Published in final edited form as:
Biol Psychol. 2009 October ; 82(2): 169–175. doi:10.1016/j.biopsycho.2009.07.004.

Auditory size-deviant detection in adults and newborn infants

Martin D. Vestergaard[a,b,*], **Gábor P. Háden**[c], **Yury Shtyrov**[b], **Roy D. Patterson**[a], **Friedemann Pulvermüller**[b], **Sue L. Denham**[e], **István Sziller**[d], and **István Winkler**[c,f]

[a] Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom

[b] MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom

[c] Institute for Psychology, Hungarian Academy of Sciences, Hungary

[d] First Department of Obstetrics and Gynaecology, Semmelweis University, Hungary

[e] Centre for Theoretical and Computational Neuroscience, University of Plymouth, Hungary

[f] Institute of Psychology, University of Szeged, Hungary

Abstract

Auditory size perception refers to the ability to make accurate judgements of the size of a sound source based solely upon the sound emitted from the source. Electro-physiological and behavioural data were collected to test whether sound-source size parameters are detected from task-irrelevant sequences in adults and newborn infants. The mismatch negativity (MMN) obtained from adults indexed automatic detection of changes in size for voices, musical instruments and animal calls, regardless of whether the acoustic change indicated larger or smaller sources. Neonates detected changes in the size of a musical instrument. The data are consistent with the notion that auditory size-deviant detection in humans is an innate automatic process. This conclusion is compatible with the theory that the ability to assess the size of sound sources evolved because it provided selective advantage of being able to detect larger (more competent) suitors and larger (more dangerous) predators.

Keywords

Adults; Auditory system; Automatic source-size pre-processing; Electroencephalography (EEG); Glottal pulse rate (GPR); Mismatch negativity (MMN); Size perception; Neonates; Vocal tract length (VTL)

1. Introduction

Humans can readily perceive the size of speakers based on their voice (Collins, 2000; Ives et al., 2005; Smith and Patterson, 2005; Smith et al., 2005). The two parameters of a voice that specify the size of a speaker are pitch and resonance scale (Fant, 1970; Fitch, 2000). Resonance scale is determined by the vocal tract length (VTL) and is sometimes specified as formant dispersion (Fitch, 1997). Pitch, on the other hand, is determined by the glottal pulse rate (GPR), which is often specified in terms of the fundamental frequency (F0) of the

[*] Corresponding author at: Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom. Tel.: +44 (0)1223 333 859; fax: +44 (0)1223 333 840. mdv23@cam.ac.uk. .

harmonic series associated with the GPR. Generally, pitch and resonance scale can characterize the size of a sound source, so long as they produce pulse-resonance sounds (van Dinther and Patterson, 2006).

The phonetic identity of vowels is specified by the formant frequencies. They are determined by the filtering of the supralaryngal vocal tract, which consists of the oral and nasal cavities above the larynx. Formant frequencies are largely independent of pitch but they vary with resonance scale (Fitch, 1997; Lee et al., 1999; Peterson, 1952). To recognize speech sounds, in the sense of understanding their phonetic value, listeners therefore need to normalize for phonetically irrelevant speaker characteristics like pitch and resonance scale (Nearey, 1989). This is a trivial task for most human listeners, and it leads to a remarkable robustness in speech perception to changes in speaker characteristics (Smith et al., 2005). It has therefore been suggested that the auditory system pre-processes sound to detect and normalize for pitch and resonance scale (Irino and Patterson, 2002), and that our ability to make judgments about the size of speakers is based on parametric by-products of this pre-processing. Since most animal calls are pulse-resonance sounds, pitch and resonance scale can code for animal size in most animate communication sounds. Humans and animals can thus use the size information encoded in communication sounds to adapt their behaviour. This ability to gauge the size of the members of a species – be it a predator or a suitor (Fairchild, 1981) – probably has an evolutionary basis (Evans et al., 2006; Fitch and Reby, 2001). According to this notion, the auditory system operates an innate frequency code that associates low frequencies with large sources and high frequencies with small sources (Ohala, 1984). This is obviously true between species, but correlates of such an innate frequency code has also been found within species, such as chimpanzees (Bauer, 1987). Within a behavioural semiotic, anatomical size is interpreted as dominance (for large sources) and submissiveness (for small sources). Because such behaviour appears early within life it is often assumed that detection of auditory size is an innate automatic process. Opponents of this view may still purport that size detection could be the result of perceptual learning of a generalized association between the visual appearance of sound sources and the frequency configuration of their resonances, or they might question whether the auditory cues that accurately signify size differences between gender and species, also reliably operate within gender and species.

Here we address this question using behavioural and electro-physiological data. Pre-attentive processing of auditory features can be studied using the mismatch negativity (MMN, Näätänen et al., 1978) (for a recent review, see Kujala et al., 2007) component of the event-related brain potential (ERP). MMN is elicited by sounds violating a detected auditory regularity whether or not the sounds are task-relevant (for a review, see Sussman, 2007) and can be recorded even in sleeping newborns (Alho et al., 1990). MMN elicitation implies: (a) that some auditory regularity has been detected by the brain (e.g., the presence of an auditory feature in the majority of sounds in a sequence), and (b) that the given sound violated this regularity (e.g., one or more feature of this sound was found to be different from the common feature(s), Winkler, 2007). Therefore, the MMN method can tell whether auditory source size is extracted and represented for task-irrelevant sound sequences. To this end, we recorded the electroencephalogram (EEG) in adults and newborn infants exposed to variation in auditory size, and we tested whether occasional auditory size-deviants elicited the discriminative MMN response. The two auditory cues of size, resonance scale and pitch, were tested independently in the adult participants; in the infants, only deviation in resonance scale was tested. To examine whether detection varies with familiarity of the stimuli, both known and unknown stimuli were used. Adults were tested with familiar speech (spoken syllables), rarely perceived musical instrument sounds (French horn), and unfamiliar animal calls (the croak of a North-American Bullfrog, which is unknown to most Europeans). Neonates were only presented with sounds from the French horn, which we assume are unfamiliar to them. The adult participants also evaluated the perceived naturalness of the sounds and were asked to identify them.

2. Experiment 1

EEG was recorded from 30 (17 males) healthy right-handed native English adults (18–37 years). No subject had any history of neurological disorder, and they all had normal hearing. A handedness inventory (Oldfield, 1971) was administered to ensure that no subjects had a laterality quotient (LQ) of less than 50; the average LQ was 90. Five subjects (3 males) were excluded due to excessive recording artefacts. The average age of the remaining subjects was 24 years. The recordings were carried out at the MRC Cognition and Brain Sciences Unit, Cambridge, U.K. after informed consent was obtained from the participants. The experiment was approved by the Cambridge Psychology Research Ethics Committee (CPREC).

2.1. Stimuli

Sequences composed of short segments of speech, musical instrument sounds, or animal calls were presented to the participants (Fig. 1). Changes in source size were simulated by changing pitch and resonance scale with a voice-coder (Kawahara and Irino, 2004). There were six deviants for each kind of stimulus: 2 single-feature deviants (resonance or pitch), each in two directions (i.e. 4 oddballs) and 2 double deviants. The double deviants were the congruent combinations of pitch and resonance (see Table 1 for the configuration and relationship of the six types of deviants). The stimuli were normalized to the same root-mean-square (RMS) value, and their perceptual centre and onset time (Marcus, 1981) were controlled, to make them equally loud and perceptually isochronous, respectively. The duration of each stimulus was 683 ms and they were delivered at 783 ms stimulus onset asynchrony (SOA).

The speech stimulus was the syllable /da/ recorded from one speaker (author RDP) as previously used in psychophysical (Ives et al., 2005) and neuroimaging (von Kriegstein et al., 2006) studies. The standard pitch was 139 Hz and the deviants were 131 Hz and 147 Hz corresponding to the musical notes C3, C#3 and D3 in ascending order. Thus, the pitch deviation was one semitone (~6%) up or down from the standard value. The just-noticeable-difference (JND) for vowel pitch is approximately 2% (Smith et al., 2005), so the semitone steps between standard and oddball represent approximately 3 JNDs. The JND for discriminating changes in resonance scale for plosive consonant-vowel syllables is approximately 4% (Ives et al., 2005), so a 12% resonance-scale deviation (0.89 and 1.12 relative to the standard taken to be 1) is approximately as salient as the 6% pitch deviation. Assuming the standard relationship between height and VTL (Fant, 1970; Fitch and Giedd, 1999) for the original speaker, the frequency shifts simulated the acoustic affects equivalent to VTLs of 17.4 cm and 13.9 cm, respectively (Turner et al., 2009). With these combinations of pitch and resonance, the stimuli simulated male speakers of approximately 157 cm (5′2″), 175 cm (5′9″) and 196 cm (6′5″) height.

The musical stimulus was a note from a French horn taken from a database (Goto et al., 2003) previously described by van Dinther and Patterson (2006). The pitch values for the horn stimuli were the same as for the speech. The JND for discrimination of simulated source size for a French horn for human listeners is between 6% and 9% (van Dinther and Patterson, 2006), so for the French horn, the resonance-deviance was set at ±22% (0.82 and 1.22) assumed to be comparable to the speech oddballs.

The animal call was a croak from an American bullfrog (*Rana catesbeiana*) taken from the suite of bullfrog calls recorded by Bee and Gerhardt (2001). The croak was taken from

Bullfrog #5, cut 1, call 10, which originally had a sustained pitch of 114 Hz. The original call was about 1 s long, so a shrinkage factor of 0.66 was used for the re-synthesis, and silence was added before and after the shrunken croaks to give 683-ms signals. The pitch values for the bullfrog stimuli were the same as for the French horn and speech. Pilot work showed that the JND for the discrimination of resonance-scale changes for bullfrog calls was around 5%, so the bullfrog oddballs were chosen between ±15% of the standard (0.87 and 1.15) for compatibility with the salience of the speech and horn resonance-scale deviants.

2.2. Stimulus naturalness

While there is ample empirical support to the notion that resonance-scale manipulation as described above leads to variation in the perception of the size of the sources (Ives et al., 2005; Smith and Patterson, 2005; van Dinther and Patterson, 2006), it is unknown whether listeners can identify the sources with no prior knowledge. The participants were therefore asked to identify and to evaluate the naturalness of the stimuli on a visual categorical scale labelled: *very natural* (1), *slightly synthetic* (2), *quite synthetic* (3), *not at all natural* (4), *and don't know* (no score). For source identification one point was scored for correct and 0.5 point for correct family (e.g., 'animal' for croak and 'instrument' for the horn). Identification scores and naturalness scores were analyzed with repeated-measures analysis of variance (ANOVA) with source as grouping variable. The behavioural evaluations were based on the oddball sequences described below and took place in the pauses between the ERP recordings for each source type.

2.3. Procedure

The participants were seated in an electrically and acoustically shielded chamber, and they were instructed to watch a silent movie of their own choice and to ignore the auditory signals. The sounds were presented in a so called optimal-oddball (Näätänen et al., 2004) sequence in which every second sound was identical (standards) having 139 Hz pitch and resonance scale 1.00. Every other sound was deviant, equally distributed between the six types of deviants (2 single-feature deviants × 2 deviance directions + 2 congruent double-feature deviants). Three stimulus blocks of 1930 sounds each (one for each stimulus type: speech, musical stimulus, and animal call) were delivered with their order balanced across subjects. Stimulus blocks started with 10 standard sounds, after which deviants emerged in every second position. Thus, 160 responses were collected for each stimulus-type/deviant together with 960 standards (discarding the first 10 standards).

EEG was recorded from 65 scalp locations, with Cz as the common reference and FPz as ground electrode (according to the international 10–20 system). Eye movements were monitored by recording electro-oculogram (EOG) with a bipolar montage of two electrode pairs: one pair placed above and below the right eye and the other on the temples lateral to the outer canthi. EEG was recorded with 16 bit resolution at a sampling rate of 500 Hz by a SynAmps amplifier (Neuroscan Labs, Charlotte, NC, USA). The signals were on-line band-pass filtered between 0.01 Hz and 100 Hz. The stimuli were presented diotically using STIM1 stimulus delivery software (Neuroscan Labs, Charlotte, NC, USA) via Sennheiser HD250 headphones, and the data were analyzed with BrainVision Analyzer (Brain Products GmbH, Munich, Germany) version 1.05. The EEG was filtered off-line between 1.0 Hz and 20 Hz. For each stimulus, an epoch of 600-ms duration including a 100-ms pre-stimulus period was extracted from the continuous EEG recording. Epochs with a voltage change below 0.5 μV or above 100 μV on any EEG or EOG channel were rejected from further analysis.

2.4. ERP analyses

Epochs were baseline-corrected using the 100-ms pre-stimulus period, re-referenced to the average of the mastoid electrodes (TP9 and TP10), and averaged separately for the different stimulus types. The mean number of artefact-free trials per participant and condition was 123 deviants and 742 standards for speech, 126 deviants and 755 standards for the horn, and 126 deviants and 765 standards for the croak. The MMN response was analyzed for the FCz electrode, which had the best signal to noise ratio. MMN peaks were sought in a window between 50 ms and 200 ms from the perceptual onset of the sounds. The MMN amplitudes were then calculated by averaging the deviant-minus-standard difference signal in the ±15 ms latency range relative to the grand-average peak for the double deviant, separately for each source and size-change direction. Fig. 2 shows the topography of the MMN in the ±15 ms latency range for the larger double deviants, indicating that the largest signals are at the fronto-central location. The statistical significance of the difference signal was determined by Bonferroni-corrected two-tailed *t*-test. The effects of sound source and deviant feature were analyzed with a three-way repeated-measures ANOVA [source (horn vs. speech vs. croak) × size (larger vs. smaller) × feature (pitch vs. resonance vs. both)]. Size was a grouping variable for testing whether the deviance relates to larger or smaller sources, and feature was the variable for grouping within size testing the effect of the deviating feature(s). Within feature, the effect of pitch vs. resonance was analyzed by pairwise comparisons with Sidak adjustment for multiple comparisons. In order to reduce random effects in the ANOVA, the MMN response was normalized to the maximum value across condition for each subject and stimulus type. Greenhouse-Geisser correction of the degrees of freedom was used to compensate for lack of sphericity, and partial eta squared values (η^2_p) are quoted to report the effect sizes.

2.5. Results

The speech stimuli were recognized as human speech by all but one participant who identified them as sheep sounds. The French horn was recognized by 15 participants, and a further six identified it as some other musical instrument. Only two subjects recognized the frog croak, and a further seven identified it as some animal sound. This effect on source recognition was highly significant ($F_{2,24} = 51.0$, $p < 0.001$, $\eta^2_p = 0.65$). The speech sound scored highest on naturalness followed by horn and croak ($F_{2,24} = 10.8$, $p < 0.001$, $\eta^2_p = 0.31$). The average naturalness scores were: 2.1 (standard deviation 0.8) for the speech, 2.6 (1.0) for the horn, and 3.2 (0.8) for the croak. Furthermore, men evaluated the sounds as more natural than did women ($F_{1,24} = 6.2$, $p = 0.02$, $\eta^2_p = 0.21$).

The grand average ERPs and deviant-minus-standard difference signals are shown in Fig. 3, with the significant portions of the MMN highlighted by grey areas. All resonance deviants and all double deviants elicited the MMN, whereas only some of the pitch deviants elicited a significant MMN. The latency of the double-deviant MMN peaks relative to the perceptual onset time was: 110 ms for larger and 108 ms for smaller sources for the horn; 140 ms for larger and 118 ms for smaller sources for speech; and 98 ms for larger and 118 ms for smaller sources for the croak. In some cases, the MMN was followed by a positive difference waveform, possibly P3a components.

The amplitude of the MMN (Fig. 4) elicited for size increments was larger than for size decrements ($F_{1,23} = 14.2$, $p = 0.001$, $\eta^2_p = 0.38$). The largest MMN was elicited by the double deviants ($F_{2,46} = 27.1$, $p < 0.001$, $\varepsilon = 0.65$, $\eta^2_p = 0.54$) followed by resonance and pitch. Pairwise comparisons showed that resonance deviants elicited larger MMN than pitch deviants ($p = 0.007$), and that the double deviants elicited larger MMNs than both pitch ($p < 0.001$) and resonance ($p = 0.002$). The MMN for horn sounds was larger than that for croak ($F_{2,46} = 5.4$, $p = 0.009$, $\varepsilon = 0.86$, $\eta^2_p = 0.19$), and this effect was more pronounced for pitch and double deviants than for resonance alone as shown by the interaction between the source and feature factors ($F_{4,92} = 8.0$, $p < 0.001$, $\varepsilon = 0.95$, $\eta^2_p = 0.26$). This tendency for pitch and the double deviant to elicit larger MMN for horn than for croak (with resonance eliciting equal MMN for the three sources) was more pronounced for larger than for smaller deviants. For smaller deviants only pitch seemed to elicit larger MMN for horn, as shown by the three-way interaction of size, feature and source ($F_{4,92} = 7.0$, $p < 0.001$, $\varepsilon = 0.83$, $\eta^2_p = 0.23$). The possible P3a responses following the MMN in some conditions were only statistically significant for increments in pitch and both pitch and resonance (i.e. the double deviant) for the horn. For the croak, the positive deflection for pitch increase occurred at the latency of the negativities for all other conditions, and for that reason, it was not considered a typical P3a response.

Overall, the MMN for the horn had the best quality, and the size of the response was larger for resonance-scale than for pitch deviance. Therefore, resonance-scale deviance for French horn sounds were used in Experiment 2 administered to newborn infants.

3. Experiment 2

17 (7 males) healthy (APGAR score 9/10), full-term, newborn infants were studied on day 2 or 3 post-partum. Their gestational age was 38–42 weeks and birth weight was 2550–4150 g. 4 subjects' data (2 males) were removed from the analyses because of excessive recording artefacts. The recordings were carried out in the maternity ward of the First Department of Obstetrics and Gynaecology, Semmelweis University, Budapest, Hungary after informed consent was obtained from one or both parents. The mother of the infant was present at the recording. The experiment was approved by the Ethics Committee of Semmelweis University as well as by the Institutional Review Board of the Institute for Psychology, Hungarian Academy of Sciences.

3.1. Stimuli

Two notes of different resonance scale, 0.82 and 1.22, were played on a French horn (Goto et al., 2003). The pitch for both notes was 175 Hz corresponding to the musical note F3. Change in size was simulated by changing the resonance scale of the sounds with a voice-coder (Kawahara and Irino, 2004). For adult listeners, the JND for resonance scale for the French horn is approximately 6–9% (van Dinther and Patterson, 2006), so the size deviance was approximately 6 JNDs, corresponding to the resonance scale factors 1/1.22 and 1.22 (where 1 denotes the original sound). In order to improve the signal-to-noise ratio of the early ERP components, sound onsets were made more abrupt by removing the original 60-ms long initial period of the sounds and imposing a 10-ms long raised-cosine ramp. The total sound duration was 545 ms. The stimuli were presented diotically using E-Prime stimulus delivery software (Psychology Software Tools, Inc., Pittsburgh, PA, USA) via ER-3A loudspeakers (EtymStic Research, Inc., Elk Grove Village, IL, USA) connected via sound tubes to self-adhesive ear-couplers (Natus Medical, Inc., San Carlos, CA, USA) placed over the babies' ears.

3.2. Procedure

The sounds were presented in a conventional oddball sequence (750 ms stimulus onset asynchrony [SOA]) with 85% of them (standards) having one and 15% (deviants) the other resonance scale value. 4 stimulus blocks of 500 sounds each (a total of 1700 standard and 300 deviant sounds) were delivered. The resonance-scale value of the standard and deviant notes was reversed in half of the stimulus blocks in order to compare deviant and standard responses elicited by identical stimuli. The order of the two types of stimulus blocks was balanced in order to equalize the distribution of the different sleep stages between the two conditions. EEG was recorded from the F3, F4, C3, Cz, and C4 scalp electrodes (10–20 systems) and from electrodes placed over the left and right mastoids (Lm and Rm, respectively), with the common reference attached to the tip of the nose. The ground electrode was placed on the forehead. Eye movements were monitored by recording the EOG between two electrodes, one placed below the left and another above the right eye. EEG was recorded with 24 bit resolution at a sampling rate of 250 Hz by V-Amp amplifiers (Brain Products GmbH, Munich, Germany). The signals were low-pass filtered at 40 Hz on line. EEG was filtered off line between 1.0 Hz and 20 Hz. For each stimulus, an epoch of 600-ms duration, including a 100-ms pre-stimulus period, was extracted from the continuous EEG record. Epochs with a voltage change below 0.1 μV or above 100 μV on C3, Cz, C4 or the EOG channel were rejected from further analysis.

3.3. ERP analyses

Responses were analyzed for the central line of electrodes. Epochs were baseline-corrected using the 100-ms pre-stimulus period and averaged separately for the different stimulus types and conditions. The mean number of artefact-free trials per infant was 129 deviants and 695 standards for the 0.82 resonance scale, and 120 deviants and 752 standards for the 1.22 resonance scale. Deviant responses were compared with those elicited by the identical standard sounds delivered in the reversed stimulus blocks. For amplitude measurements, 40-ms time windows were selected from the grand-average, deviant-minus-standard, difference waveforms on the Cz recording, which showed the best signal-to-noise ratio. For the 1.22 (smaller) resonance-scale responses, one window was centred on the negative difference peak in the 56–96 ms latency range and the other on the positive difference peak in the 204–244 ms latency range. For the 0.82 (larger) resonance-scale responses, the time window was 212–252 ms. The differences between standard and deviant responses were analyzed with paired two-tailed *t*-tests, separately for the two resonance-scale deviants (larger vs. smaller).

3.4. Results

The grand average deviant and identical standard ERPs recorded at the central line of electrodes together with the deviant-minus-standard difference waveform are shown in Fig. 5. The peak latencies are comparable with those observed in several previous studies of neonates (Háden et al., 2009; Kushnerenko et al., 2007; Stefanics et al., 2007). For larger resonance-scale (negative difference peak in the 212–252 ms range), the deviants elicited a significantly lower-amplitude central (Cz) response than the standards ($t = 2.29$, df = 12, $p = 0.041$). For smaller resonance-scale, no significant difference was observed in the 56–96 ms range (negative difference peak), whereas for the positive difference peak in the 204–244 ms latency range the deviants elicited a significantly higher-amplitude central (Cz) response than standards ($t = -3.74$, df = 12, $p = 0.003$). Thus, occasional source-size increments and decrements elicited a discriminative ERP response in sleeping newborn infants, but with reversed polarity.

4. Discussion

Both adults and newborn infants showed discriminative responses to occasional deviations in sound-source size. The adults were tested with both familiar and unfamiliar stimuli. The adults were watching a silent movie whereas the neonates were asleep during the stimulation. These results support the notion that auditory source size is extracted automatically from incoming sounds even when the sounds are not task-relevant. The infant results suggest that this ability is innate.

The discriminative responses elicited in infants by infrequent source-size increments and decrements were of opposite polarity. Since, the difference waveforms were obtained by subtracting responses elicited by identical sounds, the discriminative responses were free of components specific to the different sounds. Thus, the difference for larger and smaller size deviants cannot be explained by differences in the ERPs elicited by the different sounds. Rather, it reflects a difference in processing the direction of size changes. One should note that the discriminative ERPs found in neonates in the current as well as in previous studies (e.g., Alho et al., 1990; Háden et al., 2009; Kushnerenko et al., 2007; Stefanics et al., 2007) are not directly equivalent to the MMN and the P3a response in adults. In a detailed study of neonatal ERP responses to various deviances, Kushnerenko et al. (2007) found that the response to acoustic deviance consists of three successive peaks: an early negative difference followed by a positive and another negative difference. All three difference waveforms found in neonates show features of the adult MMN response and often only one or two of them are observable in response to a given stimulus paradigm (for a while, this caused a debate in the literature as to whether mismatch response in neonates appears as a negative or a positive difference waveform). All three waveforms respond to deviation, with the early negativity also being sensitive to large spectral differences, the middle positivity to energy changes, and the late negativity possibly to qualitative (categorical) changes (Kushnerenko et al., 2007). Given that in adults, the neural populations generating MMN vary with the type of deviation (Giard et al., 1995), we speculate that earlier in the course of development, there are no pure change detectors yet in auditory cortex. That is, these neural populations respond to a combination of the given stimulus feature as well as to the direction of deviation (such as, increase vs. decrease from a given base level).

In the adults, increments in size elicited a larger discriminative response than decrements. Several studies have found asymmetries in MMN elicitation; i.e., when a change in one direction elicits a larger or earlier MMN than a change in the opposite direction (Hasting et al., 2008; Nordby et al., 1994; Sabri and Campbell, 2000; Shtyrov and Pulvermüller, 2002; Tervaniemi et al., 1994). Shtyrov and Pulvermüller reported a study on the processing of grammatical affixes and found earlier MMN for the shorter (non-affixed) stimulus and delayed MMN for the longer stimulus. They suggested that this finding reflected additional time required to process the inflectional affix. In most studies however, MMNs to increments in stimulus length and thus stimulus energy have larger amplitudes than those to decrements. In accordance with these results, it appears that an increase in the perceived size of a sound source is more intrusive, and more likely to get detected, or cause distraction from ongoing activity, than a decrease in perceived size. It should be noted, however that for the current stimuli, the increase in perceived size is not accompanied by an increase in stimulus energy, since the stimuli were equalized to the same RMS level for all sounds. Moreover, the spectral effect of an increase in size is a downward shift in the spectrum, and vice versa: a decrease in size is reflected in an upward shift in spectrum. Thus, it is unlikely that the preferential discrimination of a size increase simply reflects sensitivity to the acoustic features, since the spectrum shifted in the opposite direction and the levels were the same. Since size increments more often signal danger than do decrements, it seems plausible that this bias has survival value and is innate. The difference between the responses elicited by occasional increases and decreases in source size in neonates suggests that one effect of the maturation of auditory cortex is an improved separation between feature and change detectors, such that change detection in adults shows less dependence on the actual stimulus features.

The ERPs for the different sound sources were morphologically very different due to the relatively short SOA. Only the Horn elicited a conventional N1-P2 complex probably because of its short onset time (55 ms). The speech and animal sounds had much longer onsets and as a result the N1 was suppressed by the short SOA. Nevertheless, all three sources elicited MMN in the 98–140 ms latency range relative the sound onset. This result shows that an N1 response is not required for elicitation of MMN.

The magnitudes of the acoustic deviant features were chosen with respect to their just noticeable difference (JND), separately for each feature, in such a way that the magnitude of the single-feature deviants was the same number of JNDs for all sound sources. This was intended to equate the perceptual salience of a change in the features. However, the response to resonance-scale deviance was larger than to pitch deviance. Hence, of the two features that are involved in the acoustic specification of the size of a sound source (pitch and resonance), the largest response was recorded for resonance scale. The resonance scale of an animal's call is related to the length of their vocal tract whereas the pitch is related to the rate of vibration of the vocal folds. The range within which the vocal folds vibrate is determined by the size and gender of the animal. Some animals (e.g., the red deer stag) can lengthen their vocal tract during vocalizations, thereby creating the illusion of larger size, the selective advantage of which is thought to be to imitate the deeper vocalisations of larger conspecies (Fitch and Reby, 2001). Nevertheless, in most other species including humans VTL is more constant than glottal pulse rate, so resonance scale would usually be a more honest signal of size (harder to fake) than pitch.

While the selective advantage of size detection may have been the evolutionary basis of human's sensitivity to resonance-scale changes, the ability to distinguish one sound source from another also has application in verbal communication. It has long been known that perceived phonological identity relies on consistent formant dispersion within sentences (Broadbent et al., 1956). It has also been demonstrated that human listeners derive an advantage from resonance-scale differences when disambiguating competing speech signals (Darwin et al., 2003; Vestergaard et al., 2009). Hence, resonance-scale discrimination and normalization serve as prerequisites for the robustness of speech perception.

Our results support the notion that there is specific neural machinery for size processing in the human auditory system, and the processing is automatic. The function of this mechanism is compatible with an evolutionary advantage of detecting dangerous (larger) predators and strong (larger) suitors. We have previously shown that newborns can extract pitch differences independently of resonance scale (Háden et al., 2009), and here we have shown that resonance-scale detection is also innate and pre-attentive. Therefore, these results are compatible with the assumption that auditory source-size processing is an innate function of the human auditory system.

Acknowledgments

We thank Clare Cook and Erika Váradiné Józsa for help with collecting the data, and Dr. János Rigó for supporting the study of neonates at the First Department of Obstetrics and Gynaecology, Semmelweis University, Budapest, Hungary. The adult experiment was funded by the UK Medical Research Council (G0500221, G9900369, U1055.04.003.00001.01, U1055.04.003.00003.01) and by the European Community under the "Information Society Technologies" (IST-2001-35282) and the "New and Emerging Science and Technologies" Programmes (NEST, Nestcom Project). The infant experiment was supported by the European Commission's 6th Framework Programme for "Information Society Technologies" (project title: EmCAP Emergent Cognition through Active Perception, contract no.: 013123).

References

Alho K, Sainio K, Sajaniemi N, Reinikainen K, Näätänen R. Event-related brain potential of human newborns to pitch change of an acoustic stimulus. Electroencephalography and Clinical Neurophysiology. 1990; 77(2):151–155. [PubMed: 1690117]

Bauer HR. Frequency code: orofacial correlates of fundamental frequency. Phonetica. 1987; 44(3): 173–191. [PubMed: 3452836]

Bee MA, Gerhardt HC. Neighbour–stranger discrimination by territorial male bullfrogs (*Rana catesbeiana*): I. Acoustic basis. Animal Behaviour. 2001; 62(6):1129–1140.

Broadbent DE, Ladefoged P, Lawrence W. Vowel sounds and perceptual constancy. Nature. 1956; 178(4537):815–816. [PubMed: 13369552]

Collins SA. Men's voices and women's choices. Animal Behaviour. 2000; 60(6):773–780. [PubMed: 11124875]

Darwin CJ, Brungart DS, Simpson BD. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. Journal of the Acoustical Society of America. 2003; 114(5):2913–2922. [PubMed: 14650025]

Evans S, Neave N, Wakelin D. Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice. Biological Psychology. 2006; 72(2):160–163. [PubMed: 16280195]

Fairchild L. Mate selection and behavioral thermoregulation in Fowler's Toads. Science. 1981; 212(4497):950–951. [PubMed: 17830192]

Fant, GCM. Acoustic Theory of Speech Production. Mouton; The Hague: 1970.

Fitch WT. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. Journal of the Acoustical Society of America. 1997; 102(2 Pt 1):1213–1222. [PubMed: 9265764]

Fitch WT. The evolution of speech: a comparative review. Trends in Cognitive Sciences. 2000; 4(7): 258–267. [PubMed: 10859570]

Fitch WT, Giedd J. Morphology and development of the human vocal tract: a study using magnetic resonance imaging. Journal of the Acoustical Society of America. 1999; 106(3 Pt 1):1511–1122. [PubMed: 10489707]

Fitch WT, Reby D. The descended larynx is not uniquely human. Proceedings of the Royal Society B: Biological Sciences. 2001; 268(1477):1669–1675.

Giard MH, Lavikainen J, Reinikainen K, Perrin F, Bertrand O, Pernier J, et al. Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. Journal of Cognitive Neuroscience. 1995; 7(2):133–143.

Goto, M.; Hashiguchi, H.; Nishimura, T.; Oka, R. RWC music database: music genre database and musical instrument database. Proceedings of 4th International Conference on Music Informational Retrieval (ISMIR 2003); 2003. p. 229-230.

Háden GP, Stefanics G, Vestergaard MD, Denham SL, Sziller I, Winkler I. Timbre-independent extraction of pitch in newborn infants. Psychophysiology. 2009; 46(1):69–74. [PubMed: 19055501]

Hasting AS, Winkler I, Kotz SA. Early differential processing of verbs and nouns in the human brain as indexed by event-related brain potentials. European Journal of Neuroscience. 2008; 27(6): 1561–1565. [PubMed: 18364028]

Irino T, Patterson RD. Segregating information about the size and shape of the vocal tract using a time-domain auditory model: the stabilised wavelet-mellin transform. Speech Communication. 2002; 36:181–203.

Ives DT, Smith DR, Patterson RD. Discrimination of speaker size from syllable phrases. Journal of the Acoustical Society of America. 2005; 118(6):3816–3822. [PubMed: 16419826]

Kawahara, H.; Irino, T. Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In: Divenyi, PL., editor. Speech Separation by Humans and Machines. Kluwer Academic; Massachusetts: 2004. p. 167-180.

Kujala T, Tervaniemi M, Schroger E. The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. Biological Psychology. 2007; 74(1):1–19. [PubMed: 16844278]

Kushnerenko E, Winkler I, Horvath J, Näätänen R, Pavlov I, Fellman V, et al. Processing acoustic change and novelty in newborn infants. European Journal of Neuroscience. 2007; 26(1):265–274. [PubMed: 17573923]

Lee S, Potamianos A, Narayanan S. Acoustics of children's speech: developmental changes of temporal and spectral parameters. Journal of the Acoustical Society of America. 1999; 105(3): 1455–1468. [PubMed: 10089598]

Marcus SM. Acoustic determinants of perceptual center (P-center) location. Perception & Psychophysics. 1981; 30(3):247–256. [PubMed: 7322800]

Näätänen R, Gaillard AW, Mantysalo S. Early selective-attention effect on evoked potential reinterpreted. Acta Psychologia. 1978; 42(4):313–329.

Näätänen R, Pakarinen S, Rinne T, Takegata R. The mismatch negativity (MMN): towards the optimal paradigm. Clinical Neurophysiology. 2004; 115(1):140–144. [PubMed: 14706481]

Nearey TM. Static, dynamic, and relational properties in vowel perception. Journal of the Acoustical Society of America. 1989; 85(5):2088–2113. [PubMed: 2659638]

Nordby H, Hammerborg D, Roth WT, Hugdahl K. ERPs for infrequent omissions and inclusions of stimulus elements. Psychophysiology. 1994; 31(6):544–552. [PubMed: 7846215]

Ohala JJ. An ethological perspective on common cross-language utilization of F0 of voice. Phonetica. 1984; 41(1):1–16. [PubMed: 6204347]

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971; 9(1):97–113. [PubMed: 5146491]

Peterson GE. The information-bearing elements of speech. Journal of the Acoustical Society of America. 1952; 24(6):629–637.

Sabri M, Campbell KB. Mismatch negativity to inclusions and omissions of stimulus features. Neuroreport. 2000; 11(7):1503–1507. [PubMed: 10841366]

Shtyrov Y, Pulvermüller F. Memory traces for inflectional affixes as shown by mismatch negativity. European Journal of Neuroscience. 2002; 15(6):1085–1091. [PubMed: 11918667]

Smith DR, Patterson RD. The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex and age. Journal of the Acoustical Society of America. 2005; 118(5):3177–3186. [PubMed: 16334696]

Smith DR, Patterson RD, Turner R, Kawahara H, Irino T. The processing and perception of size information in speech sounds. Journal of the Acoustical Society of America. 2005; 117(1):305–318. [PubMed: 15704423]

Stefanics G, Háden G, Huotilainen M, Balázs L, Sziller I, Beke A, et al. Auditory temporal grouping in newborn infants. Psychophysiology. 2007; 44(5):697–702. [PubMed: 17532802]

Sussman ES. A new view on the MMN and attention debate: the role of context in processing auditory events. Journal of Psychophysiology. 2007; 21(3–4):164–175.

Tervaniemi M, Saarinen J, Paavilainen P, Danilova N, Naatanen R. Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. Biological Psychology. 1994; 38(2–3):157–167. [PubMed: 7873700]

Turner RE, Walters TC, Monaghan JJM, Patterson RD. A statistical formant-pattern model for estimating vocal-tract length from formant frequency data. Journal of the Acoustical Society of America. 2009; 125(4):2374–2386. [PubMed: 19354411]

van Dinther R, Patterson RD. The perception of size in musical instruments. Journal of the Acoustical Society of America. 2006; 120(4):2158–2176. [PubMed: 17069313]

Vestergaard MD, Fyson NRC, Patterson RD. The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. Journal of the Acoustical Society of America. 2009; 125(2):1114–1124. [PubMed: 19206886]

von Kriegstein K, Warren JD, Ives DT, Patterson RD, Griffiths TD. Processing the acoustic effect of size in speech sounds. Neuroimage. 2006; 32(1):368–375. [PubMed: 16644240]

Winkler I. Interpreting the mismatch negativity (MMN). Journal of Psychophysiology. 2007; 21(3–4): 147–163.

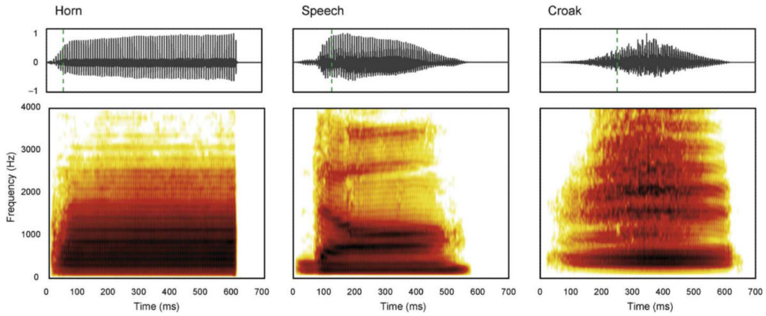

Fig. 1.
Waveforms (top) and spectrograms (bottom) of the three stimulus types for the standard sounds used in the adult experiment. The vertical dashed lines in waveforms indicate the perceptual onset of the sounds. The distinct formant structure of the sounds is discernable in the horizontal dark ridges in the spectrograms. All sounds have substantial energy in the range up to 4 kHz but they differ in their temporal envelopes.

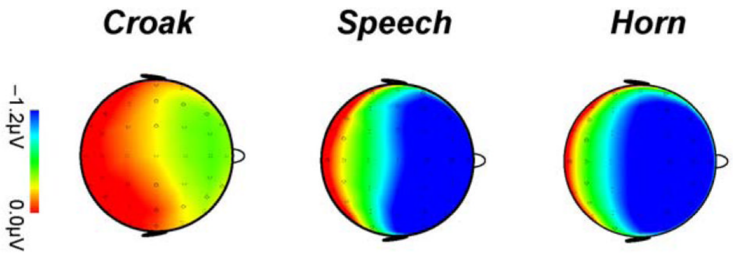

Fig. 2.

Scalp distributions of deviant-minus-standard difference signals for the larger double-deviants in the adults for each sound source at the peak latency relative to perceptual onset: horn (95–125 ms), speech (128–158 ms), croak (83–113 ms).

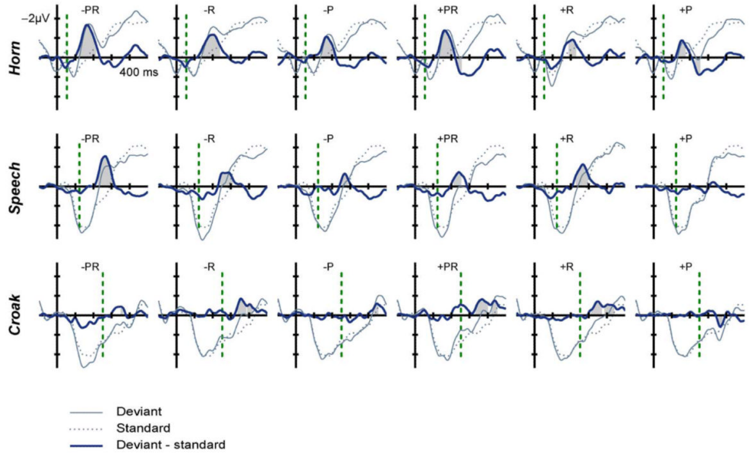

Fig. 3.

Grand-average (25 adults) fronto-central (FCz) ERPs elicited by standard (thin dotted lines) and deviant (thin solid lines) sounds for horn (top panels), speech (middle panels), and frog croak (bottom panels). The deviant-minus-standard difference signal is shown in thick lines with the significant MMN responses marked by grey shading. The deviant feature(s) are indicated in each panel: pitch (P), resonance (R), both pitch and resonance (PR), and their direction (+/−). The panels in the three left-most columns are for larger-source deviants (−) and the three right-most are for smaller-source deviants (+). The vertical dashed lines indicate the perceptual onset of the sounds.

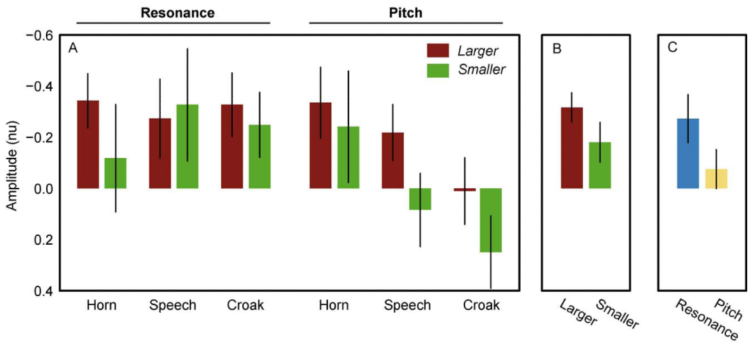

Fig. 4.

MMN amplitudes in normalized units for the interaction of size, acoustic feature and source (A), and for the main effects of size (B) and acoustic feature (C). The vertical bars indicate 95% confidence intervals of the estimated marginal means.

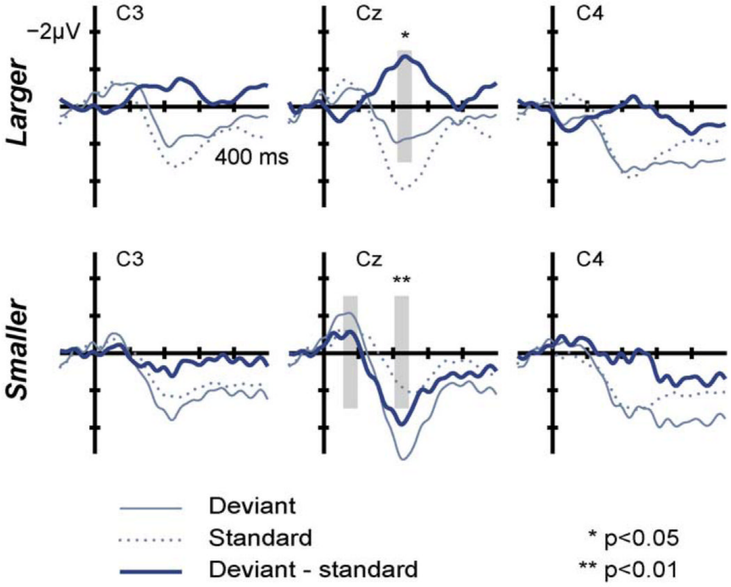

Fig. 5.
Grand-average (13 neonates) ERPs elicited by standard (thin dotted lines) and deviant (thin solid lines) sounds for larger (top row) and smaller (bottom row) resonator-size at three electrode locations (C3, Cz, and C4). Deviant-minus-standard difference ERPs are shown in thick lines. Stimulus onset is at the crossing of the axes. Time windows for the amplitude measurements are indicated by grey shading with asterisks marking significant differences between the standard and deviant responses (*$p < 0.05$; **$p < 0.01$).

Table 1

Configuration of auditory deviant features.

	Pitch		
Resonance		+R	+RP
	−P	standard	+p
	−RP	−R	

Lower (−) pitch and resonance are associated with larger sources, and higher (+) pitch and resonance are associated with smaller sources. Only the congruent double features were used.