

SCIENTIFIC REPORTS



OPEN

Human Brainstem Exhibits higher Sensitivity and Specificity than Auditory-Related Cortex to Short-Term Phonetic Discrimination Learning

Stefan Elmer, Marcela Hausheer, Joëlle Albrecht & Jürg Kühnis

Phonetic discrimination learning is an active perceptual process that operates under the influence of cognitive control mechanisms by increasing the sensitivity of the auditory system to the trained stimulus attributes. It is assumed that the auditory cortex and the brainstem interact in order to refine how sounds are transcribed into neural codes. Here, we evaluated whether these two computational entities are prone to short-term functional changes, whether there is a chronological difference in malleability, and whether short-term training suffices to alter reciprocal interactions. We performed repeated cortical (i.e., mismatch negativity responses, MMN) and subcortical (i.e., frequency-following response, FFR) EEG measurements in two groups of participants who underwent one hour of phonetic discrimination training or were passively exposed to the same stimulus material. The training group showed a distinctive brainstem energy reduction in the trained frequency-range (i.e., first formant), whereas the passive group did not show any response modulation. Notably, brainstem signal change correlated with the behavioral improvement during training, this result indicating a close relationship between behavior and underlying brainstem physiology. Since we did not reveal group differences in MMN responses, results point to specific short-term brainstem changes that precede functional alterations in the auditory cortex.

The phonetic repertoire of a language is constituted by individual vowels and consonants as well as by the permissible combinations of these sounds¹ to form consonant-vowel (CV) syllables. From a psychoacoustic perspective, a difference between the perception of consonants and vowels is that the former is rather dependent on the discrimination of temporal features like, for example, the voice-onset time (i.e., VOT)², whereas vowel's identity is more strongly mediated by frequency information (i.e., first and second formants, F1 and F2)³. The ability to meticulously discriminate spectral and temporal phonetic variations represents a ubiquitous prerequisite for the acquisition and establishment of a variety of language-related functions, including speech processing⁴, reading skills^{5,6}, and foreign language competence^{7,8}. Furthermore, it is noteworthy to mention that vowels are particularly interesting because they occur more often than consonants in any given language, and their repertoire varies quite strongly across languages. Consequently, an accurate perception and discrimination of vowels constitutes one of the most important phonetic constraints for learning a foreign language. The second argument that makes vowels particularly interesting is related to methodological approaches that enable to measure the neural responses to single formants in a highly specific manner^{9,10}. In this context, brainstem responses to harmonics sounds have previously been shown to encode the fundamental frequency (f_0) and the formants of vowels with high fidelity, therefore providing a window into the specificity of brain changes as a function of learning^{11,12}.

During phonetic discrimination learning, acoustic information can be decoded at multiple levels along the auditory pathway, however with remarkable differences in terms of spectral and temporal resolution¹⁰. The brainstem and the auditory cortex constitute the two main computational entities of the auditory system, and exhibit

Auditory Research Group Zurich (ARGZ), Division Neuropsychology, Institute of Psychology, University of Zurich, Zurich, Switzerland. Correspondence and requests for materials should be addressed to S.E. (email: s.elmer@psychologie.uzh.ch)

a complex intertwining of reciprocal bottom-up and top-down projections that operate in a serial and parallel manner in order to refine how sounds are transcribed into neural codes^{10,13}. In this intertwining of processes, the auditory brainstem mimics the spectrotemporal characteristics of an auditory event with remarkable fidelity by interacting with the auditory cortex. The auditory cortex, in turn, integrates the incoming information, enables its transcription into higher cognitive representations, and subordinates the brainstem by providing direct modulatory influence via corticofugal top-down projections¹⁴. Currently, it is generally acknowledged that both the two main actors of the auditory system are highly prone to undergo functional changes^{15–19}. However, until now only a few EEG studies have made use of test-training-retest procedures in order to track causal electrophysiological changes along the auditory pathway that emerge from short- (i.e., <1 hour) or long-term (i.e., >1 hour) phonetic discrimination training. At the cortical level, both short-^{20,21} and long-term^{22–24} training protocols consisting in discriminating phonemes manipulated in formant transitions^{20,22,23} or syllables varying in VOT^{21,24}, have reliably been shown to induce causal changes in the auditory cortex, as reflected by a modification of auditory-evoked potentials' (AEPs) strength. However, the physiological mechanisms underlying short- and long-term training are less clear. In fact, long-lasting training protocols have more often been associated with neural facilitation (i.e., increased AEP amplitudes)^{22–24}, whereas short-term ones have been shown to induce both facilitation²⁰ and adaptation (i.e., reduced amplitudes)^{18,20,21}. The latter discrepancy possibly reflects different neural signatures of the acquisition and consolidation processes²⁵. Finally, it is noteworthy to mention that the molecular, cellular, and physiological mechanisms beyond neural adaptation and facilitation as a function of training are not yet fully understood²⁶. In fact, even though increased or reduced EEG responses have previously been related to the degree of synchronization of cell assemblies²⁷, there are several other explanations that might account for these effects, including synaptic plasticity, short-term depression and facilitation, post-tetanic potentiation and hyperpolarization²⁶, changes in neural tuning²⁸, attention^{29,30}, as well as reward and motivation^{31–33}.

To the best of our knowledge, until now only two EEG studies have examined causal functional changes at the processing level of the brainstem induced by phonetic discrimination training^{11,12}. Russo *et al.*¹¹ as well as Song and colleagues¹² made use of long-term training protocols consisting of learning to distinguish vowels manipulated in terms of pitch, and consistently revealed increased phase-locking to the fundamental frequency (i.e., f_0) of the trained stimulus. Even though these results fundamentally contribute to a better understanding of the subcortical neural computations underlying phonetic discrimination learning, there are nevertheless several open questions that need to be addressed more deeply. First of all, it is unclear whether the human brainstem will show a more robust encoding after a single short-term training session. Second, it is still a matter of debate whether short-term functional changes in the brainstem will be manifested in terms of neural facilitation (i.e., increased amplitudes) or adaptation (i.e., decreased amplitudes)^{28,34}. Moreover, previous EEG studies focusing on vowel discrimination learning at the processing level of the brainstem^{11,12} exclusively focused on pitch (i.e., f_0 manipulations) and not on timbre (i.e., F1 and F2). Finally, it is important to mention that until now there are not so many studies that combined cortical and subcortical measurements^{15,35–37} for attempting to describe putative cortical-subcortical coupling mechanisms between auditory cortex and brainstem while processing vowels or CV syllables, and none of them addressed training-related changes (for an overview also consider^{38,39}).

The present work aimed at contributing to a better understanding of the neural operations underlying short-term phonetic discrimination learning at both the processing level of the brainstem and the auditory cortex. In order to objectify signal-changes at the processing level of the brainstem we made use of FFRs, whereas modulations of the auditory cortex were assessed by means of MMN responses. The FFR is supposed to be generated by the inferior colliculus and the cochlear nucleus¹³, and has previously been shown to be synchronized to the periodicity of the sound with each cycle faithfully representing time-varying f_0 and harmonics (i.e., f)^{9,10}. Furthermore, FFRs are highly replicable across test sessions and sensitive enough to capture subtle brain changes induced by training^{11,12}. By contrast, the MMN is an event-related potential (i.e., ERP) that is elicited by infrequent auditory events deviating in a specific physical dimension from a frequently presented standard stimulus⁴⁰. Since the MMN is commonly elicited during passive listening paradigms, this ERP is supposed to be generated by an automatic stimulus-driven change detection process that is relatively unaffected by attention⁴¹. According to previous studies that combined EEG and fMRI measurements, the auditory-evoked MMN is mainly generated by the auditory cortex^{42,43}.

In the present work, we used EEG and performed repeated cortical and subcortical measurements (i.e., within the same day) in two groups of participants who underwent one hour of training consisting of discriminating CV syllables manipulated in F1 or were passively exposed to the same stimuli while watching a silent movie. The purpose of this study was to evaluate whether short-term speech discrimination training may suffice to induce a modulation of cortical and subcortical brain responses as well as to increase the functional interplay between the two computational entities. In addition, we re-evaluated the influence of short-term phonetic discrimination training on neural facilitation and adaptation.

Results

Psychometric and behavioral data. The two groups did not differ in terms of general cognitive capability (KAI $t_{(22)} = 0.423$, $p = 0.676$; MWT $t_{(22)} = 0.642$, $p = 0.528$) or alertness ($t_{(19)} = 0.617$, $p = 0.545$). Otherwise, the generalized linear mixed model (i.e., 2 groups \times 2 time points) revealed a main effect of time point ($z = -2.391$, $p = 0.0403$) as well as group \times time point interaction effect ($z = 2.016$, $p = 0.0438$). As visible in Fig. 1A, the main effect of time point originated from a better discrimination at T1 (mean correct responses = 63.04%) compared to T0 (mean correct responses = 48%), whereas the group \times time point interaction was related to a higher performance of the TG compared to the PG at T1 (mean correct responses, TG T0 = 45.09%, PG T0 = 50.46%, TG T1 = 74.09%, PG T1 = 53.69%).

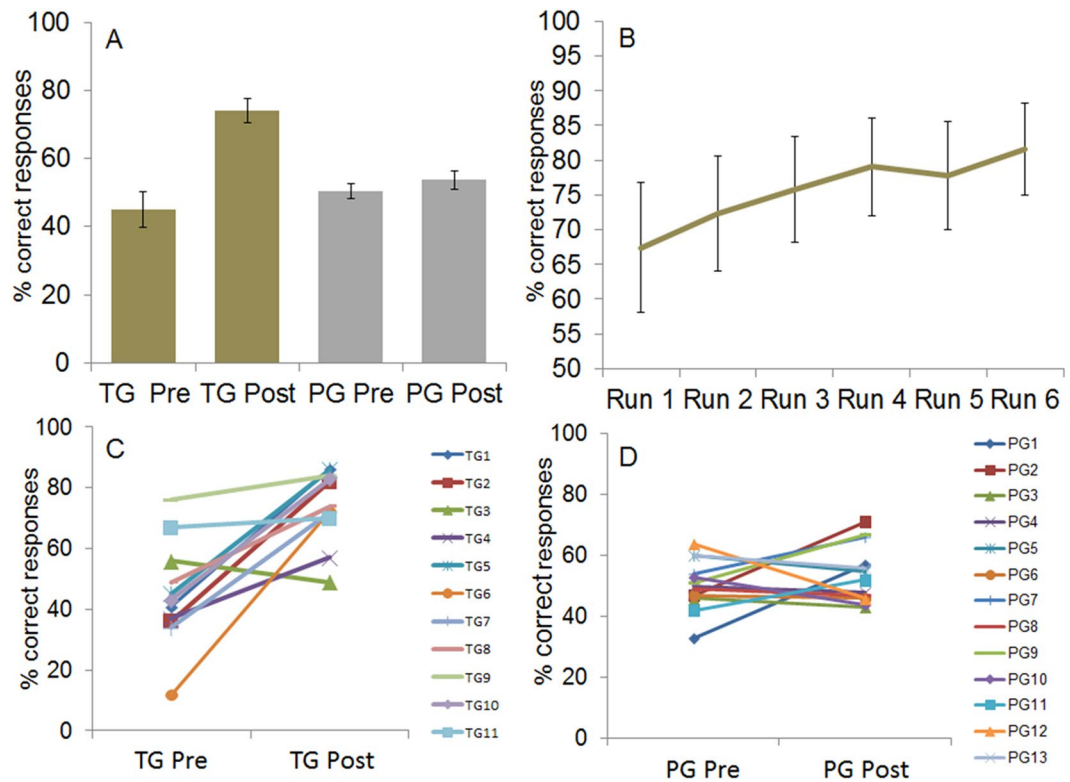


Figure 1. (A) Mean percent correct responses during the discrimination test at T0 (pre) and T1 (post) for the TG (gold bars) and the PG (silver bars). (B) Behavioral performance (mean percent correct responses) of the TG across 6 training runs. (C) Percent correct responses of the single participants of the TG in the discrimination test at T0 (pre) and T1 (post). (D) Percent correct responses of the single participants of the PG in the discrimination test at T0 (pre) and T1 (post). Error bars indicate standard error of the mean.

FFRs responses. A one-sample t-test computed against zero (i.e., no lag) across all participants in order to exclude electromagnetic interference induced by the headphones (Bonferroni corrected p value for two tests = 0.025) yielded significant results at both T0 ($t_{(23)} = 7.825$, $p < 0.001$) and T1 ($t_{(23)} = 8.475$, $p < 0.001$). These results are in line with previous literature¹³ and indicate the presence of genuine FFRs (Fig. 2) characterized by a mean delay of about 8 ms (i.e., T0, mean = 7.825 ms, sd = 1.917 ms; T1, mean = 8.475 ms, sd = 1.468 ms) reflecting signal transfer time from the ear to rostral brainstem structures¹³.

The evaluation of between-group f1 peak amplitudes (Fig. 3) by means of a t-test (i.e., percent signal change) yielded a significant group difference ($t_{(22)} = -2.147$, $p = 0.043$). Post-hoc t-tests against zero calculated separately for the two groups (i.e., Bonferroni corrected p value for two tests = 0.025) revealed that the TG was characterized by a significant signal reduction ($t_{(10)} = -2.704$, $p = 0.022$; mean % signal change = -21.36, neural adaptation, Fig. 4), whereas brain activity did not change within the PG ($t_{(12)} = 0.349$, $p = 0.733$, mean % signal change = 2.74). Finally, even though we did not have any a priori-hypotheses, for reasons of completeness, we also evaluated percent signal change in f_0 (i.e., added responses) and higher harmonics (i.e., subtracted responses, f_2 , f_3 , and f_4) between the two groups. Since we did not reveal group differences in these additional parameters (f_0 , $t_{(22)} = -0.193$, $p = 0.849$; f_2 , $t_{(22)} = -0.881$, $p = 0.388$; f_3 , $t_{(22)} = -0.586$, $p = 0.564$; f_4 , $t_{(22)} = -0.035$, $p = 0.972$), results indicate a specificity of brainstem responses to the trained stimulus attribute (i.e., f_1).

FFR: stimulus-response correlations. Potential group differences in stimulus-response correlations (i.e., stimulus tracking and lag) as a function of treatment were evaluated by contrasting the percent signal change between the two groups by means of t-tests (Bonferroni corrected p value for two tests = 0.025). These analyses did not reveal significant group differences in signal tracking ($t_{(22)} = 0.508$, $p = 0.617$) nor in lag ($t_{(22)} = -0.182$, $p = 0.857$).

FFR: brain-behavior relationships. In order to provide further evidence for the specificity of the functional changes observed within the TG at the processing level of the brainstem, we correlated percent f1 signal change with the learning performance during the training session (i.e., Δ percent correct responses between run 6 and run 1 of the training session). Results revealed a significant negative correlation (i.e., see Fig. 4D) between the two variables ($r = -0.607$, $p = 0.024$, one-tailed).

MMN responses. Between-group differences in MMN area and latency in response to spectral (i.e., early MMN) and temporal (i.e., late MMN) manipulations were evaluated by means of separate t-tests (i.e., percent

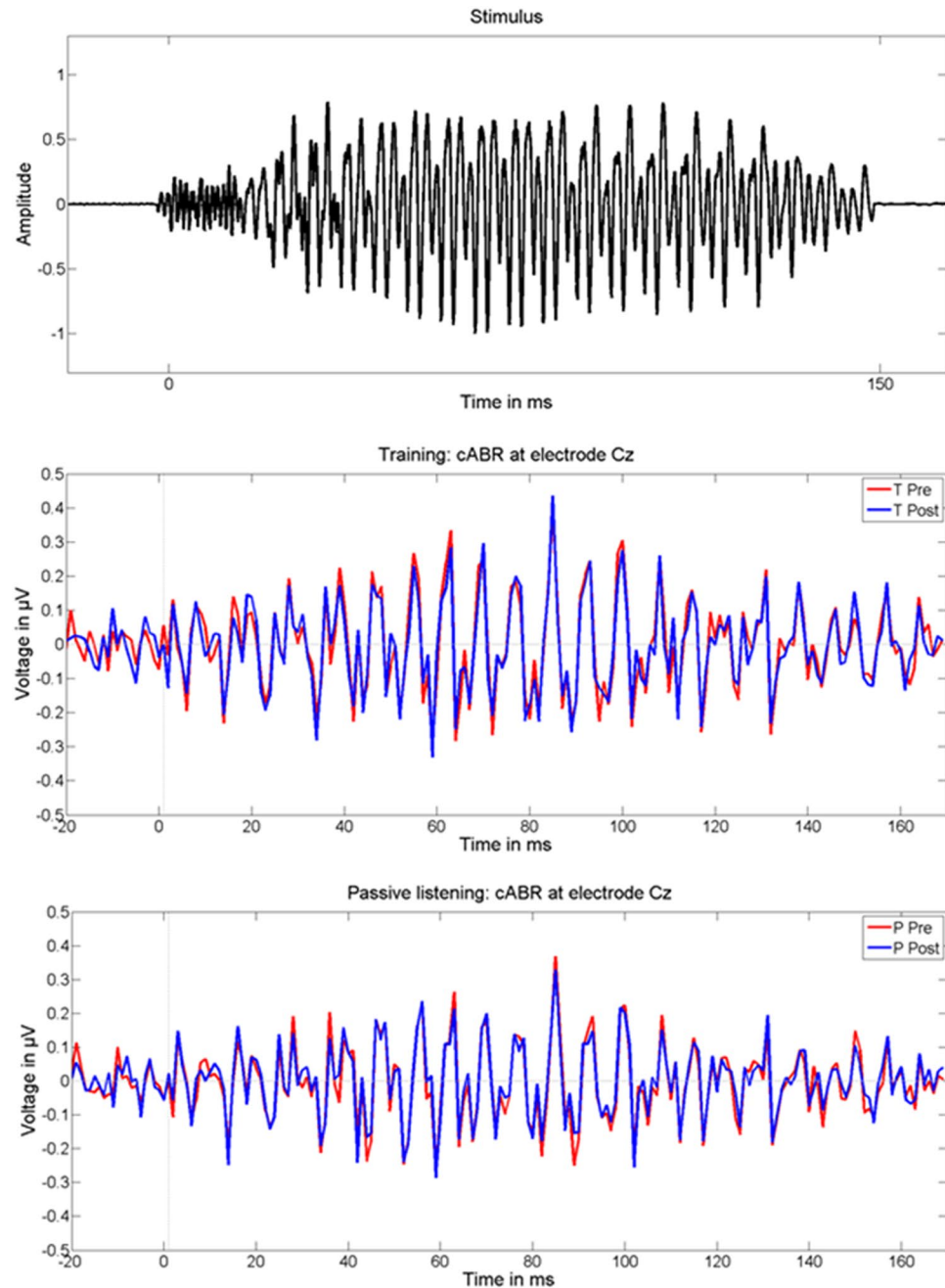


Figure 2. Stimulus waveform and FFRs (subtracted polarities). Top = stimulus waveform; middle = FFRs at T0 (pre, red line) and T1 (post, blue line) within the TG; bottom = FFRs at T0 (pre, red line) and T1 (post, blue line) within the PG.

signal change; Bonferroni corrected p value for 4 tests = 0.0125). These statistical analyses did not reveal significant group differences (spectral area $t_{(22)} = -1.167$, $p = 0.256$; temporal area $t_{(22)} = 1.656$, $p = 0.112$; spectral latency $t_{(22)} = 1.085$, $p = 0.29$; temporal latency $t_{(22)} = -0.514$, $p = 0.613$). Furthermore, in order to rule out the possibility that a general adaptation of the auditory cortex (i.e., see Fig. 5) as a consequence of repeated auditory stimulation between the two measurements points (i.e., T0 and T1) may have accounted for the lack of group differences, we performed additional post-hoc analyses within the two groups (one sample t -test against zero, two-tailed, Bonferroni corrected p value for 4 tests = 0.0125). These supplementary analyses did not reach significance (TG MMN area early, $t_{(10)} = -1.536$, $p = 0.155$; TG MMN area late, $t_{(10)} = -1.493$, $p = 0.166$; PG MMN area early, $t_{(12)} = 1.173$, $p = 0.264$; PG MMN area late, $t_{(12)} = -2.466$, $p = 0.030$).

MMN sources. LORETA source estimation (Table 1) consistently revealed MMN maxima originating from posterior superior temporal areas, irrespective of group affiliation (i.e., TG and PG), time point (i.e., T0 and T1),

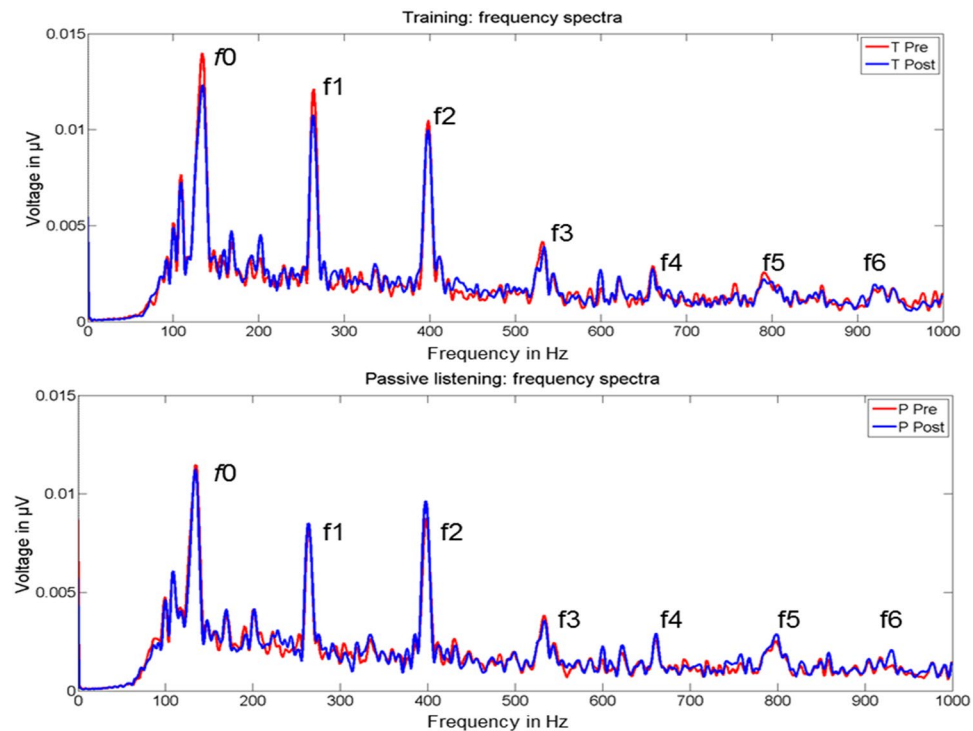


Figure 3. Grand-average FFR power spectra (i.e., subtracted polarities) at T0 (pre, red line) and T1 (post, blue line) of the TG (upper part) and PG (lower part). f_0 = fundamental frequency, f_1 – f_6 = harmonics.

and condition (i.e., spectral and temporal). These findings point to a main contribution of the auditory cortex to MMN responses.

Training-related cortical-subcortical relationships. Putative changes in cortical-subcortical interactions within the TG were evaluated by correlating (according to Pearson's r , two-tailed) the percent signal change of early MMN area and latency with f_1 signal change (Bonferroni corrected p value for two tests = 0.025). These correlational analyses did not reach significance ($r_{\text{MMN area}_f1 \text{ amplitude}} = 0.155$, $p = 0.65$; $r_{\text{MMN latency}_f1 \text{ amplitude}} = 0.355$, $p = 0.285$).

Discussion

General discussion. In the present work, we used a test-training-retest procedure in two groups of participants who performed one hour of phonetic discrimination training, or were passively exposed to the same stimulus material, with the aim to (1) infer putative changes in the brainstem and auditory cortex as a function of short-term training, (2) estimate whether these short-term changes are reflected in neural facilitation or adaptation, (3) and to describe mutual interdependences between auditory cortex and brainstem. Results demonstrated that the brainstem but not the auditory cortex distinctively altered its response properties after short-term training. Most notably, this functional change was manifested in terms of neural adaptation and restricted to the frequency range (i.e., f_1) corresponding to the trained stimulus attribute (i.e., F_1). Since this frequency-specific neural adaptation was negatively correlated with the behavioral improvement of the participants during training, results point to a close relationship (~36% explained variance) between behavior and the underlying brainstem physiology.

Brainstem responses. Nowadays, it is generally acknowledged that the human brainstem constitutes a highly plastic entity¹³ that can alter its response properties as a function of both long-⁸ and short-term training³⁴. For example, Carcagno and Plack⁴⁴ evaluated the FFR before and after ten hours of pitch discrimination training consisting of differentiating complex tones with a static-, raising-, or falling pitch contour, and found a more robust phase locking of the FFR to the static and dynamic f_0 after training. Furthermore, neural activity in the brainstem has previously been shown to be specifically modulated as a function of long-term language experience as reflected by increased f_0 magnitudes in Chinese compared to English speakers in response to iterated rippled noise with Mandarin pitch contours⁴⁵ or high rising Mandarin lexical tones⁴⁶. However, until now, only two EEG studies specifically addressed causal changes in the brainstem induced by speech discrimination training^{11, 12}. In a first study, Russo and co-workers¹¹ reported that after long-term training (i.e., 35–40 sessions of one hour each) children suffering from learning disabilities exhibited brainstem responses that were more resistant to the detrimental effect of background noise than before treatment. Similarly, Song and colleagues¹² demonstrated that native English-speaking participants who learned to incorporate foreign lexical pitch patterns varying in f_0 (i.e., 8 sessions à 30 minutes, accomplished in 14 consecutive days) were characterized by a more faithful representation of f_0 stimulus contour.

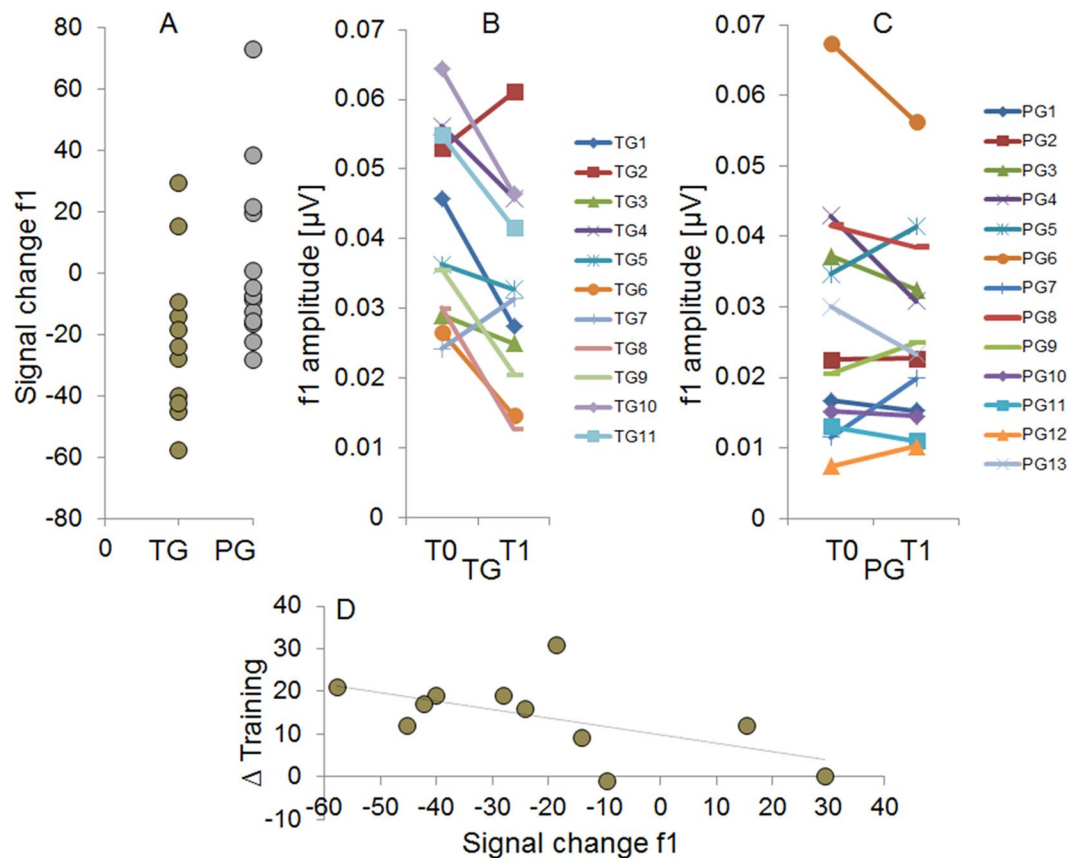


Figure 4. (A) Percent f1 signal change for each subject of the TG (gold) and of the PG (silver). (B) f1 amplitudes at T0 (pre) and T1 (post) of each participant of the TG. (C) f1 amplitudes at T0 (pre) and T1 (post) of each participant of the PG. (D) Significant correlation within the TG between percent f1 signal change and learning performance (Δ Training = percent correct responses run 6 minus run 1).

In the present work, we provide evidence for short-term changes in the human brainstem after only one hour of phonetic discrimination training. However, contrary to previous studies that used professional musicians as a model for long-term training^{8,47}, results revealed functional changes that were manifested in terms of neural adaptation and not facilitation. Interestingly, a similar neural adaptation at the processing level of the brainstem has previously been reported by Slabu and colleagues⁴⁸ in the context of a passive oddball paradigm. Thereby, the authors revealed a reduction of FFRs to deviant stimuli compared to standard ones, leading to suggest that the brainstem is able to encode statistical regularities³⁴ by suppressing responses to rare stimulus events. Even though in the present study the “deviant” stimulus (i.e.,/go/) presented during brainstem measurements occurred with a low probability during the training session, the experimental manipulation we used precludes that results were driven by stimulus statistics³⁴ or even by repetition suppression²⁸. In fact, the PG was passively exposed to the same stimulus material as the TG, however, without showing a modulation of brainstem responses in pre-post comparisons. In addition, since brainstem changes were restricted to the solely discriminative physical attribute enabling to distinguish the trained stimuli, namely F1, results clearly point to feature-specific changes possibly reflecting increased neural efficiency²⁸. This perspective is further supported by the negative correlation we revealed within the TG between percent f1 signal change and behavioral improvement during the training session.

Neural adaptation constitutes an intrinsic organizational property of the auditory system across the entire hierarchical tree, ranging from the periphery to the auditory cortex (for a review consider⁴⁹). In this context, it is noteworthy to mention a previous EEG study targeting at evaluating the encoding of statistical regularities while participants learned to segment complex tone patterns embedded in concatenated sound sequences. Interestingly, the authors revealed decreased brainstem responses to the patterned compared- to a pseudo-random condition after only fifteen minutes of task³⁴. However, by looking at brain responses of the single participants, Skoe and colleagues³⁴ noticed that neural adaptation and facilitation can go hand in hand with remarkable inter-individual differences. Furthermore, the authors revealed a positive relationship between brainstem physiology and behavior, such that better performance was related to greater neural enhancement. Notably, our results are comparable with those of Skoe and colleagues³⁴ in that the TG demonstrated decreased f1 magnitudes after short-term learning compared to the PG. Otherwise, in contrast to Skoe and co-workers, we revealed a negative instead of a positive relationship between the magnitude of brainstem responses and behavioral improvement. From a physiological perspective, the adaptation we revealed at the processing level of the brainstem can be explained at least

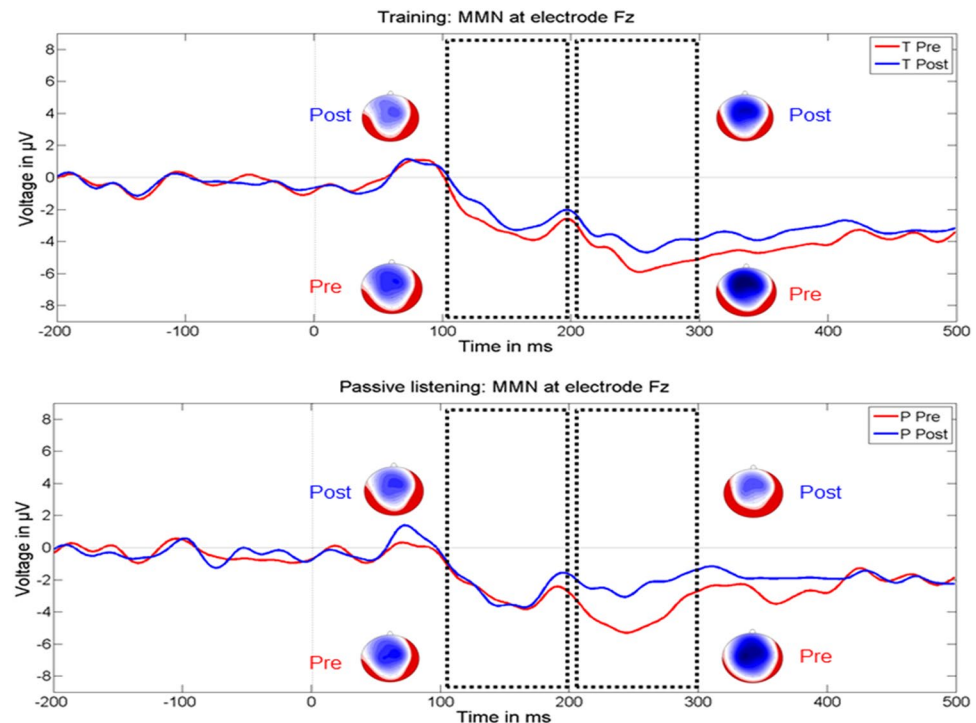


Figure 5. Grand-average MMN responses to spectral (early) and temporal (late) manipulations within the TG (upper part) and the PG (lower part) at T0 (pre, red line) and T1 (post, blue line). The black boxes indicate the time windows of analyses for both the early (100–200 ms) and the late (200–300 ms) MMN with the respective topographies. All waveforms are depicted at electrode FCz.

Group	Early MMN		Late MMN		Talairach Coordinates	Brodmann Area
	Pre	post	Pre	Post		
TG	×				X = -59, Y = -32, Z = 8	BA 42
TG			×		X = -59, Y = -32, Z = 15	BA 42
TG		×			X = -59, Y = -32, Z = 8	BA 42
TG				×	X = -59, Y = -32, Z = 15	BA 42
PG	×				X = -59, Y = -32, Z = 15	BA 42
PG			×		X = -59, Y = -32, Z = 15	BA 42
PG		×			X = -59, Y = -32, Z = 15	BA 42
PG				×	X = -59, Y = -25, Z = 15	BA 40

Table 1. Reconstructed source maxima of the MMN responses in the time windows between 100–200 (spectral manipulation, early MMN) and 200–300 ms (temporal manipulation, late MMN) for the two groups (TG and PG) and the two time points (pre = T0 and post = T1).

by three different phenomena. The first possibility is that short-term training may have altered the response properties of brainstem neurons by uncoupling neural entities that were not relevant for discriminating task-specific acoustic features, resulting in activation of fewer neurons, and consequently neural adaptation²⁸. A second possibility is that the observed brainstem changes may have been indirectly mediated by performance feedback. In fact, since in the present study only the TG received such a feedback, it is thoroughly possible that reward and motivation may have modulated brainstem activity. This perspective is supported by previous work showing that the human reward system is responsive to high-order rewards (i.e., intellectual, artistic, or altruistic pleasures)³¹ and that feedback confirming reward expectation can modulate activity in auditory-related brain regions^{32,33}. Finally, since active learning requires a stronger engagement of attention functions compared to passive listening, we cannot rule out that this variable may have played a role in mediating neural adaptation^{29,30}. Such an influence of attention could, for example, have been mediated by the cortex through corticofugal projections. In fact, such a contribution of the cortex to auditory learning mechanisms via the corticofugal system has previously been demonstrated in animals by using both ablation and pharmacological interventions^{49,50}.

A disadvantage of the EEG technique is that it does not enable to exactly determine the specific origin of the brainstem signal measured. However, currently there is evidence showing that neurons situated in the inferior colliculi are highly frequency-selective^{51,52} as well as sensitive to the direction of frequency modulation^{53,54}. Since

the TG was specifically trained to recognize subtle F1 signal changes only in one direction (i.e., always from/gu/to/go/, in the range between 364–480 Hz), we may speculate whether this specific experimental manipulation may have altered the response properties of neurons being selective to the direction of frequency modulation or rather frequency-selective neurons per se. In addition, since we did not reveal group differences in stimulus-response cross-correlations (i.e., lag and signal tracking), results suggest that during short-term training the brainstem is more likely prone to change its response properties to the spectrum of the trained stimulus attribute than to the waveform periodicity. This result is somehow in opposition with those previously reported by Russo and colleagues¹¹ who revealed an increased temporal alignment of FFRs after training, as reflected by increased quiet-to-noise inter-response correlations. However, in this previous work the authors measured children with learning disabilities that were trained for a much longer period of time (namely 35–40 hours) compared to the present work. The same is true for the work of Anderson and colleagues⁵⁵ where the authors evaluated the impact of an 8 weeks computer-based auditory training program in elderly subjects, and reported earlier brainstem peak latencies in both quiet and noise conditions after treatment. Taken together, these previous results substantiate the suspicion that brainstem changes in timing parameters may necessitate longer training periods.

MMN responses. A further goal of this study was to evaluate the functional malleability of the auditory cortex as indexed by altered MMN responses. In addition, based on previous studies indicating that neuronal entities which are sensitive to temporal and spectral acoustic attributes lie side by side in the auditory cortex³⁶, we evaluated putative transfer effects⁵⁷ from phonetic discrimination training to temporal aspects of speech processing. Reconstructed sources revealed MMN maxima originating from posterior superior temporal areas across groups (i.e., TG and PG), conditions (i.e., spectral and temporal manipulation), and time points (i.e., T0 and T1). This finding is in line with previous literature⁵⁸ and points to a main contribution of the auditory cortex to MMN responses. In the present work, we did not reveal group differences in the modulation of MMN responses (i.e., MMN area and latency) as a function of treatment, leading to suggest that the auditory cortex was not specifically modulated by training. Interestingly, previous training studies consistently revealed increased MMN responses that were accompanied by an improved behavioral performance, however, especially after multiple training sessions lasting several days or weeks^{22, 59, 60}. In particular, Ylinen *et al.*⁶⁰ measured native Finnish (i.e., quantitative language) and English speakers before and after 10 training session of 25 minutes each consisting of learning to discriminate spectral and temporal cues of English vowels. As a main result the authors reported that after training the Finnish speakers were better able to discriminate spectral vowel cues, as reflected by increased MMN responses. In a further EEG study, Tamminen and colleagues⁵⁹ applied a three-day phonetic-listen- and repeat training in a sample of Finnish speakers who learned voicing contrasts in fricative sounds (i.e., fricatives are not differentiated by voicing in Finnish) and revealed significantly increased MMN responses after the second but not the first training day. Taken together, these previous results lead to suggest that functional changes in the auditory cortex can most reliably be induced by multiple training sessions. Therefore, we may speculate whether a consolidation period is necessarily required for inducing detectable plastic changes in the auditory cortex^{61, 62}.

An alternative explanation that may account for the apparent insensitivity of MMN responses to training is that the constant serial order of the cortical and subcortical EEG measurements (i.e., FFRs were always collected first) may possibly have blurred neural facilitation through a superimposed signal adaptation. However, since between the two measurement points (i.e., T0 and T1) the two groups were additionally exposed to acoustic stimulation for one hour, we should have observed such an effect in pre-post comparisons (i.e., a significant percent MMN change against zero), irrespective of group affiliation. Finally, based on the fact that phonetic discrimination learning is an active perceptual process that operates under the influence of attentive functions, future training studies should evaluate short-term changes in the auditory cortex by combining active and passive oddball paradigms.

Cortical-subcortical coupling mechanisms. To the best of our knowledge, until now only four studies conjointly recorded FFRs and AEPs while participants were exposed to CV syllables^{36, 37} or vowels^{15, 35}. In particular, Musacchia and colleagues³⁶ measured musicians and non-musicians while participants were repeatedly exposed to the syllable/da/, and reported a positive relationship between subcortical f_0 amplitude and cortical P1-to-N1 slope. Otherwise, Bidelman and colleagues³⁵ measured young and older adults while the participants categorized vowels that spanned a perceptual continuum from/u/to/a/and revealed that older adults were characterized by slower and more variable speech classification performance than younger listeners. This differential behavioral performance was reflected by reduced brainstem amplitudes, increased cortical AEPs, as well as by a negative relationship between f_1 and cortical N1/P2 amplitudes. In a second study of the same group¹⁵, the authors recorded cortical and subcortical brain responses in older adults with and without music training while the participants categorized vowels along a continuum. Even though the authors did not find between-group differences in terms of cortical (i.e., P1-N1-P2 complex) or subcortical (i.e., f_0 amplitude) brain responses, musicians showed a closer relationship between neural activity and behavioral performance. Finally, Parbery-Clark *et al.*³⁷ investigated the effect of background noise on both brainstem and auditory cortex activity, and reported a relationship between subcortical response fidelity and cortical N1 magnitude that was predictive of speech-in-noise perception. In the present work, we did not find evidence for a relationship between auditory cortex and brainstem changes as a function of training. However, this may rather be a byproduct of unmodulated MMN responses as a function of training rather than an evidence for the inexistence of cortical-subcortical coupling mechanisms. In this context, it is also important to mention that our experimental design profoundly differed from the previous studies mentioned above. In fact, Musacchia and colleagues³⁶ as well as Bidelman *et al.*¹⁵ measured musicians, a specific group of subjects that has previously repeatedly been shown to constitute a suitable model for evaluating the influence of long-term training on auditory processing^{16, 63, 64}. Otherwise, the group of Parbery-Clark³⁷ evaluated cortical-subcortical coupling mechanisms in normal hearing young adults while

performing a speech-in-noise perception task, an experimental condition which is well known to place stronger demands on cognitive control mechanisms that have a modulatory influence on brainstem activity through the corticofugal system³⁴.

Conclusions

In summary, our results highlight causal and feature-specific changes in the human brainstem after only one hour of phonetic discrimination training, as reflected by neural adaptation in the frequency-range (i.e., f_1) corresponding to the trained acoustic feature (i.e., F_1). Since these brainstem changes correlated with the behavioral improvement of the participants during the training session, results are interpreted as reflecting neural efficiency induced by short-term phonetic discrimination training.

Materials and Methods

Participants. We evaluated the EEG data of two groups of subjects who were repeatedly measured within the same day (i.e., at time point 0 (T0, pre) and time point 1 (T1, post)) by using EEG protocols that enable the collection of both cortical and subcortical brain responses. Between the two measurement points, one group underwent active phonetic discrimination training (i.e., training group, TG, 11 subjects, 3 men, mean age = 23.54 years, SD = 3.04 years), whereas the second one was passively exposed to the same stimulus material while watching a silent movie (passive group, PG, 13 subjects, 3 men, mean age = 23.84 years, SD = 2.44 years). All participants were in the age range of 20–30 years, of German mother tongue, non-bilinguals (i.e., did not grow up with more than one language before school), non-musicians, and consistently right-handed⁶⁵. None of the participants reported a history of neurological, psychiatric or audiological disorders. In addition, all participants were tested with pure-tone audiometry (MAICO Diagnostic GmbH, Berlin) in the frequency-range of 250–8000 Hz (MAICO Diagnostic GmbH, Berlin). According to this procedure, all participants demonstrated an unremarkable audiological status (i.e., all tested frequencies could be heard below a threshold of 30 dB). The participants were paid for participation, the local ethics committee (i.e., Kantonale Ethikkommission Zurich) approved the study (in accordance with the Helsinki declaration), and written informed consent was obtained from all participants.

Cognitive capability. In order to test for group differences in cognitive capability, each participant performed two German intelligence tests, namely the MWT-B and the KAI (MWT-B, Mehrfachwahl-Wortschatz Intelligenz Test; KAI, Kurzttest für allgemeine Basisgrößen der Informationsverarbeitung). The MWT estimates crystallized intelligence, and has previously been shown to correlate fairly well ($r = 0.72$) with the global intelligence quotient in healthy adults⁶⁶. This specific test consists of 37 items which are ordered as a function of difficulty level. For each item, the participants have to choose the unique word with a meaning out of five pseudo-words. By contrast, the KAI estimates fluid intelligence, and is based on short-term memory (i.e., number- and digit span forward) and speed of information processing (i.e., reading aloud rows of random letters as fast as possible). Finally, tonic arousal, a variable which is known to have an influence on learning mechanisms, was assessed by using a subtest of the TAP test battery (Testbatterie zur Aufmerksamkeitsprüfung). During this test, participants were instructed to react as fast as possible whenever a white cross appeared randomly on a black screen. Due to a bug in the software, three participants (i.e., one of the TG and two of the PG) could not be tested on this task.

Stimulus material. The auditory stimuli consisted of two semi-artificial German CV syllables, namely/gu/ and/go/, which were created using PRAAT. The original syllable/go/was spoken by a male speaker, and recorded at a sampling rate of 44.1 kHz. In a first processing step, the consonant/g/was separated from the vowel/o/by identifying the time period between consonant burst onset and the onset of periodic oscillation taken to indicate vocal fold vibration. Afterwards, the vowel was replaced by fully artificial ones with identical fundamental frequency (i.e., $f_0 = 130$ Hz) and second formant (i.e., $F_2 = 860$ Hz) but different first formant (i.e., F_1) (i.e., $F_1/u = 364$ Hz, $F_1/o = 480$ Hz). The artificial vowels were inserted at the same temporal location as the original ones. In a successive step, pitches and amplitudes of the original vowels were convolved to the semi-artificial ones, resulting in a syllable duration of 153 ms. This procedure is particularly fruitful in that it enables to produce semi-artificial CV syllables with fully controlled physical attributes, such as duration, pitch, timbre, and harmonics⁶⁷.

Phonetic discrimination training. In the time period between repeated EEG measurements (i.e., at T0 and T1, see the experimental procedure), the TG performed one hour of phonetic discrimination training consisting in judging whether pairs of CV syllables (i.e., a continuum between/gu/and/go/) were acoustically identical or not by pressing the corresponding mouse buttons (i.e., two alternatives forced-choice task with emphasis on accuracy and not on speed). Thereby, participants received visual trial-by-trial feedback (i.e., red or green circle presented in the middle of the screen) as well as cumulative feedback during each block (i.e., % correct and incorrect responses, red and green bars at the left and right side of the screen). The two original CV syllables/gu/(i.e., $f_0 = 130$ Hz, $F_1 = 364$ Hz, and $F_2 = 860$ Hz) and/go/(i.e., $f_0 = 130$ Hz, $F_1 = 480$ Hz, and $F_2 = 860$ Hz) were further manipulated by shifting F_1 in steps of 4-Hz between 360–480 Hz, resulting in 30 acoustically different stimuli. All stimuli had a duration of 153 ms, the SOA of the syllable pairs was of 700 ms, and ITI corresponded to 1300 ms. During the training session, the stimuli were presented pairwise in a fully randomized order (i.e., 6 blocks of 10 minutes each). During every block each of the 30 stimuli was presented 8 times, resulting in 240 stimulus pairs per block. The first CV syllable of the pairs was always/gu/(i.e., $f_0 = 130$ Hz, $F_1 = 364$ Hz, and $F_2 = 860$ Hz), whereas the second one was one out of the 30 variations.

Discrimination test. In order to compare phonetic discrimination performance across the two groups (i.e., TG and PG) before and after training (i.e., training or passive exposure), all subjects performed a short phonetic discrimination test at T0 and T1 (i.e., before EEG at T0 and after EEG at T1) consisting of judging whether pairs

of CV syllables are identical or not. Thereby, subjects heard exactly the same CV syllables (i.e., see next paragraph) that were presented during the training- (i.e., TG) and passive exposure (i.e., PG) sessions (i.e., continuum between /gu/ and /go/, the first syllable was always /gu/). The sole difference is that here we used a reduced pool of stimuli (i.e., in the F1 range between 364–476 Hz, steps of 8 Hz, totally 15 acoustically different CV syllables, duration = 153 ms) that were presented in a randomized order (i.e., SOA = 700 ms, ITI = 1300 ms). Each stimulus was presented 4 times, resulting in a total of 60 trials. The phonetic discrimination test had a duration of about 2 minutes and was evaluated according to mean percent correct trials at T0 and T1.

Experimental procedure. The volunteers were randomly assigned to two groups, namely to the TG or PG. Prior to the EEG session, all subjects performed the psychometric tests as well as pure tone audiometry in order to exclude any hearing problems. Afterwards, participants underwent the phonetic discrimination test (i.e., for quantifying phonetic discrimination at T0 and T1) and started with EEG measurements (i.e., T0, FFR followed by MMN). The stimuli were delivered via headphones (Sennheiser, CX-350, Colchester, Essex, UK) while watching a silent movie. Subsequently, the TG performed one hour of phonetic discrimination training, whereas the PG was passively exposed to the same stimulus material while watching a silent movie. At the end of the treatment, the two groups underwent the second EEG session (i.e., T1, FFR followed by MMN) and accomplished the second part of the phonetic discrimination test. The entire experiment lasted about four hours.

EEG data acquisition. Continuous EEG (i.e., 32 electrodes + 2 eye channels, provided by Easy Cap, forehead ground) was recorded with a sampling rate of 5 kHz and a high pass filter of 0.1 Hz by using an EEG-amplifier (Brainproducts, Munich, Germany). This specific device has previously been shown to reliably enable the collection of both cortical and subcortical brain responses⁶⁸. The electrodes (i.e., sintered silver/silver-chloride) were located at frontal, temporal, parietal and occipital scalp sites according to the international 10–10 system. Data were collected by using linked earlobes- (i.e., FFR) or nose (i.e., MMN) references, and electrode impedance was reduced to <5 k Ω by using electrogel conductant. For all pre-processing steps, we used the Brain Vision Analyzer software package (Version 2.01, Brainproducts, Munich, Germany) and MATLAB (version 2013b). Stimulus presentation and the collection of behavioral responses were controlled by the “Presentation” software (Neurobehavioral Systems, Albany, California).

FFR: data acquisition and processing. FFTs were evaluated in response to 3000 CV syllables (i.e., /go/, $f_0 = 130$ Hz, $F_1 = 480$ Hz, $F_2 = 860$ Hz, duration = 153 ms, SPL = 85 dB, SOA = 217 ms) of each polarity (i.e., a total of 6000 presentations). During EEG measurements, the audio waveform was recorded as an additional EEG channel, and triggers were recomputed offline by using thresholding functions in MATLAB (version 2013b). The data were filtered offline between 100–1000 Hz (i.e., butterworth filter, 48 dB/oct), and artefacts exceeding ± 50 μ V were automatically rejected. Furthermore, responses were segmented into single sweeps of 173 ms (including a pre-stimulus baseline of 20 ms), baseline corrected, and averaged separately for each polarity. In order to bias higher-frequency components by maximizing the spectral response, waveforms to positive and negative polarities were subtracted¹⁰, and peak amplitudes of the harmonics were extracted for each participant by applying fast Fourier transformation (i.e., FFT). FFRs were computed over the steady-state portion of the response and for each participant spectral response amplitudes at electrode Cz were calculated over 1 Hz-wide bins surrounding f_0 and harmonics. In addition, in order to demonstrate that the FFRs were not an artefact of electromagnetic interference induced by the headphones, we performed stimulus-response correlations (i.e., cross-correlation over all sample points) and expected a lag in the range of 6–10 ms¹³. Stimulus-response cross-correlations were also computed to evaluate training-related changes (i.e., training or passive exposure) in response fidelity to the stimulus periodicity (i.e., maximal correlation between the two signals). The collection of FFRs had a duration of about 21 minutes.

MMN: data acquisition and processing. MMN responses were collected in order to test the hypothesis that short-term phonetic discrimination training may have an influence on the response properties of the auditory cortex. The same paradigm was also used to address a second more speculative research question, namely the putative influence of phonetic discrimination training on the processing of temporal speech information. Such a relationship was tackled based on previous work indicating a high degree of interaction between spectral and temporal parameters in the auditory-cortex⁶⁹. Accordingly, we used a double-deviant MMN paradigm⁷⁰ consisting of simultaneously varying F1 and duration of the deviants. Previous work has shown that double-deviant stimuli elicit a MMN with two peaks that correspond to the MMNs elicited by the two single deviations presented in isolation⁷⁰.

The stimulus material consisted of 840 standards (i.e., /gu/, duration = 153 ms, $f_0 = 130$ Hz, $F_1 = 364$ Hz, $F_2 = 860$ Hz) and 120 deviants (i.e., /go/, duration 153 ms, $f_0 = 130$ Hz, $F_1 = 480$ Hz, $F_2 = 860$ Hz) which were presented with a presentation level of 70 dB SPL and a SOA of 700 ms. Stimuli were presented in a pseudorandomized order, whereby the deviant syllable was followed by at least one standard. EEG data were filtered offline between 1–30 Hz, and artefacts (i.e., eye movements and blinks) were eliminated by using an independent component analysis (i.e., ICA)⁷¹ in association with an automatic raw data inspection (± 100 μ V). Afterwards, the data were segmented separately for standards and deviants into single sweeps of 700 ms, including a pre-stimulus baseline of 200 ms. The single sweeps were baseline corrected, and single-subject averages were computed separately for standards and deviants. MMN waveforms were calculated by subtracting AEPs in response to standards from those elicited by deviants. Based on grand average waveforms, single-subjects inspection, and topographies, MMN areas and latencies were extracted in the time range of 100–200 (i.e., MMN in response to spectral manipulation) and 200–300 (i.e., MMN in response to temporal manipulation) ms post stimulus onset at electrode FCz (i.e., showing maximal amplitudes). The MMN paradigm had a duration of about 11 minutes.

MMN: Source estimation. In order to corroborate that MMN responses essentially originated from the auditory cortex⁵⁸, intra-cortical maxima were evaluated by using a source estimation approach (i.e., LORETA, ref. 72). Thereby, we estimated the cortical origin of MMN responses separately for the two groups (i.e., TG and PG), the two time points (i.e., T0 and T1), and spectral and temporal manipulations in the time range between 100–200 (i.e., spectral manipulation) and 200–300 ms (i.e., temporal manipulation). This approach, unlike conventional dipole fitting, does not require a-priori assumptions about the number and the localization of the dipoles. LORETA calculates the three dimensional distribution of electrically active neuronal generators in the brain as a current density value (i.e., $\infty\mu\text{A}/\text{mm}^2$), and provides a solution for the inverse problem by assuming that the smoothest of all possible activity distributions is the most plausible one for explaining the data. The characteristic feature of this particular inverse solution approach is the low spatial resolution which conserves the location of maximal activity but with a certain degree of dispersion⁷³. Here, we determined the current density distribution for epochs of brain electrical activity on a dense grid of 2394 voxels at 7 mm spatial resolution. The localization error of LORETA's source identification may vary between 7⁷² and 14 mm⁷⁴. LORETA refers to a three-shell spherical model registered to the Talairach human brain atlas and source estimations are provided as x, y, z coordinates situated relative to the inter-commissural line (AC-PC line) in horizontal (x), the anterior/posterior (y), and vertical (z) directions. The solution space is confined to the grey matter portion of the human cortex, which rules out the option that subcortical tissue and white matter contribute to the solution.

Statistical analyses. All statistical analyses of the EEG data (i.e., group comparisons and correlations, FFRs and MMN responses) were performed by using normalized percent signal change values, according to the following formula: % signal change = [(T1 value – T0 value)/T0 value] × 100. This procedure is particularly fruitful in that it enables to control for inter-individual variability of cortical⁷⁵ and subcortical brain responses³⁴. Psychometric- and electrophysiological data (i.e., FFR spectral f1 peak, MMN area, and MMN latency) were evaluated by using t-tests for independent samples (two-tailed), whereas the behavioral data of the discrimination test were evaluated according to a generalized linear mixed model for binomially distributed outcome with group (TG and PG) and time point (T0 and T1) as fixed factors, and participants as random factors. Furthermore, in order to exclude that FFRs were a simply artefact of electromagnetic interference induced by the headphones, the lag of the FFRs at T0 was tested against zero. Finally, within the TG putative relationships between cortical and subcortical responses as well as between subcortical signal change and learning performance (i.e., Δ percent correct responses between run 6 and run 1 of the training session) were assessed by using Pearson's correlation. Based on the results clearly showing neural adaptation at the processing level of the brainstem after training, the relationship between training-related subcortical brain changes and learning improvement was tested in a one-tailed fashion. Otherwise, since we did not have clear a-priori hypotheses about the direction of cortical-subcortical coupling mechanisms, the correlation between auditory cortex and brainstem signal change was tested in a two-tailed manner. All omnibus tests as well as post-hoc analyses were corrected for multiple comparisons by using the Bonferroni procedure.

References

- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S. & Dupoux, E. Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience* **23**, 9541–9546 (2003).
- Jancke, L., Wustenberg, T., Scheich, H. & Heinze, H. J. Phonetic perception and the temporal cortex. *Neuroimage* **15**, 733–746, doi:10.1006/nimg.2001.1027 (2002).
- Lindblom, B. E. & Studdert-Kennedy, M. On the role of formant transitions in vowel recognition. *J Acoust Soc Am* **42**, 830–843 (1967).
- Bellis, T. J., Nicol, T. & Kraus, N. Aging affects hemispheric asymmetry in the neural representation of speech sounds. *Journal of Neuroscience* **20**, 791–797 (2000).
- Goswami, U. The neural basis of dyslexia may originate in primary auditory cortex. *Brain* **137**, 3100–3102 (2014).
- Goswami, U., Power, A. J., Lallier, M. & Facoetti, A. Oscillatory “temporal sampling” and developmental dyslexia: toward an overarching theoretical framework. *Frontiers in Human Neuroscience* **8** (2014).
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O. & Besson, M. Influence of Musical Expertise on Segmental and Tonal Processing in Mandarin Chinese. *Journal of Cognitive Neuroscience* **23**, 2701–2715 (2011).
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T. & Kraus, N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience* **10**, 420–422 (2007).
- Johnson, K. L., Nicol, T. G. & Kraus, N. Brain stem response to speech: A biological marker of auditory processing. *Ear and Hearing* **26**, 424–434 (2005).
- Skoe, E. & Kraus, N. Auditory Brain Stem Response to Complex Sounds: A Tutorial. *Ear and Hearing* **31**, 302–324 (2010).
- Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A. & Kraus, N. Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research* **156**, 95–103 (2005).
- Song, J. H., Skoe, E., Wong, P. C. M. & Kraus, N. Plasticity in the adult human auditory Brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience* **20**, 1892–1902 (2008).
- Chandrasekaran, B. & Kraus, N. The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology* **47**, 236–246 (2010).
- Luo, F., Wang, Q. Z., Kashani, A. & Yan, J. Corticofugal Modulation of Initial Sound Processing in the Brain. *Journal of Neuroscience* **28**, 11615–11621, doi:10.1523/Jneurosci.3972-08.2008 (2008).
- Bidelman, G. M., Weiss, M. W., Moreno, S. & Alain, C. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience* **40**, 2662–2673 (2014).
- Elmer, S., Meyer, M. & Jancke, L. Neurofunctional and Behavioral Correlates of Phonetic and Temporal Categorization in Musically Trained and Untrained Subjects. *Cerebral Cortex* **22**, 650–658 (2012).
- Intartaglia, B. *et al.* Native language shapes automatic neural processing of speech. *Neuropsychologia* **89**, 57–65, doi:10.1016/j.neuropsychologia.2016.05.033 (2016).
- Seppanen, M., Hamalainen, J., Pesonen, A. K. & Tervaniemi, M. Music training enhances rapid neural plasticity of N1 and P2 source activation for unattended sounds. *Frontiers in Human Neuroscience* **6** (2012).
- White-Schwoch, T., Carr, K. W., Anderson, S., Strait, D. L. & Kraus, N. Older Adults Benefit from Music Training Early in Life: Biological Evidence for Long-Term Training-Driven Plasticity. *Journal of Neuroscience* **33**, 17667–17674 (2013).

20. Alain, C. & Snyder, J. S. Age-related differences in auditory evoked responses during rapid perceptual learning. *Clin Neurophysiol* **119**, 356–366 (2008).
21. Ben David, B. M., Campeanu, S., Tremblay, K. L. & Alain, C. Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. *Psychophysiology* **48**, 797–807 (2011).
22. Kraus, N. *et al.* Central Auditory-System Plasticity Associated with Speech-Discrimination Training. *Journal of Cognitive Neuroscience* **7**, 25–32 (1995).
23. Reinke, K. S., He, Y., Wang, C. H. & Alain, C. Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research* **17**, 781–791 (2003).
24. Tremblay, K. L., Shahin, A. J., Picton, T. & Ross, B. Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clin Neurophysiol* **120**, 128–135 (2009).
25. Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K. & Collet, G. Is the auditory evoked P2 response a biomarker of learning? *Frontiers in systems neuroscience* **8**, 1–13 (2014).
26. Friauf, E., Fischer, A. U. & Fuhr, M. F. Synaptic plasticity in the auditory system: a review. *Cell Tissue Res* **361**, 177–213, doi:[10.1007/s00441-015-2176-x](https://doi.org/10.1007/s00441-015-2176-x) (2015).
27. Jackson, A. F. & Bolger, D. J. The neurophysiological bases of EEG and EEG measurement: a review for the rest of us. *Psychophysiology* **51**, 1061–1071, doi:[10.1111/psyp.12283](https://doi.org/10.1111/psyp.12283) (2014).
28. Chandrasekaran, B., Kraus, N. & Wong, P. C. M. Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology* **107**, 1325–1336 (2012).
29. Delano, P. H., Elgueda, D., Hamame, C. M. & Robles, L. Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. *Journal of Neuroscience* **27**, 4146–4153, doi:[10.1523/Jneurosci.3702-06.2007](https://doi.org/10.1523/Jneurosci.3702-06.2007) (2007).
30. Srinivasan, S., Keil, A., Stratis, K., Carr, K. L. W. & Smith, D. W. Effects of Cross-Modal Selective Attention on the Sensory Periphery: Cochlear Sensitivity Is Altered by Selective Attention. *Neuroscience* **223**, 325–332, doi:[10.1016/j.neuroscience.2012.07.062](https://doi.org/10.1016/j.neuroscience.2012.07.062) (2012).
31. Berridge, K. C. & Kringelbach, M. L. Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology (Berl)* **199**, 457–480, doi:[10.1007/s00213-008-1099-6](https://doi.org/10.1007/s00213-008-1099-6) (2008).
32. Ripolles, P. *et al.* The Role of Reward in Word Learning and Its Implications for Language Acquisition. *Current Biology* **24**, 2606–2611, doi:[10.1016/j.cub.2014.09.044](https://doi.org/10.1016/j.cub.2014.09.044) (2014).
33. Weis, T., Brechmann, A., Puschmann, S. & Thiel, C. M. Feedback that confirms reward expectation triggers auditory cortex activity. *Journal of Neurophysiology* **110**, 1860–1868, doi:[10.1152/jn.00128.2013](https://doi.org/10.1152/jn.00128.2013) (2013).
34. Skoe, E., Krizman, J., Spitzer, E. & Kraus, N. The Auditory Brainstem Is a Barometer of Rapid Auditory Learning. *Neuroscience* **243**, 104–114, doi:[10.1016/j.neuroscience.2013.03.009](https://doi.org/10.1016/j.neuroscience.2013.03.009) (2013).
35. Bidelman, G. M., Villafuerte, J. W., Moreno, S. & Alain, C. Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiology of Aging* **35**, 2526–2540, doi:[10.1016/j.neurobiolaging.2014.05.006](https://doi.org/10.1016/j.neurobiolaging.2014.05.006) (2014).
36. Musacchia, G., Strait, D. & Kraus, N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Res* **241**, 34–42 (2008).
37. Parbery-Clark, A., Marmel, F., Bair, J. & Kraus, N. What subcortical-cortical relationships tell us about processing speech in noise. *European Journal of Neuroscience* **33**, 549–557 (2011).
38. Picton, T. Between Reality and Prophecy - a Citation Classic Commentary on Human Auditory Evoked-Potentials 1. Evaluation of Components 2. Effects of Attention by Picton, T. W., Hillyard, S. A., Krausz, H. I., and Galambos, R. *Cc/Life Sci* **25**–25 (1990).
39. Picton, T. W., Hillyard, S. A., Galambos, R. & Schiff, M. Human auditory attention: a central or peripheral process? *Science* **173**, 351–353 (1971).
40. Naatanen, R. & Picton, T. The N1 Wave of the Human Electric and Magnetic Response to Sound - A Review and An Analysis of the Component Structure. *Psychophysiology* **24**, 375–425 (1987).
41. Naatanen, R., Paavilainen, P., Rinne, T. & Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin Neurophysiol* **118**, 2544–2590, doi:[10.1016/j.clinph.2007.04.026](https://doi.org/10.1016/j.clinph.2007.04.026) (2007).
42. Alho, K., Rinne, T., Herron, T. J. & Woods, D. L. Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. *Hearing Res* **307**, 29–41, doi:[10.1016/j.heares.2013.08.001](https://doi.org/10.1016/j.heares.2013.08.001) (2014).
43. Szyck, G. R., Stadler, J., Brechmann, A. & Munte, T. F. Preattentive Mechanisms of Change Detection in Early Auditory Cortex: A 7 Tesla Fmri Study. *Neuroscience* **253**, 100–109, doi:[10.1016/j.neuroscience.2013.08.039](https://doi.org/10.1016/j.neuroscience.2013.08.039) (2013).
44. Carcagno, S. & Plack, C. J. Subcortical plasticity following perceptual learning in a pitch discrimination task. *J Assoc Res Otolaryngol* **12**, 89–100, doi:[10.1007/s10162-010-0236-1](https://doi.org/10.1007/s10162-010-0236-1) (2011).
45. Krishnan, A., Gandour, J. T. & Suresh, C. H. Language-experience plasticity in neural representation of changes in pitch salience. *Brain Res* **1637**, 102–117, doi:[10.1016/j.brainres.2016.02.021](https://doi.org/10.1016/j.brainres.2016.02.021) (2016).
46. Krishnan, A., Suresh, C. H. & Gandour, J. T. Changes in Pitch Height Elicit Both Language-Universal and Language-Dependent Changes in Neural Representation of Pitch in the Brainstem and Auditory Cortex. *Neuroscience* **346**, 52–63, doi:[10.1016/j.neuroscience.2017.01.013](https://doi.org/10.1016/j.neuroscience.2017.01.013) (2017).
47. Bidelman, G. M., Gandour, J. T. & Krishnan, A. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition* **77**, 1–10, doi:[10.1016/j.bandc.2011.07.006](https://doi.org/10.1016/j.bandc.2011.07.006) (2011).
48. Slabu, L., Grimm, S. & Escera, C. Novelty Detection in the Human Auditory Brainstem. *Journal of Neuroscience* **32**, 1447–1452, doi:[10.1523/Jneurosci.2557-11.2012](https://doi.org/10.1523/Jneurosci.2557-11.2012) (2012).
49. Perez-Gonzalez, D. & Malmierca, M. S. Adaptation in the auditory system: an overview. *Front Integr Neurosci* **8**, 19, doi:[10.3389/fnint.2014.00019](https://doi.org/10.3389/fnint.2014.00019) (2014).
50. Bajo, V. M., Nodal, F. R., Moore, D. R. & King, A. J. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat Neurosci* **13**, 253–260, doi:[10.1038/nn.2466](https://doi.org/10.1038/nn.2466) (2010).
51. Andoni, S., Li, N. & Pollak, G. D. Spectrotemporal receptive fields in the inferior colliculus revealing selectivity for spectral motion in conspecific vocalizations. *Journal of Neuroscience* **27**, 4882–4893, doi:[10.1523/Jneurosci.4342-06.2007](https://doi.org/10.1523/Jneurosci.4342-06.2007) (2007).
52. Brimjoin, W. O. & O'Neill, W. E. On the prediction of sweep rate and directional selectivity for FM sounds from two-tone interactions in the inferior colliculus. *Hearing Res* **210**, 63–79, doi:[10.1016/j.heares.2005.07.005](https://doi.org/10.1016/j.heares.2005.07.005) (2005).
53. Williams, A. J. & Fuzessery, Z. M. Facilitatory Mechanisms Shape Selectivity for the Rate and Direction of FM Sweeps in the Inferior Colliculus of the Pallid Bat. *Journal of Neurophysiology* **104**, 1456–1471, doi:[10.1152/jn.00598.2009](https://doi.org/10.1152/jn.00598.2009) (2010).
54. Zhang, L. I., Tan, A. Y., Schreiner, C. E. & Merzenich, M. M. Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature* **424**, 201–205, doi:[10.1038/nature01796](https://doi.org/10.1038/nature01796) (2003).
55. Anderson, S., White-Schwoch, T., Parbery-Clark, A. & Kraus, N. Reversal of age-related neural timing delays with training. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 4357–4362 (2013).
56. Santoro, R. *et al.* Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. *Plos Computational Biology* **10**, doi:[10.1371/journal.pcbi.1003412](https://doi.org/10.1371/journal.pcbi.1003412) (2014).
57. Tremblay, K., Kraus, N., Carrell, T. D. & McGee, T. Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America* **102**, 3762–3773, doi:[10.1121/1.420139](https://doi.org/10.1121/1.420139) (1997).
58. Alho, K. Cerebral Generators of Mismatch Negativity (Mmn) and Its Magnetic Counterpart (Mmm) Elicited by Sound Changes. *Ear and Hearing* **16**, 38–51, doi:[10.1097/00003446-199502000-00004](https://doi.org/10.1097/00003446-199502000-00004) (1995).

59. Tamminen, H., Peltola, M. S., Kujala, T. & Naatanen, R. Phonetic training and non-native speech perception - New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *Int J Psychophysiol* **97**, 23–29, doi:10.1016/j.ijpsycho.2015.04.020 (2015).
60. Ylinen, S. *et al.* Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English. *Journal of Cognitive Neuroscience* **22**, 1319–1332, doi:10.1162/jocn.2009.21272 (2010).
61. Aienza, M., Cantero, J. L. & Dominguez-Marin, E. The time course of neural changes underlying auditory perceptual learning. *Learning & Memory* **9**, 138–150 (2002).
62. Ross, B. & Tremblay, K. Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Res* **248**, 48–59 (2009).
63. Chobert, J., Francois, C., Velay, J. L. & Besson, M. Twelve Months of Active Musical Training in 8- to 10-Year-Old Children Enhances the Preattentive Processing of Syllabic Duration and Voice Onset Time. *Cerebral Cortex* **24**, 956–967, doi:10.1093/cercor/bhs377 (2014).
64. Elmer, S., Hanggi, J., Meyer, M. & Jancke, L. Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex* **49**, 2812–2821, doi:10.1016/j.cortex.2013.03.007 (2013).
65. Annett, M. A Classification of Hand Preference by Association Analysis. *British Journal of Psychology* **61**, 303–321 (1970).
66. Lehrl, S., Triebig, G. & Fischer, B. Multiple-Choice Vocabulary-Test Mwt as a Valid and Short Test to Estimate Premorbid Intelligence. *Acta Neurol Scand* **91**, 335–345 (1995).
67. Kuhn, J., Elmer, S., Meyer, M. & Jancke, L. The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia* **51**, 1608–1618 (2013).
68. Bellier, L. *et al.* Topographic recordings of auditory evoked potentials to speech: subcortical and cortical responses. *Psychophysiology* **1–5**, doi:10.1111/psyp.12369 (2014).
69. Schonwiesner, M. & Zatorre, R. J. Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 14611–14616 (2009).
70. Takegata, R., Paavilainen, P., Naatanen, R. & Winkler, I. Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters* **266**, 109–112 (1999).
71. Jung, T. P. *et al.* Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **37**, 163–178 (2000).
72. Pascualmarqui, R. D., Michel, C. M. & Lehmann, D. Low-Resolution Electromagnetic Tomography - a New Method for Localizing Electrical-Activity in the Brain. *Int J Psychophysiol* **18**, 49–65, doi:10.1016/0167-8760(84)90014-X (1994).
73. Mulert, C. *et al.* Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. *Neuroimage* **22**, 83–94, doi:10.1016/j.neuroimage.2003.10.051 (2004).
74. Phillips, C., Rugg, M. D. & Friston, K. J. Anatomically informed basis functions for EEG source localization: Combining functional and anatomical constraints. *Neuroimage* **16**, 678–695, doi:10.1006/nimg.2002.1143 (2002).
75. Paukkunen, A. K. O., Leminen, M. & Sepponen, R. The effect of measurement error on the test-retest reliability of repeated mismatch negativity measurements. *Clin Neurophysiol* **122**, 2195–2202, doi:10.1016/j.clinph.2011.04.004 (2011).

Acknowledgements

We are indebted to Prof. Lutz Jäncke for having supported this research, having provided the necessary infrastructure, as well as for giving us a complete topical freedom. This research was supported by the Swiss National Foundation (grants 320030-120661, 4-62341-05, and 320030B_138668/1 to LJ).

Author Contributions

S.E. and J.K. conceived and designed the study and evaluated the data. S.E. wrote the manuscript. M.H. collected the data and J.A. contributed to the scripts used for analyses and illustrations. The authors decline any kinds of competing financial and non-financial interests. The methods were carried out in accordance with the relevant guidelines and regulations.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017