Clinical Neurophysiology 135 (2022) 154–161

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

It's not what you say, it's how you say it: A retrospective study of the impact of prosody on own-name P300 in comatose patients

Estelle Pruvost-Robieux [a,d,e,*], Nathalie André-Obadia [b], Angela Marchi [a], Tarek Sharshar [c,d], Marco Liuni [g], Martine Gavaret [a,d], Jean-Julien Aucouturier [f]

[a] *Neurophysiology Department, GHU Paris Psychiatrie et Neurosciences, Sainte Anne Hospital, Paris, France*
[b] *Neurophysiology Department, Hôpital Neurologique Pierre Wertheimer, Bron, France*
[c] *Intensive care unit, GHU Paris Psychiatrie et Neurosciences, Sainte Anne Hospital, Paris, France*
[d] *Université de Paris, Institut Paris Neurosciences et Psychiatrie IPNP (INSERM / Université de Paris), Paris, France*
[e] *Science and Technology of Music and Sound Lab (IRCAM/CNRS/Sorbonne Université), Paris, France*
[f] *FEMTO-ST Institute (CNRS/Université de Bourgogne Franche Comté), Besançon, France*
[g] *Alta Voce SAS, Houilles, France*

ARTICLE INFO

Article history:
Accepted 18 December 2021
Available online 13 January 2022

Keywords:
Disorders of consciousness
Evoked potentials
Own name protocol
Acoustic properties
P300 wave
ERP (Evoked Related Potential)

HIGHLIGHTS

- Acoustic properties of own-name stimuli used in clinical practice are very variable.
- Prosody of own-name stimuli influences latencies of the P300 response when obtained.
- No evidence that the prosody of own-name stimuli influences P300 occurrence.

ABSTRACT

Objective: The acoustic characteristics of stimuli influence the characteristics of the corresponding evoked potentials in healthy subjects. Own-name stimuli are used in clinical practice to assess the level of consciousness in intensive care units. The influence of the acoustic variability of these stimuli has never been evaluated. Here, we explored the influence of this variability on the characteristics of the subject's own name (SON) P300.
Methods: We retrospectively analyzed 251 disorders of consciousness patients from Lyon and Paris Hospitals who underwent an "own-name protocol". A reverse correlation analysis was performed to test for an association between acoustic properties of own-names stimuli used and the characteristics of the P300 wave observed.
Results: Own-names pronounced with increasing pitch prosody showed P300 responses 66 ms earlier than own-names that had a decreasing prosody [$IC_{95\%}$ = 6.36; 125.9 ms].
Conclusions: Speech prosody of the stimuli in the "own name protocol" is associated with latencies differences of the P300 response among patients for whom these responses were observed. Further investigations are needed to confirm these results.
Significance: Speech prosody of the stimuli in the "own name protocol" is a non-negligible parameter, associated with P300 latency differences. Speech prosody should be standardized in SON P300 studies.

1. Introduction

In disorders of consciousness (DOC), multimodal neurophysiological testing is recommended to better predict neurological outcome (André-Obadia et al., 2018). Evoked potentials (EPs) in response to auditory stimulation are an integral part of this approach.

BAEPs, brainstem auditory evoked potentials; DOC, disorders of consciousness; ERP, evoked related potential; HNR, harmonic-to-noise ratio; ICU, Intensive care unit; MLAEPs, middle latency auditory evoked potentials; MMN, Mismatch Negativity; RMS, root-mean-square; SON, subject's own name; TBI, traumatic brain injury; GHU, Groupe Hospitalier Universitaire; dBHL, decibels Hearing Level; SWIPE, sawtooth waveform inspiredpitchestimator.

* Corresponding author at: Service de neurophysiologie Clinique, GHU Paris Psychiatrie & Neurosciences, 1 rue Cabanis, 75014 Paris, France.
E-mail address: e.pruvost@ghu-paris.fr (E. Pruvost-Robieux).

https://doi.org/10.1016/j.clinph.2021.12.015
1388-2457/

Auditory EPs include brainstem auditory evoked potentials (BAEPs) and middle latency auditory evoked potentials (MLAEPs), used to evaluate the integrity of auditory tracts, brainstem structures, and auditory primary cortices, as well as long latency auditory EPs, which aim to discriminate conscious versus unconscious processes. As an example of the latter, oddball paradigms can be used to assess a patient's ability to discriminate between frequent standard and rare deviant sounds. In such paradigms, the mismatch negativity (MMN) response, a fronto-central negative wave recorded about 100 – 250 ms post-stimulus, is thought to index an automatic/pre-attentive cognitive processing of acoustic differences between the frequent and the deviant stimuli (which can differ in duration, frequency or pitch of the stimuli (Goodin et al., 1994; Näätänen et al., 1978)). Observing this MMN, notably in acute post-anoxic DOC, is a good predictor of exiting vegetative state, with a positive predictive value of 100% in the cohort of Fischer et al. (Fischer et al., 2006, 1999). However, in other etiologies of DOC, or later in chronic DOC, predictive values are lower (Fischer et al., 2004; Naccache et al., 2005). In addition, the sensitivity of MMN is greatly variable (Fischer et al., 1999; Kane et al., 1996).

To improve the assessment of DOC patients, some authors developed further auditory paradigms, notably to elicit a P300 response, a positive wave recorded when patients focus their attention on deviant stimuli, notably when these deviants are rare and relevant (Squires et al., 1976; Sutton et al., 1965). Many studies demonstrated the ability of a patient's own-name to grab attention (Moray, 1959; Wood and Cowan, 1995), which motivated Fischer and Morlet (2008) to develop a new oddball paradigm in which frequent and deviant tone stimuli (differing by tone duration) are intermixed with rare audio recordings of a speaker uttering the patient's own-name. In DOC patients, this paradigm was shown to elicit a P300 wave in response to the patient's own-name, which correlates with a good prognosis for awakening in some studies, sometimes with a better sensibility than MMN (Cavinato et al., 2009; Fischer et al., 2010). Yet, important discrepancies persist according to the etiologies of DOC and the delay between the onset of DOC and the completion of neurophysiological tests (André-Obadia et al., 2018; Fischer et al., 2008).

In the typical clinical implementation of own-name P300 paradigms, recordings of a patient's name prior to the evaluation may be done by clinical staff in relatively uncontrolled acoustic conditions (e.g. staff room, over a computer microphone) and, over the years, clinical institutions have constituted unofficial stimulus sets of recordings of frequent first names which can be reused without needing to record them again. The original study of Fischer et al. (2008) offered no guidelines on how these audio recordings should be made. Large acoustic differences are likely to be observed: names may be recorded by male or female staff (resulting in sex-related pitch and timbre differences; Titze, 2000), and pronounced with arbitrary intonation, for example in a questioning or assertive tone, and emotional content, for example calm or alerting tone.

Here, we ask whether that the latency, amplitude and, to some extent, the detectability of own-name P300 responses in the context of DOC evaluation can be influenced to a non-negligible degree by the acoustic and expressive characteristics of the vocal recordings used as stimuli.

To look for a link between the acoustic and expressive content of the recorded name and the ability to detect a P300 response and its characteristics we retrospectively analyzed DOC evaluations at two hospitals in France (Hospices civils de Lyon and Groupe Hospitalier Universitaire (GHU) Paris Psychiatrie et Neurosciences, Sainte Anne). We collected clinical data from N = 251 patients who underwent own-name protocols each linked to the sound stimulus used for their Evoked-related-Potentials (ERP) measurement from August 2008 to May 2021. We also analyzed a set of 120 own-name sound recordings used for DOC evaluations. We then computerized acoustic analysis on the own-name stimuli and the characteristics of the P300 responses observed.

2. Material and methods

2.1. Participants

We studied a retrospective cohort of N = 251 patients (male:160) who underwent DOC assessment with the own-name protocol (Fischer et al., 2008) from August, 7th 2008 to May, 20th 2021 at two centers, the Hospices Civils de Lyon, France (Lyon, N = 132, 53%) and the GHU Paris Psychiatrie et Neurosciences Sainte-Anne, France (Paris, N = 119, 47%). The study did not modify the usual medical practices and was conform to the ethics policies of GHU Paris Psychiatrie et Neurosciences and Hospices civils de Lyon.

2.2. Own-name protocol

All patients in the retrospective cohort underwent the same own-name protocol in both centers, with similar hardware (Micromed S.p.A., Treviso, IT). Standard tones (81% of stimuli) were 80 decibels Hearing Level (dBHL), 800 Hertz (Hz) tones with 1 millisecond (ms) rising and fall time and 75 ms duration, deviant tones (15% of stimuli) were 80dBHL, 800 Hz tones with 1 ms rising and fall time and 35 ms duration, and own-name stimuli (4% of stimuli) were 80dBHL recordings with a duration inferior to 1200 ms. In both centers, P300 paradigms are passive protocols (no instruction is given to the patient before the stimulation). Recording electrodes are placed on Fz, Cz, Pz and linked mastoids (André-Obadia et al., 2018). All electrodes are referenced to the nose. High-pass filter was set at 0.1 Hz and low-pass filter at 30 Hz for cortical derivations, and amplifier sensitivity was set at + 200 μV. The temporal window of analysis is 1000 ms (100 ms pre- and 900 ms post-stimulus onset). Three stimulus series were performed in passive conditions for each patient in Paris (up to 5 if they were not reproducible) and 5 stimulus series in Lyon. All series were timed to include at least 40 own-name stimuli. As part of standard DOC evaluation protocol, the cohort also underwent BAEPs, somatosensory evoked potentials and MMN testing on the same day of evaluation – results from these additional tests were not collated nor analyzed for the present study.

2.3. Clinical and neurophysiological outcomes

For each retrospective patient, we collected the outcome of the P300 evaluation (absent/present) as recorded in the clinical registers. Additionally, a single neurophysiologist (the first author) reanalyzed all ERP files classified as "P300 present" to measure the latency (in ms) at three electrode sites (Fz, Cz, Pz) of the own-name P300 response. For each patient, the neurophysiologist visually checked for artefacts among the ERP curves of the 5 (Lyon) or 3 (Paris) stimulus series and then averaged the non-artefacted series into a global ERP curve. The latency of the P300 response was determined visually as the position of the first peak of the P300 response (when the response displayed multiple components). This procedure was similar to the one used in the original clinical evaluation of the patients, which determined if a P300 response was present or absent, and was only used here to document the latency of this response when present. This procedure was blind to the prosody of the patient's own-name audio recording used for the evaluation.

After all information was collated, data were anonymized with the exception of the patient's first name, which was used to match each patient file with the sound recording used for the evaluation.

2.4. Sound recordings

In addition to retrospective data on patient evaluation, we collected a dataset of n = 623 audio recordings of patient first names, assembled over the years by the neurophysiology teams at Hospices Civils de Lyon and GHU Paris and used for coma evaluation at the two centers considered in this study. n = 153 (male names: 87, female names: 66; all French language) of these recordings corresponded to the ones used to test the N = 251 patients in the cohort and were retained for further analysis. While the number of different speakers involved in recording the names was undocumented (and difficult to evaluate by listening to them), there was a majority of female voice recordings (149: 96.7%). The average duration of the recordings was M = 825 ms (SD = 151 ms).

Of these n = 153 recordings, 58 (37.9%) names were either used for several different evaluations for the same patient, or were namesakes which corresponded to 2 or more patients tested with the same file (shared by 8 evaluations/patients: *Philippe*; by 6: *Alain, Mustapha, Thierry*; by 5: *Frédéric*; by 4: *Jocelyne, Lionel;* by 3: *André, Anne, Bernard, Charles, Dominique, François, Françoise, Huguette, Jeanine, Jean-Pierre, Kamel, Marc, Sébastien; and by 2: Caroline, Christophe, Christelle, Claude, Denis, Domingos, Eric, Florence, Genevieve, Georges, Gérard, Ghislaine, Guy, Jean, Jean-François, Jean-Marc, Joël, Josiane, Laurent, Lucienne, Makram, Manon, Marie, Nicolas, Noël, Pascal, Patricia, Patrick, Régis, Robert, Svilen, Sylvie, Valérie, Victoria, Vincent, Viviane, Yoann).* Because they contribute identical acoustic properties in potentially different clinical groups, the existence of namesakes has the potential to obfuscate the relation between the acoustic content of the recorded name and the patient's P300 characteristics. In the P300-negative group (N = 141), 12 patients (8.5%) had at least 1 namesake in the P300-positive group. Within the P300-positive group (N = 110), 17 patients (15.4%) had at least 1 or 2 namesakes (*2: Charles, Huguette, Jocelyne, Marc; 1: Anne, Bernard, Christian, Christophe, Dominique, Françoise, Laurent, Mustapha, Pascal, Philippe, Sébastien, Thierry, Yoann)* and none of these corresponded to repeated tests by the same patient.

2.5. Acoustical analysis

For each name recording, we used the Praat software (Boersma and Van Heuven, 2001) to extract 17 different acoustic characteristics traditionally associated with emotional expression (Juslin and Laukka, 2002).

Voice is produced when the expiratory airflow from the lungs, generated by thoracic and abdominal muscles, travels through the glottis and sets the vocal folds of the larynx into oscillations (for a review, see (Arias et al., 2021). Changes in sub-glottal pressure primarily lead to modulations of voice intensity. For instance, happy, aroused voices are typically faster and louder than calm, sad voices (Ilie and Thompson, 2006). Here, we quantify variations of intensity across recordings using (1) utterance duration (how fast or slow the name is pronounced, which is also confounded here with how short or long the name is) and (2) root-mean-square (RMS) intensity (how loud or calm the voice is).

Changes in the oscillatory properties of the vocal folds, such as their length and opening, are controlled by the laryngeal muscles and lead to modulations of vocal F0, or pitch (how low or high the voice sounds). For instance, low or high average pitch may correspond e.g. to voices with negative or positive emotional valence (Ilie and Thompson, 2006), but larger pitch variations may also differentiate e.g. fearful vs sad vocalizations (Pell et al., 2011) and local intonations at the start or end of an utterance can also be found in surprised or assertive speech (Jiang and Pell, 2017). Here, we quantify variations of pitch across recordings using the recording's (3) average fundamental frequency in Hertz (Hz), as well its (4) maximum, (5) minimum and (6) standard deviation.

Finally, increased airflow, such as in cries or anger shouts, but also altered neurological (vagal) control over the laryngeal muscles such as in stress or anxiety, may drive the vocal folds into non-linear/chaotic oscillatory regimes which result in alterations of sound quality such as roughness, noisiness or breathiness. Such modulations of vocal source quality are important in emotional behaviors (Gobl and Ní Chasaide, 2010); Johnstone and Scherer, 1999) and are known, in listeners, to trigger prioritized sensory processing and an increased involvement of subcortical structures such as the amygdala (Arnal et al., 2015). Here, we quantify variations of vocal source quality across recordings using four (7–10) standardized measures of pitch perturbation quotient (jitter, measured as the % amount of frequency modulation of the fundamental frequency (Boersma and Van Heuven, 2001), five (11–15) standardized measures of amplitude perturbation quotient (shimmer, measured as the % amount of amplitude modulations of the fundamental frequency), and two (16–17) measures of harmonic-to-noise ratio (HNR, corresponding to the % amount of additive noise or breathiness in the voice). Each of the 17 characteristics extracted here contributed one (averaged) value per recording.

In addition to these 17 averaged characteristics, we also computed dynamic pitch profiles for each recording by calculating instantaneous pitch and RMS (i.e. loudness) values on successive 10 ms windows within a recording (for pitch, using the Sawtooth Waveform Inspired Pitch Estimator (SWIPE) technique - (Camacho and Harris, 2008) - and averaging these values at 8 regularly-spaced time points within the recording (every successive 103 ms on average). The resulting pitch and RMS profiles describe the prosody of the utterance, and allows to separate e.g. names pronounced with similar mean pitch by with rising vs falling intonations.

2.6. Classification image analysis

To test for a statistical association between the prosody of each name and the presence and latency of a resulting P300 response, we used the data-driven technique of "classification images" (Murray, 2011). The classification image technique reconstructs, from the data, what configuration of a stimulus is optimal to generate an outcome ("what dynamic pitch and RMS profile of an own-name recording is optimal to reduce Cz latency" or "to obtain a present or absent P300").

In more details, to compute the classification image (optimal prosody) corresponding to the presence or absence of a P300 (binary variable), we z-scored the individual pitch profile p and RMS profile r of each stimulus; averaged the z-scored profiles of patients in the P300-positive group, weighted by the relative proportion of the group (0.44); and subtracted the sum of the z-scored pitch profile of patients in the low-latency group, weighted by the relative proportion of the group (0.56).

Similarly, to compute the classification image corresponding to an early/late P300 latency (continuous variable), we separated the P300-positive patient group into high and low-latency groups based on a mean-cut (Fz: M = 367.2 ms; Cz: M = 361.7 ms; Pz: M = 363.2 ms); z-scored the individual pitch profile p and RMS profile r of each stimulus; averaged the z-scored pitch and RMS profiles of patients in the low-latency group and we subtracted the sum of the z-scored pitch profile of patients in the high-latency group. This procedure was the same as Ponsot et al. (2018) and Goupil et al. (2021).

In both cases, the resulting classification image is a pitch and RMS profile p^* and r^* which have the same format as each individual's stimulus (i.e., each, a 8-point vector), with a distribution of values around the mean at each time-point with degree of freedom d = 109 (Fig. 1). We tested the significance of the classification image with a one-sample t-test difference to zero, at each time-point (see e.g. Ponsot et al., 2018).

2.7. Distance to optimal prosody

Finally, to compute the distance between a given name's pitch and RMS profiles p and r and the group's classification image p^* and r^*, we computed the average Pearson's correlation coefficient d= ½ * (corr(p,p^*) + corr(r,r^*)).

3. Results

3.1. Characteristics of the cohort

Main DOC etiologies for this cohort were anoxia (Lyon: N = 26, 19.6%; Paris: N = 27, 22.6%), traumatic Brain Injury (TBI) (Lyon: N = 42, 31.8%; Paris: N = 18, 15.1%), stroke (Lyon: N = 10, 7.5%; Paris: N = 16, 13.4%), subarachnoid hemorrhage (Lyon: N = 21, 15.9%; Paris: N = 17, 14.2%), and non-TBI intracerebral hemorrhage (Lyon: N = 12, 9.0%; Paris: N = 24, 20.1%).

Of the N = 251 patients in the cohort, N = 110 (43.8%) had a positive own-name P300 response (with latencies: Fz: M = 367.2 ms, SD = 81.7 ms; Cz: M = 361.7 ms, SD = 84.5 ms; Pz: M = 363.2 ms, SD = 85.2 ms). Note that of these 110 patients with measures of P300 latencies, 19 had missing values (ie. no identifiable P300 response) at the Fz site, 1 at Cz and 4 at Pz. These characteristics are summarized in Table 1.

Table 1
Clinical and neurophysiological characteristics of the cohort and acoustics characteristics of audio recordings analyzed.

Whole cohort, n	251
From Paris center, n	119
From Lyon center, n	132
Male, n	160
Disorders of consciousness etiologies	
Anoxia, n (%)	53 (21%)
Traumatic brain injury, n (%)	60 (24%)
Stroke, n (%)	26 (10%)
Subarachnoid hemorrhage, n (%)	38 (15%)
Non-TBI intracerebral hemorrhage, n (%)	36 (14%)
Others, n (%)	38 (15%)
P300 positive group, n	110
P300 Median latency on Fz, (SD), ms	367.2 (81.7)
P300 Median latency on Cz, (SD), ms	361.7 (84.5)
P300 Median latency on Pz, (SD), ms	363.2 (85.2)
Audio recordings available, n	623
Audio recordings used for DOC assessment	153
Female voice	149
Recording duration, mean (SD), ms	825 (151)
Mean pitch, Hertz (SD)	269.2 (38.4)
Mean Jitter (%)	1.69
Mean shimmer (%)	7.41
Mean HNR (dB)	17.6

SD: Standard Deviation; ms: millisecond; HNR: harmonic-to-noise ratio; dB: decibel.

3.2. Acoustic characteristics of the name dataset

Average acoustic characteristics for the 153 recorded names were within the range of normal, non-pathological speech (pitch: M = 269.2 Hz, SD = 38.4 Hz; jitter: M = 1.69%; shimmer: M = 7.41%; HNR: M = 17.6 dB). Although the number of recordings by male speakers was very small (N = 4 /153), there were well-expected differences between recordings made by male and female speakers (Titze, 2000): names pronounced by male speakers were pronounced with lower pitch (M = -144.5 Hz), more jitter/rough-

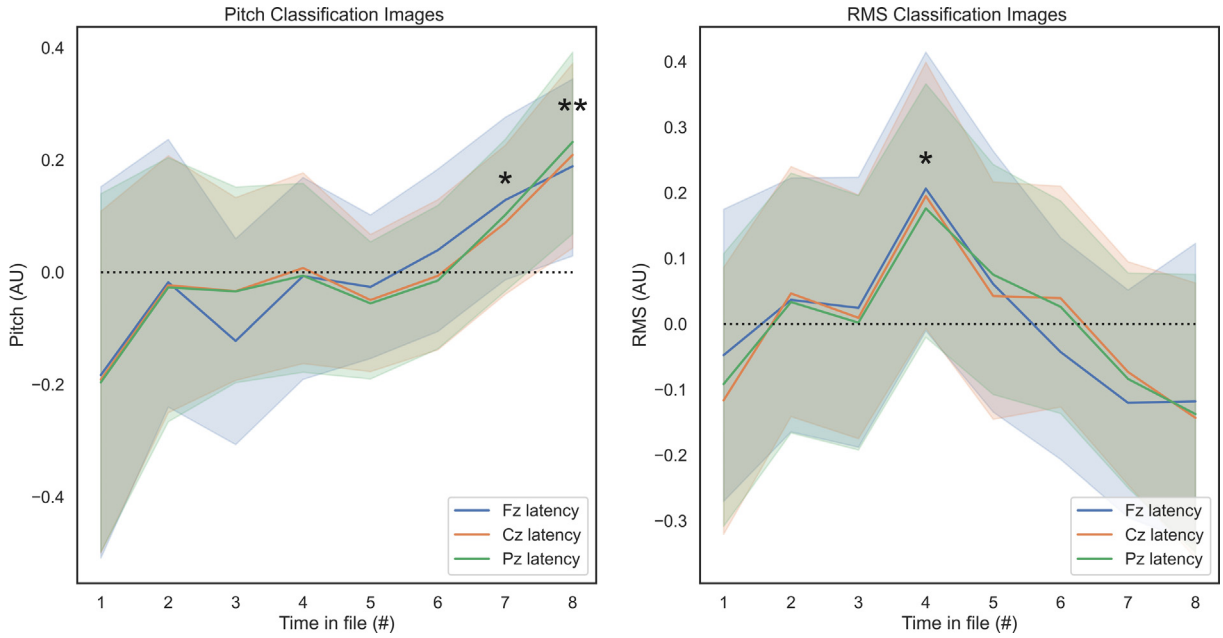

Fig. 1. Optimal pitch and RMS prosody to decrease P300 latency. Left: Pitch classification images for P300 latency on Fz, Cz and Pz. Pitch is calculated at 8 regularly-spaced time points within each own-name recording, and correlated to the latency of the P300 response in Fz (blue), Cz (red) and Pz (green). Recordings with ascendant pitch profile, notably at the end of the recording were associated to the shortest P300 latencies. Right: RMS classification images. RMS is calculated at 8 regularly-spaced time points within the recording, and correlated to the **latency** of the P300 response in Fz (blue), Cz (red) and Pz (green). Recordings with an arching RMS profile were associated to the shortest P300 latencies. RMS: root-mean-square; AU: Arbitrary Unit.

ness (M=+1.40%), more shimmer/roughness (M=+10.3%) and less harmonicity (M = -79 dB).

Less predictably, there were also systematic differences in how the female majority of speakers pronounced male and female first names, with male names (n = 84/149) being pronounced with less harmonicity (HNR: M = -1.40 dB; t(146) = -2.87, p = .004, corrected alpha = 0.0025; a voice with less harmonicity appears more noisy and breathy) than female names (n = 65/149), as well as marginally more shimmer (M=+0.74 dB, t(146) = 1.95, p = .053; a voice with more shimmer appears coarse, rough or aroused). While these differences may suggest that female caregivers address male and female patients with different attitudes, we think it plausible that they also, and perhaps primarily, reflect phonetic and morphological differences between male and female first names (e.g. in the English language, names beginning with a voiced sound – Brian, David, Gregory – are given more frequently to males, and names beginning with unvoiced sounds – Carol, Chelsea, Fiona – more frequently to females; (Slepian and Galinsky, 2016).

Beyond average characteristics, the prosody within each name recording (i.e. pitch and RMS dynamic profiles) was also highly dynamic (repeated-measure mixed ANOVA: main effect of time on z-scored pitch: F(7,1050) = 37.5, p < .001; on z-scored RMS: F(7,1050) = 84.3, p < .001), and arbitrary. In particular, neither did pitch nor RMS prosody differ between male and female first names (interaction time × patient sex on z-scored pitch: F(14,1050) = 1.02, p = .42; z-scored RMS: F(14,1050) = 0.94, p = .51).

On the whole, all such differences, which are arbitrary with respect to a patient's medical condition, highlight that ERP with own-name stimuli are not auditorily equivalent from patient to patient (some of them should be more alerting due to their acoustic differences). In the following, we therefore test whether such acoustic differences have systematic consequences on electrophysiological measures.

3.3. Association between time-averaged acoustic characteristics and electrophysiological measures

We found no statistical association between the time-averaged acoustic characteristics of each name and the P300 measures of the corresponding patients. First, to test for associations with measured P300 latency in the P300-positive group (N = 110), we conducted separate ordinary least-square (OLS) linear regressions of Fz, Cz and Pz latencies on each characteristic, Bonferroni-corrected for the number of characteristics (20; corrected alpha = 0.0025). None of the characteristics (incl. e.g. durations of the first name recordings, on Cz latency: R^2 = 0.003, t(108) = -0.51, p = .60) were significant regressors of latency at either electrode (all ps > 0.23).

Second, to test for associations with the presence/absence of a SON P300 response, we conducted separate two-sample t-tests comparing the distribution of each characteristic between the P300 positive (N = 110) and negative (N = 141) groups. None of the characteristics differed statistically between the names of the P3-positive patients and the names of P300-negative patients at the corrected level (all ps > 0.03; best, mean RMS: t(249) = 2.06, p = .039).

3.4. Association between time-varying prosody and electrophysiological measures

We did find retrospective evidence of an association between the time-varying prosody of each name with P300 latencies. Using the data-driven technique of classification images (Murray, 2011), we reconstructed, from the data, what dynamic pitch and RMS profile of an own-name recording is optimal to generate an early, rather than a late, P300 latency (see section II.6). The optimal prosody had an increasing pitch profile, with significantly positive coefficients at the end of the names for the Fz (segment 7: t(109) = 1.75, p = .083; segment 8: t(109) = 2.20, p = 0.03), Cz (segment8: t(109) = 2.50, p = .013) and Pz electrodes (segment 8: t(109) = 2.78, p = .0065). It also had an arching RMS profile, peaking in the middle of the word (Fz: t(109) = 1.81, p = .07; Cz: t(109) = 1.95, p = .05; Pz: t(109) = 1.74, p = .08), albeit non statistically (Fig. 1).

To quantify the effect of this optimal prosody on P300 latency, we ranked all N = 110 P300-positive stimuli by increasing correlation of their individual pitch profile to the Cz classification image, and found that P300 latency on Cz significantly decreased with increasing correlation to the optimal prosody (OLS regression: coef = -56.7, R^2 = 0.043, t(108) = -2.19, p = .03, Fig. 2). P300-positive patients whose first names happened to be pronounced with increasing pitch and arching RMS prosody had P300 responses that were + 66.13 ms (95% CI [6.36 ms; 125.90 ms]) earlier than patients whose names had a decreasing prosody.

Contrary to latency classification images, pitch and RMS classification images corresponding to the classification of name prosodies into P300-positive (N = 110) and P300-negative patients (N = 141) were non-significantly different from zero.

In summary, there was no evidence in our data that the prosody of own-name was related to the presence or absence of P300. However, when P300 was present, the prosody was related to P300 latency.

4. Discussion

Electrophysiological responses to audio recordings of a patient's own name are an important part of prognosis evaluation in disorders of consciousness (André-Obadia et al., 2018; Fischer et al., 2008). Yet, no general guidelines specify how own-name stimuli should be pronounced when recorded, and large acoustic differences are likely to exist between the stimuli used for different patient names, arbitrarily with respect to a patient's clinical condition. Using a data-driven reverse-correlation analysis on a cohort of 251 retrospective coma patients tested at two centers in Lyon and Paris (France), we showed here that difference in the prosody of recorded names (i.e. whether names were pronounced with a rising or falling intonation) correlate with differences in P300 latencies of 66.13 ms (95% CI [6.36 ms, 125.90 ms]) among patients for whom these responses were observed. This association appeared despite the huge variability between patient conditions (two databases with various DOC etiologies, various delays between the onset of DOC and the neurophysiological assessment, etc.), and some overlap due to namesakes.

The fact that rising intonations were associated with earlier P300 responses is coherent with a wealth of speech production data showing that pitch intonation is an important cue in vocal communication. First, rising pitch is more salient (Gordon and Poeppel, 2002), associated with questions versus declaratives sentences. In Goupil et al. (2021), rising prosodies associated with unreliable utterances were found to be automatically processed to influence words working. Rising pitch is also more affiliative, e.g. associated with trustworthiness (Ponsot et al., 2018). On the contrary, falling pitch correspond to a default speech production mode (a natural consequence of the decrease in subglottal air pressure during the exhalation phase of breathing (Gussenhoven, 2002), which could be judged as less pertinent for the subject. It is therefore striking that a similar pattern (Fig. 1) emerges from our data in a purely data-driven manner, without making any *a priori* hypothesis (other than choosing pitch and RMS as parameters of prosody).

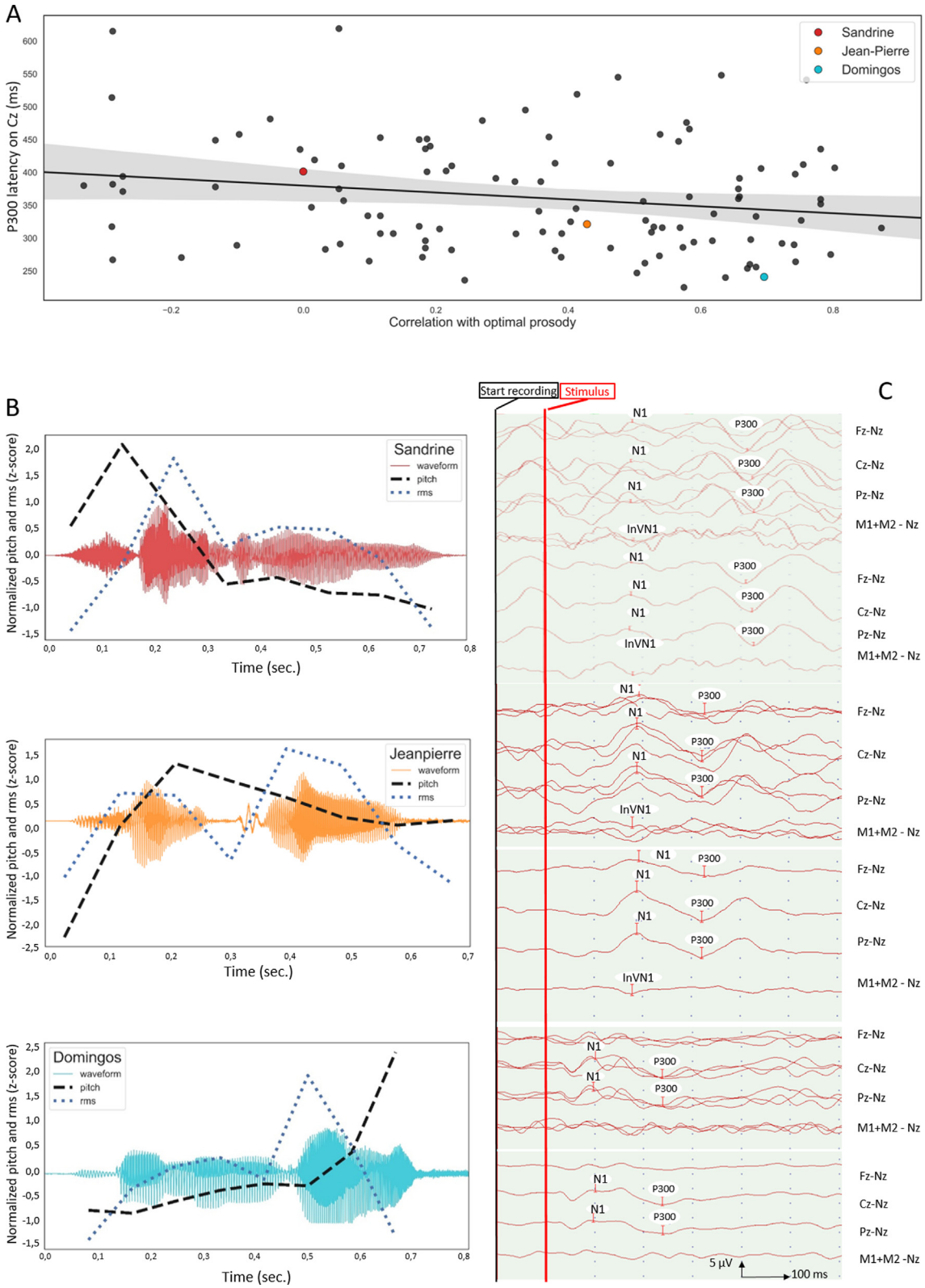

Fig. 2. Correlation between individual P300 recordings and « optimal » prosody profile on Cz to generate a reduced P300 latency. (A) Correlation of each Cz latency of P300 positive recordings on the optimal prosody profile. Optimal prosody profile was defined as the prosody who was associated to short P300 latencies according to the data-driven analyze. Three patients are colored ('Sandrine' in red, 'Jean-Pierre' in orange and 'Domingos' in blue). (B) Normalized pitch and rms of the three first-names recordings used to elicit a P300 response in the three colored patients on the panel A. Each recording has a different pitch intonation. (C) P300 and N1 waveforms of the three colored patients identified on panels (A) and (B). For each panel, the pre-stimulus period (100 ms), stimulus (red-line) and the post-stimulus period (600 ms) are displayed. Recordings were performed on Fz, Cz, Pz and linked-mastoids electrodes, all being referred to the nose (Nz). On top: The three stimulations series superimposed; On the bottom: Average of the three stimulations series. RMS: root-mean-square; ms: millisecond; sec: second; Nz: Nose.

More generally, the fact that own-name prosody is associated with variations of P300 latency is consistent with a wealth of cognitive neuroscience research on healthy participants showing that cognitive ERP are modulated by the acoustic characteristics of the stimuli. It is for instance widely established that the occurrence of a MMN (Horváth et al., 2008) increases, and its latency decreases, with greater separation between the standard and the deviant sounds in pitch (Horváth et al., 2008; Näätänen et al., 1982; Pakarinen et al., 2007), duration (Näätänen et al., 2004) or intensity (Schröger and Winkler, 1995). Similarly, in three-stimulus P300 paradigms, P300 latency for the non-target stimulus tones is shorter when their pitch deviates more strongly from standard tones (Katayama and Polich, 1998). Beyond mean pitch, the emotional content of stimuli modulates MMN, for example, modulations of emotional tone of a violin elicited MMN (Goydke et al., 2004) and, for vocal stimuli, modulation of the emotional characteristics of the vowel "a" - anger, fear, happiness, surprise, disgust or sadness - elicited MMN in healthy subjects and autism spectrum disorder patients (Charpentier et al., 2018). Processing the expressive changes of a human voice is also different whether the speaker voice was familiar or not in awake healthy subjects: larger amplitude in the late phase of P300 response with the own-name uttered by a familiar voice (Holeckova et al., 2008, 2006) and 46 ms MMN onset latency difference between familiar and non-familiar voices in Rachman, Dubal and Aucouturier (2019). It is therefore unsurprising that variations in the dynamic pitch and RMS profile of stimuli, although these had never been specifically tested in the previous literature, should be associated, here, with variations of latency of several 10 s of ms.

Finding an association of own-name prosody with P300 latency in DOC patients has important implications for clinical practice. While the occurrence of a P300 response is the clinical criteria for prognosis (Cavinato et al., 2009; Daltrozzo et al., 2007; Fischer et al., 2008), latency is also relevant because its variability can mislead the practitioner in its interpretation of the presence / absence of the response. In healthy participants, Barry et al d(2020) described the P300/Late Positive Complex (LPC) with temporal principal components analysis (PCA), found that P3a, P3b, novelty P3 and a positive slow wave appeared in that order, and that the novelty P3, or nP3 (peaking between 360–450 ms) was the only component significantly affected by the change stimulus (Barry et al., 2020). In DOC patients however this latency can be unusually delayed (Fischer et al., 2010) and clinical recommendations do not typically attempt to separate P300 sub-components (André-Obadia et al., 2018). Thus, in clinical practice, it's mainly the detection of a positive wave in a compatible temporal window which drives the decision of "P300 present versus P300 absent", and late peaks may be rejected as irrelevant. Our work shows that it is therefore important to standardize and optimize the prosody of own-name recordings to improve testing reproducibility.

Because of its retrospective design, our study has some limitations. First, due to missing values, and the relatively small size of the cohort, we couldn't test for interactions with patient etiologies, or time spent since the onset of DOC. Further work could also test for effects of voice timbre, e.g. whether responses are facilitated by female/male voices, or by voices that are familiar or not to the patient (Bekinschtein et al., 2004; Holeckova et al., 2006; Machado et al., 2007), and voice emotion, e.g. whether responses are facilitated by smiling, positive voices (Arias et al., 2018). Second, in current practice, the P300 own-name protocol is performed with recording electrodes at Fz, Cz and Pz sites and linked mastoids (André-Obadia et al., 2018). However, fundamental research demonstrated that pitch contours are encoded in the superior temporal gyrus (Hamilton et al., 2018; Tang et al., 2017; Yi et al., 2019). It should be interesting to investigate the effects of these prosody differences with a larger set of recording electrodes to improve the spatial resolution of these acquisitions.

After using this retrospective analysis to let the prosodic patterns of Fig. 1 emerge from the data, the logical next step of this work will be to confirm results in a prospective study where new patients will be evaluated with the two versions of their own name (one corresponding to Fig. 1, and the other corresponding to its mathematical opposite). In this future study, presence and characteristics of the P300 responses could be compared within-subject. The prosodic pattern found here may also end up as guidelines for recording new stimuli, by training staff to pronounce names with optimal intonations or, more practically, by providing a web service able to synthesize new stimuli, on demand, with the optimal intonation characteristics (Burred et al., 2019).

Funding

Work funded by ERC StG CREAM 335536, ANR SEPIA, and Fondation pour l'Audition *FPA RD-2018–2 (to JJA)*, and "Sauver la vie 2020" grant (to EPR).

Funders had no involvement in the study.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

André-Obadia N, Zyss J, Gavaret M, Lefaucheur J-P, Azabou E, Boulogne S, Guérit J-M, McGonigal A, Merle P, Mutschler V, Naccache L, Sabourdy C, Trébuchon A, Tyvaert L, Vercueil L, Rohaut B, Delval A. Recommendations for the use of electroencephalography and evoked potentials in comatose patients. Neurophysiol Clin 2018;48(3):143–69. https://doi.org/10.1016/j.neucli.2018.05.038.

Arias P, Belin P, Aucouturier J-J. Auditory smiles trigger unconscious facial imitation. Curr Biol 2018;28(14):R782–3.

Arias P, Rachman L, Liuni M, Aucouturier J-J. Beyond Correlation: Acoustic Transformation Methods for the Experimental Study of Emotional Voice and Speech. Emot Rev 2021;13(1):12–24. https://doi.org/10.1177/1754073920934544.

Arnal L, Flinker A, Kleinschmidt A, Giraud A-L, Poeppel D. Human Screams Occupy a Privileged Niche in the Communication Soundscape. Curr Biol 2015;25(15):2051–6. https://doi.org/10.1016/j.cub.2015.06.043.

Barry RJ, Steiner GZ, De Blasio FM, Fogarty JS, Karamacoska D, MacDonald B. Components in the P300: Don't forget the Novelty P3! Psychophysiology 2020;57(7). https://doi.org/10.1111/psyp.v57.710.1111/psyp.13371.

Bekinschtein T, Leiguarda R, Armony J, Owen A, Carpintiero S, Niklison J, et al. Emotion processing in the minimally conscious state. J Neurol Neurosurg Psychiatry 2004;75:788. https://doi.org/10.1136/jnnp.2003.034876.

Boersma P, Van Heuven V. Speak and unSpeak with PRAAT. Glot Int 2001;5:341–7.

Burred JJ, Ponsot E, Goupil L, Liuni M, Aucouturier J-J. CLEESE: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition. PLoS One 2019;14:e0205943–e0205943. 10.1371/journal.pone.0205943.

Camacho A, Harris JG. A sawtooth waveform inspired pitch estimator for speech and music. J Acoust Soc Am 2008;124(3):1638–52. https://doi.org/10.1121/1.2951592.

Cavinato M, Freo U, Ori C, Zorzi M, Tonin P, Piccione F, Merico A. Post-acute P300 predicts recovery of consciousness from traumatic vegetative state. Brain Inj 2009;23(12):973–80. https://doi.org/10.3109/02699050903373493.

Charpentier J, Kovarski K, Houy-Durand E, Malvy J, Saby A, Bonnet-Brilhault F, Latinus M, Gomot M. Emotional prosodic change detection in autism Spectrum disorder: An electrophysiological investigation in children and adults. J Neurodev Disord 2018;10(1). https://doi.org/10.1186/s11689-018-9246-9.

Daltrozzo J, Wioland N, Mutschler V, Kotchoubey B. Predicting coma and other low responsive patients outcome using event-related brain potentials: A meta-analysis. Clin Neurophysiol 2007;118(3):606–14. https://doi.org/10.1016/j.clinph.2006.11.019.

Fischer C, Dailler F, Morlet D, Fischer, Dailler, Morlet. Novelty P3 elicited by the subject's own name in comatose patients. Clin Neurophysiol 2008;119:2224–30. 10.1016/j.clinph.2008.03.035.

Fischer C, Luaute J, Adeleine P, Morlet D. Predictive value of sensory and cognitive evoked potentials for awakening from coma. Neurology 2004;63(4):669–73. https://doi.org/10.1212/01.WNL.0000134670.10384.E2.

Fischer C, Luaute J, Morlet D. Event-related potentials (MMN and novelty P3) in permanent vegetative or minimally conscious states. Clin Neurophysiol 2010;121(7):1032–42. https://doi.org/10.1016/j.clinph.2010.02.005.
Fischer C, Luauté J, Némoz C, Morlet D, Kirkorian G, Mauguière F. Improved prediction of awakening or nonawakening from severe anoxic coma using tree-based classification analysis. Crit. Care Med. 2006;34(5):1520–4. https://doi.org/10.1097/01.CCM.0000215823.36344.99.
Fischer C, Morlet D, Bouchet P, Luaute J, Jourdan C, Salord F. Mismatch negativity and late auditory evoked potentials in comatose patients. Clin Neurophysiol 1999;110(9):1601–10. https://doi.org/10.1016/S1388-2457(99)00131-5.
Gobl C, Ní Chasaide A. Voice Source Variation and Its Communicative Functions. In: Hardcastle WJ, Laver J, Gibbon FE, editors. The Handbook of Phonetic Sciences. Wiley; 2010. p. 378–423. 10.1002/9781444317251.ch11.
Goodin D, Desmedt J, Maurer K, Nuwer MR. IFCN recommended standards for long-latency auditory event-related potentials. Report of an IFCN committee. International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol 1994;91(1):18–20. https://doi.org/10.1016/0013-4694(94)90014-0.
Gordon M, Poeppel D. Inequality in identification of direction of frequency change (up vs. down) for rapid frequency modulated sweeps. Acoust Res Lett Online 2002;3(1):29–34. https://doi.org/10.1121/1.1429653.
Goupil L, Ponsot E, Richardson D, Reyes G, Aucouturier J-J. Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature. Nat Commun 2021;12:861. https://doi.org/10.1038/s41467-020-20649-4.
Goydke KN, Altenmüller E, Möller J, Münte TF. Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. Cogn Brain Res 2004;21(3):351–9. https://doi.org/10.1016/j.cogbrainres.2004.06.009.
Gussenhoven C. Intonation and Interpretation: Phonetics and Phonology. Grenoble, France: International Speech Communication Association; 2002.
Hamilton LS, Edwards E, Chang EF. A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. Curr Biol 2018;28(12):1860–1871.e4. https://doi.org/10.1016/j.cub.2018.04.033.
Holeckova I, Fischer C, Giard M-H, Delpuech C, Morlet D. Brain responses to a subject's own name uttered by a familiar voice. Brain Res 2006;1082(1):142–52. https://doi.org/10.1016/j.brainres.2006.01.089.
Holeckova I, Fischer C, Morlet D, Delpuech C, Costes N, Mauguière F. Subject's own name as a novel in a MMN design: A combined ERP and PET study. Brain Res 2008;1189:152–65. https://doi.org/10.1016/j.brainres.2007.10.091.
Horvath J, Czigler I, Jacobsen T, Maess B, Schroger E, Winkler I. MMN or no MMN: no magnitude of deviance effect on the MMN amplitude. Psychophysiology 2008;45:60–9. 10.1111/j.1469-8986.2007.00599.x.
Ilie G, Thompson WF. A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. Music Percept 2006;23:319–30. https://doi.org/10.1525/mp.2006.23.4.319.
Jiang X, Pell MD. The sound of confidence and doubt. Speech Commun 2017;88:106–26. https://doi.org/10.1016/j.specom.2017.01.011.
Johnstone T, Scherer K. The effects of emotions on voice quality. In: Proceedings of the XIVth international congress of phonetic sciences. University of California at Berkeley, Berkeley, CA. 1999, pp. 2029–32.
Juslin P, Laukka P. Impact of Intended Emotion Intensity on Cue Utilization and Decoding Accuracy in Vocal Expression of Emotion. Emotion 2002;1:381–412. https://doi.org/10.1037//1528-3542.1.4.381.
Kane NM, Curry SH, Rowlands CA, Manara AR, Lewis T, Moss T, Cummins BH, Butler SR. Event related potentials - neurophysiological tools for predicting emergence and early outcome from traumatic coma. Intensive Care Med 1996;22(1):39–46. https://doi.org/10.1007/BF01728329.
Katayama J, Polich J. Stimulus context determines P3a and P3b. Psychophysiology 1998;35(1):23–33.
Machado C, Korein J, Aubert E, Bosch J, Alvarez MA, Rodríguez R, Valdés P, Portela L, García M, Pérez N, Chinchilla M, Machado Y, Machado Y. Recognizing a Mother's Voice in the Persistent Vegetative State. Clin EEG Neurosci 2007;38(3):124–6. https://doi.org/10.1177/155005940703800306.
Moray N. Attention in dichotic listening: Affective cues and the influence of instructions. Q J Exp Psychol 1959;11(1):56–60. https://doi.org/10.1080/17470215908416289.
Murray RF. Classification images: A review. Journal of Vision 2011;11(5):2.
Näätänen R, Gaillard AWK, Mäntysalo S. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol (Amst) 1978;42(4):313–29. https://doi.org/10.1016/0001-6918(78)90006-9.
Näätänen R, Pakarinen S, Rinne T, Takegata R. The mismatch negativity (MMN): towards the optimal paradigm. Clin Neurophysiol 2004;115(1):140–4. https://doi.org/10.1016/j.clinph.2003.04.001.
Näätänen R, Simpson M, Loveless NE. Stimulus deviance and evoked potentials. Biol Psychol 1982;14(1-2):53–98. https://doi.org/10.1016/0301-0511(82)90017-5.
Naccache L, Puybasset L, Gaillard R, Serve E, Willer J-C. Auditory mismatch negativity is a good predictor of awakening in comatose patients: a fast and reliable procedure. Clin Neurophysiol 2005;116(4):988–9. https://doi.org/10.1016/j.clinph.2004.10.009.
Pakarinen S, Takegata R, Rinne T, Huotilainen M, Näätänen R. Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). Clin Neurophysiol 2007;118(1):177–85. https://doi.org/10.1016/j.clinph.2006.09.001.
Pell MD, Kotz SA. On the Time Course of Vocal Emotion Recognition. PLoS One 2011;6.
Ponsot E, Burred JJ, Belin P, Aucouturier J-J. Cracking the social code of speech prosody using reverse correlation. Proc Natl Acad Sci U S A 2018;115(15):3972–7. https://doi.org/10.1073/pnas.1716090115.
Rachman L, Dubal S, Aucouturier JJ. Happy you, happy me: Expressive changes on a stranger's voice recruit faster implicit processes than self-produced expressions. Soc Cogn Affect Neurosci 2019;14:559–68. https://doi.org/10.1093/scan/nsz030.
Schröger E, Winkler I. Presentation rate and magnitude of stimulus deviance effects on human pre-attentive change detection. Neurosci Lett 1995;193(3):185–8.
Slepian ML, Galinsky AD. The voiced pronunciation of initial phonemes predicts the gender of names. J Pers Soc Psychol 2016;110:509–27. https://doi.org/10.1037/pspa0000041.
Squires KC, Wickens C, Squires NK, Donchin E. The effect of stimulus sequence on the waveform of the cortical event-related potential. Science 1976;193(4258):1142–6.
Sutton S, Braren M, Zubin J, John ER. Evoked-potential correlates of stimulus uncertainty. Science 1965;150(3700):1187–8.
Tang C, Hamilton LS, Chang EF. Intonational speech prosody encoding in the human auditory cortex. Science 2017;357(6353):797–801.
Titze IR. Principles of voice production. Salt Lake City, Utah: National Center for Voice and Speech; 2000.
Wood N, Cowan N. The cocktail party phenomenon revisited: how frequent are attention shifts to one's name in an irrelevant auditory channel? J Exp Psychol Learn Mem Cogn 1995;21:255–60. https://doi.org/10.1037//0278-7393.21.1.255.
Yi HG, Leonard MK, Chang EF. The Encoding of Speech Sounds in the Superior Temporal Gyrus. Neuron 2019;102(6):1096–110. https://doi.org/10.1016/j.neuron.2019.04.023.