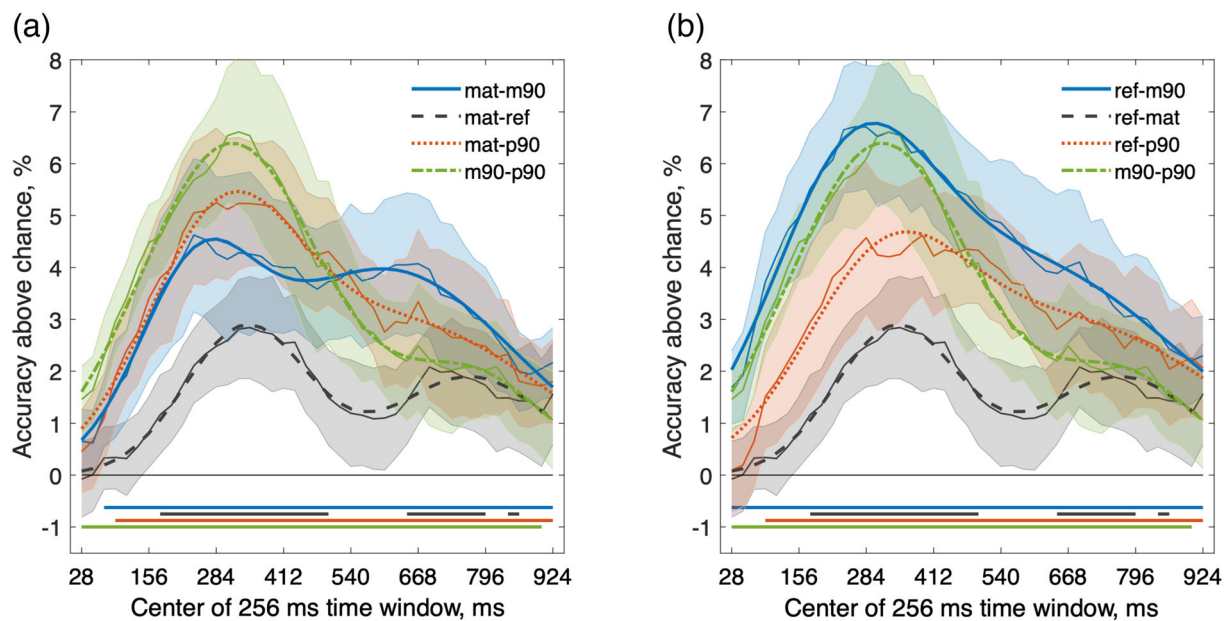


FIGURE 9 Temporal decoding results. Jagged thin lines indicate means of classification accuracy; smooth lines indicate sums of two normal distributions fitted to the data. The straight solid lines indicate statistical significance (one-tailed t test, $p(\text{FDR}) < 0.05$) above chance. (a) Pairwise decodings when classification accuracies were compared to the *mat*. (b) Pairwise decodings when classification accuracies were compared to *ref*. Note that *m90-p90* and *mat-ref/ref-mat* are identical in both plots.

$p = 0.041$) or *m90* versus *p90* ($t_7 = 2.69$, $p = 0.031$). Classification of *ref* versus *m90* was also faster than *ref* versus *p90* ($t_{15} = -2.17$, $p = 0.047$) or *m90* versus *p90* ($t_{11} = -2.77$, $p = 0.018$). Moreover, the slope of the decoding accuracy was shallower for *ref* versus *mat* than

for *ref* versus *m90* ($t_{19} = 2.79$, $p = 0.012$), *ref* vs. *p90* ($t_{19} = -2.31$, $p = 0.033$) or *m90* versus *p90* ($t_{19} = -2.51$, $p = 0.021$).

Finally, we investigated the temporal time course for the successful decoding of tone pairs. Figure 9 shows the

results of the temporal decoding analyses averaged across all channels. The peak classification accuracy was around 300 ms (time window from 150 to 450 ms). Again, there was some variation in the maximum classification accuracy and the shape of the functions. We quantified these differences by fitting a sum of two normal distributions to each participant's data (with six free parameters, coefficient, mean and standard deviation) and compared the fitted parameters of the functions. The sum of two normal distributions was used because the results clearly indicated two temporally separate peaks/processes. When classification accuracies were compared with the *mat* (Figure 9a), significantly higher classification accuracy was found for *mat* versus *p90* ($t_{19} = 3.52$, $p = 0.002$), *mat* versus *m90* ($t_{19} = -3.47$, $p = 0.003$) and *m90* versus *p90* ($t_{19} = -5.91$, $p < 0.001$) than for *mat* versus *ref*. The classification of *mat* versus *ref* was also significantly slower than all other classifications. This reached statistical significance much later (170 ms) than other classifications (30–90 ms). When classification accuracies were compared with *ref* (Figure 9b), the peak accuracy and the latency of all probe types differed significantly from each other, except between *ref* versus *m90* and *p90* versus *m90*.

3.3 | Discussion of the second experiment

Although the results of the second experiment suggest that the phenomenon of enlarged octaves is manifested in this experimental setting, there is marked inter-participant variability in performance and brain activity. In particular, the behavioural results from the active listening task of the second experiment (see Figure 4a) provide strong evidence for the existence of the phenomenon of enlarged octaves.

Similar to the first experiment with passive listening, applying active listening in the second experiment revealed significantly lower N1-P2 peak-to-peak amplitudes for the mathematical or enlarged octaves than for the compressed octaves (see Figures 3c and 5c for the results from the first and second experiment, respectively). This finding may support the adaptation paradigm with the helical model for octave-type intervals. Despite the active classification task where the participants were able to discriminate the *mat* and *ref* tone pairs from each other (see Figure 4a), there was no significant difference between mathematically exact (*mat*) and enlarged (*ref*) octaves in N1-P2 peak-to-peak measurements at Cz (Figure 5c), similar to the corresponding results of the first experiment using passive listening (Figure 3c). ERP analysis of all electrodes showed the

largest and earliest differences between the most dissonant *m90* and the most consonant *ref* and *mat* probes. Later and less prominent differences were also found for the high-pitched sounding *p90*, *mat* and *ref* probes. However, differences between ERPs to *ref* and *mat* were found only very late, over 500 ms after probe onset (all significant channels present for the first time at 578 ms; Figures 5a and 7), although these two probes were behaviourally well-distinguished (Figure 4a).

According to spatial decoding analysis, the earliest significant classification was found for the pairs *ref* versus *m90* and *mat* versus *p90*. The latest classification was for the pair *ref* versus *mat*, although this was still much earlier (around 200 ms) than in ERP analyses. In the temporal decoding analysis, the peak classification accuracy generally was around 300 ms. This finding may be related to the results from a backward masking paradigm, suggesting that perceptually obvious pitch contrasts between two tones are resolved by the auditory system about 250 ms after tone onset (Massaro, 1975; Massaro & Idson, 1977). The lack of clear differences between N1-P2 amplitudes to mathematically exact and slightly enlarged octave-type intervals in both the first and second experiments may also be due to unfinished pitch processing at the N1 and P2 latencies. According to the present ERP and decoding results, resolving the chroma and subjectively preferred enlarged octave occurs only later, continues for several hundred milliseconds and presumably requires attention and perhaps also working memory (see also (Regev et al., 2019).

The spatial decoding analysis found the earliest classification accuracy for the most distinguishable consonance-dissonance pairs (*mat* versus *m90* and *ref* versus *m90*). The highest decoding accuracy was between *mat* versus *p90* and *ref* versus *m90*. The lowest and slowest accuracy was found for *ref* versus *mat* (Figure 8). These results show that EEG contains differential information for the different probe types and that this information can be found starting at around 200 ms after probe onset.

As mentioned, the peak classification accuracy in the temporal decoding analysis was around 300 ms. The *mat* versus *ref* classification was significantly slower in time and lower in accuracy than any other classification. In temporal decoding, there was a significant difference in decoding accuracy between *mat* versus *m90* and *ref* versus *m90*. While the theoretically most distinguishable consonance-dissonance pair (*ref* vs. *m90*) showed the highest accuracy, the pair *mat* versus *m90* had a significantly lower peak. In addition, there was a second, later peak in classification around 700–800 ms, most clearly visible in pairs *mat* versus *m90* and *mat* versus *ref* (Figure 9). This may indicate the involvement of later cognitive processes.

Participants were instructed to classify the subjective octave *ref* of their choice on the listening experiment as 'good', the *m90* and *mat* as 'low' and the *p90* as 'high'. Therefore, the present classification data cannot be considered spontaneous classification. Moreover, the classification results may have been influenced by the guided training of the classification task before the second EEG experiment. Nevertheless, the participants adjusted the subjectively best four-octave interval (*ref*) without guidance.

4 | GENERAL DISCUSSION

The invasive methods applied in animal studies (McKinney & Delgutte, 1999; Ohgushi, 1983) for investigating the neural origin of the octave enlargement phenomenon at the auditory nerve level are not suitable for humans. Furthermore, the existence of the stretched-tuned tonal memory matrix in semantic memory (Terhardt, 1970, 1974) is impossible to verify by any known brain imaging technique. However, our results supported de Cheveigné's (2023) theory; according to the present results, the distinction between compressed and enlarged octaves is resolved earlier than the distinction between mathematically exact and enlarged octaves (Figures 8b and 9b). Behaviourally, none of the participants preferred compressed octaves (Figure 4a).

Previous EEG studies investigating the early processing of pitch chroma have resulted in seemingly conflicting results. Briley et al. (2013) used iterated rippled noise, known to uniformly activate the tonotopically arranged cochlear nerve neurons (Patterson et al., 1996). They observed N1-P2 adaptation effects, suggesting distinct chroma and height representations in the auditory cortex that were in concordance with the helical model of pitch perception. In contrast, Regev et al. (2019) found only a small mismatch negativity (MMN) component in ERPs to chroma deviations in unattended Shepard tones (Shepard, 1964), but even this MMN may have been due to small pitch-height change accompanying the chroma deviation. They concluded that the neuronal resources that were adapted in the experiments of Briley et al. (2013) could not be truly chroma-sensitive neurons. Instead, they argued that using complex tones to measure adaptation related to a tone's chroma causes an adaptation based on the physical similarity between tones an octave apart because their frequency spectra overlap. To combat this adaptation confounding, we also used a sinusoidal tone as an adapter to a complex tone probe, eliminating the possibility of overlapping frequency spectra between the consecutive stimuli. An alternative approach to avoid such overlap could have been applying low-pass filtered complex tones as adapters and complex tones

with a missing fundamental frequency as probes. We found that the N1-P2 adaptation results obtained with complex tone adapters in the first and second experiments were not markedly different from those obtained with sinusoidal adapters in the first experiment. The present findings suggest that the representation of pitch chroma is independent of the stimulus spectrum. Our results are also congruent with previous fMRI studies, suggesting distinct chroma and height representations (Moerel et al., 2015; Warren et al., 2003). However, due to the lack of temporal resolution and applied experimental procedures, it is unclear whether those fMRI findings reflect early or late processing. Our present results suggest that the answer may be both. A coarse distinction of chroma between two consecutive tones appears to take place relatively early and be measurable from N1-P2 adaptation, whereas finer distinctions are processed later.

The ANOVA results of the second experiment (see Figure 6) suggest that the information reflected by EEG is widely dispersed or that there may be multiple cerebral sources. The efficiency of spatial decoding (compared to ERPs) also suggests this, as it exploits all channels simultaneously. However, differences between the ERPs to the physically adjacent *ref* and *mat* probes over the right frontal scalp after 500 ms from probe onset (see Figure 7) may be related to the crucial role of the right-hemisphere auditory cortex in pitch processing (e.g., Hyde et al., 2008; Zatorre, 2001). For localization of the cortical or subcortical neural sources or networks active in the present consonance-dissonance classification task, EEG (or magnetoencephalography) and functional magnetic resonance imaging (fMRI) studies with the same experimental design would be necessary. With this approach, data could then be combined with representational similarity analysis to achieve good temporal and spatial accuracy simultaneously (Salmela et al., 2018).

5 | CONCLUSIONS

In both the present experiments, we found a significant reduction (adaptation) in the N1-P2 amplitude both for mathematically exact and enlarged octave-type intervals compared with the compressed octave (major seventh). This supports the theory that neural representations of tones sharing the same (or almost the same) chroma (octave equivalence) overlap significantly. The result was observed both in passive and active listening conditions and both with complex and sinusoidal adapter tones. However, there were no significant differences between mathematically exact and enlarged octaves.

In the second experiment, the behavioural results supported the existence of the octave enlargement

phenomenon. The exact octave and the enlarged octave were explicitly distinguishable, and the enlarged version was perceptually more consonant. In the ERP analyses of the second experiment, the earliest differences were found between the most dissonant *m90* (major seventh type interval) and the most consonant *mat* (mathematically exact octave) and *ref* (subjectively preferred enlarged octave) probes. The results also showed that detecting the octave enlargement phenomenon or distinction between *mat* and *ref* occurs later in processing and probably requires attention and working memory.

AUTHOR CONTRIBUTIONS

JJ, JV and KA conceived and designed the studies. JJ and JV gathered the first study data. JJ and JH collected the second study data. JV and JJ performed pre-processing, ERP analyses and listening experiment analyses. VS and JJ performed the decoding analyses. JJ, JH, KA and VS performed the behavioural data analysis. All authors wrote sections of the manuscript and revised it. All authors read, revised and approved the submitted version of the manuscript.

ACKNOWLEDGEMENTS

The Alfred Kordelin Foundation supported this research. The authors thank Mr. Joonas Hotti, B.A.; Mr. Tommi Makkonen, M.Sc.; Dr. Jukka Pätynen, D.Sc.; and Dr. Jaakko Kauramäki, D.Sc.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.


PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.16150>.

DATA AVAILABILITY STATEMENT

Data are available in Zenodo upon direct request to the authors: [10.5281/zenodo.7770331](https://zenodo.org/record/7770331).

ORCID

Jussi Jaatinen  <https://orcid.org/0000-0003-4976-3474>
 Viljami Salmela  <https://orcid.org/0000-0001-7218-5321>
 Kimmo Alho  <https://orcid.org/0000-0001-8563-2792>

REFERENCES

Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jääskeläinen, I. P., Pekkonen, E., & Ilmoniemi, R. J. (1998). Processing of novel sounds and frequency changes in the human auditory cortex: Magnetoencephalographic recordings.

- Psychophysiology*, 35(2), 211–224. <https://www.ncbi.nlm.nih.gov/pubmed/9529947>, <https://doi.org/10.1111/1469-8986.3520211>
- Bell, A., & Jedrzejczak, W. W. (2017). The 1.06 frequency ratio in the cochlea: Evidence and outlook for a natural musical semitone. *PeerJ*, 5, e4192. <https://doi.org/10.7717/peerj.4192>
- Bendor, D., Osmanski, M. S., & Wang, X. (2012). Dual-pitch processing mechanisms in primate auditory cortex. *The Journal of Neuroscience*, 32(46), 16149–16161. <https://doi.org/10.1523/JNEUROSCI.2563-12.2012>
- Bode, S., Feuerriegel, D., Bennett, D., & Alday, P. M. (2019). The decision decoding ToolBOX (DDTBOX)—A multivariate pattern analysis toolbox for event-related potentials. *Neuroinformatics*, 17(1), 27–42. <https://doi.org/10.1007/s12021-018-9375-z>
- Briley, P. M., Breakey, C., & Krumbholz, K. (2013). Evidence for pitch chroma mapping in human auditory cortex. *Cerebral Cortex*, 23(11), 2601–2610. <https://doi.org/10.1093/cercor/bhs242>
- Butler, R. A. (1968). Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. *The Journal of the Acoustical Society of America*, 44(4), 945–950. <https://doi.org/10.1121/1.1911233>
- Butler, R. A. (1972). Frequency specificity of the auditory evoked response to simultaneously and successively presented stimuli. *Electroencephalography and Clinical Neurophysiology*, 33(3), 277–282. [https://doi.org/10.1016/0013-4694\(72\)90154-x](https://doi.org/10.1016/0013-4694(72)90154-x)
- Cramer, E. M., & Huggins, W. H. (1958). Creation of pitch through binaural interaction. *The Journal of the Acoustical Society of America*, 30(5), 413–417. <https://doi.org/10.1121/1.1909628>
- de Cheveigné, A. (2023). Why is the perceptual octave stretched? An account based on mismatched time constants within the auditory brainstem. *The Journal of the Acoustical Society of America*, 153(5), 2600–2610. <https://doi.org/10.1121/10.0017978>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Demany, L., & Semal, C. (1990). Harmonic and melodic octave templates. *The Journal of the Acoustical Society of America*, 88(5), 2126–2135. <https://doi.org/10.1121/1.400109>
- Dowling, W. J., & Harwood, J. L. (1985). *Music cognition*. Academic Press.
- Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *NeuroImage*, 15(1), 207–216. <https://doi.org/10.1006/nimg.2001.0949>
- Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, 19(3), 576–585. <https://doi.org/10.1093/cercor/bhn108>
- Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography: Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2), 413–497. <https://doi.org/10.1103/RevModPhys.65.413>

- Hartmann, W. M. (1993). On the origin of the enlarged melodic octave. *The Journal of the Acoustical Society of America*, 93(6), 3400–3409. <https://doi.org/10.1121/1.405695>
- Honingh, A., & Bod, R. (2011). In search of universal properties of musical scales. *Journal of New Music Research*, 40(1), 81–89. <https://doi.org/10.1080/09298215.2010.543281>
- Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, 46(2), 632–639. <https://doi.org/10.1016/j.neuropsychologia.2007.09.004>
- Jaatinen, J., & Pätynen, J. (2022). Effect of inharmonicity on pitch perception and subjective tuning of piano tones. *The Journal of the Acoustical Society of America*, 152(2), 1146–1157. <https://doi.org/10.1121/10.0013572>
- Jaatinen, J., Pätynen, J., & Alho, K. (2019). Octave stretching phenomenon with complex tones of orchestral instruments. *The Journal of the Acoustical Society of America*, 146(5), 3203–3214. <https://doi.org/10.1121/1.5131244>
- Massaro, D. W. (1975). Backward recognition masking. *The Journal of the Acoustical Society of America*, 58(5), 1059–1065. <https://doi.org/10.1121/1.380765>
- Massaro, D. W., & Idson, W. L. (1977). Backward recognition masking in relative pitch judgments. *Perceptual and Motor Skills*, 45(1), 87–97. <https://doi.org/10.2466/pms.1977.45.1.87>
- McKinney, M. F., & Delgutte, B. (1999). A possible neurophysiological basis of the octave enlargement effect. *The Journal of the Acoustical Society of America*, 106(5), 2679–2692. <https://doi.org/10.1121/1.428098>
- Moerel, M., de Martino, F., Santoro, R., Ugurbil, K., Goebel, R., Yacoub, E., & Formisano, E. (2013). Processing of natural sounds: Characterization of multipeak spectral tuning in human auditory cortex. *The Journal of Neuroscience*, 33(29), 11888–11898. <https://doi.org/10.1523/JNEUROSCI.5306-12.2013>
- Moerel, M., de Martino, F., Santoro, R., Yacoub, E., & Formisano, E. (2015). Representation of pitch chroma by multi-peak spectral tuning in human auditory cortex. *NeuroImage*, 106, 161–169. <https://doi.org/10.1016/j.neuroimage.2014.11.044>
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425. <https://doi.org/10.1111/j.1469-8986.1987.tb00311.x>
- Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. N. (1988). Frequency and location specificity of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, 69(6), 523–531. [https://doi.org/10.1016/0013-4694\(88\)90164-2](https://doi.org/10.1016/0013-4694(88)90164-2)
- Ohgushi, K. (1983). The origin of tonality and a possible explanation of the octave enlargement phenomenon. *The Journal of the Acoustical Society of America*, 73(5), 1694–1700. <https://doi.org/10.1121/1.389392>
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869. <https://doi.org/10.1155/2011/156869>
- Patterson, R. D., Handel, S., Yost, W. A., & Jaysurya Datta, A. (1996). The relative strength of the tone and noise components in iterated rippled noise. *The Journal of the Acoustical Society of America*, 100(5), 3286–3294. <https://doi.org/10.1121/1.417212>
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767–776. [https://doi.org/10.1016/s0896-6273\(02\)01060-7](https://doi.org/10.1016/s0896-6273(02)01060-7)
- Picton, T. W. (2010). *Human auditory evoked potentials*. Plural Publishing. <https://play.google.com/store/books/details?id=N22CwAAQBAJ>
- Picton, T. W., Woods, D. L., & Proulx, G. B. (1978). Human auditory sustained potentials. II. Stimulus relationships. *Electroencephalography and Clinical Neurophysiology*, 45(2), 198–210. [https://doi.org/10.1016/0013-4694\(78\)90004-4](https://doi.org/10.1016/0013-4694(78)90004-4)
- Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198, 181–197. <https://doi.org/10.1016/j.neuroimage.2019.05.026>
- Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *NeuroImage*, 49(2), 1641–1649. <https://doi.org/10.1016/j.neuroimage.2009.09.045>
- Regev, T. I., Nelken, I., & Deouell, L. Y. (2019). Evidence for linear but not helical automatic representation of pitch in the human auditory system. *Journal of Cognitive Neuroscience*, 31(5), 669–685. https://doi.org/10.1162/jocn_a_01374
- Salmela, V., Salo, E., Salmi, J., & Alho, K. (2018). Spatiotemporal dynamics of attention networks revealed by representational similarity analysis of EEG and fMRI. *Cerebral Cortex*, 28(2), 549–560. <https://doi.org/10.1093/cercor/bhw389>
- Schuck, O. H., & Young, R. W. (1943). Observations on the vibrations of piano strings. *The Journal of the Acoustical Society of America*, 15(1), 1–11. <https://doi.org/10.1121/1.1916221>
- Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, 36(12), 2346–2353. <https://doi.org/10.1121/1.1919362>
- Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4), 305–333. <https://www.ncbi.nlm.nih.gov/pubmed/7134331>, <https://doi.org/10.1037/0033-295X.89.4.305>
- Sundberg, J., & Lindqvist, J. (1973). Musical octaves and pitch. *The Journal of the Acoustical Society of America*, 54(4), 922–929. <https://doi.org/10.1121/1.1914347>
- Terhardt, E. (1970). Oktavspreizung und Tonhöhenverschiebung bei Sinustönen. *Acustica*, 22, 345–351.
- Terhardt, E. (1971). Die Tonhöhe Harmonischer Klänge und das Oktavintervall. *Acustica*, 24, 126–136.
- Terhardt, E. (1974). Pitch, consonance, and harmony. *The Journal of the Acoustical Society of America*, 55(5), 1061–1069. <https://doi.org/10.1121/1.1914648>
- Walliser, K. (1969). Über die Spreizung von empfundenen Intervallen gegenüber mathematisch harmonischen Intervallen bei Sinustönen. *Frequenz*, 23(5), 139–143. <https://doi.org/10.1515/FREQ.1969.23.5.139>

- Ward, W. D. (1954). Subjective musical pitch. *The Journal of the Acoustical Society of America*, 26(3), 369–380. <https://doi.org/10.1121/1.1917806>
- Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 100(17), 10038–10042. <https://doi.org/10.1073/pnas.1730682100>
- Zatorre, R. J. (2001). Neural specializations for tonal processing. *Annals of the new York Academy of Sciences*, 930, 193–210. <https://doi.org/10.1111/j.1749-6632.2001.tb05734.x>

How to cite this article: Jaatinen, J., Vääntänen, J., Salmela, V., & Alho, K. (2023). Subjectively preferred octave size is resolved at the late stages of cerebral auditory processing. *European Journal of Neuroscience*, 58(7), 3686–3704. <https://doi.org/10.1111/ejn.16150>