Neural repetition suppression to vocal and non-vocal sounds

Camille Heurteloup, Annabelle Merchie, Sylvie Roux, Frédérique Bonnet-Brilhault, Carles Escera, Marie Gomot

Title:

Neural repetition suppression to vocal and non-vocal sounds

Author names and affiliations:

Camille Heurteloup[1*], Annabelle Merchie[1*], Sylvie Roux[1], Frédérique Bonnet-Brilhault [1], Carles Escera[2,3,4**], Marie Gomot[1**]

* Co-First-authors and **Co-Last authors

1: UMR1253 iBrain, INSERM, University of Tours, France.

2: Brainlab Cognitive Neuroscience Research Group, Department of Clinical Psychology and Psychobiology, University of Barcelona, Catalonia-Spain.

3: Institute of Neurosciences, University of Barcelona, Catalonia-Spain.

4: Institut de Recerca Sant Joan de Déu (IRSJD), Catalonia-Spain

Corresponding author: Gomot Marie, UMR1253, INSERM, Université de Tours, Centre de Pédopsychiatrie, CHRU Bretonneau 2 Boulevard Tonnellé, 37044 TOURS Cedex 09 FRANCE.

Email: gomot@univ-tours.fr

Phone: (+33)247478664

Running head: Vocal regularity encoding

Abstract

Adaptation to the sensory environment is essential in everyday life, to anticipate future events and quickly detect and respond to changes; and to distinguish vocal variations in congeners, for communication. The aim of the current study was to explore the effects of the nature (vocal/non-vocal) of the information to be encoded, on the establishment of auditory regularities. In electrophysiology, neural adaptation is measured by the 'Repetition Positivity' (RP), which refers to an increase in positive potential, with the increasing number of repetitions of a same stimulus. The RP results from the combined variation of several ERP components; the P1, the first positivity (~100ms) may reflect the onset of repetition effects. We recorded auditory evoked potentials during a roving paradigm in which trains of 4, 8 or 16 repetitions of the same stimulus were presented. Sequences of vocal and non-vocal complex stimuli were delivered, to study the influence of the type of stimulation on the characteristics of the brain responses. The P1 to each train length, and the RP responses were recorded between 90 and 200ms, reflecting adaptation for both vocal and non-vocal stimuli. RP was not different between vocal and non-vocal sequences (in latency, amplitude and spatial organization) and was found to be similar to that found in previous studies using pure tones, suggesting that the repetition suppression phenomena is somehow independent of the nature of the stimulus. However, results showed faster stabilization of the P1 amplitude for non-vocal stimuli than for vocal stimuli, which require more repetitions. This revealed different dynamics for the establishment of regularity encoding for non-vocal and vocal stimuli, indicating that the richness of vocal sounds may require further processing before full neural adaptation occurs.

Keywords: Vocal, Neural adaptation, Repetition Positivity, Habituation, Prediction, ERPs

1. Introduction

We encounter many sensory regularities in everyday life through sensory adaptation, whether simply at the level of word composition, or more complex scene schemas (Turk-Browne et al., 2009). At the cerebral level, information must be processed optimally to select the most relevant events. This requires building and continuously updating a sensory memory trace, following the presentation of a stimulus (Winkler et al., 2009). A repeated stimulus will be considered as regular, leading to the adaptation of the response, through the habituation process. Such storage is an essential property of the sensory systems, in order to process the incoming flow of information. Identifying the probability that the previous stimulus will reappear, enables the prediction of future events (Winkler et al., 2009, 2001). This makes the detection of environmental changes faster, improving our ability to react quickly and optimally. It also underlies our ability to detect modulations of social indices like gaze, facial expressions or voice intonation, and to communicate with others.

One of the most relevant auditory stimuli for our species is voice, which constitutes a very strong social input and tends to be processed as a priority (Whitten et al., 2020). Human beings are considered to be voice experts, because of our experience with and ability to decode such auditory stimuli (Latinus & Belin 2011). The voice is acoustically richer than other types of auditory stimuli in terms of harmonics, pitch and intensities, with more details to encode. Voice stimuli may therefore require further processing prior to full adaptation. Belin et al. (2004) proposed a neurocognitive model of vocal perception. In brief, part of this model suggests that all auditory stimuli are processed, in a general low-level auditory analysis, in the primary cortex A1, and that vocal stimuli then enter a specific, voice structural analysis, involving other regions close to primary auditory cortex, such as the superior temporal gyrus. This suggests that vocal stimuli require additional processing, involving additional regions, which may be reflected in different dynamics for establishing a regularity for vocal versus non-vocal stimuli.

The neural process underlying habituation is called 'Repetition Suppression' (RS). RS refers to neural adaptation - a decrease in neural activity during repeated exposure to the same stimuli. It therefore reflects the formation and continuous updating of sensory memory traces (James et al., 2000; Schacter et al., 2004; Haenschel et al., 2005) and is translated at the cerebral level by a decrease in neural response with an increasing number of repetitions of the same stimulus (neural fatigue or sharpen activity models) (Desimone et al., 1996; Grill-Spector et al., 2006; Sagaert et al., 2013). RS was initially studied in animals at the level of the individual neuron and was termed 'Stimulus Specific Adaptation' (SSA; see Escera & Malmierca, 2016). SSA refers to the decrease in the response of a single neuron with increasing repetitions of the same stimulus (standard), with no decrease in response to a rare stimulus (deviant) (Ulanovsky et al., 2003), and has been recently dissociated from prediction error (Parras et al., 2017). SSA has been observed at cortical (Miller, Li, & Desimone, 1991; Ulanovsky et al., 2004) and sub-cortical levels (Perez-Gonzalez & Malmierca, 2012; Ayala & Malmierca 2013), both in visual (Müller et al., 1999; Vogels, 2012, 2014) and auditory regions (von der Behrens et al., 2009; Nelken et al., 2013). A complementary process of RS, the Repetition Enhancement (RE), reflects the recognition of a stimulus, the anticipation of its appearance, and results in an increase in neural response with repetition (Segaert et al., 2013; Vogels 2016). These processes can also be explained by the predictive coding model (Friston et al., 2005; Bendixen et al., 2012; Auksztulewicz & Friston, 2016) which proposes that the first presentation of a stimulus triggers the generation of a prediction of future sensory input. When this stimulus is repeated, it is compared to the prediction. According to Bayesian models of perceptual inference, RS would reflect a decrease in computational demand that occurs as the prediction error decreases, due to the match between sensory inputs and expected information (Summerfield et al., 2008). In parallel an Expectation Suppression (ES) effect has been identified, which would correspond to a diminution of the neural activity while the expectation become strongest. ES would co-

exist with RS (deGardelle et al., 2013; Grotheer & Kovacs, 2015) but the two phenomena would be independent as the ES effect occurs slightly later than the RS, as shown in both the visual (Summerfield et al., 2011) and the auditory modalities (Todorovic & Lange, 2012). Finely the RE which indexes an increase in the prediction strength when expected, would also be observed with a short delay, in separate frontal brain areas (Recasens et al., 2015).

The establishment of stimulus regularity has previously been studied in both visual and auditory modalities, in electrophysiology (Haenschel et al., 2005; Costa-Faidella et al., 2011; Recasens et al., 2015; Ferrari et al., 2017) and brain imaging (Fiebach et al. al.; 2005, Gagnepain et al., 2008; Andics et al., 2013; Cacciaglia et al., 2019). It was first explored indirectly by studying the electrophysiological response to the detection of change, i.e., the 'Mismatch negativity' (MMN), a negative component, obtained by subtracting the response following the repetition of a standard stimulus from the response evoked by a changing deviant stimulus (Näätänen et al., 1978). MMN indexes deviancy detection, occurring when a stimulus is incongruent with the memory representation of the preceding repeated stimuli (Winkler et al., 2001; Näätänen et al., 2007). In the framework of the predictive coding model, MMN is considered as a marker of error detection, caused by a deviation from a learned regularity (Garrido et al., 2009).

Previous electrophysiological studies in humans, directly investigated auditory regularity encoding through the use of a "roving paradigm", in which a stimulus is repeated a number of times (n), then followed by a new stimulus, which is also repeated n times. This leads to a continuous updating of the memory trace, which is suppressed at the end of each stimulus train (Cowan et al., 1993; Baldeweg et al., 1999; 2004, Haenschel et al., 2005; Costa-Faidella et al., 2011). These studies have highlighted RS phenomena, through the modulation of electrophysiological indices. The repeated presentation of a stimulus results in a decrease of the N1 component, between 90 and 150 ms and an increase in the positive components P1 and P2, reflecting the adaptation of responses. Through comparing the responses to a new stimulus and

to the same stimulus, after a number of repetitions, it is possible to isolate an electrophysiological index of the neural adaptation. The Repetition Positivity (RP), corresponds to a positive deflection between 50 and 250 ms, increasing with the number of repetitions (Baldeweg et al., 2004; Haenschel et al., 2005; Costa-Faidella et al., 2011).

The present study aims to explore the influence of the information to be encoded on the establishment of an auditory regularity. Although adaptation to vocal sound regularities is essential to extract relevant information for social communication, the formation of auditory regularity in the context of vocal stimuli has received little research interest. To our knowledge, most previous studies focused upon pure tones repetitions. Only one study explored the establishment of regularity to more complex semi synthetized sounds of two categories: vowels sounds representing Finnish vowels, and their vowel-like equivalents with increased formants frequencies, that were unfamiliar sounds (Ylinen & Huotilainen, 2007). Using these complex non-natural sounds in a roving paradigm, this study did not show any RP, possibly because of the very low number of participants and of repetitions used (3-4) compared to a minimum of 12 repetitions in other paradigms (Costa-Faidella et al., 2011).

Another study compared RP to vocalizations with different emotional valences, and showed that repetition suppression was increased for positive vocal stimuli as compared to neutral or negative vocal sounds (Pinheiro et al., 2017). Authors concluded that positive vocalizations lead to enhanced sensory prediction. However although vocal stimuli might be considered as potent predictors, they also contain rich acoustic information. Whether we become accustomed to repeated vocal information in the same way as to non-vocal auditory stimuli thus remains to be explored.

To study brain correlates of auditory regularity in the context of vocal stimuli, trains of different lengths (4, 8, 16 repetitions) comprising either vocal or non-vocal complex sounds are compared. The different train lengths should allow us to determine the minimum number of

repetitions required to elicit RP with more social stimuli. We assume that vocal regularity encoding is related to the phenomenon of RS, and anticipate an increase in the positivity (RP) with increasing number of vocal sound repetitions. Analysis of the RP characteristics (amplitude, latency and organization of the response through brain topography analyses) to vocal and non-vocal auditory stimuli will provide information on the RS effect in each condition, that will thereby be compared independently of the sensory response specific to each stimulus category. The comparison of P1 amplitude modulation with increasing number of repetitions in each condition will allow to estimate the dynamic of the neural adaptation, by providing information about the amount of repeated information necessary to attain a plateau in the response (to reach a full neural adaptation). We hypothesize that neural adaptation is likely to be influenced by the vocal/non-vocal aspect of the stimuli, with more repetitions needed for the P1 amplitude to reach a plateau for vocal stimuli. The involvement of memory traces could be weaker and/or slower for vocal stimuli, since these contain richer acoustic complexity than non-vocal stimuli.

2. Materials and Method

We report how we determined our sample size, all data exclusions (if any), all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

Previous EEG studies rarely report all the information needed to calculate sample sizes (Larson and Carbine, 2017). The sample size for this exploratory EEG study was thus determined based on previous research in the field. We chose a conventional sample size in ERP research (20 participants), and reported all the information needed for sample size estimation in future studies.

2.1. Population

Twenty young adults aged 18 to 30 (mean = 24 ± 2 years) participated in the study (12 female). None had neurological, psychiatric, or metabolic disorders or were under medication at the time of the study. None had a hearing deficit as tested with an audiometer at different frequencies (250, 750 and 1500 Hz). Each participant signed an informed consent form, and the protocol received approval from Ethic Committee (PROSCEA2017/23; ID RCB: 2017-A00756-47).

2.2. Stimuli

A total of 8 vocal sounds and 8 non-vocal complex sounds were used to study the possible impact of vocal/non-vocal nature of the information to be encoded. The vocal sounds consisted of the vowel "a" uttered with a neutral prosody by 8 female speakers of different identities, selected from an existing database of vocal sounds validated (on the basis of valence and emotion recognition) on an independent sample of adults (n = 16) (Charpentier et al., 2018). Non-vocal stimuli were synthetic complex sounds with acoustic characteristics close to those of the vocal sounds. For this we created complex sounds with a global frequency spectrum similar to that of the voices. Firstly, using speech analysis software (Praat ®; see Boersma, 2002), we measured the values of the fundamental frequency (F0) and the first 4 formants of each of the eight voices selected as vocal stimuli. The second step consisted of synthesizing complex sounds using sinewaves of the corresponding frequency values (Adobe Audition® software). Amplitudes of each harmonic were adapted to best fit the voices frequency spectrums. To achieve the necessary attenuations (spectral slope), each time the frequency doubled we applied a decrease of 12 dB to the amplitude of the harmonic (Kreiman and Gerratt, 2012). In both categories, each sound was normalized according to the

root mean square of the amplitude, so that all stimuli had the same energy, using Matlab (The MathWorks Inc., Natick, MA, USA); fade-in and fade-out effects of 10% (30ms) were applied. Global energy was equalized at 65dB SPL. We obtained 8 vocal sounds with F0 ranging from 190 Hz to 230 Hz, with frequency steps of about 5 Hz, and 8 non-vocal sounds with the same F0s and spectral structure (Table 1; Figure 1). Sounds are controlled for their 'speech value' as the synthetized sounds mimic their natural vocal equivalents in terms of main frequencies; the two categories of sounds therefore contain the F0, F1 and F2 allowing the phoneme /a/ automatic categorization as previously demonstrated (Jacobsen et al., 2004).

Sound	Fundamental frequency F0	Formant 1	Formant 2	Formant 3	Formant 4
1	190	905	1515	2830	3716
2	199	845	1358	2732	4228
3	205	975	1326	2698	3477
4	210	705	1480	2847	4479
5	215	877	1522	2971	4094
6	220	806	1502	2322	3021
7	226	797	1483	2943	4096
8	229	886	1465	3005	3640

Table 1: Frequency composition of each non-vocal sounds based on the identification of F0, F1, F2, F3 and F4 of the corresponding vocal sounds.

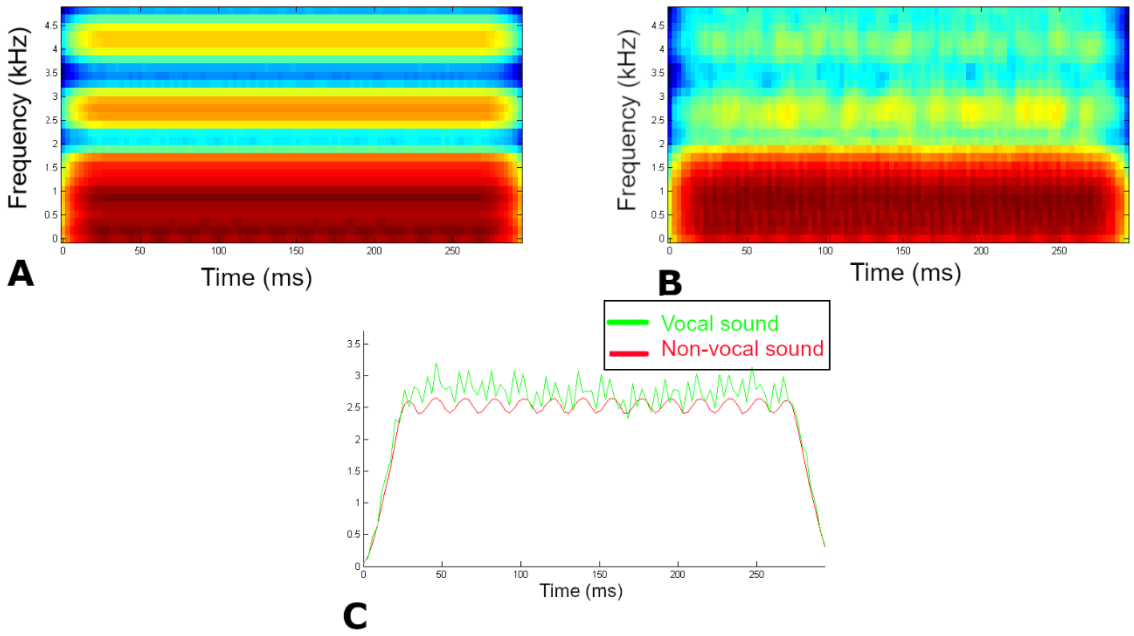

Figure 1 : Physical features of a vocal sounds and of its non-vocal counterpart. (A) spectrogram of the non-vocal sound, (B) spectrogram of the vocal sound, (C) energy of the vocal and the non-vocal sound as a function of time (green: vocal sound, red: non-vocal).

2.3. Sequences of stimulation

We used a roving paradigm (Cowan et al., 1993; Baldeweg et al., 2004) in which trains of stimulation of different frequencies/voices and different lengths were presented. Each train was

composed of the same stimulus, which was repeated n times (depending on the length of the train), and called standard (S). The first stimulus of each train was considered as deviant compared to the preceding repeated standard. We used trains of 4, 8 and 16 repetitions. These trains were delivered in a pseudo-random order, with the only constraint that the same stimulus could not be delivered in two consecutive trains.

Two sequences were presented, the first consisting of vocal stimuli and the second consisting of the non-vocal stimuli described above. The order of presentation of the sequences was counterbalanced between subjects. Each stimulus was presented for 300 ms, with a SOA of 646 ms, via two speakers located at 1.2 m from the ears of the participants (Logitech Z-2300). 120 repetitions of each train length were presented for each sequence, resulting in a total of 6720 stimuli (3360 per sequence) and a recording time of approximately one hour. A schematic representation of the roving paradigm used in this study is illustrated in Figure 2.

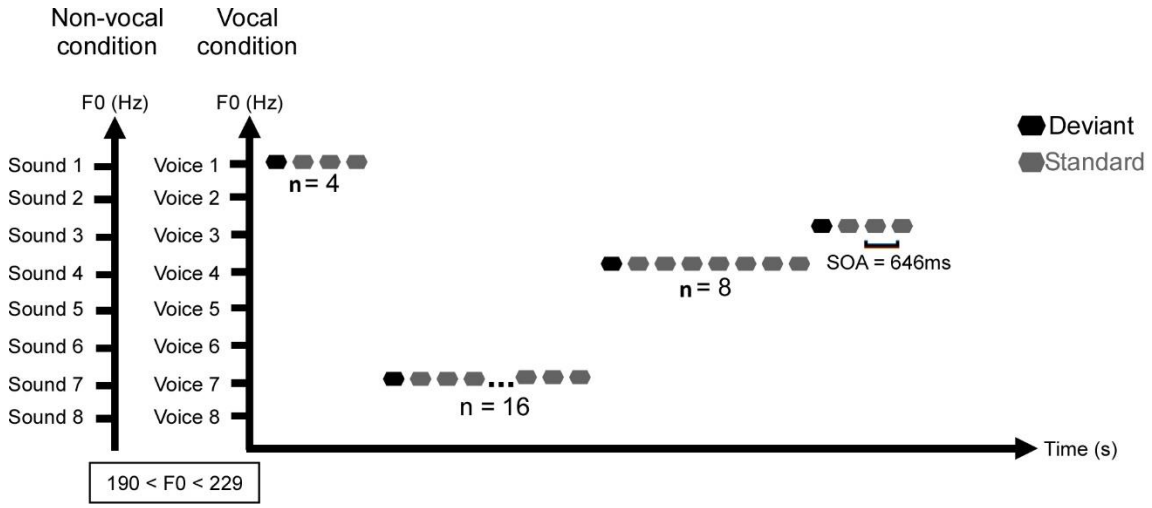

Figure 2 : Roving paradigm used in this study. Trains of four, eight or sixteen repetitions of the same stimulus were pseudo-randomly delivered. The SOA was constant at 646 ms. Vocal condition: run composed of vocal stimuli, the vowel "a", uttered by 8 different female speakers with progressive increasing of fundamental frequencies (190≤F0≤229). Non-vocal condition comprised eight synthetic complex sounds with acoustic characteristics close to those of the vocal sounds, with the same F0 progressive increase (190≤F0≤229).

2.4. EEG data acquisition and processing

During EEG recording, participants were sitting comfortably in a reclining armchair, located in a sound-attenuated room. Subjects watched a silent movie without subtitles whilst sounds were delivered; they were instructed that they would have to briefly tell the story of the movie at the end of the recording session. This procedure avoided voluntary directing of attention towards the auditory stimulation. The two stimulation sequences were delivered, with a break between, so that the subject could move and relax.

Presentation software (NeuroBehavioral Systems Inc., Berkeley, CA) was used to deliver the stimulation sequences, during which the EEG was recorded from 64 active electrodes (ActiveTwo Systems Biosemi, The Netherlands) with a sampling rate of 512 Hz. Horizontal and vertical eye movements were monitored using electrodes placed on the left and right outer canthi and below the left eye. An electrode was placed on subject's nose for offline re-referencing. The ELAN software package was used for the analysis of EEG-ERP (Aguera, et al., 2011). The EEG signal was amplified and filtered with a 0.1 Hz high-pass filter (Butterworth filter, order 1). Artefacts resulting from eye movements were removed by applying Independent Component Analysis (ICA) as implemented in EEG Lab. Blink artifacts were captured into components and selectively removed via inverse ICA transformation. Sixty-four components were examined, and one or two components were removed in each subject, to account for vertical and horizontal movements. Motion artefacts, characterized by high frequency or high amplitude signals, were discarded manually by an experimenter blind from trial type. A 30 Hz low-pass filter was applied (Butterworth filter, order 3), and ERPs were averaged over a 700 ms time window, including a 100 ms pre-stimulus baseline.

The recordings of the two stimulation sequences were carried out separately for each subject, and vocal and non-vocal stimulations were averaged separately.

2.5. Data analysis

2.5.1. Measurements

Responses to each stimulus, in positions 1 to 16, were averaged individually for each sequence, so that the effect of repetition for each position could be observed. Trials presented at the same position, but in trains of different lengths (e.g., stimulus 3 in a train of 4, 8 and 16 repetitions) were also averaged together. This resulted in three times more stimuli from 1 to 4, and twice as many stimuli from 5 to 8, than stimuli from positions 9 to 16. Responses to stimuli in positions 2 to 4 were averaged, as well as those in positions 5 to 8, those in positions 9 to 12 and finally those in positions 13 to 16, in order to obtain more robust responses by increasing the number of trials. For each group of stimuli, the mean number of artifact-free trials was: 925 ± 72 (S2-4), 820 ± 63 (S5-8), 404 ± 45 (S9-12) and 410 ± 31 (S13-16) for the non-vocal sequence and 912 ± 81 (S2-4), 807 ± 77 (S5-8), 401 ± 44 (S9-12) and 400 ± 41 (S13-16) for the vocal sequence.

For the 4 groups of repeated standards, only the P1 component was clearly identified and measured for both conditions. To isolate the RP response, the difference between the average stimuli 13 to 16 (long train) and the averaged stimuli 2 to 4 was calculated for both conditions.

In addition, to isolate MMN responses after each length of train, the differences between the first stimulus (deviant) and last stimulus (standard) of each length of train (4, 8 and 16) were calculated for both conditions. For these stimuli, the mean number of artifact-free trials were: 307 ± 27 (S1), 104 ± 9 (S4), 103 ± 10 (S8) and 102 ± 9 (S16) for the non-vocal sequence and 302 ± 29 (S1), 100 ± 9 (S4), 102 ± 11 (S8) and 100 ± 10 (S16) for the vocal sequence.

The P1 (positive peak at ~ 100 ms) was identified as the first positive deflection occurring between 70 and 120 ms. This time-window was selected on the basis of studies with similar stimuli (Sheerer et al., 2013; Charpentier et al., 2018). Amplitudes and latencies of the P1 were measured in each participant in a 70-120 ms time window, centered on the peak of the grand mean average of the group. RP and MMN measure time windows were selected on the basis of previous studies with similar stimuli (Pinheiro et al., 2017; Bishop et al. 2011; Charpentier et al., 2018). RP was identified as a positive deflection occurring between 90 and 200 ms and the MMN as a negative deflection occurring in a 130-190 ms time-window. Amplitudes and latencies of the RP and the MMN were measured in a 90-200 ms and a 130-190 ms time window, respectively, centered on the peak of the grand mean average of the group.

2.5.2. Statistical analyses

After visual inspection of the responses scalp distributions, and based on previous studies (Haenschel et al., 2005; Costa-Faidella et al., 2011), amplitudes and latencies were analyzed at Fz.

Two-way ANOVAs were performed on P1 amplitudes and latencies, with the Repetition (2-4, 5-8, 9-12 and 13-16) and Condition (vocal vs non-vocal) as within subject factors.

Two-way ANOVAs were performed on MMN amplitudes and latencies with the Repetition (4, 8 and 16) and Condition (vocal vs non-vocal) as within subject factors.

Since the factor Repetition displayed more than 2 levels, a Greenhouse-Geisser correction was applied to correct for potential violations of the sphericity assumption. Additional Bonferroni post-hoc tests were performed to examine the direction of the interactions. The effects sizes are shown as η^2p.

We performed permutation tests based on randomizations (Edgington, 1995) over a 50-300 ms time-window at each electrode and at each time point, to assess difference in the RP between conditions (vocal vs. non-vocal). These analyses provide supplementary information on condition differences, by confirming peak analyses or by affording additional topographical findings or amplitude statistical comparisons for responses whose peak is barely measurable. Each permutation test involved the random permutation of the values for the 20 pairs of data compared (corresponding to the 20 participants), then the calculation of the sum of squared sums of values in each of the two obtained samples, and finally the computation of the difference between these two statistical values. For each analysis we performed 10,000 such randomizations, to obtain an estimate of the distribution of this difference under the null hypothesis. This distribution was then compared to the actual difference between the values in the two conditions (vocal vs. non-vocal). Correction for multiple comparisons was performed using the statistical– graphical method of Guthrie and Buchwald (1991) which tabulates the minimum number of consecutive time samples that need to be significant in the ERP differences, in order to have a significant effect over a given time window. For the analyses of RP (250 ms: 50-300, i.e., 125 sampling points), the minimum number corresponded to 12 consecutive time points (i.e., 24 ms) with p values below the .05 significance level.

3. Results

The grand mean ERPs to standard after 2-4, 5-8, 9-12 and 13-16 repetitions, for vocal and non-vocal conditions are illustrated in Figure 3A. The P1 component (positive peak at ~ 100 ms) was clearly observed, and modulated by repetition, for both conditions. The mean peak amplitudes and latencies of the P1 for each condition are reported in Table 2.

Table 2: Summary of P1 mean amplitudes and latencies according to the number of repetitions (m±SEM)

Number of repetitions	Amplitude (µV ± SEM) at Fz		Latency (ms ± SEM) at Fz	
	Non-vocal	Vocal	Non-vocal	Vocal
2-4	1.13 ± 0.25	1.62 ± 0.24	91.50 ± 2.52	90.15 ± 2.31
5-8	1.74 ± 0.26	1.86 ± 0.24	95.34 ± 3.25	96.62 ± 3.52
9-12	1.98 ± 0.31	2.05 ± 0.28	96.15 ± 3.12	99.73 ± 3.03
13-16	1.70 ± 0.29	2.15 ± 0.27	101.31 ± 3.21	101.85 ± 2.33

Auditory evoked potentials were acquired over the entire scalp, revealing a large fronto-central positivity. Analyses were performed at the Fz electrode, where the responses culminated.

3.1. Difference in repetition effects between conditions

3.1.1. P1 component

A significant repetition effect on P1 amplitudes was found ($F(3, 57) = 17.635$, $p < .001$ $\varepsilon = .875$, $\eta^2p = .48$) due to an increase in P1 amplitude with increased repetition. In addition, a significant condition effect was observed ($F(1, 19) = 6.287$, $p = .021$, $\varepsilon = 1.000$, $\eta^2p = .25$) due to larger amplitudes for the vocal versus non-vocal condition. An interaction between the condition and the number of repetitions was highlighted ($F(3, 57) = 3.514$, $p = .021$, $\varepsilon = .848$, $\eta^2p = .16$).

Post-hoc analyses (Bonferroni) revealed different repetition effects between the two conditions, with a significant increase in P1 amplitude between 2-4 and the three other groups of stimuli (2-4 vs 5-8, 9-12 and 13-16 respectively: $p < .001$, $p < .001$, $p < .001$) but not between 5-8 and 9-12, nor between 5-8 and 13-16 repetitions and between 9-12 and 13-16 repetitions ($p = 0.45$) for the non-vocal condition. For the vocal condition, a significant increase in P1 amplitude was

revealed between 2-4 and 9-12 repetitions (p = 0.011) and between 2-4 and 13-16 repetitions (p < .001) but not between 2-4 and 5-8 repetitions, 5-8 and 9-12 nor between 9-12 and 13-16. Statistical difference in the P1 amplitude was found for vocal versus non-vocal conditions, for 2-4 (p = .003) and 13-16 (p = .008) repetitions, with larger amplitudes observed in the vocal condition.

A significant repetition effect was found for P1 latencies (F (3, 57) = 7.18, p = .001, Ɛ = .743, η^2p = .27) due to an increase in P1 latencies as repetitions increased. There was no condition effect (F (1, 19) = .18, p = .680, Ɛ = 1.000, η^2p = .01 and no interaction between the condition and the number of repetitions (F (3, 57) =, 69, p = .526, Ɛ = .776, η^2p = .04).

Statistical analyses performed on a larger set of electrodes surrounding Fz (FCz, F1, F2, FC1, FC2), revealed similar results for both P1 amplitude and P1 latency.

The scalp distribution of the P1 remained stable across the two conditions, corresponding to a large fronto-central positivity, increasing with repetition.

An example of the P1 scalp potential distribution at 100 ms after 13-16 for non-vocal and vocal conditions with the statistical maps of the topographical differences is illustrated in Figure 3C.

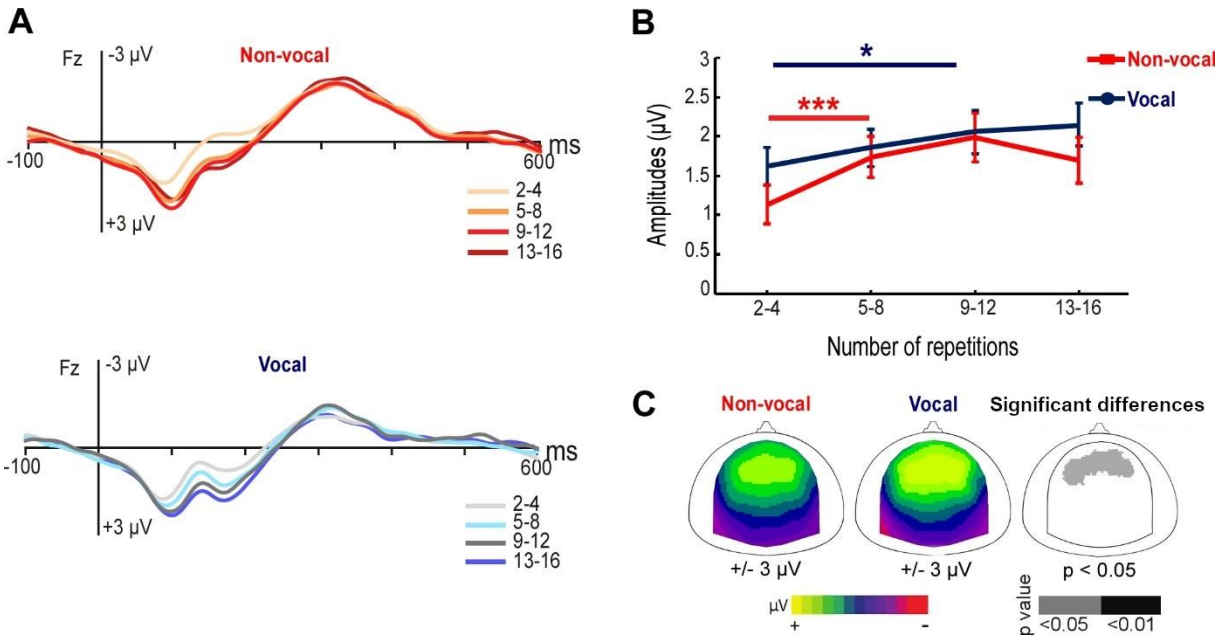

*Figure 3 : (A) Standard ERPs at Fz electrode after 2-4, 5-8, 9-12 and 13-16 repetitions in vocal (top) and non-vocal (below) conditions. (B) P1 amplitude values plotted for 2-4, 5-8, 9-12 and 13-16 repetitions in vocal (blue) and non-vocal (red) conditions (error bars represent SEM) ; * $p < 0.05$, ** $p < 0.01$. (C) P1 scalp potential distribution at 100 ms after 13-16 for non-vocal and vocal conditions and Statistical map of the topographical differences calculated using permutation analyses.*

3.1.2. Repetition Positivity

To better understand the activity associated with stimulus repetition we isolated the RP by subtracting the average responses to stimuli 2-4 (short trains) from the average response to stimuli 13-16 (long trains). Positive deflections corresponding to the RP were observed between 90 and 200ms over fronto-central electrodes in both vocal and non-vocal conditions. The RP grand-average difference waveforms (13-16 minus 2-4 repetitions) for vocal and non-vocal conditions are illustrated in Figure 4A.

Permutation analyses performed between the two conditions, revealed no difference between RP for vocal and non-vocal sounds, in the reported time-window. The scalp distribution of RP, characterized by a large fronto-central positivity, remained stable across the two conditions. Scalp distributions of RP, for vocal and non-vocal conditions, based on the results of the permutation analyses, are shown in Figure 4B. Comparison of the responses obtained on mastoid sites, where the negative activity associated with the fronto-central RP is observed, was also performed. Results revealed no significant differences in the amplitude of the mastoid negativity, between vocal and non-vocal conditions. This reverse polarity between the Fz electrode and the mastoids, confirms the location of RP generators in the supra temporal plane, at the level of the auditory regions.

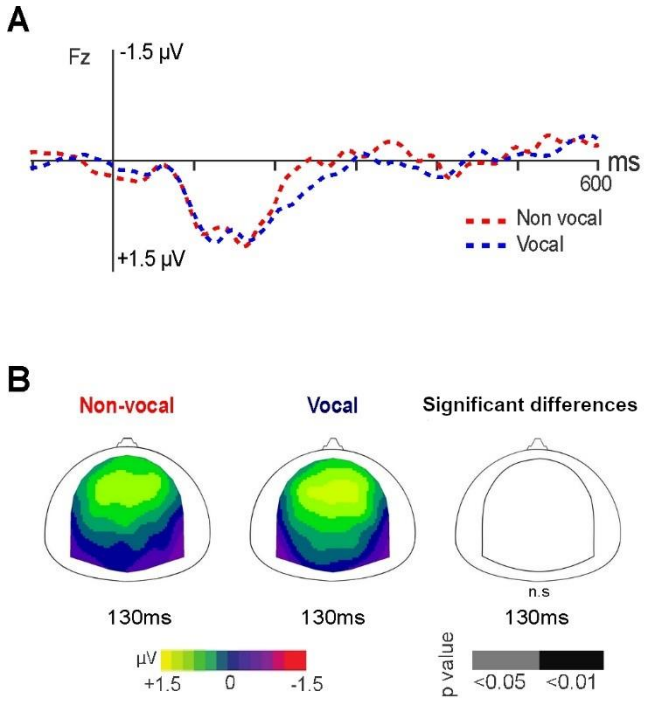

Figure 4 : (A) Repetition Positivity (RP) resulting from the difference wave between 13-16 and 2-4 repetitions for vocal (blue) and non-vocal (red) conditions. (B) RP scalp potential distribution at 130 ms for vocal (right) and non-vocal (left) conditions and Statistical map of the topographical differences calculated using permutation analyses.

The repetition effect on deviance detection was analyzed after 4, 8 and 16 repetitions. The mean peak amplitudes and latencies of the MMN for each condition are reported in Table 3.

Table 3: Summary of MMN mean amplitudes and latencies according to the number of repetitions (m±SEM)

Number of repetitions	**Amplitude (µV ± SEM) at Fz**		**Latency (ms ± SEM) at Fz**	
	Non-vocal	Vocal	Non-vocal	Vocal
4	-1.60 ± 0.47	-1.01 ± 0.30	155.56 ± 2.93	168.75 ± 3.48
8	-2.41 ± 0.41	-1.45 ± 0.44	157.81 ± 3.13	174.71 ± 3.17
16	-2.91 ± 0.42	-1.81 ± 0.34	156.84 ± 2.47	170.01 ± 3.47

The response to the repeated stimuli in positions 4, 8 and 16 was subtracted from the responses to the deviant stimulus (first stimulus of a train). The MMN grand-average difference waveforms after 4 (S1 minus S4), 8 (S1 minus S8) and 16 (S1 minus S16) repetitions are illustrated in Figure 5A. The standard ERPs for 4, 8 and 16 repetitions, and the deviant ERPs after 4, 8 and 16 repetitions of the preceding standard, at Fz electrode, for non-vocal and vocal conditions are shown in Figure 5B.

A significant repetition effect was found on MMN amplitudes ($F (2, 38) = 12.251$, $p < .001$, $\varepsilon = .922$, $\eta^2p = .39$) due to an increase in MMN amplitude with increasing repetitions. No effect of condition ($F (1, 19) = 1.901$, $p = .184$, $\varepsilon = 1.000$, $\eta^2p = .09$) and no interaction between condition and repetition ($F (2, 38) = .530$, $p = .556$, $\varepsilon = .845$, $\eta^2p = .03$) were found.

A significant condition effect was found for MMN latencies ($F (1, 19) = 8.56$, $p = .009$, $\varepsilon = 1.000$, $\eta^2p = .31$) due to an earlier MMN for non-vocal versus vocal stimuli. No effect of repetition ($F (2, 38) = .208$, $p = .807$, $\varepsilon = .973$, $\eta^2p = .01$) and no interaction between condition and repetition ($F (2, 38) = .627$, $p = .537$, $\varepsilon = .984$, $\eta^2p = .03$) were observed for MMN latencies. The scalp distribution of the MMN, corresponding to a large fronto-central negativity, remained stable between the two conditions. Topographical distributions of MMN observed after 4, 8 and 16 repetitions are presented in Figure 5C.

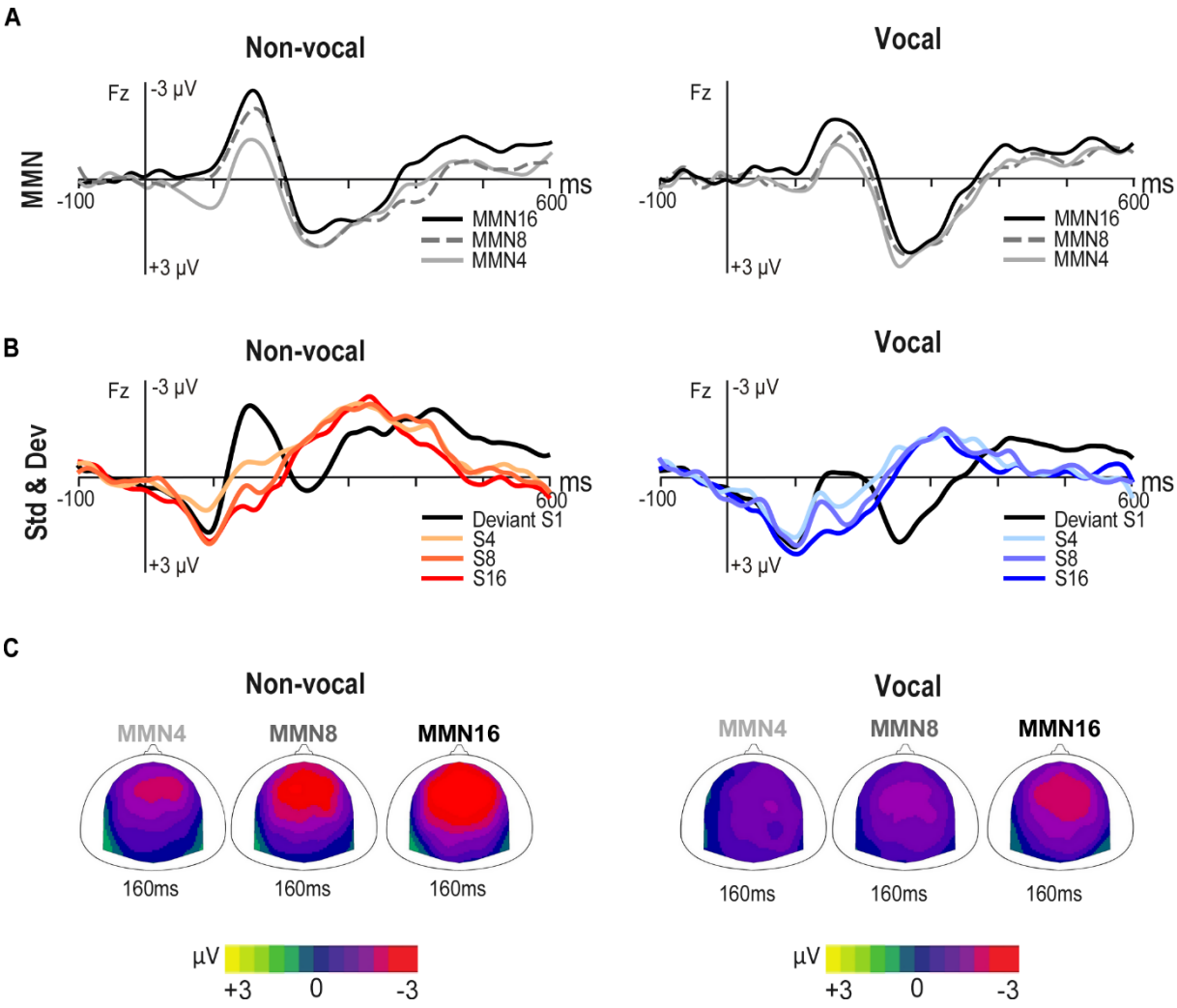

Figure 5 : (A) Mismatch Negativity (MMN) difference waves after 4 (grey), 8 (dotted line) and 16 (black) repetitions for non-vocal (left) and vocal (right) conditions. (B) Standard after 4, 8 and 16 repetitions and deviant ERPs at Fz electrode for non-vocal (left) and vocal (right) conditions. (C) MMN scalp potential maps at 160 ms for non-vocal (top) and vocal (below) conditions.

4. Discussion

This study aimed to explore the possible influence of the information to be encoded (vocal versus non-vocal) on the establishment of auditory regularity indexed by the repetition positivity.

An increase in the number of repetitions generated a significant increase in positivity at the fronto-central electrodes, between 90 and 200 ms, for both vocal and non-vocal stimuli. Results are consistent with previous studies using pure tones, highlighting an RP between 50 and 250 ms (Baldeweg et al., 2004, Haenschel et al., 2005; Costa-Faidella et al., 2011; Recasens et al., 2015). It should be noted that the latencies observed are later than expected from the literature, a delay that might be due to the nature of the stimuli used. Indeed responses to vocal stimuli have later components latencies than pure tones, possibly explained by a later encoding due to their complexity. Both the scalp topographies; characterized by a large positivity spreading over fronto-central regions, and the shape of the RP curves obtained, were very similar in the two conditions. A permutations analysis revealed no significant differences between conditions. This might be explained by the very close spectral composition between the two types of stimuli used, since the non-vocal stimuli were based on the vocal stimuli. Alternatively, the brain activity associated with RS may display the same organization, regardless of the stimuli used. This perspective is reinforced by the fact that the RP obtained in the present work is also similar to the RP observed in response to pure tones (Haenschel et al., 2005; Costa-Faidella et al., 2011). However, the only other study which attempted to explore regularity encoding by using a roving paradigm and comparing more complex stimuli like vocal and non-vocal sounds (Ylinen & Huotilainen, 2007), did not reveal any RP in response to stimuli repetition. Contrary to our initial assumptions, Ylinen and Huotilainen's negative results would not be related to the low number of repetitions used (2-3 versus 5-6 repetitions) since we showed that 5-8 repetitions are sufficient to observe an effect. The contradictory results could instead, be explained by the use of divers vowels instead of a single vowel - as in our study. Previous studies have shown differences in brain responses between vowels, due to the differences in their acoustic compositions (Obleser et al., 2003; Shestakova et al., 2004), which could have led to inconsistent results.

The observation of a polarity reversal at mastoid electrodes, confirmed the involvement of temporal auditory areas in the RS phenomenon, in response to vocal and non-vocal complex sounds repetition. This is similar to previous observations in response to pure tones (Cooper et al., 2013). Although our methodological approach did not aim to accurately locate the regions generating the observed responses, findings are consistent with Recasens et al. (2015) who used MEG to identify the regions involved in RP to pure tones repetition. Their study allowed to target the regions involved in RS and RE separately, and highlighted the contribution of both temporal regions and anterior frontal insula in these two phenomena.

A few studies have highlighted that RP results from the joint modulation of the P1, N1 and P2 components (increase of P1 and P2 with a decrease of N1) (Baldeweg et al., 1999; Haenschel et al., 2005; Costa-Faidella et al., 2011). Due to the specific pattern of brain potentials elicited by our stimuli, results did not allow for the isolation of these three individual components, for the vocal and non-vocal conditions. However, the latency and morphology of the RP obtained, confirms the joint modulation of components in the same latency range.

Although our results revealed similar RPs for vocal and non-vocal conditions, different dynamics were observed in terms of stimulus adaptation, with different profiles of variation in responses amplitude as repetitions increase for each condition. Statistical analysis of the P1 component revealed an increase in amplitude with repetition, just as previous studies have shown modulation of the canonical ERPs in response to pure tones (Haenschel et al., 2005; Costa-Faidella et al., 2011) and to more complex auditory stimuli (Jacobsen & Schröger, 2001). Moreover, we showed that the effect of repetition no longer changes after 5-8 repetitions for the non-vocal condition, but remains stable only after 9-12 repetitions for the vocal condition. Based on these observations, the establishment of regularity appears to be faster for non-vocal stimuli, with only 5-8 repetitions required, than for vocal stimuli, where regularity is only fully established after 9 to 12 repetitions. This is consistent with our initial assumption, that

adaptation to vocal stimuli would be slower. Moreover it seems that the complexity of voice, which is acoustically richer, requires further processing steps, and subsequently leads to slower neural adaptation. At the basic processing level, the spectral richness of the sounds could explain differences in neural adaptation. Although they have been controlled on several physical features, the sounds used in the present study do not contain exactly the same acoustic information. One mechanistic explanation for the repetition effect is stimulus-specific adaptation, i.e. neurons specialized in processing specific physical features (e.g. frequency) have higher activity when processing identical features of the stimulus and are therefore subject to the so-called "refractory process" (Näätänen et al., 2005). The acoustic richness of vocal sounds in terms of spectral frequencies compared to non-vocal sounds with more focused frequency components would explain the observed difference in electrophysiological responses, as the complexity of the physical features to be encoded could recruit a larger number of neurons and ultimately lead to reduced dimension-specific neural adaptation (or given the observed data, to a neural adaptation that would be slower to take place). This is in line with previous NIRS studies on complex speech structure detection in infants, which found that stimulus complexity influences repetition effects (Gervain et al., 2008; Bouchon, Nazzi, & Gervain, 2015).

On the other hand, the human brain is efficient in making predictions and in automatically detecting slight changes in voice, due to its expertise (Latinus & Belin 2011). And indeed voices are processed faster and more automatically compared to synthetic sounds (Whitten et al., 2020). In the same line, studies have shown that the familiarity of a stimulus has an impact on automatic auditory processing (Jacobsen et al., 2005). Social cues are more familiar to us because they contain more relevant information than simpler stimuli such as the non-vocal sounds used in this study. According to this voice preference, one could have expected a faster or larger neural adaptation for vocal sounds than for non-vocal synthetized unfamiliar sounds.

Analyses of ERPs and time-frequency features to repeated vocal emotions (Pinheiro et al., 2017) showed that both induced pre stimulus beta power and RP amplitude were increased for happy vocalizations, suggesting that positive vocalizations lead to strong sensory prediction. Moreover, as described earlier, the repetition positivity includes a predictive coding component (Cacciaglia, et al., 2019) and vocal sounds, which involve larger neural networks (those encoding for speech, language, social context) than non-vocal sounds, in turn will elicit stronger predictions. These may have affected the RP more strongly as suggested by Costa-Faidella et al. (2011) who showed larger RP in the predictable than in the unpredictable condition. Yet in the present study, although comparison of vocal and non-vocal repetitions showed that neural adaptation takes longer for vocal stimuli, RPs were similar in amplitude.

From the predictive coding perspective, the RP and the P1 modulation we recorded to repeated stimuli, probably reflect the combined effects of both RS, ES and RE. The classical roving paradigm and the ERPs method used do not allow to spatially and temporally distinguish between these different phenomena. However, as RS is thought to be low-level and automatic (Kouider et al., 2009) and ES more sensitive to attention and stimulus familiarity (Grotheer & Novacs, 2014), one can assume that even if vocal stimuli represent potent predictors compared to non-vocal stimulation (which should increase ES), their acoustic complexity would slow down the RS effect, resulting in a decreased global response. Alternatively, the impact of the ES effect might also depend on the social relevance of the vocalization. Based on the present results, we were able to draw conclusions on these different hypotheses. Similar RP were recorded for both non-vocal and vocal sounds, but stimulus adaptation was different, with the latter being faster for non-vocal sounds than for vocal sounds. These observations may indicate that regarding neural adaptation, the acoustic simplicity of a sound takes precedence over its familiarity and/or social value.

According to the present results, a maximum of 9 to 12 repetitions is sufficient to establish auditory regularity, regardless of the type of sound. This is consistent with the results obtained by Costa-Faidella et al. (2011), who highlighted a maximum effect at 12 repetitions, with pure sounds. It would therefore not be necessary to use trains of stimulation with 36 repetitions, as done by Haenschel et al. (2005). This is confirmed by the RP response, which shows the same morphology found in previous studies (Costa-Faidella et al., 2011; Haenschel et al., 2005), and also displays a similar amplitude (about 1.5μV), suggesting that the effect no longer evolves between 12 and 36 repetitions.

In addition to the RP analysis, which directly correlates with regularity encoding, we studied the effect of repetition on deviance detection by analyzing the MMN response to each train change. For that, MMN difference waveforms were measured after 4, 8 and 16 repetitions in each condition. Results showed an increase in MMN amplitude with repetition for both vocal and non-vocal complex sounds, as previously shown for tones (Baldeweg et al., 2004; Garrido et al., 2009). These observations support the fact that deviance detection is impacted by previous regularity encoding. However, one study using pure tones in a roving paradigm, did not reveal an increase in MMN amplitude with an increasing number of preceding standard repetitions (Cooper et al., 2013). This could be due to the deviant used, which was the last stimulus of each train, varying in duration. Few studies have focused on the impact of long-term memory traces on short-term memory traces formation. A previous MMN study, presented native-vowel and non-native vowel deviants to highlight the existence of language-dependent memory traces (Näätänen et al., 1997). Results showed that the MMN elicited by native deviant sounds was larger than the MMN elicited by non-native deviant sounds. Another study using a roving paradigm, studied the impact of long-term memory trace related to native language characteristics on the short-term memory trace formed by phonemes, according to the number of preceding stimulus repetitions (Huotilainen et al., 2001). Three different types of non-vocal

stimuli were used: (1) *pure tones* (2) *prototype-vowels*: complex synthetic sounds based on the frequency composition of 8 vowels and (3) *non-prototype vowels*: complex synthetic sounds based on the same vowels, with their first two formants shifted upwards. Results showed that a reduced number of repetitions was required to produce the most prominent MMN for prototype-vowels than for non-prototype vowels, probably due to stronger long-term memory traces for the former. In the present study we used non-vocal, complex, synthetic sounds based on vocal vowel composition, but also natural-vocal sounds, in order to study the impact of the vocal component. We found no statistical differences in MMN amplitudes, between the conditions. This could be due to inter-individual variability, or to the close composition of the two groups of stimuli. Regarding latencies, Huotilainen et al. (2001) showed a longer MMN latency for non-prototype than for prototype vowels and pure tones, but only after few repetitions. In the present study, an earlier MMN was obtained for the non-vocal condition than the vocal condition, with no effect of repetition. Again, such results could be due to the specificity of vocal sounds, which are more complex to encode and process. In the present study, comparing vocal and non-vocal regularity encoding, findings do not strengthen the link between long-term and short-term memory traces, as in previous studies, but support early discrimination of simplest sounds deviancy. This result is consistent with the earlier establishment of regularity in response to non-vocal than to vocal sounds, and reinforces the link between the process of regularity encoding and deviance detection (Baldeweg et al., 2004; Garrido et al., 2009).

5. Conclusion

The purpose of this study was to explore the possible impact of the nature of the information (vocal versus non-vocal) to be encoded, on the establishment of auditory regularity. Once established, auditory regularity leads to adaptation, a phenomenon known as repetition suppression. We highlighted a Repetition Positivity (RP) between 90 and 200ms, following

repeated trains of vocal and non-vocal complex stimulations, confirming the results of previous studies using simple auditory stimuli. As predicted, repetition suppression seems to be sensitive to the nature/complexity of the stimuli to be encoded. Indeed, there are different underlying dynamics for establishing regularity for vocal and non-vocal stimulations, with more repetition required for responses to vocal sounds to stabilize. Adaptation to the auditory environment would therefore follow a different pattern, depending on the type of stimuli, with voice requiring additional processing due to the more complex composition of vocal stimuli. In terms of social communication, this adaptation is essential to quickly detect vocal variations and ultimately, to identify the emotional state of others. This high predictive role of vocal emotion is suggested by the study of Pinheiro et al. (2017) which highlighted the impact of positive emotions on neural adaptation. Moreover, one can assume that this adaptation to sensory environment may increase throughout development to become optimal. Deficits in such adaptation may lead to difficulties in everyday life, particularly in detecting relevant environmental changes or in developing adjusted social interactions. Carrying out similar studies in neuro-typical children may help to establish the optimal developmental pathway of repetition suppression, and related neural adaptation. In neurodevelopmental disorders such as Autism Spectrum Disorders, where adaptation deficits have been recently reported at the subcortical level (Font-Alaminos et al., 2020), repetition suppression studies would allow us to determine whether difficulties in social interaction and adaptation to change, result from deficits in establishing a regular context.

Ethics Statement

The protocol received approval from Ethics Committee (PROSCEA2017/23; ID RCB: 2017-A00756-47). Each participant signed an informed consent form.

Author Contributions

Camille Heurteloup, Carles Escera, Sylvie Roux & Marie Gomot designed the study. Camille Heurteloup & Annabelle Merchie performed data acquisition. Camille Heurteloup, Annabelle Merchie & Marie Gomot were responsible for data and statistical analyses. All authors were involved in preparing and reviewing the manuscript.

Funding

This work was supported by the INSERM (National Institute for Health and Medical Research) and the ANR (ANR-19-CE37-0022-01). C.H was supported by a grant of the "Ministère de la recherche". The funding sources had no involvement in the study design, data collection and analysis, or the writing of the report.

Conflict of interest

All authors declare that they have no conflicts of interest.

Acknowledgments

We thank all the volunteers for their time and participation in this study and Remy Magne for technical support.

Availability of data and material

The conditions of our ethics approval do not permit public archiving of the data supporting this study. Readers seeking access to this data stimuli should contact the lead author Marie Gomot at the UMR 1253 Inserm, University of Tours. Access will be granted unconditionally to named individuals in accordance with ethical procedures governing the reuse of sensitive data.

Stimuli, procedures and scripts for main analyses are available in open access at osf.io/afr4c.

Preregistration

No part of the study was pre-registered prior to the research being conducted.

References

Aguera, P.E., Jerbi, K., Caclin, A. & Bertrand, O. (2011). ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals. Comput Intell Neurosci, 2011, 158970. doi: 10.1155/2011/158970.

Andics, A., Gál, V., Vicsi, K., Rudas, G. & Vidnyánszky, Z. (2013). FMRI repetition suppression for voices is modulated by stimulus expectations. Neuroimage, 69:277–283. doi: 10.1016/j.neuroimage.2012.12.033.

Auksztulewicz, R. & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. Cortex, 80, 125-140. doi: 10.1016/j.cortex.2015.11.024.

Baldeweg, T., Klugman, A., Gruzelier, J. & Hirsch, S.R. (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. Schizophr Res, 69(2-3), 203-217. doi: 10.1016/j.schres.2003.09.009.

Baldeweg, T., Williams, J.D. & Gruzelier, J.H. (1999). Differential changes in frontal and sub-temporal components of mismatch negativity. Int J Psychophysiol, 33(2), 143-148. doi : 10.1016/s0167-8760(99)00026-4.

Belin, P., Fecteau, S. & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. Trends Cogn Sci, 8(3), 129-135. doi: 10.1016/j.tics.2004.01.008.

Bendixen, A., SanMiguel, I. and Schröger,. E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. Int Psychophysiol, 83, 120-131. doi: 10.1016/j.ijpsycho.2011.08.003.

Bishop, D.V.M., Hardiman, M.J. & Barry, J.G. (2011). Is auditory discrimination mature by middle childhood? A study using time-frequency analysis of mismatch responses from 7 years to adulthood. Dev Sci, 14(2), 402-416. doi: 10.1111/j.1467-7687.2010.00990.x.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. Glot International 5, 341-345.

Bouchon C., Nazzi T. & Gervain J. (2015). Hemispheric asymmetries in repetition enhancement and suppression effects in the newborn brain. PLoS ONE 10(10): e0140160. doi:10.1371/journal.pone.0140160.

Cacciaglia, R., Costa-Faidella, J., Zarnowiec, K., Grimm, S., & Escera, C, (2019). Auditory predictions shape the neural rsponses to stimulus repetition and sensory change. Neuroimage, 186, 200-210. doi: 10.1016/j.neuroimage.2018.11.007.

Charpentier, J., Kovarski, K., Houy-Durand, E., Malvy, J., Saby, A., Bonnet-Brilhault, F., Latinus, M. & Gomot, M. (2018). Emotional prosodic change detection in autism Spectrum disorder: an electrophysiological investigation in children and adults. J Neurodev Dis,10:28. doi: 10.1186/s11689-018-9246-9.

Cooper, R., Atkinson, R., Clark, R.A. & Michie, P.T. (2013). Event-related potentials reveal modelling of auditory repetition in the brain. Int Psychophysiol, 88, 74–81. doi: 10.1016/j.ijpsycho.2013.02.003.

Costa-Faidella, J., Baldeweg, T., Grimm, S. & Escera, C. (2011). Interactions between "what" and "when" in the auditory system: temporal predictability enhances repetition suppression. J Neurosci, 31(50), 18590-18597. doi: 10.1523/JNEUROSCI.2599-11.2011.

Cowan, N., Winkler, I., Teder, W. & Naatanen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). J Exp Psychol Learn Mem Cogn, 19(4), 909-921. doi: 10.1037//0278-7393.19.4.909.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. Proc Natl Acad Sci U S A, 93(24), 13494-13499. doi: 10.1073/pnas.93.24.13494.

Edgington, E. S. (1995). Randomization tests (3rd ed.). New York, NY: Marcel Dekker.

Escera, C. & Malmierca, M.S. (2014). The auditory novelty system: An attempt to integrate human and animal research. Psychophysiology, 51, 111-123. doi: 10.1111/psyp.12156.

Ferrari, V., Codispoti, M. & Bradley, M. (2017). Repetition and ERPs during emotional scene processing: A selective review. Int J Psychophysiol, 111, 170–177. doi : 10.1016/j.ijpsycho.2016.07.496.

Fiebach, C.J., Gruber, T. & Supp, G.G. (2005). Neuronal mechanisms of repetition priming in occipitotemporal cortex: spatiotemporal evidence from functional magnetic resonance imaging and electroencephalography. J Neurosci, 25(13), 3414-3422. doi: 10.1523/JNEUROSCI.4107-04.2005.

Font-Alaminos, M., Cornella, M., Costa-Faidella, J., Hervás, A., Leung, S., Rueda, I., & Escera, C. (2019). Increased subcortical neural responses to repeating auditory stimulation in children with autism spectrum disorder. Biological Psychology, 149, 107807. doi: 10.1016/j.biopsycho.2019.107807.

Friston, K. (2005). A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci, 360(1456), 815-836. doi: 10.1098/rstb.2005.1622.

Gagnepain, P., Chetelat, G., Landeau, B., Dayan, J., Eustache, F. & Lebreton, K. (2008). Spoken word memory traces within the human auditory cortex revealed by repetition priming and functional magnetic resonance imaging. J Neurosci, 28(20), 5281-5289. doi: 10.1523/JNEUROSCI.0565-08.2008.

de Gardelle, V., Waszczuk, M., Egner, T and Summerfield, C. (2013). Concurrent Repetition Enhancement and Suppression Responses in ExtrastriateVisual Cortex. Cerebral Cortex, 23, 2235-2244. doi: 10.1093/cercor/bhs211.

Garrido, M., Kilner, J., Kiebel, S.,Stephan, K., Baldeweg, T. &Friston, K. (2009). Repetition suppression and plasticity in the human brain. NeuroImage, 48, 269–279. doi: 10.1016/j.neuroimage.2009.06.034.

Gervain, J., Macagno, F., Cogoi, S., Peña, M. & Mehler, J. (2008). The neonate brain detects speech structure. Proc Natl Acad Sci U S A, 105(37), 14222–14227. doi: 10.1073/pnas.0806530105.

Grill-Spector, K., Henson, R. & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci, 10(1), 14-23. doi: 10.1016/j.tics.2005.11.006.

Grotheer, M. & Kovacs, G. (2014). Repetition Probability Effects Depend on Prior Experiences. J Neurosci, 34(19), 6640-6646. doi: 10.1523/JNEUROSCI.5326-13.2014.

Grotheer, M. & Kovacs, G.(2015). The relationship between stimulus repetitions and fulfilled expectations. Neuropsychologia, 67, 175-82. doi: 10.1016/j.neuropsychologia.2014.12.017.

Guthrie, D. & Buchwald, J.S. (1991). Significance testing of difference potentials. Psychophysiology, 28(2), 240-244. doi: 10.1111/j.1469-8986.1991.tb00417.x.

Haenschel, C., Vernon, D.J., Dwivedi, P., Gruzelier, J.H. & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. J Neurosci, 25(45), 10494-10501. doi: 10.1523/JNEUROSCI.1227-05.2005.

Huotilainen, M., Kujala, A. & Alku, P. (2001). Long-term memory traces facilitate short-term memory trace formation in audition in humans. Neurosci Lett, 310(2-3), 133-136. doi: 10.1016/s0304-3940(01)02096-1.

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology, 41*(4), 654–659. https://doi.org/10.1111/1469-8986.2004.00175.x

Jacobsen, T. & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? Psychophysiology, 38(4), 723-727.

Jacobsen, T., Schröger, E., Winkler, I., & Horváth, J. (2005). Familiarity Affects the Processing of Task-irrelevant Auditory Deviance. Journal of Cognitive Neuroscience, 17(11), 1704–1713. doi:10.1162/089892905774589262

James, T.W., Humphrey, G.K., Gati, J.S., Menon, R.S. & Goodale, M.A. (2000). The effects of visual object priming on brain activation before and after recognition. Curr Biol, 10(17), 1017-1024. doi: 10.1016/s0960-9822(00)00655-2.

Kouider, S., Eger, E., Dolan, R., & Henson, R. N. (2009). Activity in face-responsive brain regions is modulated by invisible, attended faces: evidence from masked priming. Cereb Cortex, 19(1), 13-23. doi: 10.1093/cercor/bhn048.

Kreiman, J. & Gerratt, B. (2012). Perceptual interaction of the harmonic source and noise in voice. J Acoust Soc Am, 131(1), 492–500. doi: 10.1121/1.3665997.

Latinus, M. & Belin, P. (2011). Human voice perception. Curr Biol, 21(4), R143-145. doi: 10.1016/j.cub.2010.12.033.

Larson, M.J. & Carbine, K.A. (2017) Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor. Int J Psychophysiol, 111, 33-41. doi: 10.1016/j.ijpsycho.2016.06.015

Näätänen, R., Gaillard, A.W. & Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. Acta Psychol (Amst), 42(4), 313-329. doi: 10.1016/0001-6918(78)90006-9.

Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature, 385(6615), 432-4. doi : 10.1038/385432a0.

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. Psychophysiology, 42(1), 25–32. doi:10.1111/j.1469-8986.2005.00256.x

Näätänen, R., Paavilainen, P., Rinne, T. & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol, 118(12), 2544-2590. doi: 10.1016/j.clinph.2007.04.026.

Obleser, J., Elbert, T., Lahiri, A. & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. Brain Res Cogn Brain Res, 15(3), 207-213. doi: 10.1016/s0926-6410(02)00193-3.

Parras, G.G., Nieto-Diego, J., Carbajal, G.V., Valdés-Baizabal, C., Escera. C., & Malmierca, M.S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. Nature Communications, 8, 2148. doi: 10.1038/s41467-017-02038-6.

Pinheiro, A.P., Barros, C., Vasconcelos, M., Obermeier, C. & Kotz, S.A. (2017). Is laughter a better vocal change detector than a growl ? Cortex, 92, 233-248. doi: 10.1016/j.cortex.2017.03.018

Recasens, M., Leung, S., Grimm, S., Nowak, R. & Escera, C. (2015). Repetition suppression and repetition enhancement underlie auditory memory-trace formation in the human

brain: an MEG study. Neuroimage, 108, 75-86. doi :10.1016/j.neuroimage.2014.12.031.

Schacter, D.L., Dobbins, I.G. & Schnyer, D.M. (2004). Specificity of priming: a cognitive neuroscience perspective. Nat Rev Neurosci, 5(11), 853-862. doi: 10.1038/nrn1534.

Segaert, K., Weber, K., de Lange, F.P., Petersson, K.M. & Hagoort, P. (2013). The suppression of repetition enhancement: a review of fMRI studies. Neuropsychologia, 51(1), 59-66. doi: 10.1016/j.neuropsychologia.2012.11.006.

Scheerer, N.E, Behich, J., Liu, H. & Jones JA. (2013).). ERP correlates of the magnitude of pitch errors detected in the human voice. Neuroscience, 240,176-85. doi: 10.1016/j.neuroscience.2013.02.054.

Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V. & Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. Brain Res Cogn Brain Res, 21(3), 342-350. doi: 10.1016/j.cogbrainres.2004.06.011.

Summerfield, C., Monti, J. Trittschuh, E., Mesulam, M.M. & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci, 11(9), 1004–1006. doi: 10.1038/nn.2163.

Summerfield, C., Wyart, V., Mareike Johnen, V & de Gardelle, V. (2011). Human scalp electroencephalography reveals that repetition suppression varies with expectation. Front Hum Neurosci, 5, 67. doi: 10.3389/fnhum.2011.00067.

Todorovic, A., van Ede, F., Maris, E., & de Lange, F.P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J Neurosci, 31(25):9118 –9123. doi: 10.1523/JNEUROSCI.1425-11.2011.

Todorovic, A. & Lange, F.P. (2012). Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields. J Neurosci, 32(39), 13389-13395. doi: 10.1523/JNEUROSCI.2227-12.2012.

Turk-Browne, N.B., Scholl, B.J., Chun, M.M. & Johnson, M.K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. J Cogn Neurosci, 21(10), 1934–1945. doi: 10.1162/jocn.2009.21131.

Ulanovsky, N., Las, L., Farkas, D. & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. J Neurosci, 24(46), 10440-10453. doi: 10.1523/JNEUROSCI.1905-04.2004.

Vogels, R. (2016). Sources of adaptation of inferior temporal cortical responses. Cortex, 80, 185-195. doi: 10.1016/j.cortex.2015.08.024.

Whitten, A., Key, A.P., Mefferd, A.S. & Bodfish, J.W. (2020). Auditory event-related potentials index faster processing of natural speech but not synthetic speech over nonspeech analogs in children. Brain Lang, 207: 104825. doi: 10.1016/j.bandl.2020.104825.

Winkler, I., Denham, S.L. & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci, 13(12), 532-540. doi: 10.1016/j.tics.2009.09.003.

Winkler, I., Schroger, E. & Cowan, N. (2001). The role of large-scale memory organization in the mismatch negativity event-related brain potential. J Cogn Neurosci, 13(1), 59-71. doi: 10.1162/089892901564171.

Ylinen, S. & Huotilainen, M. (2007). Is there a direct neural correlate for memory-trace formation in audition? Neuroreport, 18(12), 1281-1284. doi: 10.1097/WNR.0b013e32826fb38a.