ELSEVIER

Neuroscience Letters 265 (1999) 179–182

Neuroscience Letters

Neuronal populations in the human brain extracting invariant relationships from acoustic variance

Petri Paavilainen[a,*], Maria Jaramillo[a], Risto Näätänen[a], István Winkler[a, b]

[a]Cognitive Brain Research Unit, Department of Psychology, P.O.B. 13, FIN-00014 University of Helsinki, Helsinki, Finland
[b]Institute for Psychology, Hungarian Academy of Sciences, P.O.B. 398, H-1394 Budapest, Hungary

Received 5 February 1999; received in revised form 1 March 1999; accepted 1 March 1999

Abstract

The ability to extract invariant relationships from physically varying stimulation is critical for example to categorical perception of complex auditory information such as speech and music. Human subjects were presented with tone pairs randomly varying over a wide frequency range, there being no physically constant tone pair at all. Instead, the invariant feature was either the direction of the tone pairs (ascending: the second tone was higher in frequency than the first tone) or the frequency ratio (musical interval) of the two tones. The subjects ignored the tone pairs, and instead attended a silent video. Occasional deviant pairs (either descending in direction or having a different frequency ratio) elicited the mismatch negativity (MMN) of the event-related potential, demonstrating the existence of neuronal populations which automatically (independently of attention) extract invariant relationships from acoustical variance.

Keywords: Cognitive neuroscience; Mismatch negativity; Event-related potentials; Sensory memory; Invariance detection; Speech perception; Music perception

The ability of the brain to extract invariant relationships from the physically varying auditory stimulation is of critical importance to higher perceptual functions such as speech and music perception. This is exemplified by our ability to recognize the same melody irrespective of the key in which it is played. Certain invariant relationships between sequentially presented tones, such as those involving the directions and magnitudes of frequency changes (musical intervals), form an essential factor in creating any given melody, which in turn can be transposed in different keys (i.e. it is independent of the absolute physical parameters). Similarly, although the physical features of the speech signal considerably vary depending on the speaker, we nevertheless usually categorize the phonemes correctly. For instance, we hear the same phoneme /i/ irrespective of whether it is spoken by a male or a female voice. Here we present neurophysiological evidence for the existence of neuronal populations in the human brain which are able to extract aforedescribed invariant relationships from acoustically varying stimulation.

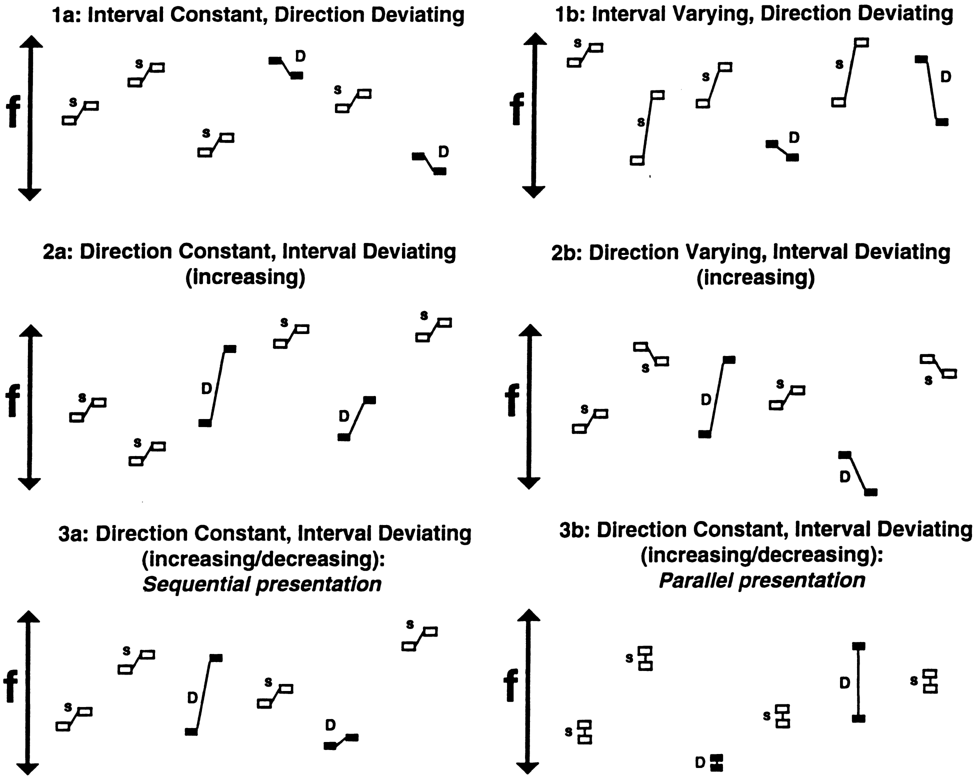
The mismatch negativity (MMN) is an automatic (attention-independent) brain response (a component of the event-related potential, ERP) to sound change. MMN studies have revealed that the physical sound features (e.g. frequency, intensity) are automatically analyzed and encoded in short-duration sensory-memory traces in the auditory cortex [3,5,6]. In the basic MMN paradigm, the subject is presented with a sequence of physically constant 'standard' stimuli with short inter-stimulus intervals. Infrequently replacing the standard stimulus with an occasional deviant stimulus (e.g. a tone of a different frequency) results in MMN elicitation, indicating a 'mismatch' between the new sensory input and the representation of the stimulus encoded in sensory memory. Thus the contents and properties of auditory sensory memory can be probed with MMN [8]. The MMN data to be reported in the present article demonstrate that also higher-order, 'abstract', invariant features of continuously varying acoustic stimulation are encoded in the memory traces [7,10]. The invariant features included in the present study were the direction of frequency change within a tone pair and the frequency ratio of either the two members of a tone pair or of two tones presented in parallel (forming a complex tone). The results have potential implications for the understanding of the neural processing of

---

* Corresponding author. Tel.: +358-9-1912-3496; fax: +358-9-1912-2924.

E-mail address: petri.paavilainen@helsinki.fi (P. Paavilainen)

PII: S0304-3940(99)00237-2



1a: Interval Constant, Direction Deviating

1b: Interval Varying, Direction Deviating

2a: Direction Constant, Interval Deviating (increasing)

2b: Direction Varying, Interval Deviating (increasing)

3a: Direction Constant, Interval Deviating (increasing/decreasing): *Sequential presentation*

3b: Direction Constant, Interval Deviating (increasing/decreasing): *Parallel presentation*

Fig. 1. A schematic illustration showing examples of stimuli used in the various conditions. In each condition, the position of the stimulus pairs on the frequency scale randomly varied over a large frequency range (*y*-axis). The common feature of the standard pairs was either the direction of the pair (ascending, descending, Condition 1) or the relative magnitude of the within-pair frequency interval (Conditions 2 and 3). S, standard pair; D, deviant pair.

complex acoustically varying sound information such as speech perceived in a categorical way.

Ten subjects (age range 22–37 years; six males) participated in the first part of the study (Conditions 1 and 2). They were presented with eight blocks of 640 tone pairs (two blocks per each condition, order of the blocks randomized) binaurally via headphones while watching a self-selected silent video with subtitles. Tone pairs consisted of two 50-ms tones (including rise and fall times of 10 ms each; intensity about 80 dB SPL) of different frequencies separated by a silent gap of 40 ms. The silent inter-pair period was 400 ms. Each block consisted of two types of tone pairs, standard ($P = 0.875$) and deviant pairs ($P = 0.125$), delivered in a random order. The tone pairs randomly varied over a wide frequency range, as the tones for each pair were selected from 11 tones, ranging in frequency with one whole-note steps from 523 (C_5) to 1661 Hz (Gis_6).

Thus, there were no physically constant standard stimuli at all. Instead, the common feature of the standard pairs was of higher-order or 'abstract' nature: all standard pairs either had the same direction of frequency change (ascending, i.e. the second tone was higher than the first tone, Condition 1) or they had the same within-pair frequency interval (i.e. the frequency ratio between the two tones forming a pair was constant, Conditions 2 and 3). The occasional deviant pairs violated these rules (i.e. were either descending in direction or had a different frequency ratio, respectively). For schematic illustration of the different conditions, see Fig. 1.

In Condition 1a, standard pairs were ascending (low-high) and deviant pairs descending (high-low). The within-pair frequency interval was always one whole step on the musical scale. In Condition 1b, the within-pair frequency interval was also varied, so that the standard pairs were composed of random combinations of all the possible different ascending intervals that could be formed from the 11 tones. Correspondingly, the deviant pairs were random variants of all the possible descending intervals.

In Condition 2a, both the standard and deviant pairs were ascending. The standard pairs had a constant within-pair frequency interval (one whole-note step), whereas the deviant pairs had a larger within-pair interval (selected at random from any of the larger-than-one-step intervals that could be formed from the 11 tones). In Condition 2b, also the direction of both types of pairs was randomly varied (half of the pairs ascending, half descending), the deviant pairs thus differing from the standard pairs only by having a proportionally larger frequency change (in either direction).

Conditions 3a,b were conducted with 15 new subjects (age range 20–33 years; six males). As in Condition 2a, both the standard and deviant pairs were ascending. The standard pairs again had a constant within-pair frequency interval (now five steps on a musical scale). In each stimulus

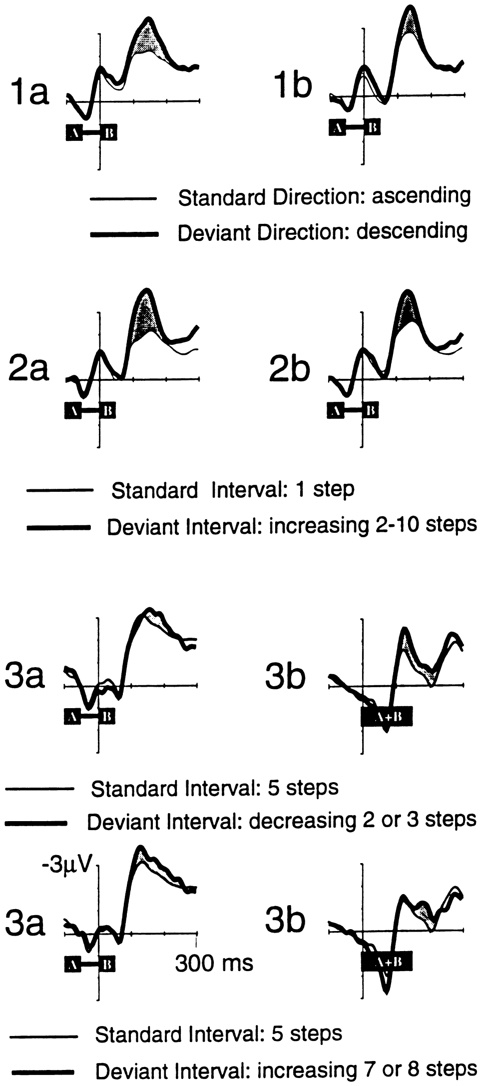


Fig. 2. ERPs to standard (thin line) and deviant pairs (thick line) at the electrode site Fz in the different conditions. Shaded area = MMN. The black rectangles indicate the timing of the stimuli. In Conditions 1, 2 and 3a, the two tones were presented sequentially, pair-wise, and in 3b as a single 150-ms complex tone consisting of two parallel frequency components. The 100-ms period preceding the onset of the second tone of a pair (or the onset of the complex tone in Condition 3b) was used as the 0 μV baseline.

block, there were two types of deviant pairs ($P = 0.075$ each), ones having either a larger (seven or eight steps) or a smaller (two or three steps) within-pair frequency interval than the standards. In Condition 3a, the two tones were presented sequentially (pair-wise), similarly as in the previous conditions. In Condition 3b, the two tones were presented in parallel as a single 150-ms complex tone. The two (simple) tones again varied over a wide frequency range as in the other conditions, the deviant tones having either a larger or a smaller frequency interval between the two frequency components than the standards.

In all conditions, ERPs were recorded to standard and deviant pairs with Ag-AgCl electrodes at Fpz, Fz, Cz, Pz, F3, F4 and left and right mastoids (frequency limits of DC – 40 Hz; sampling rate 250 Hz). The horizontal electro-oculogram (EOG) was monitored with an electrode placed at the outer canthus of the right eye. The nose served as a reference for all the electrodes. EEG epochs with EEG or EOG artifacts exceeding ±150 μV at any electrode were discarded from ERP averaging. ERPs were digitally filtered (bandpass 0.1–30 Hz, slope 24 db/octave) and arithmetically re-referenced to the average of the left and right mastoids.

A statistically significant MMN was elicited by the deviant pairs in all conditions (Fig. 2, Table 1). The results of Condition 1a replicated our previous findings [7,10] indicating that the brain can extract an abstract attribute corresponding to a simple concept ('rise', 'fall'), i.e. to derive a common invariant feature from a set of individual varying physical events and automatically detect violations against this abstract feature. The results of Condition 1b further extended these findings, demonstrating that the brain mechanisms underlying MMN can genuinely extract the direction of the pair, independently of both the absolute frequency level of the pair and the within-pair frequency ratio.

The results of Conditions 2a,b showed that the brain can also extract the magnitude of a frequency interval irrespective of its position in the absolute frequency scale and of its direction (ascending/descending). Furthermore, the results of Condition 3a confirmed that the brain can detect both increasing and decreasing within-pair frequency-interval violations.

The results of Condition 3b showed that results analogous to those in Condition 3a could be obtained even if the two tones were presented in parallel (rather than sequentially as in Conditions 1–3a). This result is especially interesting as both in music and speech, a multitude of the critical information of invariant relationships is presented also in parallel [2,9], rather than sequentially as in the afore-presented tone-pair conditions (1–3a). For example, chords consist of several simultaneous frequency components. In a similar way, vowels are composed of several simultaneous formants, with frequencies and relative intensities that determine the identity of a given vowel. It is obvious that there must be certain invariances within the phonemes that provide the necessary stimulus basis for correctly recognizing them irrespective of the speaker's sex, tone of the voice, and other factors which affect the acoustical 'surface structure' of the speech signal. Such invariances might include, for example, the direction of the frequency glides in the fundamental formant of the speech signal (intonation) and especially the ratio between the frequencies of certain formants forming a given vowel [4,9]. The results obtained in Condition 3b with the complex stimuli, simulating a kind of simple 'phoneme' with two formants, demonstrate the existence of neuronal populations capable of detecting such invariances.

Table 1
Mean MMN amplitudes (μV) at Fz during the designated post-baseline time windows (standard-stimulus mean amplitudes are subtracted from the deviant-stimulus mean amplitudes)[a]

Condition	Time window (ms)	MMN amplitude at Fz (μV)	SEM	*F*-value	d.f.	*P* <
1a	100–200	−1.00	0.35	7.91	1,9	0.05
1b	112–160	−0.94	0.29	7.72	1,9	0.05
2a	100–200	−1.28	0.21	32.11	1,9	0.001
2b	100–200	−0.73	0.19	9.75	1,9	0.05
3a						
Increase	260–272	−0.56	0.24	4.64	1,14	0.05
Decrease	204–244	−0.57	0.28	5.64		0.05
3b						
Increase	156–204	−0.71	0.12	41.52	1,14	0.001
Decrease	104–156	−0.73	0.26	8.16		0.05

[a] Two-way ANOVAs (Electrode [Fz, F3, F4] × stimulus (standard, deviant); Greenhouse–Geisser correction applied) were used to determine whether the mean amplitudes of the deviant- and standard-stimulus ERPs significantly differed from each other.

Our results indicate that sensory information automatically encoded in sensory representations reflected by MMN could in the present conditions only have been produced by higher-order neural processes extracting complex invariant relationships from the sequence of individual, varying auditory events. The distinct MMNs obtained in our different conditions to the specific types of deviant pairs implicate, of course, the existence of neuronal populations capable of detecting such abstract invariances (as opposed to registration of purely physical stimulus features which would have been demonstrated if our deviant pairs had simply deviated on a single physical dimension such as duration or intensity). The present results thus reveal critical fundamental neural mechanisms involved in the processing of complex auditory information such as speech and music [1,6,11].

This work was supported by the Academy of Finland and the University of Helsinki. We thank Jaana Simola and Rika Takegata for their help in data collection.

[1] Aulanko, R., Hari, R., Lounasmaa, O.V., Näätänen, R. and Sams, M., Phonetic invariance in the human auditory cortex. NeuroReport, 4 (1993) 1356–1358.

[2] Deutsch, D., The Psychology of Music, Academic Press, New York, 1982, p. 542.

[3] Giard, M.H., Perrin, F., Pernier, J. and Bouchet, P., Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. Psychophysiology, 27 (1990) 627–640.

[4] Miller, J.D., Auditory-perceptual interpretation of the vowel J. Acoust. Soc. Am., 85 (1989) 2114–2134.

[5] Näätänen, R., Attention and Brain Function, Erlbaum, Hillsdale, NJ, 1992, p. 494.

[6] Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J. and Alho, K., Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature, 385 (1997) 432–434.

[7] Paavilainen, P., Jaramillo, M. and Näätänen, R., Binaural information can converge in abstract memory traces. Psychophysiology, 35 (1998) 483–487.

[8] Ritter, W., Deacon, D., Gomes, M., Javitt, D.J. and Vaughan Jr., H.G., The mismatch negativity of event-related potential as a probe of transient auditory memory: a review. Ear Hear., 16 (1995) 51–67.

[9] Rosner, B.S. and Pickering, J.B., Vowel Perception and Production, Oxford Science, Oxford, 1994, p. 432.

[10] Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M. and Näätänen, R., Representation of abstract attributes of auditory stimuli in the human brain. NeuroReport, 3 (1992) 1149–1151.

[11] Sams, M., Aulanko, R., Aaltonen, O. and Näätänen, R., Event-related potentials to infrequent changes in synthesized phonetic stimuli. J. Cogn. Neurosci., 2 (1990) 344–357.