

MEG Studies of Vowel Processing in Auditory Cortex

Colin Phillips
U. of Delaware

Krishna Govindarajan
Alec Marantz
M.I.T.

David Poeppel
Tim Roberts
Howard Rowley
U.C. San Francisco

Elron Yellin
Mark of the Unicorn Inc.

1. Cues to Speaker and Category in Vowels

Vowels are acoustically complex sounds, consisting of bands of energy at a fundamental frequency (F0) and a series of prominent harmonics of the fundamental frequency (formants).

- The fundamental frequency carries information about prosody and speaker voice.
- The formants carry information about the phonetic category of the vowel.

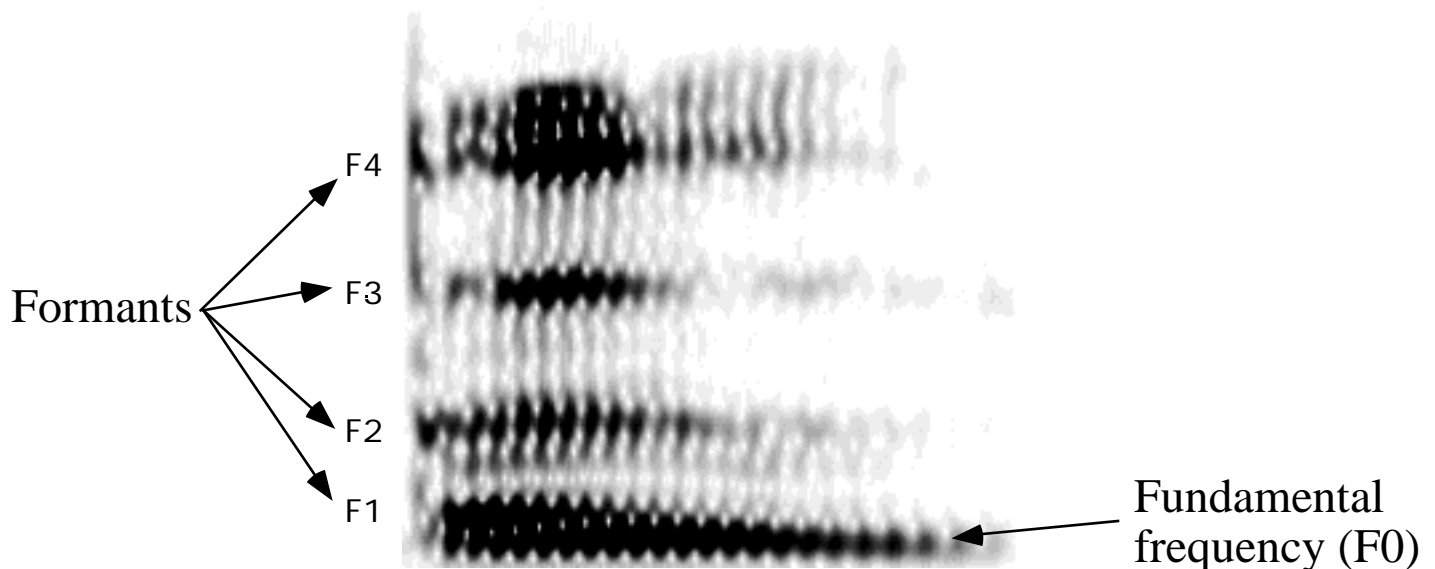


Figure 1: Spectrogram of the vowel /u/ taken from the word 'boot'

For example, the first formant (F1) carries information about the *height* of the vowel: high vowels (e.g. /i/, /u/) have a low F1 (~300Hz); low vowels (e.g. /a/, /æ/) have a high F1 (~700Hz). The second formant carries information about the *frontness/backness* of the vowel: front vowels (e.g. /i/, /e/) have high F2 values (~2000Hz), and back vowels (e.g. /u/, /a/) have lower F2 values (~1000Hz).

The studies described here demonstrate an evoked response in auditory cortex which appears to track the formant structure of vowels rather than the fundamental frequency.

2. Pure tones: pitch affects M100 latency

We focus in this poster on analyses of the auditory M100 response, a well-known evoked response generated in supratemporal auditory cortex approx. 100ms after the onset of any acoustic stimulus. Figure 2a illustrates the characteristic M100 waveform, recorded with a 37-channel MEG array, and the stages involved in obtaining a source-localization from the magnetic waveforms (the whole-head reconstruction in Figure 2d shows somatosensory rather than auditory responses).

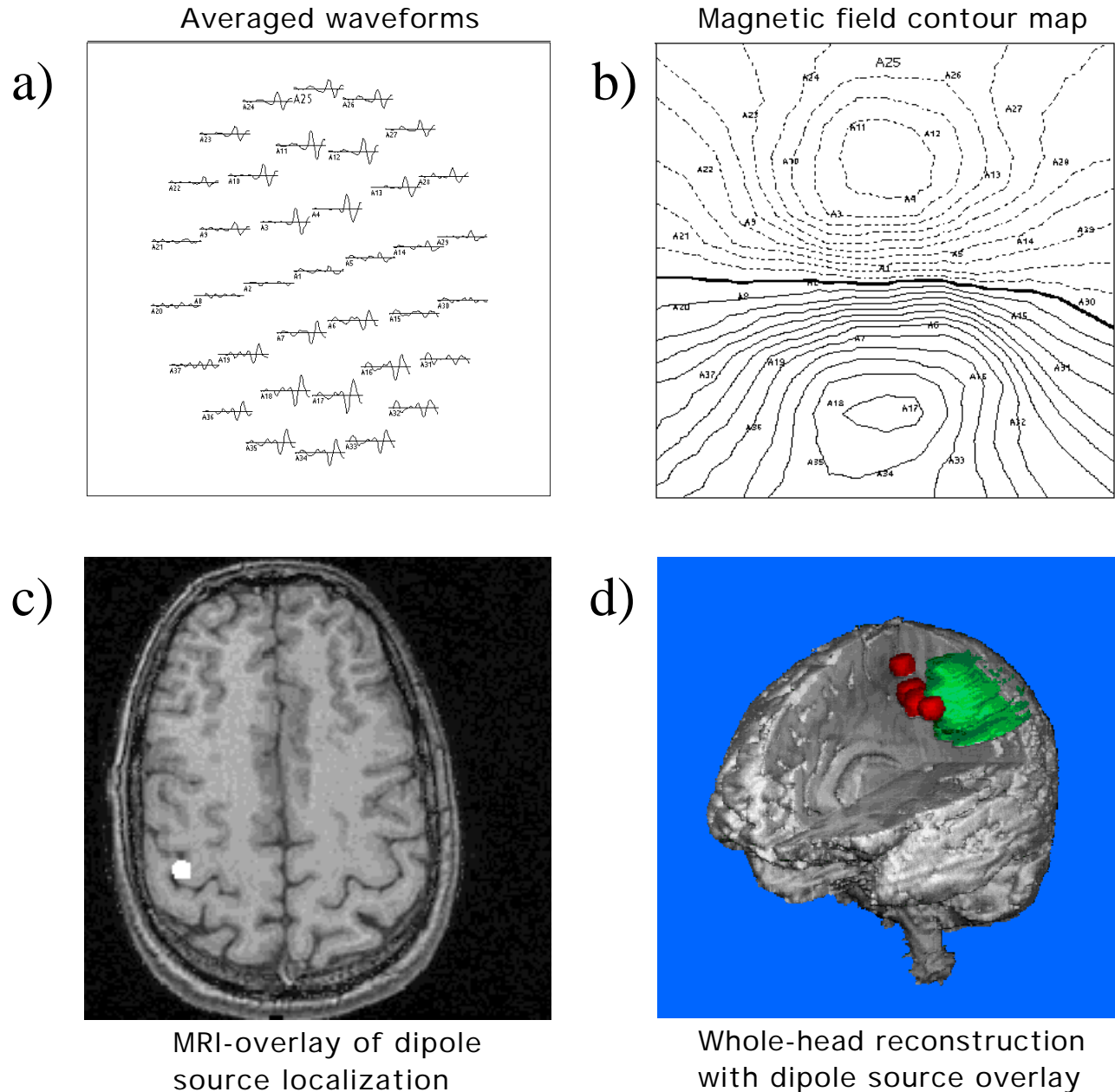


Figure 2: Measurement and localization of evoked brain responses using Magnetic Source Imaging

The latency of evoked M100 responses varies systematically as a function of the pitch of tone stimuli (Roberts & Poeppel 1996; Ragot & Lepaul-Ercole 1996). Frequencies at 1000Hz or above evoke the shortest latency M100s; lower frequencies evoke progressively longer latency M100s.

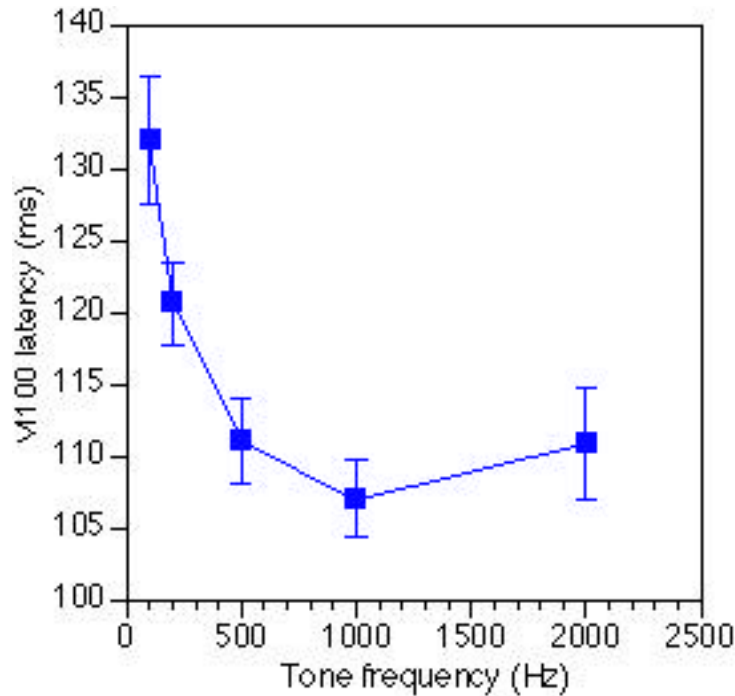


Figure 3: Effect of tone-frequency on M100 latency (n=5)

The latency of M100 in response to a tone can be predicted by the following empirical formula:

$$L = 3/f + k$$

where f is the tone frequency, and k is a time constant that varies across subjects from 90-120 ms. This formula can be used to predict the M100 latency difference for two tones of different frequencies.

We use this finding as a tool for understanding the cause of the systematic M100 latency changes in response to different stimuli in our vowel experiments.

3. Three-formant vowels

To investigate responses to vowels as a function of pitch (F0), phonetic category (formants) and hemisphere, we presented 3 synthetic vowels (/a/, /i/ and /u/) at two fundamental frequencies (100Hz, male voice (m) and 200Hz, female voice (f)) to 6 subjects while recording the evoked neuromagnetic field from the left or right temporal cortices using a 37-channel biomagnetometer. The formant values are shown in Table 1.

**Table 1:
Vowel formants–spectral composition (in Hz)**

Stimulus	F0	F1	bandwidth	F2	bandwidth	F3	bandwidth
a(m)	100	710	90	1100	100	2540	90
a(f)	200	850	90	1220	100	2810	90
i(m)	100	280	90	2250	100	2890	90
i(f)	200	310	90	2790	100	3310	90
u(m)	100	310	90	870	100	2250	90
u(f)	200	370	90	950	100	2670	90

Procedure

Vowel stimuli were synthesized using a Klatt formant synthesizer (SenSyn, Sensimetrics Inc.). Each vowel was presented 100 times in pseudorandom order to the ear contralateral to the MEG dewar. ISIs varied randomly from 750–1250ms.

We measured the extracranial magnetic field using a 37-channel first-order gradiometer (Magnes, BTi, San Diego, CA). Recordings took place in a shielded room. Subjects were lying down, and the dewar containing the sensor array was positioned to record from auditory cortex. The appropriate position was determined by first localizing the generator of the M100 response to series of 1000Hz and 200Hz tones.

Responses were recorded with a sampling rate of 1kHz and a bandwidth of 400Hz. Responses were then averaged according to event type and filtered using a low-pass filter from 1-20Hz. For each of the averaged and filtered responses a single-dipole source localization model was computed using an iterative least-squares algorithm. M100 latencies were identified as the time interval between 80 and 150ms post-stimulus onset with the largest root mean square (RMS) of the field strength across the 37 channels.

Results

Measured M100 latencies varied systematically as a function of vowel category differences ($F(2,5)=27.2$, $p<0.0001$), but there were no significant changes due to F0 differences (figure 4). This result was significant in both hemispheres.

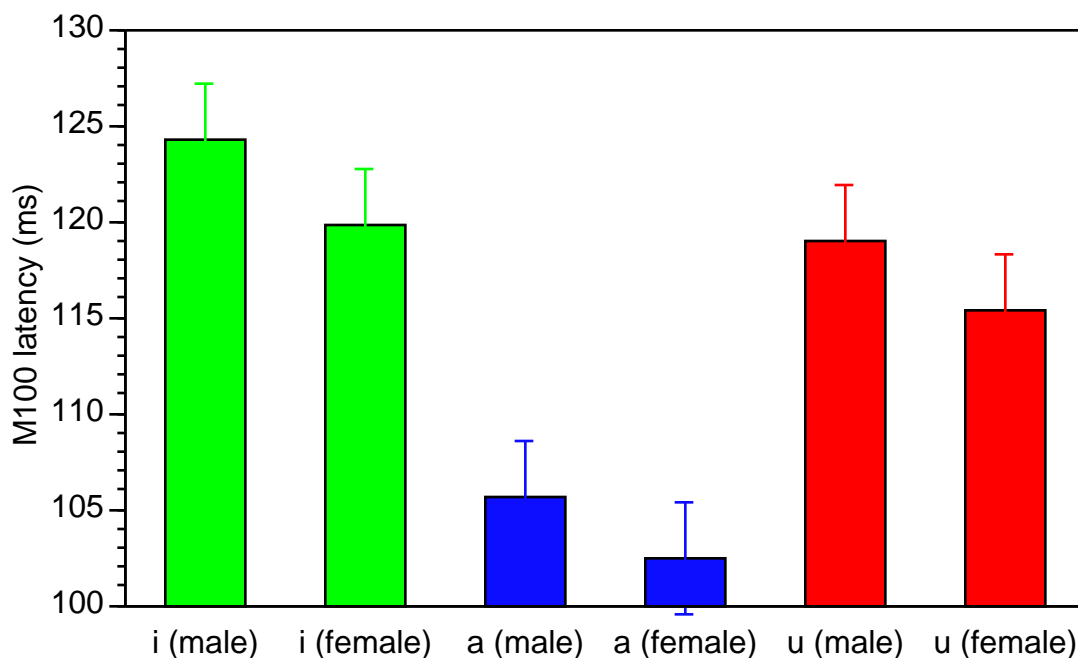


Figure 4: Latency variation due to category (n=6)

This result is consistent with Roberts & Poeppel's findings about effects of tone-frequency on M100 latency. The fastest responses were to vowels with the most energy around 1000Hz.

More specifically, M100 latencies tracked variation in the first formant (F1) of the vowels extremely closely ($r = -0.97$). The consistently earlier M100 to female vowels may be due to the higher F1 values in the female vowels. However, this could also be an effect of the higher F0 in the female vowels. The second experiment addresses this uncertainty.

There was no significant effect of F0 or formant values on the localization of M100 responses; responses to all vowels localized reliably to the same region of supratemporal auditory cortex (figures 5–6).

Male /a/ ■
 Female /a/ □

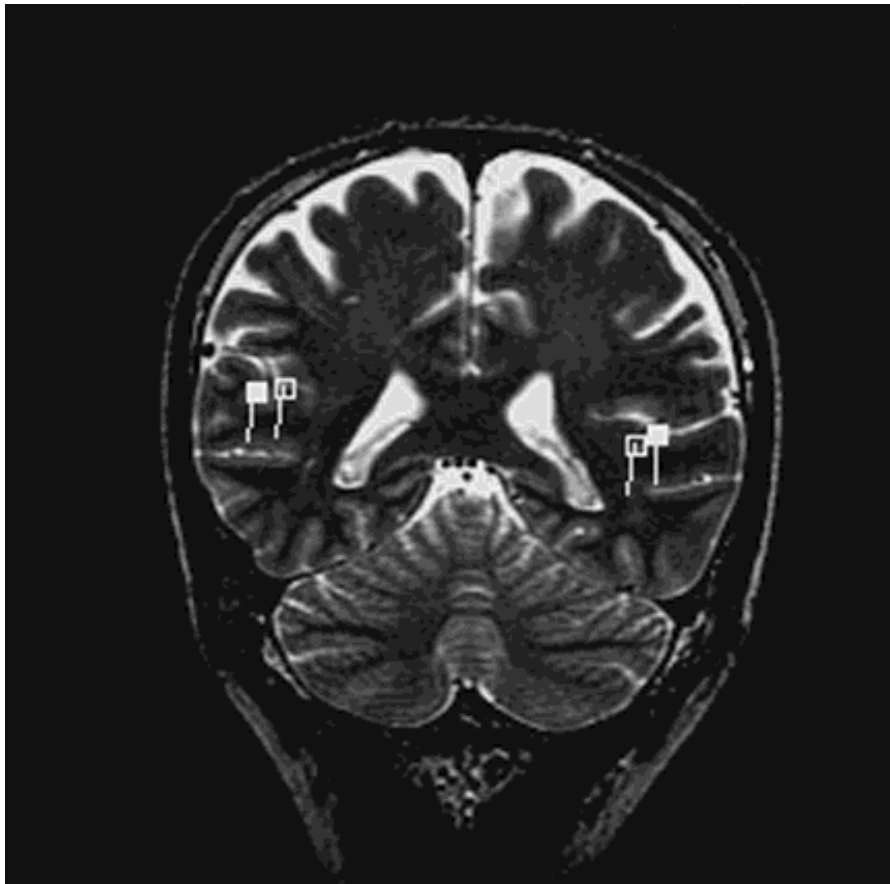


Figure 5: Localizations of M100 responses to 3-formant /a/'s in left and right auditory cortices

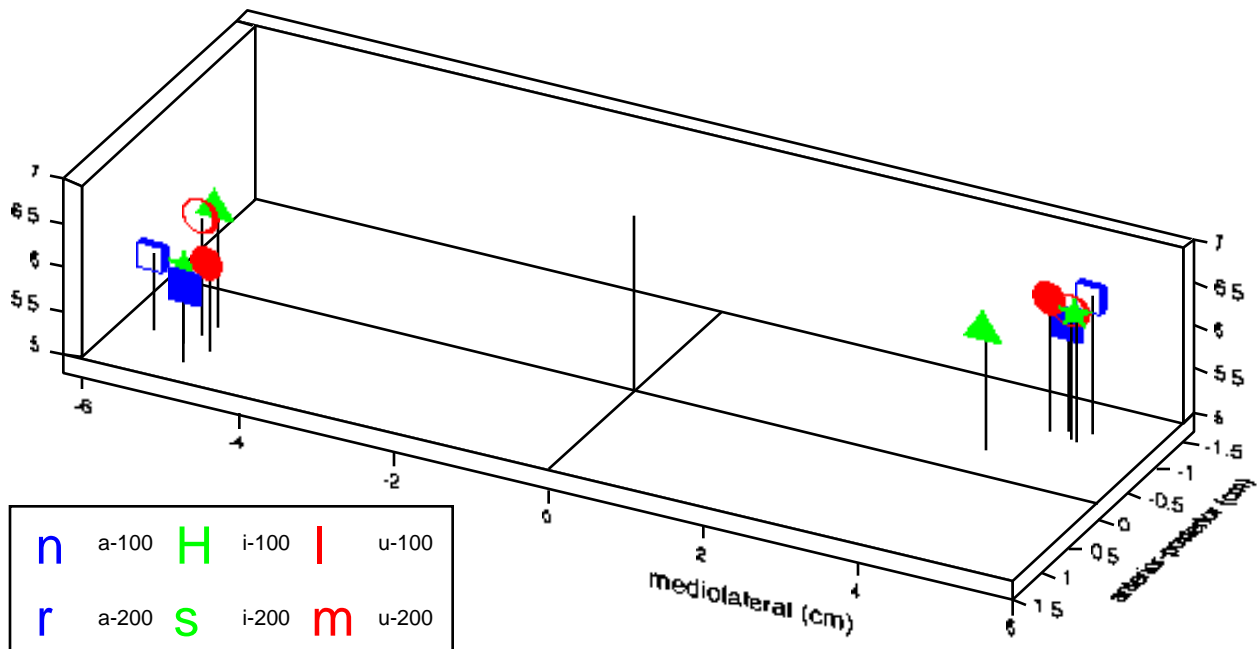


Figure 6: Source localizations of M100 response to vowels (same subject as Figure 5)

The results of this experiment indicate that latency differences in M100 responses to vowels are determined by the energy bands that carry category information (formants) rather than the energy band that carries prosody and voice information (F0).

4. Single-formant vowels

The first vowel experiment was intended to determine the relative contribution of the fundamental frequency and the formants to early responses to vowel sounds. The results suggested that the value of the first formant, which carries information about vowel height, is the main factor determining M100 latencies. In a subsequent experiment we set out to test this correlation further, by isolating the contribution of F1.

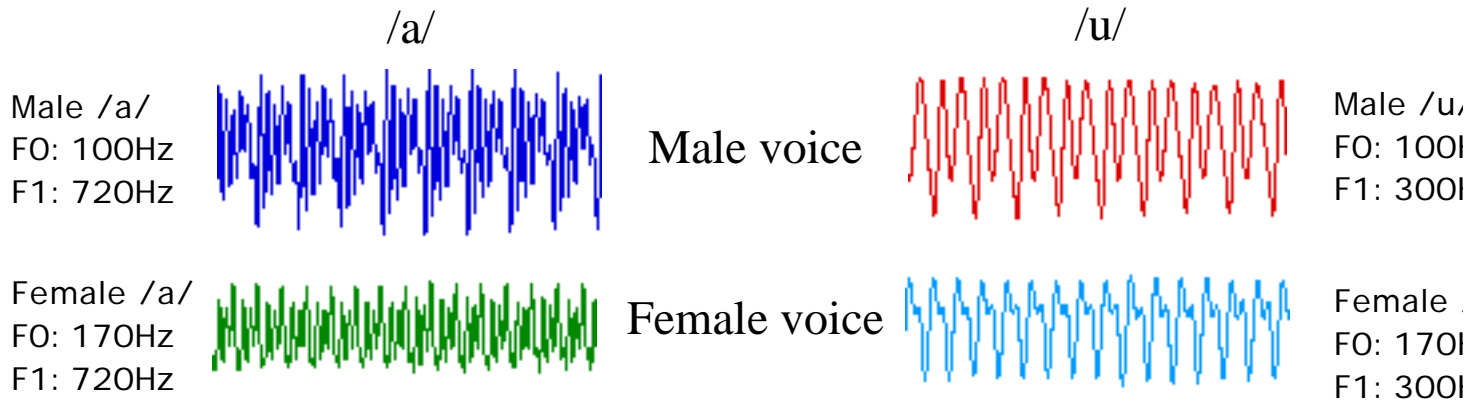


Figure 7: Waveforms of single-formant vowel stimuli

Although naturally produced vowels contain at least 3 strong formants, it is possible to synthetically generate vowels with just a single formant in addition to the fundamental frequency. In the case of the vowels /a/ and /u/ the sounds sound somewhat unnatural but are quite easily identified by subjects.

For this study we used single-formant versions of /a/ and /u/ at each of two fundamental frequencies. In this experiment we did not adjust the value of F1 in the female versions to accommodate the higher fundamental frequency. Thus F0 and F1 varied entirely independently.

MEG recordings were made from left temporoparietal cortex of 9 subjects with normal hearing while 100 tokens of each of the four single-formant vowels was presented to the contralateral ear.

Pretests using pure tones on each of the subjects in this study replicated the effect of frequency on M100 latency, showing almost identical effect size and means to Roberts & Poeppel (1996).

The different predictions, according to whether F0 or F1 determines M100 latency, are shown in Figures 9–10. If latency variation is due to F0 then female vowels are predicted to show approx. 10ms faster M100s; if latency variation in vowels is due to F1 then /a/ is predicted to show approx. 5ms faster M100s. If both F0 and F1 affect response latencies, then we expect responses to /a/ to be faster than responses to /u/, with responses to female vowels additionally being faster for both categories.

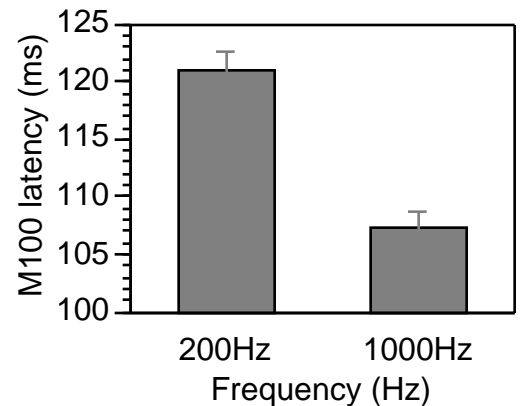


Figure 8: Effect of tone frequency on M100 latency replicated for subjects in this experiment (n=9)

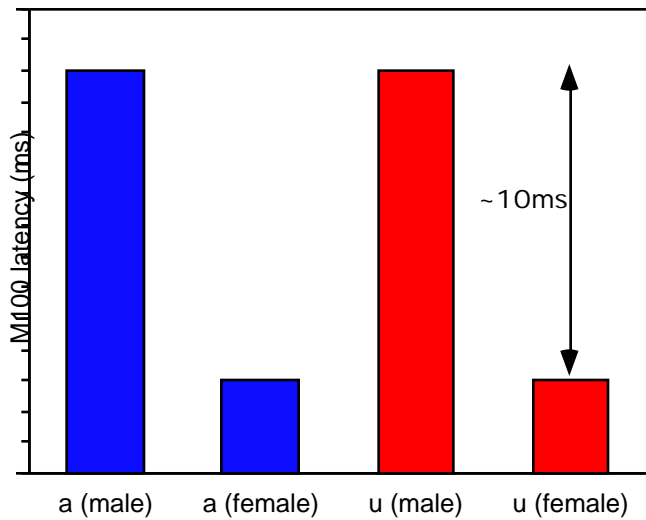


Figure 9: Prediction if F0 determines M100 latency

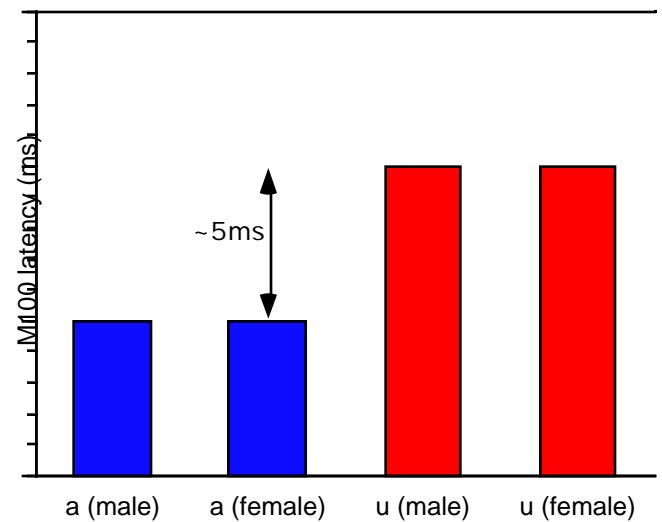


Figure 10: Prediction if F1 determines M100 latency

The results showed a significantly shorter latency M100 in response to single formant /a/ ($F(1,8)=4.2$, $p < 0.05$), with an effect size of 5.2ms. This result strongly supports the claim that M100 varies as a function of the F1 value of a vowel.

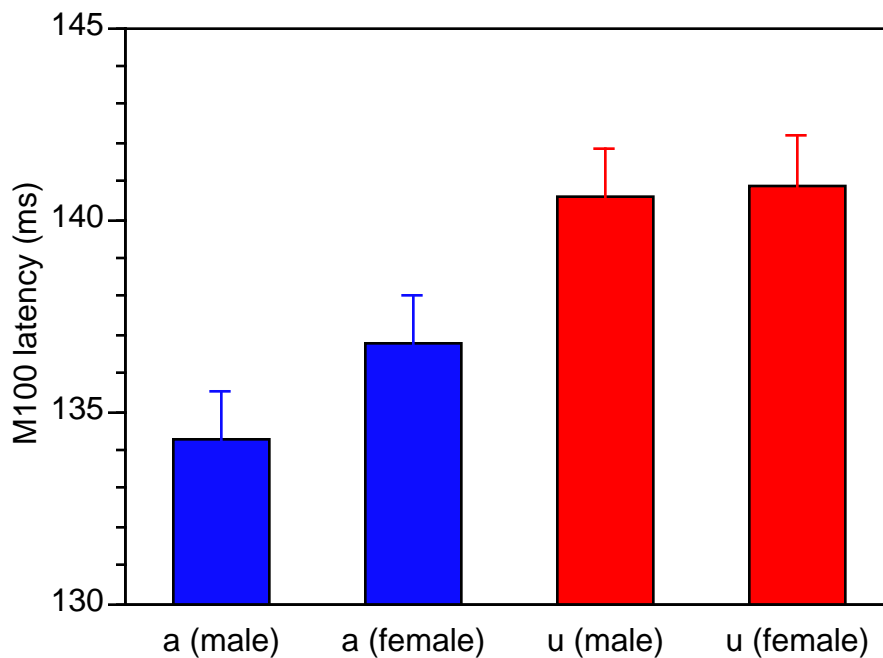


Figure 11: F1 determines M100 latency (n=9)

There was no evidence for an effect of F0 on M100 latencies, and no effect of either category or voice on localizations of these evoked responses.

5. Why formants determine M100 variation

Having established that M100 latencies vary as a function of the formants of vowels, and are not affected by variations in F0, we now need to ask why this is the case. Two possible explanations are:

- **Speech specific account:** formant values are the major determinant of response latencies because in speech sounds they contain the information that enables speakers to identify phoneme categories.
- **General auditory account:** the auditory system is more sensitive to energy in the frequency range where formants occur, regardless of whether the energy is part of a speech sound or not. Energy in F0 is just not so salient to the auditory system.

Although we know of no direct support for the speech-specific account, an ERP study by Ragot & Lepaul Ercole (1996) does challenge our conclusion that the formant frequencies rather than F0 determine the latency of the M100. They independently varied the fundamental and spectrum of tone complexes and found N100 latency variation due to changes in the fundamental rather than to changes in the spectrum. Several properties of their stimuli make their results difficult to compare with ours. Most crucially, the energy in the Ragot & Lepaul-Ercole sounds was heavily concentrated in the F2 range, rather than the F1 range.

However, results of a pilot MEG study with tone complexes in our lab are more consistent with the general auditory account. We created complexes of pairs of pure tones, with frequencies and amplitudes matching those of the stimuli in the single-formant vowels experiment. In one subject tested, the tone complexes did not sound remotely speechlike, but variation in the “pseudo-F1” caused the greatest changes in M100 latency.

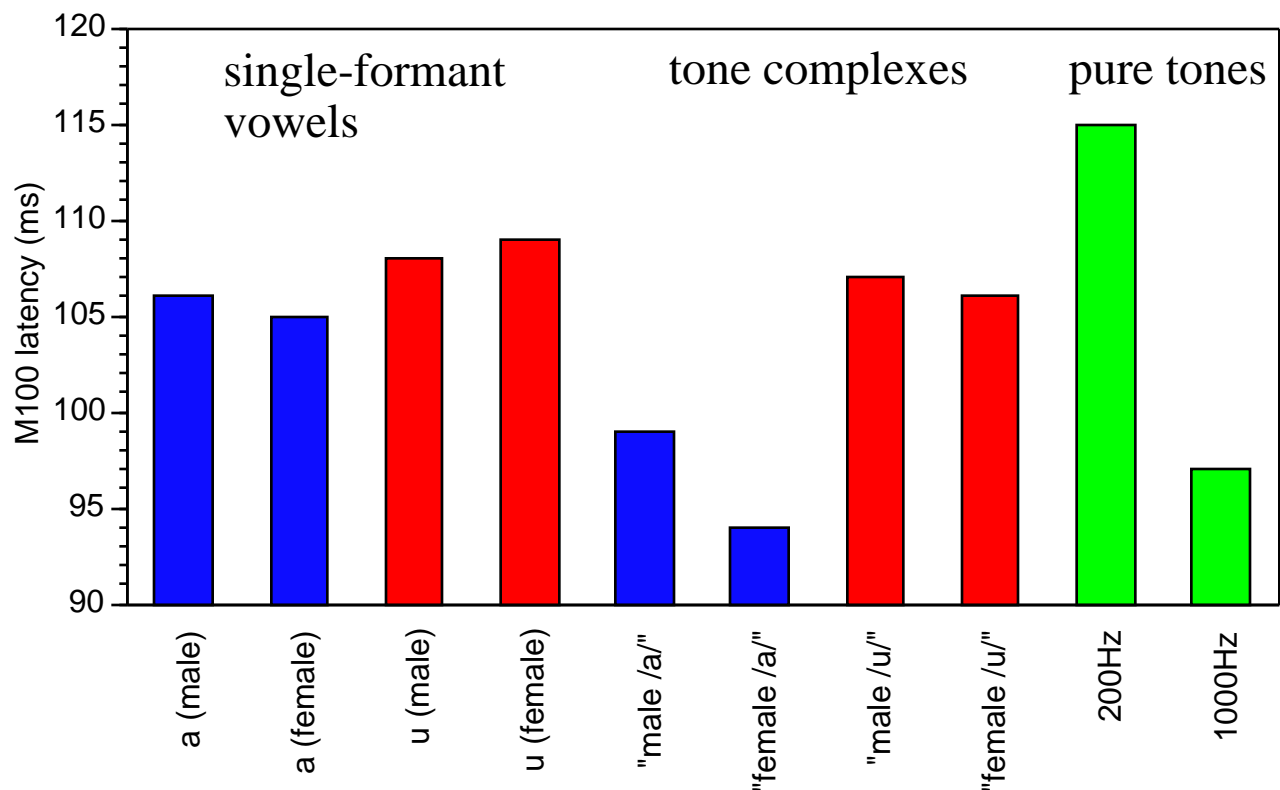


Figure 12: Pilot comparison of single-formant vowels and matching tone complexes (n=1)

6. Conclusion

Vowel sounds contain energy at a number of different frequencies, which between them convey both speaker and category information. Our results show that the auditory M100 response varies as a function of category-relevant information (formants) rather than voice-relevant information (fundamental frequency); specifically, the latency of M100 closely tracks the value of the first formant of a vowel, which contains information about the height of the vowel.

Although we have shown that M100 latency correlates closely with linguistically relevant properties of vowels, it remains for future work to determine whether this response pattern is particular to speech sounds, or whether it reflects general properties of auditory processing.

Acknowledgment

We would like to acknowledge the support of the following organizations for this research: NEC Corporation, NSF SGER grant #SBR-9423242, the McDonnell-Pew Center for Cognitive Neuroscience at MIT, the NSF-funded RTG Program 'Language: Acquisition and Computation' (#DIR9113607), NIH grant DC00194-02, and the JST funded project 'Mind Articulation' (co-PIs W. O'Neil, MIT and Y. Miyashita, U. of Tokyo).