CORTEX 121 (2019) 385–398

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research report

Electrophysiological evidence of phonemotopic representations of vowels in the primary and secondary auditory cortex

Anna Dora Manca [a,b], *Francesco Di Russo* [c,d], *Francesco Sigona* [a,b] *and Mirko Grimaldi* [a,b,*]

[a] *Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL), University of Salento, Lecce, Italy*
[b] *Laboratorio Diffuso di Ricerca interdisciplinare Applicata alla Medicina (DReAM), Lecce, Italy*
[c] *Dipartimento di Scienze Motorie, Umane e della Salute, University of Rome "Foro Italico", Rome, Italy*
[d] *IRCCS Fondazione Santa Lucia, Rome, Italy*

ARTICLE INFO

Article history:
Received 3 December 2018
Reviewed 22 January 2019
Revised 18 May 2019
Accepted 20 September 2019
Action editor Alessandro Tavano
Published online 13 October 2019

Keywords:
N1
Primary auditory cortex
Superior temporal gyrus
Electroencephalography
Distinctive features

ABSTRACT

How the brain encodes the speech acoustic signal into phonological representations is a fundamental question for the neurobiology of language. Determining whether this process is characterized by tonotopic maps in primary or secondary auditory areas, with bilateral or leftward activity, remains a long-standing challenge. Magnetoencephalographic studies failed to show hierarchical and asymmetric hints for speech processing. We employed high-density electroencephalography to map the Salento Italian vowel system onto cortical sources using the N1 auditory evoked component. We found evidence that the N1 is characterized by hierarchical and asymmetrical indexes in primary and secondary auditory areas structuring vowel representations. Importantly, the N1 was characterized by early and late phases. The early N1 peaked at 125–135 msec and was localized in the primary auditory cortex; the late N1 peaked at 145–155 msec and was localized in the left superior temporal gyrus. We showed that early in the primary auditory cortex, the cortical spatial arrangements—along the lateral-medial and anterior-posterior gradients—are broadly warped by phonemotopic patterns according to the distinctive feature principle. These phonemotopic patterns are carefully refined in the superior temporal gyrus along the inferior-superior and anterior-posterior gradients. The dynamical and hierarchical interface between primary and secondary auditory areas and the interaction effects between Height and Place features generate the categorical representation of the Salento Italian vowels.

* *Corresponding author*. Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL), University of Salento, Lecce, Italy.
E-mail address: mirko.grimaldi@unisalento.it (M. Grimaldi).
https://doi.org/10.1016/j.cortex.2019.09.016

1. Introduction

How does the brain convert the speech acoustic signal into abstract (phonological) representations? We want to address this issue within the neurobiology of language perspective. In doing this, we need to coherently link linguistic primitives together with neurophysiological primitives traditionally assumed to be at the core of the computation and representation of speech sounds (Embick & Poeppel, 2015; Grimaldi, 2012).

According to linguistic theory (Halle, 2002; Stevens, 2002), the most relevant representational linguistic primitives are not phonemes, but rather smaller units: i.e., *distinctive features*. Distinctive features are universal representational links between articulatory plans and acoustic outputs and must have correlates in terms of both articulation and audition. Bundles of distinctive features, characterized by polar oppositions (binary values), form the consonant and vowel segments. For instance, vowel features identify binary contrasts for tongue height and backness/frontness in the mouth or lip rounding. Distinctive features, then, specify the phonemic contrasts that are used in the language, such that a change in the value of a feature can contrastively generate a new word: e.g., English /æ/ [+low] in ['bæg] *bag* versus /e/ [-low] in ['beg] *beg*. The auditory pathways decode the speech signal structures and ensure the identification of acoustic landmarks that provide evidence for the action of specific articulators and contrastive features marking phonemes. Vowels are characterized by the first two peaks of their spectral envelopes (F1 and F2 values in Hz): F1 inversely correlates with tongue height (low F1 is consistent with high vowels), while F2 correlates with tongue frontness in the mouth (high F2 values are consistent with front vowels) and lip rounding (lip rounding lowers the F2 values) (Peterson & Barney, 1952; Stevens, 2002).

From the neurophysiological point of view, we may assume that the acoustic structures map directly onto clusters of neurons within the auditory cortex thanks to the specific sensitivity of nerve cells to spectral properties of sounds (Ohl & Scheich, 1997; Romani, Williamson, & Kaufman, 1982; Saenz & Langers, 2014): i.e., the so-called *tonotopic principle*. This place coding of acoustic frequencies is ensured by the selective activation of the cochlear neurons regularly positioned along the basilar membrane. Then, the neural signals emitted by cochlear neurons are transmitted in the brainstem and preserved up along the auditory cortex (Mesgarani, Cheung, Johnson, & Chang, 2014; Talavage et al., 2004). Additionally, the temporal mechanism of auditory encoding, known as the *tonochrony principle*, might augment or supplement the tonotopic strategy in the frequency range critical to human speech: this means that the latency of auditory evoked components is sensitive to some stimulus properties (Roberts, Ferrari, Stufflebeam, & Poeppel, 2000). In this respect, the distinctive features would be real, in the sense of being universal neuronal mechanisms for perceiving and producing sounds of speech (Teuber, 1967). If this is true, different clusters of neurons should be selectively activated depending on the kind of features computed and represented. As a result, a sort of *phonemotopic* map should dynamically emerge within the auditory cortex. Pursuing this line of research, discoveries regarding the structure and functional organization of the brain may explain the neurobiological properties of the computations and representations theorized by linguists.

Strictly connected to this topic, there is the question whether speech is processed bilaterally, or whether the left hemisphere plays a more dominant role. In line with the *asymmetric sampling in time* (AST) model (Poeppel, 2003), the input speech signal has a bilateral neural representation at the A1, but phonological computations are left-lateralized in the ~20–50 msec temporal integration window while syllabic computation is right-lateralized in the ~150–250 msec integration window in secondary auditory areas. This view is better specified in the *dual-stream model*: at an early stage, a spectro-temporal analysis is carried out bilaterally in the superior temporal gyrus (STG). The categorical (phonological) processing, instead, involves the middle-to-posterior portion of the superior temporal sulcus (STS) bilaterally, although some indications of left-lateralization may emerge (Hickok & Poeppel, 2007; Peelle, 2012). However, the issue remains controversial (Scott & McGettigan, 2013), and a meta-analytic investigation of fMRI data revealed the left hemisphere (in particular, the left mid-STG) to be dominant in phoneme processing (DeWitt & Rauschecker, 2012).

Further than with neuroimaging techniques, this issue has been extensively investigated through magnetoencephalography (MEG), thanks to event-related magnetic fields (ERMFs) and the auditory N1m component (Manca & Grimaldi, 2016). As the N1 is not a unitary event (Näätänen & Picton, 1987; Woods, 1995), the major challenge is to find a correlation between the temporal events contributing to the N1, their hierarchical generation from the primary to the secondary auditory areas, and the bilateral or left hemispheric activation. MEG offers optimal temporal resolution and is thought to perform better than electroencephalography (EEG) in localizing neural activity from the scalp (Ahlfors, Han, Belliveau, & Hämäläinen, 2010; Baillet, 2017). MEG investigations of speech failed to show both hierarchical involvement of auditory areas and clear effects of hemispheric lateralization (Manca & Grimaldi, 2016). When cortical sources of N1m responses are reported, the supratemporal plane—an area that includes the A1 and the STG (Obleser, Elbert, Lahiri, & Eulitz, 2003; Poeppel et al., 1997)—the planum temporale (Obleser, Lahiri, & Eulitz, 2004a) or the area around the STS (Eulitz, Obleser, & Lahiri, 2004) are suggested as the bilateral centers of speech processing. One possible limitation may be due to the fact that MEG is particularly sensitive to tangential (i.e., parallel to the scalp) neuronal sources. Conversely, EEG is sensitive to both radial (i.e., towards or away from the scalp) and tangential sources, although the signal is dominated by radial sources (Malmivuo, Suihko, & Eskola, 1997). Thus, in principle, EEG and the Event-Related Potential (ERP) N1 component should be responsive to a larger range of cortical sources and permit investigators to pick up the dynamical and spatially-distributed neuronal activity involved in speech processing. Experiments on the replicability of MEG and EEG measures showed only a minor advantage for MEG. It seems that EEG localization may be more accurate than MEG localization for the same number of sensors (Liu, Dale, & Belliveau, 2002). Furthermore, advances in high-density electrode montages and EEG source analysis have improved the ability to

accurately localize EEG signals (Cohen & Halgren, 2003, pp. 615–622). A second limitation is intrinsic to the practice of modeling the N1 sources by a single equivalent current dipole (ECD) in each hemisphere, restricting the source analysis to the rising slope and peak of the N1m (Lütkenhöner, Krumbholz, & Seither-Preisler, 2003).

MEG investigations left open the question whether speech sound maps are solely determined by bottom-up acoustic information or modulated by top-down information relying on abstract representation of distinctive features (Manca & Grimaldi, 2016). A solid piece of evidence is that the acoustic distance between the first two formants of a vowel is preserved in the auditory cortex and is directly reliable in sensor and source data along the Talairach 3D coordinate system: lateral medial (x), anterior-posterior (y), and inferior-superior (z) gradients. At the same time, amplitudes, latencies, and spatial gradients in the auditory cortices tentatively suggest that acoustic–articulatory properties are affected by top-down features such as Height, Place and Round (Manca & Grimaldi, 2016). Clues of orderly cortical representations of abstract features emerge when more than one pair of vowels are investigated (Obleser et al., 2004a; Ohl & Scheich, 2004) or when an entire phonological system has been studied with appropriate statistical analyses able to discern different levels of auditory brain operations (Scharinger, Idsardi, & Poe, 2011). Thus, it is hard to disambiguate N1m evidence suggesting pure acoustic patterns from evidence indicating abstract phonological features.

We recorded ERPs from 16 subjects and analyzed N1 amplitude, latency, and topography. Contrary to previous studies, we modeled the N1 using two bilateral mirror symmetric pairs of ECDS concurrently. In order to capture the multiple N1 cortical generators correlated to vowel features, we also included analysis of the whole N1 duration. We investigated the five-vowel system characterizing the Salento Italian (SI) variety spoken in Southern Apulia: /i, ɛ, a, ɔ, u/. This simple set of vowels results in the most common vowel system in the world's languages (de Boer, 2001). Thus, the findings of N1 modulations may provide evidence on the neural computations involved in representing the distinctive properties of this typology of vowel systems. The five-vowel system under investigation is marked by three contrasts for Height (referring to F1 and the vertical tongue position in the mouth): [+high] /i/, /u/; [-high, -low] /ɛ/, /ɔ/; [+low] /a/ and one contrast for Place (referring to F2 and the horizontal tongue position in the mouth): [-back] /i/, /ɛ/; [+back] /a/, /ɔ/, /u/ (Fig. 1, see also Table 1). The [±round] feature is redundant since /ɔ/, /u/ are both [+back] and [+round], and the vowel /a/ is contrastively only [+low], so its features [+back, -round] are predictable by means of this specification (Calabrese, 1995).

With the aim of assessing hierarchical involvement of auditory areas, the effects of hemispheric lateralization, and the acoustic/abstract representation of SI vowels, we employed two linear mixed-effects statistical models. The acoustic model included the predictors F1 and F2 as fixed effects and Subject as random intercept, while the phonological model included the phonological predictors Height (three contrasts) and Place (one contrast) as fixed effects and subject as random intercept. In this way, we want to ascertain whether spatial arrangement of neuronal sources are a pure

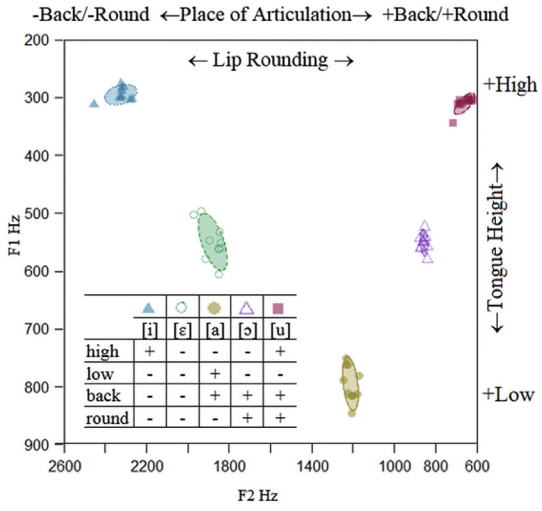

Fig. 1 – F1–F2 representation in Hz of the Salento Italian (SI) vowels and their specification in term of distinctive features. 68.27% confidence ellipse corresponding to ±1 SD from the bivariate mean. F1 is inversely correlated with articulatory tongue height, while F2 reflects the place of articulation in the horizontal dimension.

bottom-up reflection of spectro-temporal differences between vowels, or whether they are simultaneously warped by top-down information relying on polar oppositions determined by abstract distinctive features information.

2. Materials and methods

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

2.1. Subjects

Sixteen volunteer students of the University of Salento (eight men, eight women; mean ± SD, 23 ± 3 years) participated in the experiment after providing written informed consent. All subjects were consistently right-handed according to the Handedness Edinburgh Questionnaire (Oldfield, 1971), and none of them had any known neurological disorders or other significant health problems. The Ethical Committee of the Vito Fazzi Hospital in Lecce approved the experimental procedure. The study was carried out in accordance with the guidelines of the Declaration of Helsinki. The data were acquired in the *Centro di Ricerca Interdisciplinare sul Linguaggio* (CRIL) in Lecce (Italy).

2.2. Stimuli and procedure

The stimuli consisted of the five stressed SI vowels and a pure tone. A native Italian male speaker (age 32) realized ten repetitions of each vowel in isolation, at a normal rate. The speech signal was recorded in a soundproof room with CSL 4500 and a Shure SM58-LCE microphone with a sampling rate

Table 1 – Pitch (F0), Formant Frequency (F1, F2, F3 in Hz) mean values and rise and fall-times (msec) of the vowels used as stimuli (SD is given in parenthesis). The parameters F2–F1 are also given.

Vow.	F0	F1	F2	F3	F2–F1	Rise	Fall
[i]	145	294 (±11)	2325 (±50)	2764 (±28)	2031 (±47)	21	28
[ɛ]	145	549 (±32)	1880 (±46)	2489 (±60)	1330 (±71)	40	32
[a]	140	794 (±30)	1231 (±24)	2528 (±95)	418 (±46)	33	27
[ɔ]	140	550 (±14)	856 (±13)	2551 (±54)	306 (±21)	29	22
[u]	130	310 (±12)	660 (±33)	2437 (±49)	349 (±25)	22	23

of 44.1 kHz and an amplitude resolution of 16 bits. The stimuli were edited and analyzed using the speech analysis software Praat 5.2 (Boersma & Weenink, 2011). All stimuli were normalized for duration (200 msec), for the F0 values according to the values of a representative sample of SI vowels (Grimaldi, 2009)—i.e., 130 Hz for /i/, 140 Hz for /ɛ, a, ɔ/, and 145 Hz for /u/—and for intensity (70 dB/SPL). The F0–F3 formant values were measured in the vowel steady tract (.025 sec) centered at the midpoint. The ramp for rise/fall-times was not edited to preserve natural-sounding speech (Table 3), as it has been shown that the rise- and fall-times of vowels do not affect the relative N1 latencies and amplitudes (Gage, Poeppel, Roberts, & Hickok, 1998; Grimaldi, Manca, & Di Russo, 2016). A pure tone of 1000 Hz and duration of 200 msec was created by Praat software. In the experimental protocol, the best five exemplars of each vowel type and the pure tone were binaurally transmitted to the subjects through two loudspeakers (Creative SBS 2.1 350) at a comfortable loudness (about 70dB/SPL) with Presentation software 2.0. Before the EEG recordings, participants were familiarized with the stimuli. All of the subjects were able to identify each of the vowels with an accuracy of 100%.

2.3. *Experimental design*

During the experiment, the participants were seated in front of a computer monitor in a shielded room. They were asked to listen to the vowels and to push a button with their left index finger whenever they heard a pure tone of 1 KHz used as distractor stimulus (Fig. 2). Two blocks of 1000 vowel stimuli each were presented. Each block consisted of 200 tokens per vowel category and 70 distractor stimuli. Stimuli were randomly presented with a variable inter-stimulus interval that ranged between 1000 and 1400 msec. The distractor stimulus was interspersed with a probability between 6% and 7% in the train of the vowel sounds. To reduce excessive eye movements, participants were asked to fixate on a white crosshair located in the center of the monitor. The experiment lasted approximately one hour.

Table 2 – Mean amplitude (μV), latency (msec), and SD (±) values of the five SI vowels for the early and late N1.

Vowel	Early N1		Late N1	
	Latency	Amplitude	Latency	Amplitude
[i]	128 (±7.6)	−3.20 (±1.5)	145 (±6.7)	−2.56 (±.8)
[ɛ]	127 (±7.7)	−2.19 (±.6)	146 (±6.7)	−2.31 (±.9)
[a]	134 (±11.6)	−2.45 (±1.6)	151 (±9.3)	−2.07 (±.9)
[ɔ]	130 (±9.3)	−2.45 (±1.1)	148 (±8.5)	−1.98 (±.8)
[u]	131 (±8.6)	−3.17 (±1.6)	152 (±7.3)	−2.69 (±1.1)

2.4. *Data acquisition and preprocessing*

Continuous EEG was recorded using a 64-channel ActiCap™ (Brain Products GmbH, Germany) and Brain Vision Recorder 1.20 (Brain Products GmbH, Germany) at a sampling rate of 250 Hz, an online band pass filter of .16–80 Hz, and a notch filter at 50 Hz. Vertical eye movements were monitored using Fp2 and an additional electrode attached below the right eye. FT9 and FT10 were used for horizontal movements. The online reference was at FCz, the ground was AFz, and the impedance was kept under 5 KΩ. Off-line signal processing was carried out using Brain Vision Analyzer 2.0.1 (Brain Products GmbH, Germany). The EEG was segmented in relation to the onset of the five vowels; thus, the distractor and the following stimulus were left out of analyses. ERP epochs of 1200 msec (including 200 msec pre-stimulus baseline) were extracted, digitally filtered by a .5–50 Hz band pass filter (48 db) and re-referenced to the average of the left and right mastoids (M1/2). Ocular artifacts were removed by applying an ICA algorithm that, on average, removed three components. Additionally, rejection criteria for trials were set to 120 μV maximum absolute difference. On average, 9.2% of trials were rejected and 3 ICA components were excluded. Artifact-free segments were separately averaged for each vowel, and a baseline correction was executed over the applied pre-stimulus portion. Finally, grand averages were computed across all subjects and for each vowel type. Analyses were focused on the N1 component in the 80–160 msec range.

2.5. *Statistical analysis*

Looking at the N1 wave, it was evident the presence of two peaks clearly separated in time. The earliest N1 peaked between 90 and 125 msec. The later N1 peaked between 135 and 160 msec. These two peaks were also different in terms of

Table 3 – Talairach coordinates of the bilateral source locations of the five vowels (and relative mean) for the early and late N1 waves. The ± symbols before the X values indicate that sources were constrained to be symmetric in both hemispheres.

Vowel	Early N1			Late N1		
	X	Y	Z	X	Y	Z
[i]	±46	−24	13	±52	−4	1
[ɛ]	±48	−26	12	±49	−8	−2
[a]	±54	−30	15	±60	−25	−2
[ɔ]	±51	−22	9	±56	−17	−7
[u]	±50	−17	10	±46	−11	−10
Mean	**±50**	**−24**	**12**	**±53**	**−13**	**−4**

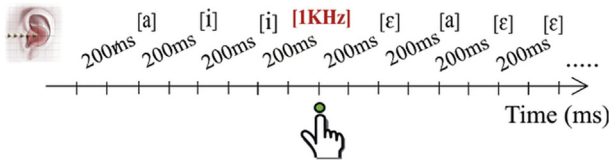

Fig. 2 – Scheme of the experimental design. Participants had to press a response button when they heard a pure tone (occurring with a probability between 6% and 7%), represented as [1 Khz].

topography, with the earlier peak (early N1) focusing on medial electrodes and the later peak (late N1) focusing on the more lateral electrodes over the left scalp. For these reasons, the N1 peak amplitudes and latencies were measured at electrodes with large amplitude in the two individuated intervals: e.g., FCz, Cz or CPz for the early N1 and FC3, C3 or CP3 for the late N1. The latency and amplitude values were analyzed separately for each early and late N1 component with two linear mixed effects models using R (R Core Team, 2015), *lme4* (Bates, Maechler, & Bolker, 2015), and *multcomp* (Hothorn, Bretz, & Westfall, 2008) (with Tukey post-hoc). The acoustic model included the acoustic predictors F1 and F2 as fixed effects and Subjects as random intercept; the phonological model included the phonological predictors Height and Place as fixed effects and subjects as random intercept. Specifically, we defined three contrasts for Height ([+high] /i, u/, [-high, -low] /ɛ, ɔ/, and [+low] /a/) and one for Place ([-back] /i, ɛ/, [+back] /u, ɔ, a/). We separated spectro-temporal and phonological predictors in two different models as, notwithstanding the existing correlation between formants and distinctive features, we wanted to determine whether the N1 amplitudes, latencies and ECD sources are better accounted for by acoustic gradient predictors or by distinctive features predictors.

Furthermore, we tested the hemispheric asymmetries for the late N1 on the mean amplitudes for the four strongest electrodes in each hemisphere, i.e., C3-FC5-F3-FC1 for the left and C4-FC6-F4-FC2 for the right hemisphere. The acoustic and phonological models were built by using Hemisphere and the acoustic (F1 and F2) or the Height and Place predictors as fixed effects and subjects as random effect. Visual inspection of residual plots did not reveal any evidence of deviations from homoscedasticity or normality. *p* values were obtained by likelihood ratio tests of the full model with effect in question against the model without that effect. We performed model comparison analysis on the base of previous literature (Baayen, 2008; Pinheiro & Bates, 2000) to investigate what model exhibits the best fit for the data. The best model will be the one with lower values of the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC), while the statistical significance ($\alpha = .05$) was evaluated using likelihood ratios (which provided as logarithm units, logLR) associated with a *p*-value.

2.6. *Source analysis*

Tridimensional topographical maps and an estimation of the early and late N1 intracranial sources were conducted using BESA 2000. We used the spatiotemporal source analysis of BESA that estimates location, orientation, and time course of the equivalent dipolar sources (ECD) by calculating the scalp distribution obtained for a given model (forward solution). This distribution was then compared to that of the actual AEPs. Interactive changes in source location and orientation led to the minimization of residual variance (RV) between the model and the observed spatiotemporal AEP distribution. The three-dimensional coordinates of each ECD in the BESA model were determined with respect to the Talairach axes. BESA assumed a realistic approximation of the head (based on the MRI template based on 24 subjects). The possibility of interacting ECDs was reduced by selecting solutions with relatively low ECD moments with the aid of an "energy" constraint (weighted 20% in the compound cost function, as opposed to 80% for the RV). The optimal set of parameters was found in an iterative manner by searching for a minimum in the compound cost function. Initially, on the grand average ERP for all vowels, a single dipole pair (symmetric in the left and right hemisphere) was fitted to the entire N1 range (90–160 msec). Then, it was compared with a two-dipole pair solution chosen to minimize overlap between the early and late N1 phase, fitting one dipole pair in the 90–125 msec range and the other at 135–160 msec. The two-pair solution gave the lower RV, and the addition of a third pair did not substantially decrease the RV (about .1%). For these reasons, the two-pair model was used for further analysis. To obtain a reliable and stable model of the early and late N1, a first model was made on the grand average AEP for all vowels using the two bilateral mirror symmetric pairs of ECDs on the basis of the topographical maps obtained here and in previous studies (McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2003; Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005, 2002) showing bilateral distribution on the auditory N1 component. Then, to compare statistically the N1 source localizations across vowels, the model was used as a starting point to model the AEP of each subject, fitting the source locations and orientations on the individual data. Only the source location (not the orientation) was used for the analyses. The accuracy of the source model was evaluated by measuring its RV as a percentage of the signal variance as described by the model and by applying residual orthogonality tests (ROT) (Bocker, Cornelis, Brunia, & Van den Berg-Lenss, 1994). The resulting individual time series for the ECD moments (the source waves) were subjected to an orthogonality test, referred to as a source wave orthogonality-test (SOT) (52). All t-statistics were evaluated for significance at the 5% level.

3. Results

3.1. *Waveforms and topographical maps*

In the N1 range, an early and a late peak were detected. The early N1 peaked at 125–135 msec on medial electrodes around the vertex (FCz, Cz, CPz), while the late N1 peaked at 145–155 msec over lateral electrodes of the left hemisphere (FC3, C3, CP3). Fig. 3 shows AEP waveforms of representative electrodes where activity was prominent. Fig. 4 represents the topographical mapping of the N1 elicited by each vowel. The early N1 topography coincided with the classical N1

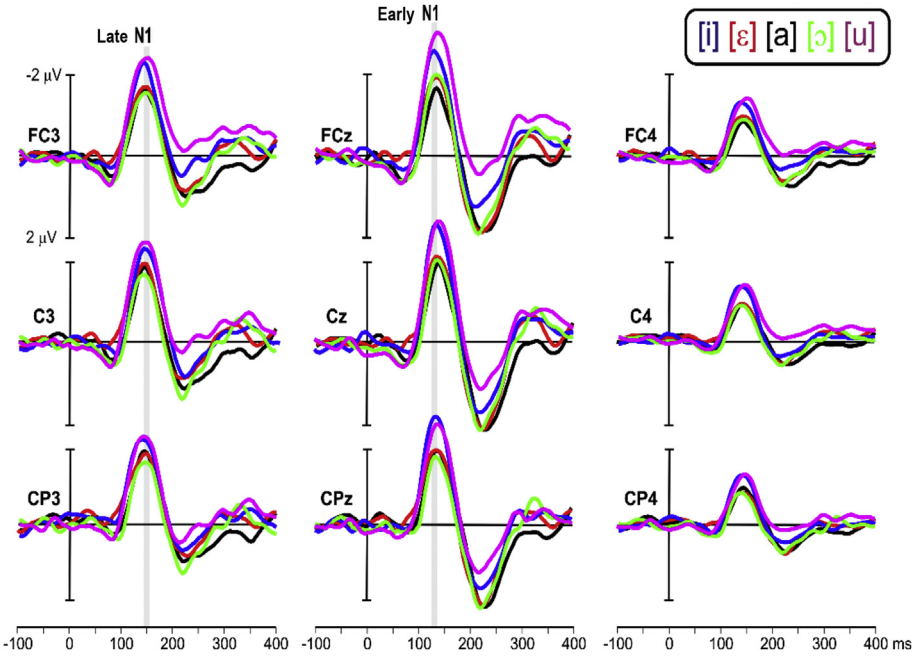

Fig. 3 – Grand average (N = 11) of the early and late N1 phases at the most representative electrode sites. The five vowels are superimposed using different colors. A vertical gray bar marks the time windows in which the N1 peaks.

distribution, focusing on medial central scalp areas with a tangential distribution (negative on the vertex and positive on bilateral temporal sites). This component was posterior for /a/ and /ɔ/ to /ɛ/, /i/, /u/. The late N1 was observed over the left central scalp with a radial distribution at the skull (its positive counterpart was not detectable from the scalp). The late N1 was more lateral and posterior for the [+low] /a/ and the [-high -low] /ɛ, ɔ/ to the [+high] /i, u/.

3.2. *Hemispheric asymmetry*

To test hemispheric asymmetries on the late N1 amplitudes, the Hemisphere effect was added in the models. Both acoustic [$\chi^2(1) = 91.1$, $p < .001$] and phonological [$\chi^2(1) = 93.9$, $p < .001$] models showed a leftward laterality. On average, late N1 amplitudes were −2.32 μV (SD ± .5) for the left and −1.11 μV (SD ± .3) for the right. The main effects of F1 [$\chi^2(1) = 4.6$,

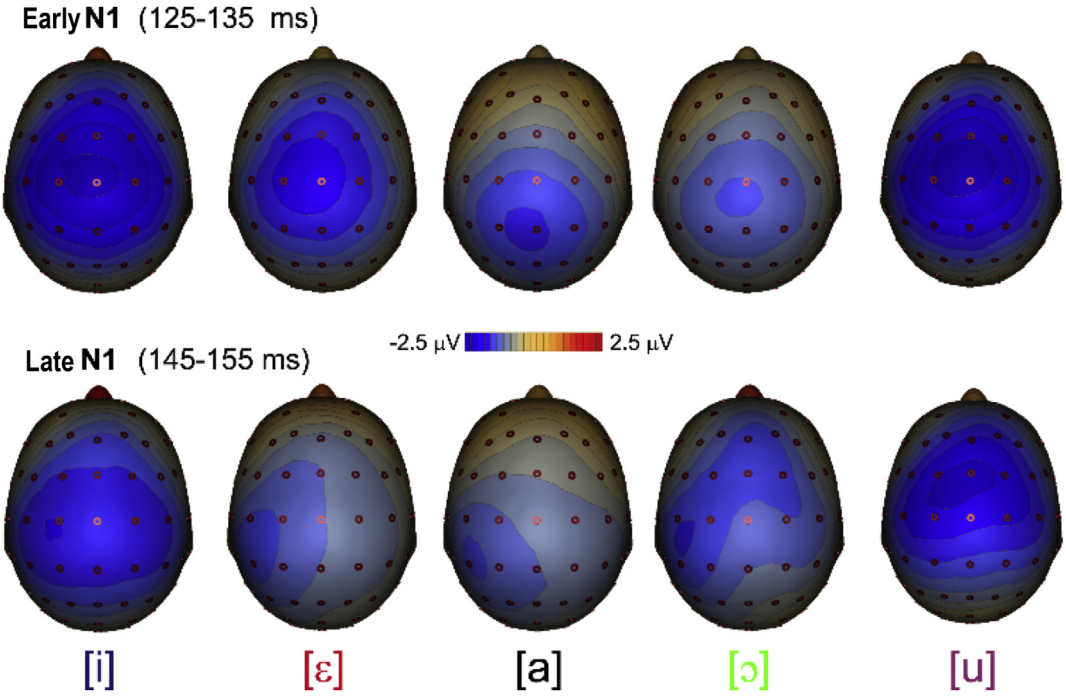

Fig. 4 – Early and late N1 topographical three-dimensional maps displayed from above.

$p = .031$] and Height [$\chi^2(2) = 8.9$, $p = .011$] and the interactions Hemisphere x F1 [$\chi^2(1) = 5.8$, $p = .016$] and Hemisphere x Height [$\chi^2(2) = 6.0$, $p = .048$] were statistically relevant. In the left hemisphere, the [+high] /i, u/ vowels, with low F1, elicited greater responses than the [-high -low] /ɛ, ɔ/ and [+low] /a/ vowels ($p < .001$). F2 [$\chi^2(1) = 21.5$, $p = .643$] and Place [$\chi^2(1) = 96.1$, $p = .327$] were not statistically relevant. Model comparison revealed that the phonological model provides a better fit for the data to the acoustic model (logLR = 2.444).

3.3. Amplitudes and latencies

Table 2 and Fig. 5(A, B) show early and late N1 amplitude and latency values. Fig. 5C represents the leftward laterality of the late N1 in respect to the early N1. For both early and late N1 amplitudes, the acoustic model showed a main effect for F1 [early N1: $\chi^2(1) = 10.5$, $p = .001$; late N1: $\chi^2(1) = 7.4$, $p = .006$] and the phonological model showed a main effect for Height [early N1: $\chi^2(2) = 1.2$, 1 $p = 0.$; late N1: $\chi^2(2) = 8.3$, $p = .015$]. That is, the amplitudes increase with decreasing F1 values of vowels (cf. Fig. 1).

In the phonological model, the early N1 responses to the [+high] /i, u/ elicited greater amplitudes than the [-high, -low] /ɛ, ɔ/ ($p = .003$) and [+low] /a/ ($p = .020$); however, the /ɛ, ɔ, a/ vowels did not statically differ ($p > .001$). These findings were partially paralleled by the late N1: responses to /i, u/ elicited greater amplitude than /ɛ, ɔ/, but responses to /i, u/ were not different from /a/ responses ($p = .081$). Again, the vowels /ɛ/, /ɔ/, and /a/ did not statically differ ($p > .992$). F2 and Place were not statistically relevant [early N1: F2 ($\chi^2(1) = 966$, $p = .756$; Place ($\chi^2(2) = 52.9$, $p = .467$; late N1: F2 ($\chi^2(1) = 5090$, $p = .943$; Place ($\chi^2(2) = 3.7$, $p = .540$]. The phonological model provided a better fit for the early N1 data (logLR = 1.176), whereas the acoustic model provided a better fit for the late N1 data (logLR = .837).

As for latency, the acoustic model did not show significant effects for the early N1 data [F1: ($\chi^2(1) = 3.1$, $p = .077$; F2: ($\chi^2(1) = 3.5$, $p = .059$]. The phonological model revealed a better goodness of fit (logLR = 3.708), showing a significant effect for Place [$\chi^2(1) = 4.8$, $p = .028$): the [+back] /a, ɔ, u/ were, on average, 3.12 msec later than the [-back] /i, ɛ/. Height was not statistically relevant ($\chi^2(1) = 5.9$, $p = .050$]. Statistics for late N1 values showed a main effect for F2 [$\chi^2(1) = 9.0$, $p = .003$] and Place [$\chi^2(1) = 7.7$, $p = .005$], confirming that the [+back] vowels with low F2 values were later than the [-back] vowels (on average, 4.9 msec). The F1 and Height predictors were not statistically relevant [F1: ($\chi^2(1) = .0413$, $p = .839$; Height: ($\chi^2(2) = .59.7$, $p = .742$]. The acoustic model fitted slightly better than the phonological model (logLR = .151).

3.4. ECD localization

Table 3 shows the source coordinates for the five vowels and the two components. The intracranial localization of the N1 sources for the five vowels is shown in Fig. 6. The waves represent the time course of those sources in both hemispheres (averaged across vowels). For all vowels, the early N1 was bilaterally localized within the primary auditory cortex in the Brodmann area (BA41). The early N1 time-course showed that this component initiated at 80 msec and peaked at 130 msec with equal intensity in the two hemispheres. The late N1 was localized more ventrally and anteriorly within the STG in the BA22. The late N1 time-course revealed that this component initiated at 110 msec and peaked at 150 msec, and that it was much larger in the left hemisphere than in the right [t(10) = 23.4, $p < .0001$]. The early N1 dipole orientation was mostly radial, pointing towards the vertex, while the late N1 dipole orientation was more radial, also showing foci over the bilateral temporal areas.

In Fig. 7(A–D), the early and late N1 bivariate source distribution along the lateral-medial/anterior-posterior and superior-inferior/anterior-posterior axes are represented in two-dimensional planes.

3.4.1. Lateral-medial dimension (x)

The acoustic model for the early N1 ECD showed F1 [$\chi^2(1) = 9.82$, $p = .002$] and F2 [$\chi^2(1) = 6.49$, $p = .011$] effects. This suggests that the /a, ɔ, u/ vowels close in the F2–F1 dimension [i.e., [+back] vowels]] elicited lateral source locations to the /ɛ, i/vowels with larger inter-format distances [i.e., [-back] vowels]. In the phonological model, effects for Height [$\chi^2(2) = 7.07$, $p = .029$] and Place [$\chi^2(1) = 8.40$, $p = .004$] emerged. The Height effect indicated that the sources of the [-low] /a/ were, on average, 4 mm lateral to the [+high] /i, u/ ($p = .022$), whereas the other vowels were not cortically distinguished ($p > .001$); the Place effect evinced that the [+back] vowels were more lateral than the [-back] vowels. The phonological model fitted better for the early N1 data (logLR = 1.619).

For the late N1 data, the acoustic model showed a main effect for F1 [$\chi^2(1) = 31.5$, $p < .001$) and F2 ($\chi^2(1) = 1.26$, $p = .262$]:

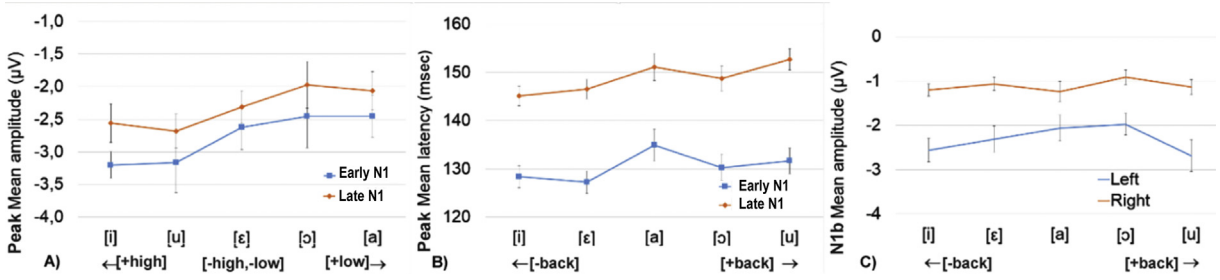

Fig. 5 – Amplitudes, latencies, and hemispheric asymmetries of the N1. A) Mean amplitudes of the early and late N1 phases for [+high] /i/, /u/, [-high, -low] /ɛ/, /ɔ/, and [+low] /a/ vowels. B) Mean latency of the early and late N1 phases for [+back] /a, ɔ, u/ and [-back] /i, ɛ/, vowels. C) Hemispheric asymmetries in the amplitude of the late N1 phase as functions of the perceived vowels.

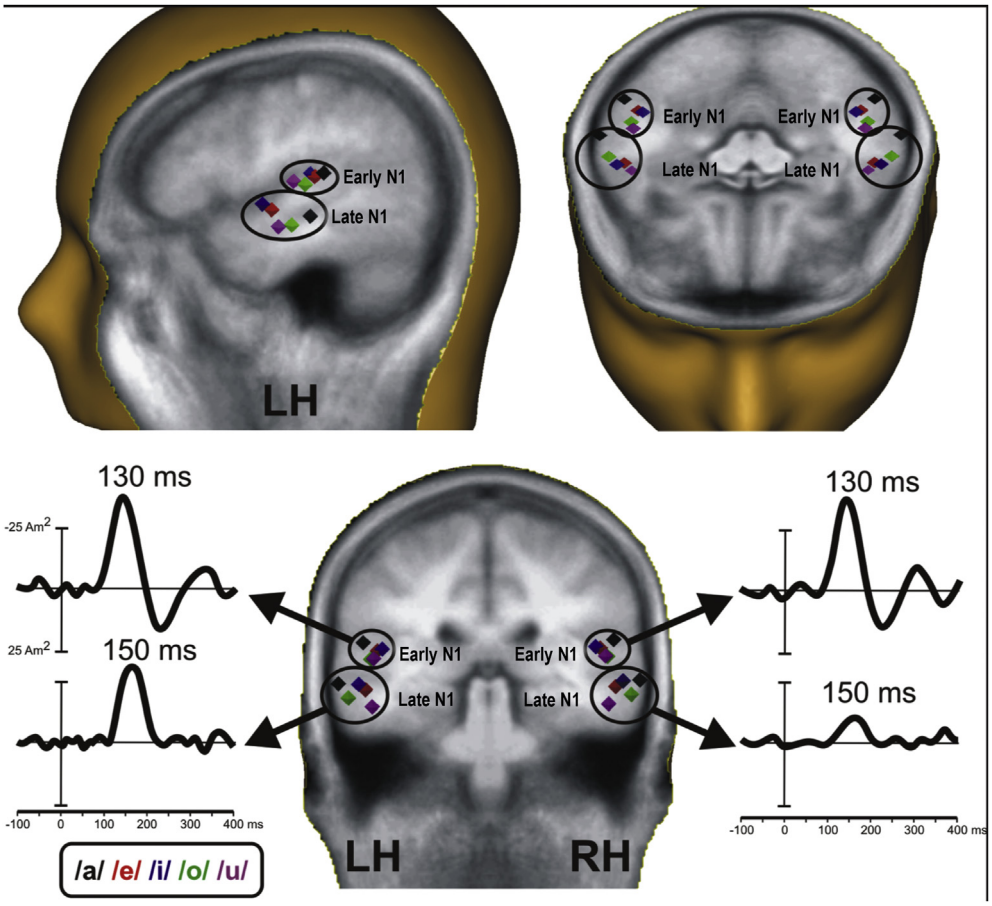

Fig. 6 – Early and late N1 source locations and time-course. Vowel representation is coded by colored dots: /i/ blue, /ɛ/ red, /a/ black, /ɔ/ green, /u/ purple. LH = left hemisphere, RH = right hemisphere.

the highest F1 values (e.g., /a/) were collocated anteriorly and the lowest F2 values (e.g., /u/) medially. The phonological model evidenced effects for Height [$\chi^2(2) = 27$, $p < .001$]. Post-hoc comparisons showed that the [+low] /a/ was at the most lateral position—on average, 7 mm to [-high, -low] /ɛ, ɔ/ and 12 mm to [+high] /i, u/—and that the [-high, -low] vowels were 4 mm lateral to the [+high] vowels ($p < .001$). The interaction Height × Place [$\chi^2(1) = 29.2$ $p < .001$] showed that: (i) /a/ and /ɔ/ were not cortically distinguished; (ii) within the [-high -low] vowels, the [-back] /ɛ/ was medial to the [+back]/ɔ/; (iii) within the [+high] vowels, the [-back] /i / was lateral to the [+back] /u/. The phonological model better fits the late N1 data (logLR = 14.5).

3.4.2. *Anterior-posterior dimension (y)*

The early N1 acoustic model showed effects for F1 [$\chi^2(1) = 76.1$, $p < .001$] and F2 [$\chi^2(1) = 51.7$, $p < .001$]: they indicated that vowels with high F1 (i.e., /a/) and vowels with high F2 (i.e., /ɛ/ and /i/) tended to elicit ECDs at posterior locations in respect of /ɔ/ and /u/. In the phonological model, a main effect for [$\chi^2(2) = 84.8$, $p < .001$] and [$\chi^2(1) = 45.9$, $p < .001$] emerged. On average, the [+low] /a/ was at the most posterior position – on average 8 mm to the [-high, -low] /ɛ, ɔ/ and 11 mm to the [+high] /i, u/ ($p < .001$); in their turn, the [-high, -low] /ɛ, ɔ/ were 3 mm posterior to the [+high] /i, u/. Moreover, the [+back] vowels /ɔ, u/ were anterior to the [-back] vowels /ɛ, i/. The interaction Height × Place [$\chi^2(1) = 4.2$, $p = .038$] revealed Place effects within the [-high, -low] and [+high] vowels, so that /ɛ/ was posterior to /ɔ/ and /i/ was posterior to /u/. The phonological model better described the source data (logLR = 2.86). With regard to the late N1 data, the acoustic model revealed effects for the F1 [$\chi^2(1) = 96.4$, $p < .001$] and F2 [$\chi^2(1) = 47.8$, $p < .001$], which means that vowels close in the F2–F1 dimension (i.e., the [+back] /a, ɔ, u/) elicited posterior ECDs to vowels with larger inter-formant distances (i.e., [-back] /ɛ, i/). In the phonological model, we found that Height [$\chi^2(2) = 111$, $p < .001$] and Place [$\chi^2(1) = 89.2$, $p < .001$] predictors affected the ECD patterns. On average, the [+low] /a/ was at the most posterior location—about 8 mm to [-high, -low] /ɛ, ɔ/ and 13 mm to [-high] /i, u/; in their turn, /ɛ, ɔ/ were 3 mm posterior to /i, u/ ($p < .001$). Crucially, contrary to the early N1 sources, the [+back] /a, ɔ, u/ were, on average, 8 mm posterior to the [-back] /ɛ, i/. In addition, the interaction of Height × Place [$\chi^2(1) = 5.3$, $p = .021$] indicated the Place effects within [-high, -low] and [+high] vowels. Contrary to the N1a sources, the [-back] /ɛ/ was anterior to the [+back] /ɔ/ and the [-back] /i/ was anterior to the [+back] /u/. The phonological model fitted the data better than the acoustic model (logLR = 22.0).

3.4.3. *Inferior-superior dimension (z)*

The early N1 was affected by F1 [$\chi^2(1) = 18.5$, $p < .001$] and F2 [$\chi^2(1) = 23.9$, $p < .001$] in the acoustic model. Vowels with higher F1 and higher F2 values (e.g., /a/, /ɛ/, and /i/) tended to

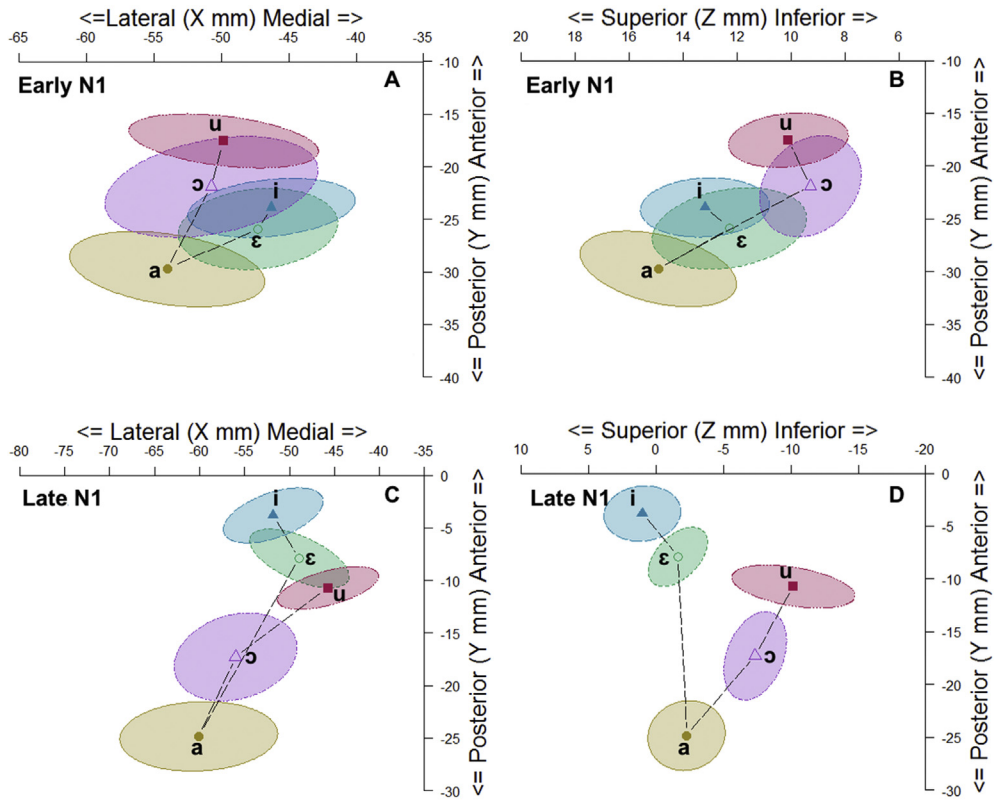

Fig. 7 – (A–D). Early N1 and late N1 source locations in the two-dimensional plane determined by the Lateral-Medial/ Posterior-Anterior and Superior-Inferior/Posterior-Anterior axes. Mean values are represented by dots. 68.27% confidence ellipses of the source location for each vowel (corresponding to ±1 SD from the bivariate mean).

be generated in the superior ECDs. The phonological model provided a better fit for the data (logLR = 16.13), evincing a main effect for Height [$\chi^2(2) = 59.9$, $p < .001$]: the [+low] /a/ was situated at the most superior location ($p < .001$), but the [-high, -low] /ɛ, ɔ/ were not cortically distinguished from the [+high] /i, u/ ($p = .111$). Also, an effect for Place was evident: the [+back] /a, ɔ, u/ were, on average, 3 mm inferior to the [-back] /i, ɛ/ [$\chi^2(1) = 37.1$, $p < .001$]. Moving to the late N1, the acoustic model highlighted clear effects for F1 [$\chi^2(1) = 25.7$, $p < .001$] and F2 [$\chi^2(1) = 96.2$, $p < .001$]. Again, this suggests that the vowels /a, ɔ, u/, with F2–F1 close dimension, were inferior to the vowels /i, ɛ/ with larger F2–F1 distances. In the phonological model, the Place effect was significant [$\chi^2(1) = 79.1$, $p < .001$]: the [+back] vowels were inferior to the [-back] vowels by 8 mm on average. Moreover, a significant interaction Height x Place was noticeable [$\chi^2(1) = 22.4$, $p < .001$]. Within the [+back] vowels, /ɔ/ is superior to /u/; within the [-back] vowels, /ɛ/ is inferior to /i/. The vowels /a/ and /ɛ/ were not cortically separated ($p = .92$), while /a/ was statistically different from /ɔ/. The phonological model better fitted the data (logLR = 3.47).

4. Discussion

Three are the novel findings of the present study. First of all, we found evidence for different hierarchical indexes structuring vowel representation within the N1 component: the early N1 peaking at 125–135 msec in the A1 (BA41) and the late N1 peaking at 145–155 msec in the STG (BA22). Secondly, these components are characterized by hemispheric asymmetries: the early N1 shows a bilateral activity, while the late N1 shows a leftward preponderance. Finally, hierarchical and hemispheric modulation of the early and late N1 shed light on the encoding of spectro-temporal properties of vowels into distinctive feature representations through the tonotopic activation of lateral-medial, anterior-posterior, and inferior-superior gradients. These N1 source localizations should be taken with caution, because of the well-known EEG low spatial resolution; however, previous studies successfully localized the N1 sources within the A1 and the STG (e.g., McDonald et al., 2003; Teder-Sälejärvi et al., 2005, 2002; Weise, Schröger, & Horváth, 2018).

4.1. *Early and late N1 phase*

According to the literature, the N1 is not a unitary event. Scalp distribution of the N1 responses to clicks, noise, bursts, and tones hint at least three components (Näätänen & Picton, 1987). Not all N1 components are, however, tonotopically organized (Woods, 1995). The first component is maximally recorded from the fronto-central scalp, peaks between 85 and 110 msec, is generated by tangentially oriented currents in both A1, and shows tonotopic sources (Hari, Aittoniemi, Järvinen, Katila, & Varpula, 1980; Wood & Wolpaw, 1982). The second component is detectable at approximately

150 msec in the mid-temporal scalp regions and is generated by radially oriented neuronal sources in the STG with tonotopic distribution. Due to the radial orientation of the underlying current dipole, this component is not picked up with MEG (Näätänen & Picton, 1987, p. 386). The third component is a negative wave at the vertex at 100 msec whose generators are not known (Inui, Okamoto, Miki, Gunji, & Kakigi, 2006; Picton, Campbell, Baribeau-Braun, & Proulx, 1978; Wolpaw & Penry, 1975; Wood & Wolpaw, 1982). The early and late N1 we found present patterns that are in accordance with the first and second components previously hypothesized (the N1′/P90 and N1c according to Woods's, 1995 classification). However, our early N1 shows slightly longer peak latency than previous data (we will turn to this finding in the following paragraph). As far as we know, this is the first study that found clear evidence for these hypothesized components in humans for speech sounds. Probably, early MEG studies failed to report different intracranial origins of the N1 events, as the N1m sources were generally modeled by a single ECD in each hemisphere and, more importantly, the source analysis was confined to the rising slope and peak of the N1m component, without taking solutions at the N1m peak or after it (Obleser et al., 2004, 2003; Sharinger et al., 2011). Conversely, we modeled the N1 using two bilateral mirror symmetric pairs of ECDs simultaneously, and, crucially, we included the whole N1 duration (from 100 to 160 msec) in the analysis, which permitted us to capture the multiple (temporally-differentiated) N1 cortical generators. Also, it is very likely that the EEG sensitivity to radial and tangential ECDs – as compared to MEG, which is blind to radially oriented ECDs (Eulitz et al., 2004; Seppo, Han, Belliveau, & Hämäläinen, 2010) – permitted us to separate the activity related to vowel encoding. Although, as we noted in the Introduction, EEG localization may be more accurate than MEG localization (Liu et al., 2002) and achieve even more localized source analysis than with whole-head planar gradiometer MEG devices (Malmivuo et al., 1997), our findings need further investigations in order to be corroborated. Overall, our data suggest that combining EEG with MEG would represent the ideal approach to an in-depth investigation of speech processing. To deeply understand the hemispheric modulation of the early and late N1 phases we need to discuss amplitudes, latency and especially source data.

4.2. *Amplitudes, latency and source data*

As observed before, amplitudes show broad F1 and Height encoding processes in both early and late N1: amplitudes increase with decreasing F1 values so that the [+high] vowels/i, u/elicited greater amplitudes than non-high vowels (Obleser et al., 2003; Scharinger et al., 2011; Shestakova, Brattico, Soloviev, Klucharev, & Huotilainen, 2004). For latencies, the acoustic model revealed a significant effect only for the late N1: vowels with low F2 (/a, ɔ, u/) were later than vowels with high F2 (/i, ɛ/). Instead, the phonological model shows effects for both early and late N1, revealing that the [+back] vowels /a, ɔ, u/ peaked later than the [-back] /i, ɛ/ (Obleser et al., 2004a, b). Crucially, the phonological model better fitted the early N1, while the acoustic model better fitted the late N1 amplitudes and latencies, suggesting that Height and Place distinctive features are encoded early in the A1.

This finding is confirmed by the source data, which offer a fine-grained picture. Previous studies showed that N1m ECDs are dependent on both spectro-temporal cues and distinctive features (Manca & Grimaldi, 2016). The lateral-medial axis showed medial locations for sounds with high frequencies or lateral positions for close F2–F1 distances, so that the [+back] vowels (with small F2–F1 intervals) are, as a result, more lateral than the [-back] vowels (Eulitz et al., 2004; Obleser, Elbert, & Eulitz, 2004b). Also, it has been found that the [+round] vowels (with low F2) elicit more lateral sources (Scharinger et al., 2011). The anterior-posterior plane seems responsive to F1 and F2 values associated with Height and Place, so the [+high] vowels are more anterior than the [-high] vowels, and the [+back] vowels are more posterior than the [-back] vowels (Obleser et al., 2004a; Scharinger et al., 2011). The inferior-superior axis showed sensitivity to F1 and Height; it has been found that low vowels are superior to high vowels (Obleser et al., 2003) but the reverse pattern seems true only for [-back] vowels (Scharinger, Monahan, & Idsardi, 2012). Yet the sources of rounded vowels turn out to be inferior to non-rounded vowels (Scharinger et al., 2011). We replicated these tonotopic data, adding new representational patterns thanks to the early and late N1 hierarchical-hemispheric modulation.

4.3. *The hierarchical-hemispheric modulation: from acoustic to distinctive features*

Of note is the fact that the phonological model provides a better fit for both early and late N1 ECDs along all tonotopic gradients. This suggests that: (i) distinctive features are better predictors than acoustic F1 and F2 patterns for vowel representations; (ii) abstract processes begin early in the A1 in both hemispheres and carry on in the left STG. In fact, recent studies have progressively eroded the paradigm that considers the A1 as only "sensory analytic" and therefore ruled out from cognitive processes (see Weinberger, 2015; Bernal & Ardila, 2016 for a review). The extensive research on the A1 over the past 10 years is incompatible with the view that its function is limited to the analysis of acoustic stimuli independent of their acquired cognitive significance. In brief, responses to A1 neurons reflect both the physical and cognitive properties associated with learning and memory processes. Also, damage in the A1 is associated with so-called "pure word-deafness." Patients with this syndrome have difficulties discriminating phonemic contrasts as, for example, voiceless to voiced stop consonants (Brody, Nicholas, Wolf, Marcinkevich & Artz, 2013). We will further discuss this problem in the next paragraph.

The acoustic model provides a general result concerning the encoding of the F2–F1 relation that specifies vowels for Place: tongue retraction lowers F2 frequencies, reducing F2–F1 distances. As in early studies (Eulitz et al., 2004; Obleser et al., 2004b), the [+back] vowels /a, ɔ, u/, with close F2–F1, are more lateral in the early N1 (Fig. 7A) and more posterior and more inferior in the late N1 (Fig. 7B) than the [-back] vowels /i, ɛ/. This finding is also in line with an ECoG study that investigated the phonological American English system (Mesgarani et al., 2014): the STG electrodes showed a selective response to

F2–F1 differences separating low-back, low-front, and high-front vowels. In our study, the ECD patterns found that the F2–F1 parameters are preserved within the phonological model, which shows that the [+back] vowels are more lateral in the early N1 and more posterior and more inferior in the late N1. However, in both our study and previous studies, the acoustic model fails to adequately capture the tonotopic mapping of the three tongue heights marking the SI vowel system (as well as the American English vowel system which also differentiates between tense and lax vowels). Our phonological model caught these contrasts and, more importantly, elucidated the dynamical nature of features representation thanks to Height × Place interactions affecting source ECDs. These interaction effects between Height and Place were not found even when the same statistical models adopted here were employed with MEG to study the Turkish vowel system (Scharinger et al., 2011).

The early and late N1 anterior-posterior gradients highlight that the [+low] /a/ results at the most posterior position and the [-high, -low] /ɛ, ɔ/ are significantly posterior to the [+high] /i, u/. Crucially, the interaction Height x Place effects reveal source ECDs selectivity within the [-high, -low] and [+high] vowels: in the early N1, the [-back] /ɛ/ is posterior to the [+back] /ɔ/ and the [-back] /i/ is posterior to the [+back] /u/. In the late N1, the reverse pattern holds, because the /ɛ, i/ vowels reach a more anterior position during early and late N1 hierarchical-hemispheric modulation (Fig. 7D): /ɛ/results anterior to /ɔ/and /i/anterior to /u/(as already found for the Turkish system; Scharinger et al., 2011). A further modulation Height x Place is noticeable in the inferior-superior gradient, where again, moving from early to late N1, the /ɛ, i/ vowels reach a superior position (Fig. 7D). This modulation selectively separates ECDs for Height contrasts within the [+back] /ɔ, u/ and [-back] /ɛ, i/, so that the [-high, -low] /ɔ/ is superior to the [+high] /u/ and the [-high, -low] /ɛ/ is inferior to the [+high] /i/. Overall, our results suggest that computational processes leading to abstract representations of SI vowels warp the spatial arrangement of neuronal sources according to distinctive features in such a way that interaction effects between Height and Place play a crucial role in phonemotopic encoding of vowels.

4.4. *The acoustic and the phonological models in light of the analysis by synthesis theory*

However, we have to note that feature variables, because they are characterized by binary values, are per se discrete compared to continuous acoustic variables, leading the phonological model to better fit in statistical comparison. An in-depth interpretation of these facts may be reached in light of the Analysis by Synthesis theory (Halle & Stevens, 1962; Stevens, 2002). This theory assumes that cues from the input signal trigger guesses about the identity of phonemes; then the internal synthesis of predicted phonological representation is compared with the input spectrum generated by the auditory periphery. In this way, perceptual analysis contains a step of synthetically generated candidate representations in linked bottom-up/top-down fashion. In this view, the goal of speech perception is to convert the acoustic features into distinctive features. The model implies that the same system implicated for production is involved in speech perception, suggesting that the motor system contributes dynamically in a predictive manner (Poeppel & Monahan, 2011). From this perspective, the statistical approach we employed catches the continuous process that recodes acoustic patterns into neurophysiological patterns at the cortical level. That is, phonemotopic representations dynamically emerge between primary and secondary auditory areas in terms of distinctive features. On the contrary, acoustic representations are prevalently confined in the auditory periphery. This would lead the phonological model to better fit the data. Contrary to previous data, we found another interesting result, i.e., that the early N1 in the A1 shows slightly longer peak latency. As a reviewer pointed out, this would suggest that interconnections with other areas are already ongoing at this point, inhibiting different regions of A1 to enhance those frequencies that are more expected. According to the Analysis by Synthesis model, we hypothesize that this may be due to the activity of the motor areas recruited at this stage to predictively generate audio-motor representations that are compared with long-term memory representations (i.e., distinctive features). In fact, it has been clearly shown that large distributed motor regions are dynamically recruited for specific computational reasons during speech perception (Hickok & Poeppel, 2007; Rampinini et al., 2017; Schwartz, Basirat, Menárd, & Sato, 2012; Skipper, Devlin, & Lametti, 2017), although we cannot exclude that the active task employed in the present experiment has contributed to more active audio-motor processes. In brief, the input signal in the auditory periphery leads to predictive processing that compares acoustic representations with phonological representations via audio-motor representations: in the A1 (early N1), a preliminary phonemotopic map of audio-motor features is generated along the lateral-medial gradient for Place and the anterior-posterior gradient for Height and Place; this map is better refined in the STG (late N1) along the anterior-posterior and inferior-superior gradients.

Another possible interpretation of these findings might be that clusters of neurons in the A1 generate phonetic representation of vowels, which are neither fairly analog to the acoustic signal nor yet abstract phonological representations. At this stage, it is likely that auditory and motor areas are jointly recruited in searching for audio-motor representations to be linked with distinctive feature representations: this activity probably contributed to longer peak latency in the A1. This kind of phonetic maps would allow organizing speech sounds into linguistically-relevant categories leading to a provisional representation of distinctive features (as suggested by Fig. 7A–B where the cortical generators overlap). The dynamical computations of the signal in the left STG produces discretized representations of vowels according to a finer interaction between Height and Place features (as suggested by Fig. 7C where the cortical generators define discrete neuronal spaces). Overall, our data suggest that it is necessary to increase investigations in this direction in order to understand in depth how vowels and consonants are computed and represented by the brain.

The dynamical and hierarchical conversion of acoustic features onto phonemotopic maps is represented in Fig. 8, where multidimensional scaling shows how the relational organization between vowel centroids in the acoustic space (see Fig. 1) is mirrored in the neural dimension. It is evident

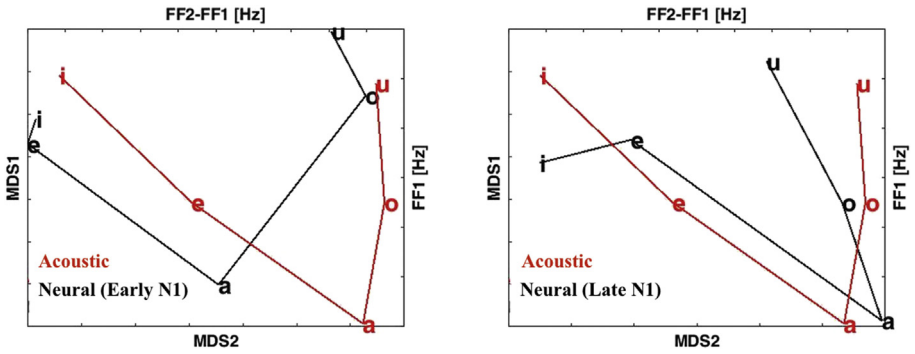

Fig. 8 – Multidimensional scaling (MDS) of acoustic and neural space (realized with Matlab Statistics Toolbox, 2012).

how the initial stage of phonemotopic (or phonetic) mapping in the A1 is well refined in the STG.

4.5. *Vowel encoding and hemispheric lateralization*

Our findings have important implications for current theories on speech hemispheric lateralization. The data discussed above contrast with models hypothesizing that spectro-temporal analysis of speech sounds is bilaterally performed in the A1 (Poeppel, 2003) and then phonological computations are left lateralized, but also with models suggesting that the spectro-temporal analysis is carried out bilaterally in the STG, while phonological processing is carried out in the left STS (Hickok & Poeppel, 2007) or with other points of view stressing the exclusive contribution of the left STG in phonological representations (DeWitt & Rauschecker, 2012; Scott, Blank, Rosen, & Wise, 2000; Scott & Johnsrude, 2003; Scott & McGettigan, 2013). We maintain that the initial stage of speech encoding is bilaterally performed in the A1: however, we showed that at this level the cortical spatial arrangement is already warped by phonemotopic or at least phonetic patterns according to distinctive features principles. So, it is probable that spectro-temporal analysis, previously attributed to the A1, is peculiar to the cochlea–brainstem pathways (until the proximity of the A1): here, properties of the speech waveform are mirrored with remarkable fidelity (Bidelman, Moreno, & Alain, 2013). Conversely, late cortical evoked activities, from the A1 to the STG, progressively encode the phonetic-phonological features necessary to generate categorical speech percepts. Indeed, our data suggest that along the bilateral A1 and the left STG multiple (parallel) representations of vowels are formed, leading to the progressive conversion of the acoustic signal into categorical patterns through hierarchical reshaping of neuronal maps along the lateral-medial, anterior-posterior, and inferior-superior gradients. This dynamical interface between the A1 and the STG generates the encoding of Place and Height features for SI vowels.

5. Conclusions

Overall, the findings of the present study suggest that vowel discretization is the result of a continuous process that converts the incoming acoustic signal into neurophysiological signal. In particular, we hypothesize that the spectro-temporal states characterizing vowels are continuously converted into appropriate neurophysiological states (Grimaldi, 2018). In this way, properties of the spectro-temporal states undergo changes interacting with the neurophysiological states until synchronized synapses, distributed within the A1 and the STG, are generated. From this perspective, the classical distinction between bottom–up processes reflecting acoustic differences and top–down processes reflecting distinctive feature representations should be reinterpreted as a continuous-dynamical process involving changes of physical states (spectro-temporal states into neurophysiological states) where progressive structure and property rearrangements result in categorical representation of vowels according to distinctive features specifications.

Open practices

The study in this article earned Open Materials and Open Data badges for transparent practices. Materials and data for the study are available at https://osf.io/qwa42/view_only=b6377f9def604d7faa4977fc9264580d/.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any competing financial, commercial and non-financial interests that might be perceived to influence the results and/or discussion reported in this paper.

CRediT authorship contribution statement

Anna Dora Manca: Methodology, Validation, Investigation, Data curation, Writing - review & editing. **Francesco Di Russo:** Methodology, Validation, Data curation, Writing - review & editing. **Francesco Sigona:** Software, Formal analysis, Data curation, Writing - review & editing. **Mirko Grimaldi:** Conceptualization, Resources, Supervision, Project administration, Funding acquisition, Methodology, Validation, Data curation, Writing - original draft, Writing - review & editing.

Acknowledgements

We wish to thank two anonymous reviewers for their stimulating comments on the manuscript permitting us to improve the interpretation of the data. MG is very grateful to Andrea Calabrese for illuminating discussions on many issues of this research: he taught him to generalize starting from complex data. We would like to thank also Philip Monahan, Elvira Brattico, and Giovanna Marotta for their useful comments on previous versions of the manuscript. This research was supported by the Italian Ministry of Education, University and Research. Grant. No. 20128YAFKB006.

REFERENCES

Ahlfors, S. P., Han, J., Belliveau, J. W., & Hämäläinen, M. S. (2010). Sensitivity of MEG and EEG to source orientation. *Brain Topography, 23*, 227–232.

Baayen, R. H. (2008). *Analyzing Linguistic Data. A practical introduction to statistics*. Cambridge: Cambridge University Press.

Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience, 20*, 327–339.

Bates, D., Maechler, M., & Bolker, B. (2015). Walker S fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bernal, B., & Ardila, A. (2016). From hearing sounds to recognizing phonemes: Primary auditory cortex is A truly perceptual language area. *AIMS Neuroscience, 3*(4), 454–473. https://doi.org/10.3934/Neuroscience.2016.4.454.

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage, 79*, 201–212.

Bocker, K. B. E., Cornelis, H. M., Brunia, C. H. M., & Van den Berg-Lenss, M. M. C. (1994). A Spatiotemporal dipole model of the stimulus preceding negativity prior to feedback stimuli. *Brain Topography, 7*, 71–88.

Boersma, P., & Weenink, D. (2011). *Praat: Doing phonetics by computer (computer program), Version 5.2*. http://www.praat.org/.

Brody, R. M., Nicholas, B. D., Wolf, M. J., Marcinkevich, P. B., & Artz, G. J. (2013). Cortical deafness: A case report and review of the literature. *Otology and Neurotology, 34*, 1226–1229.

Calabrese, A. (1995). Constraint-based theory of Phonological markedness and simplification procedures. *Linguistic Inquiry, 2*(26), 373–463.

Cohen, D., & Halgren, E. (2003). Magnetoencephalography (neuromagnetism). In *Encyclopedia of neuroscience*. Amsterdam: Elsevier.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences United States of America, 109*(8), 505–514.

de Boer, B. (2001). *The origins of vowel systems*. Oxford: Oxford University Press.

Embick, D., & Poeppel, D. (2015). Towards a computational(ist) neurobiology of language: *Correlational, integrated*, and *explanatory* neurolinguistics. *Language and Cognitive Neuroscience, 30*(4), 357–366. https://doi.org/10.1080/23273798.2014.980750.

Eulitz, C., Obleser, J., & Lahiri, A. (2004). Intra-subject replication of brain magnetic activity during the processing of speech sounds. *Cognitive Brain Research, 19*, 82–91.

Gage, N., Poeppel, D., Roberts, T. P. L., & Hickok, G. (1998). Auditory evoked M100 reflects onset acoustics of speech sounds. *Brain Research, 814*(1), 236–239.

Grimaldi, M. (2009). Acoustic correlates of phonological microvariations. The case of unsuspected highly diversified metaphonetic processes in a small area of Southern Salento (Apulia). In D. Tock, & W. L. Wetzels (Eds.), *Romance languages and linguistic theory 2006* (pp. 89–109). Amsterdam: Benjamins.

Grimaldi, M. (2012). Toward a neural theory of language: Old issues and new perspectives. *Journal of Neurolingusitics, 25*(5), 304–327. https://doi.org/10.1016/j.jneuroling.2011.12.002.

Grimaldi, M. (2018). The phonetics-phonology relationship in the neurobiology of language. In P. Petrosino, P. Cerrone, & H. van der Hulst (Eds.), *From sounds to structures: Beyond the veil of maya* (pp. 66–104). Berlin: Mouton De Gruyter.

Grimaldi, M., Manca, A. D., & Di Russo, F. (2016). Electroencephalography evidence of vowels computation and representation in auditory cortex. In A. M. Di Sciullo (Ed.), *Biolinguistic investigations on the language faculty* (pp. 80–100). Amsterdam: Benjamins.

Halle, M. (2002). *From memory to speech and back: Papers on phonetics and phonology 1954–2002*. Berlin: Mouton de Gruyter.

Halle, M., & Stevens, K. N. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory, 8*, 155–159. https://doi.org/10.1109/TIT.1962.1057686.

Hari, R., Aittoniemi, K., Järvinen, M. L., Katila, T., & Varpula, T. (1980). Auditory evoked transient and sustained magnetic fields of the human brain localization of neural generators. *Experimental Brain Research, 40*(2), 237–240.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Neuroscience, 8*, 393–402.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biomedical Journal, 50*(3), 346–363.

Inui, K., Okamoto, H., Miki, K., Gunji, A., & Kakigi, R. (2006). Serial and parallel processing in the human auditory cortex: A magnetoencephalographic study. *Cerebral Cortex, 16*, 18–30.

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping, 16*, 47–62.

Lütkenhöner, B., Krumbholz, K., & Seither-Preisler, A. (2003). Studies of tonotopy based on wave N100 of the auditory evoked field are problematic. *Neuroimage, 19*, 935–949.

Malmivuo, J., Suihko, V., & Eskola, H. (1997). Sensitivity distributions of EEG and MEG measurements. *Biomedical Engineering IEEE Transactions, 44*, 196–208.

Manca, A. D., & Grimaldi, M. (2016). Vowels and consonants in the brain: Evidence from magnetoencephalographic studies on the N1m in normal-hearing listeners. *Frontiers in Psychology, 7*, 1413. https://doi.org/10.3389/fpsyg.2016.01413.

MATLAB and statistics Toolbox release. (2012). Natick, Massachusetts, United States: The MathWorks, Inc.

McDonald, J. J., Teder-Sälejärvi, W. A., Di Russo, F., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by crossmodal spatial attention. *Journal of Cognitive Neuroscience, 15*, 10–19.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science, 343*, 1006–1010.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*, 375–425.

Obleser, J., Elbert, T., & Eulitz, C. (2004b). Attentional influences on functional mapping of speech sounds in human auditory cortex. *BMC Neuroscience, 5*, 24. https://doi.org/10.1186/1471-2202-5-24.

Obleser, J., Elbert, T., Lahiri, A., & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity

determined by formant frequencies. *Cognitive Brain Research, 15*, 207–213.
Obleser, J., Lahiri, A., & Eulitz, C. (2004a). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience, 16*, 31–39. https://doi:10.1162/089892904322755539.
Ohl, F. W., & Scheich, H. (1997). Orderly cortical representation of vowels based on formant interaction. *Proceedings of the National Academy of Sciences United States of America, 94*, 9440–9444.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychology, 9*(1), 97–113.
Peelle, J. E. (2012). The hemispheric lateralization of speech processing depends on what "speech" is: A hierarchical perspective. *Frontiers in Human Neurosciences, 6*, 309. https://doi.org/10.3389/fnhum.2012.00309.
Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of Acoustical Society of America, 242*, 175–184.
Picton, T. W., Campbell, K. B., Baribeau-Braun, J., & Proulx, G. B. (1978). The neurophysiology of human attention: A tutorial review. In J. Requin (Ed.), *Attention and performance VII* (pp. 429–467). New Jersey: Erlbaum, Hillsdale.
Pinheiro, J., & Bates, D. M. (2000). *Linear and nonlinear mixed-effects models in S and S-Plus*. New York: Springer-Verlag.
Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*, 245–255.
Poeppel, D., & Monahan, P. (2011). Feedforward and feedback in speech perception: Revisiting analysis by sysnthesis. *Language and Cognitive Processes, 26*(7), 935–951.
Poeppel, D., Phillips, C., Yellin, E., Rowley, H. A., Roberts, T. P., & Marantz, A. (1997). Processing of vowels in supratemporal auditory cortex. *Neuroscience Letters, 221*, 145–148. https://doi.org/10.1016/S0304-3940(97)13325-0.
R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.
Rampinini, A. C., Handjaras, G., Leo, A., Cecchetti, L., Ricciardi, E., Marotta, G., et al. (2017). Functional and spatial segregation within the inferior frontal and superior temporal cortices during listening, articulation imagery, and production of vowels. *Scientific Report, 7*, 17029.
Roberts, T. P., Ferrari, P., Stufflebeam, S. M., & Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: Stimulus dependence and insights toward perception. *Journal of Clinical Neurology, 17*, 114–129.
Romani, G. L., Williamson, S. J., & Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science, 216*, 1339–1340.
Saenz, M., & Langers, D. R. M. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research, 307*, 42–52.
Scharinger, M., Idsardi, W. J., & Poe, S. A. (2011). Comprehensive three-dimensional cortical map of vowel space. *Journal of Cognitive Neuroscience, 23*, 3972–3982.
Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2012). Asymmetries in the processing of vowel height. *Journal of Speech Language and Hearing Research, 55*(3), 903–918.
Schwartz, J.-L., Basirat, A., Menárd, L., & Sato, M. (2012). The perception-for-action-Control theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics, 25*(5), 336–354. https://doi.org/10.1016/j.jneuroling.2009.12.004.
Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*(12), 2400–2406. https://doi.org/10.1093/brain/123.12.2400.
Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences, 26*, 100–107.
Scott, S. K., & McGettigan, C. (2013). Do temporal processes underlie left hemisphere dominance in speech perception? *Brain and Language, 127*, 36–45.
Seppo, P. A., Han, J., Belliveau, J. W., & Hämäläinen, M. S. (2010). Sensitivity of MEG and EEG to source orientation. *Brain Topography, 23*, 227–232.
Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., & Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. *Brain and Cognitive Research, 21*, 342–350.
Skipper, J. I., Devlin, J. T., & Lametti, R. D. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language, 164*, 77–105.
Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of Acoustical Society of America, 111*, 1872–1891.
Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology, 91*, 1282–1296. https://doi.org/10.1152/jn.01125.2002.
Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial Congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience, 17*, 1396–1409.
Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event related potential ERP) recordings. *Cognitive Brain Research, 14*, 106–114.
Teuber, H.-L. (1967). Lacunae and research approaches to them: I. In C. H. Millikan, & F. L. Darley (Eds.), *Brain mechanisms underlying Speech and language (204–216)*. New York: MIT Press.
Weinberger, N. M. (2015). New perspective in the auditory cortex: Learning and memory. In G. Hickok (Series Ed.) & G. C. Celesia (Vol. Ed.), *Handbook of Clinical Neurology* (Vol. 129, pp. 117–147). Amsteram: Elesevier.
Weise, A., Schröger, E., & Horváth, J. (2018). The detection of higher order acoustic transitions is reflected in the N1 ERP. *Psychophysiology*, 55(7), e13063. https://doi.org/10.1111/psyp.13063.
Wolpaw, J. R., & Penry, J. K. A. (1975). Temporal component of the auditory evoked response. *Electroencephalography and Clinical Neurophysiology, 39*, 609–620.
Woods, D. L. (1995). The component structure of the N1 wave of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology, 44*, 102–109.
Wood, C. C., & Wolpaw, J. R. (1982). Scalp distribution of human auditory evoked potentials. II. Evidence for multiple sources and involvement of auditory cortex. *Electroencephalography and Clinical Neurophysiology, 54*, 25–38.