Vowel categorization induces departure of M100 latency from acoustic prediction

Timothy P. L. Roberts,[CA] Elissa J. Flagg and Nicole M. Gage[1]

Department of Medical Imaging, University of Toronto, 150 College St., Toronto, Ontario, M5S 3E2, Canada; [1]Mail Code 5100, 3151 Social Sciences Plaza A, University of California, Irvine, CA 92697-5100, USA

[CA]Corresponding Author: tim.roberts@utoronto.ca

Received 16 April 2004; accepted 26 May 2004

DOI: 10.1097/01.wnr.0000134928.96937.10

MEG studies have shown that the timing (latency) of the evoked response that peaks ~100 ms post-stimulus onset (M100) decreases as frequency increases for sinusoidal tones. We investigated M100 latency using a continuum of synthesized vowel stimuli in which the dominant formant frequency increases from 250 Hz (perceived /u/) to 750 Hz (perceived /a/) in 50 Hz steps. While M100 latency did vary inversely with formant frequency overall, frequency modulation was flattened within each vowel category. However, for mid-continuum ambiguous tokens (i.e. those with increased reaction time/decreased accuracy in the concurrent behavioral identification task), M100 reverted to formant frequency differences, agreeing with previous findings of frequency-dependence. A theory is proposed in which phonological categorization emerges from specific spatial distribution of frequency-tuned neurons. *NeuroReport* 15:1679–1682 © 2004 Lippincott Williams & Wilkins.

Key words: Auditory cortex; Auditory perception; Categorical perception; Language; Magnetoencephalography; M100; Speech perception

INTRODUCTION

MEG uses a high spatial density biomagnetometer to record extracranial magnetic fields generated by intracranial electrical activity. MEG not only offers a noninvasive measure of brain function, but its submillisecond temporal resolution represents a distinct advantage over hemodynamic imaging (cf. PET, fMRI). Thus, MEG is capable of revealing aspects of neuromagnetic activity for which timing is critical. Previous studies have shown that the latency of the auditory evoked neuromagnetic component that peaks about 100 ms post-stimulus onset is sensitive to stimulus attributes like intensity and frequency [1], leading to the suggestion that this tonochronic encoding mechanism might augment or supplement a tonotopic strategy in the frequency range critical to human speech. It has also been shown that age influences M100 latency for children and teenagers [2].

The latency of the auditory M100 exhibits frequency sensitivity for vowels as well as pure tone stimuli. In this study, M100 latency is increased for the vowel /u/ relative to the vowel /a/ [3]. Tonochrony predicts the existence and direction of this latency differential – first formant (F1) frequency is lower in /u/ (~300 Hz) than /a/ (~700 Hz). However, /u/ and /a/ differ not only in F1 frequency (an acoustic feature), but also in phonological category membership. While F1 frequency may exert a dominant effect on M100 latency, as supposed by Roberts *et al.* [3], vowel category membership could also contribute to the M100 latency (especially given human tolerance for a range of within-category F1 values [4]). This experiment seeks to determine whether stimulus category membership exerts an influence on M100 latency beyond that predictable from an acoustic model [3].

Using a continuum of synthesized vowels with clear exemplars of /u/ and /a/ at each end and ambiguous mid-continuum tokens, we ask whether M100 characteristics cluster according to stimulus phonological category or distribute evenly according to F1 frequency? If phonological category membership contributes to the M100, some leveling of peak latency and/or amplitude values would be observed within stimuli of the same category; however, if category membership has no effect as early as M100, physical stimulus attributes (spectral energy distribution, dominantly around F1) are predicted to dominate M100 latencies, resulting in a characteristic 1/f shape curve with latency prolongation at lower frequencies, similar to that seen for tones [1,3].

MATERIALS AND METHODS

Eleven vowel-like stimuli along the /u/-/a/ continuum were created using the Klatt synthesizer [5] implemented in Sensyn software (Sensimetrics, MA). Only F1 varied between tokens, from 250 to 750 Hz in 50 Hz increments. To retain plausibility across this continuum, F2 values of 1000 Hz and F3 values of 2500 Hz were allowed broader bandwidth (300 Hz) than is typical in vowel synthesis. Stimuli were 400 ms in duration with 50 ms linear ramps.

Stimuli were presented monaurally (right ear) via Etymotic ER3A transducers and air-hose to an inner ear tip after amplification (TDT Inc., Series II, FLA). Auditory detection threshold values were determined for all stimuli for subsequent presentation at 40 dB SL. MEG recording was made using a 37-channel biomagnetometer (Magnes 4-D Neuroimaging, San Diego, CA) positioned over the contralateral temporal area. Positioning of the sensor array was adjusted until the response across sensors to presentation of a 1 kHz reference tone formed an antisymmetric evoked response distribution. The 11 stimuli were randomly interleaved (ISI 1000 ± 100 ms) and at least 200 trials of each stimulus were presented. Subjects lay on their side during the procedure, inside the magnetically shielded room. During stimulus presentation, subjects performed a two alternative forced choice classification (with button box response, using fingers of the left hand). Reaction time and classification were monitored. Categorization was determined from the proportion of responses to the 200 elements of each stimulus token. After 50% of the trials, the role of the buttons was reversed to compensate for potentially confounding mechanical/logistical differences in making one response over another.

MEG data was sampled at a rate of 1041 Hz/channel and grouped into 600 ms epochs time-locked to stimulus presentation (incl. 100 ms pretrigger baseline). Epochs were averaged by stimulus type for each subject, filtered at 1–40 Hz, and M100 peak latency and amplitude were determined. The M100 was defined as that response occurring in a latency window of 80–150 ms post-stimulus with maximum r.m.s. field strength across sensors (at least a factor of 6 greater than pretrigger noise), satisfying a single equivalent dipole model with a correlation coefficient of $r > 0.95$.

Subjects: Six healthy right-handed adult volunteers participated with the approval of the institutional committee on human research; written informed consent was obtained. All were competent English speakers, indistinguishable with respect to performance on behavioral classification: all exhibited categorical perception for first three stimuli (F1: 250–350 Hz; 97% /u/ response) and last three stimuli (F1: 650–750 Hz; 98% /a/ response), and variable categorization of mid-continuum stimuli (/a/ response 10% at 400 Hz, 27% at 450 Hz, 49% at 500 Hz, 72% at 550 Hz, 90% at 600 Hz).

Tone control: Subjects underwent a second MEG recording session in which stimuli were 400 ms sinusoidal tones with frequencies matching the F1 position for the vowel continuum, with addition of extrema tokens at 100 Hz and 1 kHz. One hundred trials of each token were presented pseudorandomly. All other experimental details were similar to those described above; there was no behavioral task.

An incidental behavioral observation led to the following additional analysis methodology: in two subjects, the mid-continuum stimulus of the vowel continuum (F1: 500 Hz) exhibited chance categorization characteristic of a categorical perception boundary. Examining categorization of this boundary token according to preceding context probed perceptual modulation [6]. When preceded by one of the clear /u/ stimuli, proportional categorization of the boundary token shifted towards /a/; when preceded by one of the clear /a/ stimuli, proportional categorization of the boundary token shifted towards /u/. This raised the issue of whether category status of the percept affects M100 latency when physical stimulus properties are held constant. For these two subjects, MEG epochs from presentations of the boundary token were additionally selectively averaged according to preceding context to see whether M100 latency of the same token increased after presentation of /a/ due to the induced perceptual bias towards /u/ or decreased after presentation of /u/ due to the induced perceptual bias towards /a/.

RESULTS

Figure 1 shows that M100 latency across the vowel continuum does not form a smooth 1/f function as for pure sinusoids [1], but rather forms latency plateaus at each extrema. Of note, the M100 latency to vowel tokens is dominated by F1 frequency – latencies to /u/ stimuli (lower F1) are longer than in response to /a/ (higher F1), as predicted by F1 frequency [3].

In contrast to the smooth 1/f form of the M100 latency to sinusoid tones, latencies to vowel-like sounds along the continuum formed plateaus that correspond to continuum endpoint vowel categories. Thus M100 latency was prolonged for /u/ relative to /a/, but within /u/ and /a/ stimuli, F1 frequency did not determine M100 latency precisely; within each endpoint vowel category frequency-dependence is leveled i.e. within-category latency differences are minimized. As such, stimuli with F1: 250–350 Hz and similarly with F1: 650–750 Hz cluster together in neuromagnetic activity, just as they clustered together in the behavioral identification task (Fig. 2). Conversely mid-continuum tokens that show variable categorization also show M100 latency variability. While this discrepancy in the mid-continuum range (F1: 400–600 Hz) potentially represents either reversion to acoustic prediction in absence of a true corresponding category or the formation of an intermediate pseudo-vowel category, the lack of clustering in either behavioral or M100 latency values speaks to reversion to acoustically-driven frequency-dependence.

A further observation is that the amplitude (relating to the number of synchronously activated neurons) of the evoked response is significantly diminished for ambiguous tokens compared to tokens with clear category membership

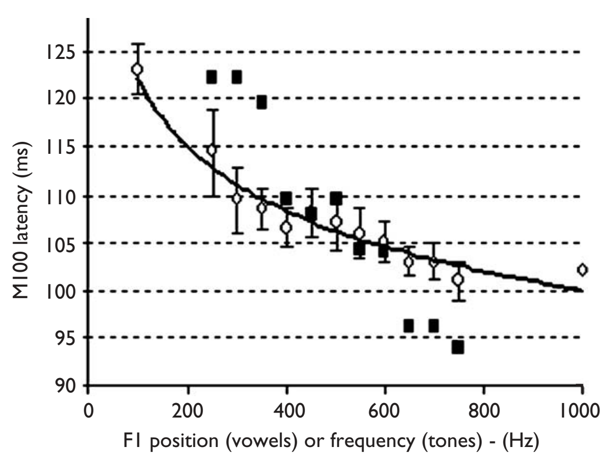

Fig. 1. M100 latency for vowel continuum (solid squares) and sinusoidal tones (open circles, with 1 / f curve fit).

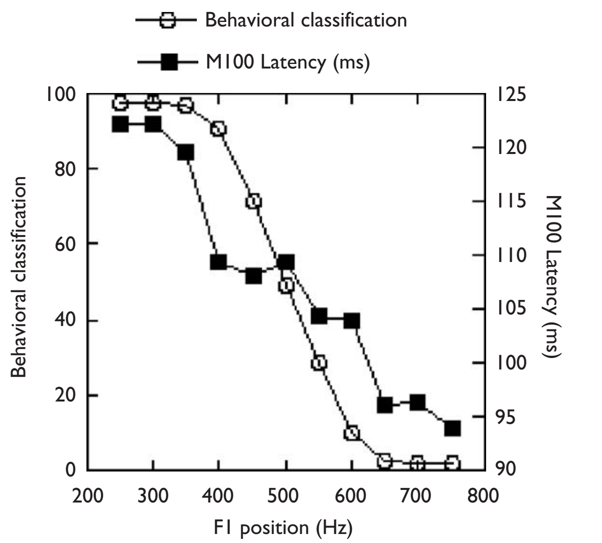

Fig. 2. Correlation between behavioral classification and M100 latency averaged across all subjects.

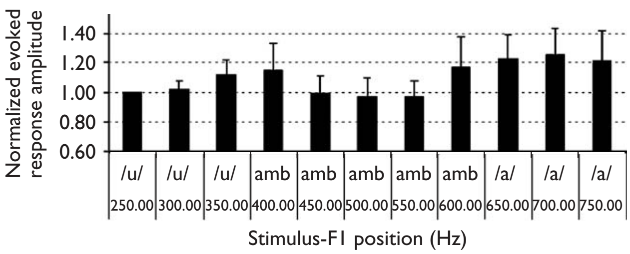

Fig. 3. Evoked response amplitude as a function of vowel continuum position (mean ± s.e.m. across subjects).

($p<0.05$; Fig. 3). By implication, the M100 to mid-continuum tokens is generated by fewer (or a less coherent set of) neurons.

If we accept the putative tonotopic organization principle of auditory cortex [7–9], discrete spatially contiguous groups of neurons selectively react to specific frequencies. One is led to speculate that the range of F1 frequencies captured by the category /u/ (250–350 Hz) could correspond to a circumscribed neuronal population, such that the frequency range defines the vowel category by virtue of forming an area of dedicated neurons that get activated in concert by any of the frequencies within the range covered [10,11]. However, the lower amplitude M100s from mid-continuum stimuli result from frequency content falling near but not within a range that engages a circumscribed neuronal cohort for a given category. Because mid-continuum stimuli do not strongly activate either category's neuronal cohort, there is no within-category M100 latency leveling.

In two subjects, one token (F1: 500 Hz) elicited almost symmetric /a/ *vs* /u/ responses (49/51%), signaling an ambiguous behavioral percept. Selectively averaging the behavioral response to this stimulus according to whether its predecessor was a clear /u/ or a clear /a/, categorization was found to be influenced by preceding stimulus. Although acoustic properties remained unaltered, categorization as /u/ rose to 58% when preceded by an unequivocal /a/ and conversely 57% categorization as /a/ when preceded by /u/ (Fig. 4). Preceding context did not affect M100 latency for the ambiguous token; M100 remained true to the physical spectral attributes.

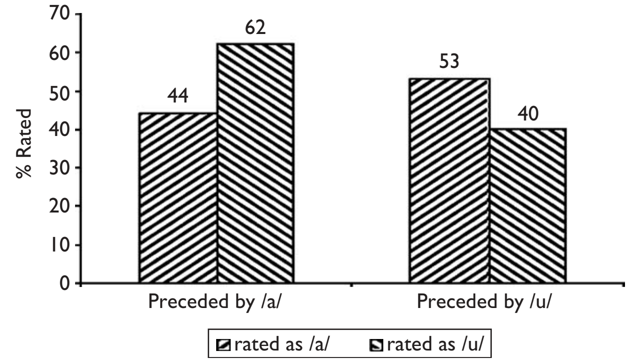

Fig. 4. Perceptual modulation of ambiguous boundary token by context for two subjects with a boundary tone. Without regard for preceding context result came out by chance (~50% /a/, ~50% /u). When preceded by an unequivocal /a/, the number of /u/ rating increased to 57%. When preceded by an unequivocal /u/, the number of /a/ ratings increased to 58%.

Disregarding preceding context, a true boundary token gives rise to an ambiguous percept categorized at chance. Perceptual modulation applies when acoustics alone underdetermine category membership, i.e. when category membership is equivocal. Within the design of this experiment, context biases classification of the boundary token towards the opposite category – the percept is recognized as different from the clear category exemplar that preceded it. This is a simple manifestation of selective adaptation [6].

Holding acoustics constant while varying percept shows that the dominant effect on M100 characteristics is still physical stimulus properties; when context biases perception, M100 latency and amplitude do not shift to the values characteristic of those stimulus categories. A further observation is a striking asymmetry in the behavioral responses to the unequivocal tokens at either end of the continuum, despite across the board 97% categorization accuracy. Reaction time for clear /u/ tokens (683 ± 54 ms) is greater than that to clear /a/ tokens by nearly 100 ms (591 ± 41 ms, $p<0.05$). While M100 latency differences were found to be prolonged by ~20 ms for clear tokens of /u/ *vs* clear tokens of /a/, there was no difference in the accuracy of classification: virtually all of the first three tokens of the continuum were classified as /u/ and virtually all of the final three tokens of the continuum were classified as /a/; nonetheless, there was a significant delay in reaction time for classification of /u/ compared with /a/. This is unlikely to arise because of additional uncertainty as one would expect this to impact accuracy as well. Thus we are led to speculate that the 20 ms delayed M100 response is expanded during the processes of categorization and labeling to almost 100 ms by the time of the behavioral response.

DISCUSSION

The results of this study are that (1) for a continuum of vowel-like sounds of varying F1, the M100 latency forms plateaus analogous to the discrete categories defined by categorical perception; (2) only between categories does the M100 latency revert to the 1/f form predicted by a simple acoustic model; (3) ambiguous tokens can have their percept manipulated under appropriate circumstances, but the M100 retains a latency characteristic of actual spectral attributes, rather than the context-biased percept; and (4) the prolonged latency of tokens clearly perceived as /u/ *vs* those perceived as /a/ is correlated with a nearly 100 ms reaction time delay in identification.

Evidence of categorical effects (latency leveling within categories) as early as ~100 ms post stimulus reveals the relevance of mentally represented categories to early perceptual processing. Taken together, the frequency range of both M100 latency plateaus and amplitude plateaus suggest that activation of distinct tonotopic cohorts is responsible for leveling of within-category variation in the neural response, and for concomitant categorical perception of vowels, a manifestation of the perceptual magnet effect [12], with minimization of within-category differences in both perception and, in this case M100 latency. We speculate that M100 latency does not resolve within-category F1 variation, because the neural response to all vowels in a category is generated by the same neuronal cohort. However, latency values for tone stimuli show that the auditory M100 latency can resolve 50 Hz frequency increments. Mid-continuum stimuli show that neuronal responses to tokens with the spectral complexity of vowels can in principle resolve frequency differences of this magnitude, but this sensitivity only emerges when the frequencies involved are not captured within a neuronal cohort defining a particular phonological category.

The absence of M100 latency plateaus for tone stimuli suggests that the apparent lack of sensitivity to within vowel category frequency differences relates specifically to membership in a category, as opposed to anything inherent or epiphenomenal about the relationships between F1 frequencies (e.g. 250 Hz, 300 Hz, 350 Hz) or the way the stimuli group in terms of their physical attributes. Since speech sounds ultimately map onto mental representations of discrete phonological units, vowel stimuli must come to be treated as belonging to one category or another. The increased acoustic complexity of vowels compared with tones does not challenge attributing latency plateau effects to categorical perception, since M100 latency does resolve F1 frequency variation between mid-continuum vowel stimuli, which have the same spectral characteristics as endpoint tokens.

In summary, we suggest that the ultimate perception is parasitic on M100 encoding of physical stimulus properties when they unequivocally correspond to a native language category/neuronal cohort. For tokens that are not members of a distinct native language category, physical stimulus properties dominate M100 latency; in such cases, the role of context emerges as a biasing factor (Fig. 5). A caveat exists: the particular context effect is seen here under circumstances of forced choice between two alternatives and is not necessarily predictive of effects on vowel identification when category choices are unconstrained.

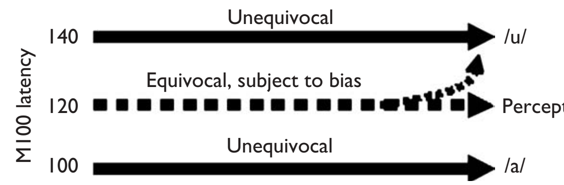

Fig. 5. A simple model suggests that if the M100 latency is characteristic of a particular vowel percept, the behavioral response will not deviate (e.g. M100=100 ms or M100=140 ms). If the M100 response is not characteristic of vowel category membership, perception may be susceptible to bias.

CONCLUSION

Vowel category membership influences M100 latency, forming latency plateaus analogous to classification and departing from the strictest prediction of a 1/f model. We suggest that the difference between M100 latency behavior to stimuli continua with inherent categorical differences (i.e. vowels) *vs* stimuli with no inherent categorical differences (i.e. tones) provides a potentially objective neuronal measure of intact native language phonological representations. We predict, for example, that individuals with certain language impairments could exhibit a vowel-continuum latency slope with tone-like 1/F1 form, suggesting a problem with development of neuronal cohorts appropriate for phonological categorization that could lead to impaired native language phonological category representations.

REFERENCES

1. Roberts TPL and Poeppel D. Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport* 1996; **7**:1138–1140.
2. Gage NM, Siegel B and Roberts TPL. Cortical auditory system maturational abnormalities in children with autism disorder: and MEG investigation. *Dev Brain Res* 2003; **144**:201–209.
3. Roberts TPL, Ferrari P, Stufflebeam SM and Poeppel D. Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *J Clin Neurophysiol* 2000; **17**:114–129.
4. Fry DB. Acoustic features of English sounds. In: Fry DB (ed.). *The Physics of Speech*. Cambridge: Cambridge University Press; 1979, pp. 110–114.
5. Klatt D. Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am* 1980; **67**:971–975.
6. Eimas PD and Corbit JD. Selective adaptation of linguistic feature detectors. *Cogn Psychol* 1973; **4**:99–109.
7. Romani GL, Williamson SJ and Kaufman L. Tonotopic organization of the human auditory-cortex. *Science* 1982; **216**:1339–1340.
8. Pantev C, Bertrand O, Eulitz C, Verkindt C, Hampson S, Schuierer G *et al*. Specific tonotopic organizations of different areas of the human auditory-cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalogr Clin Neurol* 1995; **94**:26–40.
9. Wessinger CM, Buonocore MH, Kussmaul CL and Mangun GR. Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Hum Brain Mapp* 1997; **5**:18–25.
10. Mäkelä AM, Alku P and Tiitinen H. The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. *Neurosci Lett* 2003; **353**:111–114.
11. Obleser J, Lahiri A and Eulitz C. Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 ms post syllable onset. *Neuroimage* 2003; **20**:1839–1847.
12. Kuhl PK. Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Percept Psychophys* 1991; **50**:93–107.

Acknowledgements: The authors thank Susanne Honma, Mary Mantle and Jeff Walker for technical assistance. This research was supported by the National Alliance for Autism Research.