The neural representation of segments and features

William Idsardi
University of Maryland at College Park

CUNY Phonology Forum, January 12, 2012
Collaborators

Phil Monahan (now BCBL)

Mathias Scharinger (now MPI)
3 Recent books

- J Schnupp, I Nelken and A King, *Auditory Neuroscience: Making Sense of Sound*
- PC Hansen, ML Kringelbach and R Salmelin, *MEG: An Introduction to Methods*
- BCJ Moore, LK Tyler, WD Marslen-Wilson, *The Perception of Speech*
Outline

• The problem
• MEG apparatus and methods
• Some experiments
Figure 1.16
Spectrograms of the words “hot,” “hat,” “hit,” and “head,” spoken in a high-pitched (top row) and a low-pitched (bottom row) voice.

from Schnupp et al 2010
Figure 2.3
Approximate best frequencies of various places along the basilar membrane, in hertz.

from Schnupp et al 2010
Cochleagram

Figure 2.6
Spectrogram of, and basilar membrane response to, the spoken word “head” (compare figure 1.16).

from Schnupp et al 2010
Ascending pathway

from Schnupp et al 2010
Cortical areas

Figure 2.18
Drawings showing identified auditory cortical areas in the ferret (A), the cat (B) and the rhesus macaque (C). Primary areas are shown in dark gray, higher-order (belt and parabelt) areas in light gray.

from Schnupp et al 2010
Figure 4.5
(A) Spectrotemporal receptive field of a neuron recorded in the auditory thalamus of the cat. BW: spectral bandwidth, BF: best frequency, Lat: response latency. (B) Modulation transfer function for the same neuron. BSM, best spectral modulation, BTM, best temporal modulation. Reproduced from figure 1 of Miller et al. (2002) with permission from the American Physiological Society.
Overall
A model
MEG apparatus
UMD MEG machine

(157 channels, KIT, Japan)
MEG Limitations

• cannot measure action potentials, instead post-synaptic changes

• poor capture of gyral changes, instead mostly sulcal

• cannot measure individual neurons, instead coordinated activity of tens of thousands of neurons
Speech and MEG
Auditory responses

- M100 (= N1m) an evoked response around 100 ms
- Mismatch field (MMF/MMN) a derived response between about 150-300 ms
- deviant response - standard response
What to measure?

- Amplitude
- Timing
- Location
Knowns (tones, noise)

- **Amplitude:**
  MMF is stronger for larger differences divergence from acoustic distance?
  (e.g. Chang English [u] vs. French [y], [u]?)

- **Timing:**
  M100 is later the further from ~1000 Hz

- **Location:**
  Tonotopy: lower frequencies more lateral
Speech is hard

• Not one frequency but many
• Organized into harmonics (laryngeal source) and formants (vocal tract filter)
• period-to-period micro-variations
Spectral integration for pitch harmonics
Neuromagnetic Evidence for Early Auditory Restoration of Fundamental Pitch

Philip J. Monahan¹*, Kevin de Souza¹, William J. IJsarda¹,²

¹Department of Linguistics, University of Maryland, College Park, Maryland, United States of America, ²Neuroscience and Cognitive Science Program, University of Maryland, College Park, Maryland, United States of America

Abstract

Background: Understanding the time course of how listeners reconstruct a missing fundamental component in an auditory stimulus remains elusive. We report MEG evidence that the missing fundamental component of a complex auditory stimulus is recovered in auditory cortex within 100 ms post stimulus onset.

Methodology: Two outside tones of four-tone complex stimuli were held constant (1200 Hz and 2400 Hz), while two inside tones were systematically modulated (between 1300 Hz and 2300 Hz), such that the restored fundamental (also known as “virtual pitch”) changed from 100 Hz to 600 Hz. Constructing the auditory stimuli in this manner controls for a number of spectral properties known to modulate the neuromagnetic signal. The tone complex stimuli only diverged on the value of the missing fundamental component.

Principal Findings: We compared the M100 latencies of these tone complexes to the M100 latencies elicited by their respective pure tone (spectral pitch) counterparts. The M100 latencies for the tone complexes matched their pure sinusoid counterparts, while also replicating the M100 temporal latency response curve found in previous studies.

Conclusions: Our findings suggest that listeners are reconstructing the inferred pitch by roughly 100 ms after stimulus onset and are consistent with previous electrophysiological research suggesting that the inferential pitch is perceived in early auditory cortex.
Previous findings: pitch

Roberts et al 2004 *NeuroReport*
Figure 1. A composite spectrogram of the seven complex tones used in the experiment. The duration of each complex tone was 70 ms, including 10 ms rise and decay time. Each complex tone included shoulder tones of 1200 Hz and 2400 Hz. Internal sidebands were synthesized in 100 Hz steps inward from the shoulder tones in six of the seven stimuli to induce the inferred fundamental components.

doi:10.1371/journal.pone.0002900.g001
<table>
<thead>
<tr>
<th>$F_{\text{inf}}$</th>
<th>$F_1$</th>
<th>$F_2$</th>
<th>$F_3$</th>
<th>$F_4$</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure sinusoids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>8.82</td>
<td>0.82</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>9.14</td>
<td>0.78</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>300</td>
<td>9.24</td>
<td>0.76</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>400</td>
<td>9.30</td>
<td>0.75</td>
<td>123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>600</td>
<td>600</td>
<td>9.35</td>
<td>0.74</td>
<td>179</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200</td>
<td>1200</td>
<td>9.40</td>
<td>0.73</td>
<td>337</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2400</td>
<td>2400</td>
<td>9.43</td>
<td>0.72</td>
<td>606</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone Complexes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 1200</td>
<td>1300</td>
<td>2300</td>
<td>2400</td>
<td>1800</td>
<td>552</td>
<td>−0.000020</td>
<td>−1.97</td>
<td></td>
</tr>
<tr>
<td>100 1200</td>
<td>1700</td>
<td>1900</td>
<td>2400</td>
<td>1800</td>
<td>430</td>
<td>0.000020</td>
<td>−1.10</td>
<td></td>
</tr>
<tr>
<td>200 1200</td>
<td>1400</td>
<td>2200</td>
<td>2400</td>
<td>1800</td>
<td>510</td>
<td>−0.000010</td>
<td>−1.85</td>
<td></td>
</tr>
<tr>
<td>300 1200</td>
<td>1500</td>
<td>2100</td>
<td>2400</td>
<td>1800</td>
<td>474</td>
<td>−0.000002</td>
<td>−1.64</td>
<td></td>
</tr>
<tr>
<td>400 1200</td>
<td>1600</td>
<td>2000</td>
<td>2400</td>
<td>1800</td>
<td>447</td>
<td>0.000009</td>
<td>−1.36</td>
<td></td>
</tr>
<tr>
<td>600 1200</td>
<td>1800</td>
<td>1800</td>
<td>2400</td>
<td>1800</td>
<td>490</td>
<td>−0.000002</td>
<td>−1.50</td>
<td></td>
</tr>
<tr>
<td>1200 1200</td>
<td>2400</td>
<td>1800</td>
<td>600</td>
<td>−0.000002</td>
<td>−2.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$F_{\text{inf}}$ = Inferred Fundamental (in Hz); $F_1$ = First Harmonic (in Hz); $F_2$ = Second Harmonic (in Hz); $F_3$ = Third Harmonic (in Hz); $F_4$ = Fourth Harmonic (in Hz); $M_1$ = Spectral Centre of Gravity (in Hz); $M_2$ = Standard Deviation (in Hz); $M_3$ = Skewness; $M_4$ = Kurtosis.

doi: 10.1371/journal.pone.0002900.t001
Figure 2. Comparison of the MEG waveforms to a pure sinusoid (in this case, 600 Hz) and tone complex with the corresponding inferred fundamental (in this case, 12-18-24) for a representative subject. Data is the RMS from 10 channels (five sink, five source) in the left hemisphere. The peak around 100 ms post-onset of the target (0 ms represents the onset of the target) is the M100. The peak latency of the M100 to the pure sinusoid and its corresponding tone complex were closely matched. The head-models represent the magnetic field contours for the M100. The red regions represent the source of the dipole and the blue regions represent the sink of the dipole.

doi:10.1371/journal.pone.0002900.g002
Figure 3. M100 RMS latencies to single sinusoid tones, tone complexes (plotted by their inferred fundamental component), and the 12-17-19-24 kHz tone complex, whose fundamental component is 100 Hz. Error bars refer to ±1 standard error of the group mean.
doi:10.1371/journal.pone.0002900.g003
Fixed Effect Tests

<table>
<thead>
<tr>
<th>Source</th>
<th>Nparm</th>
<th>DF</th>
<th>DFDen</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>tone type</td>
<td>1</td>
<td>1</td>
<td>85</td>
<td>3.4573</td>
<td>0.0664</td>
</tr>
<tr>
<td>Frequency</td>
<td>1</td>
<td>1</td>
<td>85</td>
<td>43.2134</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>tone type*Frequency</td>
<td>1</td>
<td>1</td>
<td>85</td>
<td>1.1715</td>
<td>0.2822</td>
</tr>
</tbody>
</table>

Graph showing M100 (ms) vs. Frequency (Hz) with different line types and colors for tone type categories.
Discussion

• Neural response shows a combination of harmonics into an integrated representation of pitch

• Higher order invariant

• What is the mechanism? Comb filter?
Spectral integration for formants
Auditory sensitivity to formant ratios: Toward an account of vowel normalisation

Philip J. Monahan\textsuperscript{1} and William J. Idsardi\textsuperscript{2}

\textsuperscript{1}Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain, \textsuperscript{2}Department of Linguistics, Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD, USA

A long-standing question in speech perception research is how listeners extract linguistic content from a highly variable acoustic input. In the domain of vowel perception, formant ratios, or the calculation of relative bark differences between vowel formants, have been a sporadically proposed solution. We propose a novel formant ratio algorithm in which the first (F1) and second (F2) formants are compared against the third formant (F3). Results from two magnetoencephalographic experiments are presented that suggest auditory cortex is sensitive to formant ratios. Our findings also demonstrate that the perceptual system shows heightened sensitivity to formant ratios for tokens located in more crowded regions of the vowel space. Additionally, we present statistical evidence that this algorithm eliminates speaker-dependent variation based on age and gender from vowel productions. We conclude that these results present an impetus to reconsider formant ratios as a legitimate mechanistic component in the solution to the problem of speaker normalisation.
Previous findings: F1

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

Roberts et al 2004 NeuroReport
Formants

![Graph showing formants F1, F2, F3, and f0 for adults and children, distinguishing between males and females.](image-url)
Formant ratios

- **Male**
- **Female**

Graph showing:
- **F1/F3**
- **F2/F3**
Figure 1. Vowel space normalised against F3. Note: Traditional vowel space plotted in the proposed normalised vowel space (F3 as the normalising factor). Formant values from which ratios were computed are from Hillenbrand et al. (1995) and averaged across age and gender per vowel category (except for /a/ which was not collected in Hillenbrand et al. (1995); instead, those values were computed from the frequency values used in the experiments reported here (before F3 modulation)).
Figure 3. Evoked M100 temporal waveform and magnetic field contour. Note: Temporal waveform from 10 left hemisphere channels (RMS: thick black line superimposed) and the magnetic field distribution at peak latency of the M100 for a representative subject. For the magnetic field distributions, the left anterior and right posterior channels are measuring the ingoing magnetic field sinks and the left posterior and right anterior channels are measuring the outgoing magnetic field sources.
Figure 4. Experiment 1: spectral slices of vowel tokens. Note: LPC-based spectral envelopes of the vowel sounds used in Experiment 1. The solid line indicates the token with a higher F3 (smaller F1/F3 ratio) and the dashed line indicates tokens with a lower F3 (larger F1/F3) ratio. Spectral envelopes smoothed with six-pole LPC filter.

<table>
<thead>
<tr>
<th>Vowel type</th>
<th>F3 height</th>
<th>F1 Centre frequency</th>
<th>F1 Bandwidth</th>
<th>F2 Centre frequency</th>
<th>F2 Bandwidth</th>
<th>F3 Centre frequency</th>
<th>F3 Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>Low</td>
<td>500</td>
<td>80</td>
<td>1500</td>
<td>90</td>
<td>2040</td>
<td>150</td>
</tr>
<tr>
<td>/a/</td>
<td>High</td>
<td>500</td>
<td>80</td>
<td>1500</td>
<td>90</td>
<td>3179</td>
<td>150</td>
</tr>
<tr>
<td>/æ/</td>
<td>Low</td>
<td>580</td>
<td>80</td>
<td>1712</td>
<td>90</td>
<td>2156</td>
<td>150</td>
</tr>
<tr>
<td>/æ/</td>
<td>High</td>
<td>580</td>
<td>80</td>
<td>1712</td>
<td>90</td>
<td>3247</td>
<td>150</td>
</tr>
</tbody>
</table>

Note: The centre frequency and bandwidth for each of the first three formants are provided in Hertz. The stimuli were synthesised using KLSyn (Stevens & Bickley, 1991), a user interface for the HLSyn speech synthesiser. The formant ratio calculations were performed in Mel frequency space and then converted back into Hertz for the speech synthesis.
M100 for \( \varepsilon \) and \( \theta \)
Figure 6. Experiment 2: spectral slices of vowel tokens. Note: LPC-based spectral envelopes of the vowel sounds used in Experiment 2. The solid line indicates the token with a high F3 (smaller F1/F3 ratio) and the dashed line indicates the token with a low F3 (larger F1/F3 ratio). Spectral envelopes smoothed with six-pole LPC filter.

<table>
<thead>
<tr>
<th>Vowel type</th>
<th>F3 height</th>
<th>F1 Centre frequency</th>
<th>F1 Bandwidth</th>
<th>F2 Centre frequency</th>
<th>F2 Bandwidth</th>
<th>F3 Centre frequency</th>
<th>F3 Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ə/</td>
<td>Low</td>
<td>500</td>
<td>80</td>
<td>1500</td>
<td>90</td>
<td>2040</td>
<td>150</td>
</tr>
<tr>
<td>/ə/</td>
<td>High</td>
<td>500</td>
<td>80</td>
<td>1500</td>
<td>90</td>
<td>3179</td>
<td>150</td>
</tr>
<tr>
<td>/o/</td>
<td>Low</td>
<td>497</td>
<td>80</td>
<td>938</td>
<td>90</td>
<td>2011</td>
<td>150</td>
</tr>
<tr>
<td>/o/</td>
<td>High</td>
<td>497</td>
<td>80</td>
<td>938</td>
<td>90</td>
<td>3118</td>
<td>150</td>
</tr>
</tbody>
</table>

Note: The centre frequency and bandwidth for each of the first three formants are provided in Hertz. The stimuli were synthesised using KLSyn (Stevens & Bickley, 1991), a user interface for the HLSyn speech synthesiser. The formant ratio calculations were performed in Mel frequency space and then converted back into Hertz for the speech synthesis.
M100 for o and ø
Discussion

• Early neural responses show vowel normalization consistent with formant ratios (or bark differences)

• Higher order invariant

• Effect is modulated by density of vowel space (existing vowel categories?)
Maps for vowels
A Comprehensive Three-dimensional Cortical Map of Vowel Space

Mathias Scharinger\textsuperscript{1,2}, William J. Idsardi\textsuperscript{1},
and Samantha Poe\textsuperscript{1}

Journal of Cognitive Neuroscience 2011
Previous findings

Fig. 2. Mean spatial ECD coordinates of waves M100 and M200 with their standard errors (crossed bars) for the different stimulus signals. The arrows represent amplitude and orientation (with respect to the horizontal axis) of the ECD moment. ●, Virtual pitch (250 Hz); ■, spectral pitch (250 Hz); and ▼, spectral pitch (1000 Hz).

Pantev et al 1989 Science
# Turkish vowels

<table>
<thead>
<tr>
<th></th>
<th>front</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-round</td>
<td>round</td>
</tr>
<tr>
<td>high</td>
<td>i</td>
<td>ü (y)</td>
</tr>
<tr>
<td>low</td>
<td>e (ɛ)</td>
<td>ö (oe)</td>
</tr>
</tbody>
</table>
A. Experimental design

\[ p = 0.11 \]

[Sound representation with symbols and time axis labeled 'Time']

Response
C. Acoustic maps and articulatory classification of Turkish vowels

Front vowels:
- [i]
- [y]
- [ɛ]
- [æ]

Back vowels:
- [o]
- [ʊ]
- [ʌ]

F1 [Hz] vs. F2 [Hz]
- High vs. Low
- Rounded vs. Unrounded
D. Cortical maps
Discussion

- **Two** separate perpendicular maps: front vowels and back vowels
- In both maps, rounding is represented on medial-lateral axis (following Pantev)
- Height is represented differently in the two maps (or, distance from intersection?)
- Maps show warping from acoustic values
Maps for consonants
Categorical perception of /s/ and /ʃ/? An MMN study

Sol Lago, Yakov Kronrod, Mathias J. Scharinger, William J. Idsardi ---- University of Maryland, USA
Behavioral results

Figure 1: Identification and discrimination results from the 10-step fricative continuum. Across-category discrimination was significantly better than within-category discrimination (logistic regression % accuracy ~ category type; coefficient = -0.776, Wald-z = -3.55, p < 0.05).
Figure 2: Fricative continuum, defined by the 5th and 6th frication formants, F5 and F6. A, B, C, and D represent fricative categories that were used in the MMN study. Each category was represented by 3 variable tokens, 0.125 Barks apart in F5 and F6.

Figure 3: Experimental design of the modified optimum-2 passive-odd-ball paradigm. There were 4 blocks, one for each token A, B, C, D as standard.
Results - Dipoles

Right

Left

Wednesday, January 18, 12
Discussion

• lower frequencies again more lateral

• category-based warping of cortical distance (within category sounds closer together)

• need category information to explain results (acoustics isn’t sufficient)

• need to do matched noise, other languages (Thai, Korean)
Mismatch for dialects
You had me at “Hello”: Rapid extraction of dialect information from spoken words

Mathias Scharinger a,c,* , Philip J. Monahan b , William J. Idsardi a

a Department of Linguistics, University of Maryland, College Park, MD, USA
b Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastián, Spain
c Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
Figure 1: Experimental setup
# Material

<table>
<thead>
<tr>
<th>Dialect</th>
<th>Segment</th>
<th>Length [ms]</th>
<th>Intensity [dB]</th>
<th>Pitch [Hz]</th>
<th>F1 [Hz]</th>
<th>F2 [Hz]</th>
<th>F3 [Hz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAE</td>
<td>[h]</td>
<td>117</td>
<td>47</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>[ε]</td>
<td>74</td>
<td>70</td>
<td>106</td>
<td>474</td>
<td>1311</td>
<td>2664</td>
</tr>
<tr>
<td></td>
<td>[l]</td>
<td>62</td>
<td>74</td>
<td>105</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>[o]</td>
<td>270</td>
<td>70</td>
<td>100</td>
<td>612</td>
<td>1602</td>
<td>2961</td>
</tr>
<tr>
<td>SAE</td>
<td>[h]</td>
<td>100</td>
<td>45</td>
<td>98</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>[ε]</td>
<td>66</td>
<td>68</td>
<td>110</td>
<td>561</td>
<td>1592</td>
<td>2978</td>
</tr>
<tr>
<td></td>
<td>[l]</td>
<td>60</td>
<td>74</td>
<td>104</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>[o]</td>
<td>257</td>
<td>70</td>
<td>110</td>
<td>573</td>
<td>1580</td>
<td>3014</td>
</tr>
</tbody>
</table>

*Table 1: Averaged acoustic characteristics of the “Hello” tokens*
Real deviant (across dialect)

Sham deviant (within dialect)

SAE

AAE

[ft]

* n.s.

n.s. n.s.

[ms]

[ms]
Figure 3: Equivalent Current Dipole (ECD) locations for deviants in the left hemisphere for the early (200 ms) and late (400 ms) time window. Dipole locations differed in the anterior-posterior and in the lateral-medial dimension.
Discussion

• Find significant mismatch responses for both SAE and AAE (but not symmetrical)
• ‘bleached’ token: “hello”
• need dialect category to explain responses, acoustics alone isn’t sufficient
More experiments

• Allophonic vs. phonemic t/d (PNAS 2006)
• English w, j, v, ʒ (Brain & Lg 2011)
• English æ, ε, ɪ (JSLHR 2012)
• Turkish ü, œ, ε (J. of Lab Phon 2011) (compare Nimz)
• Vowels compared with musical chords
General Conclusions

• neural evidence for “higher order” or integrated information

• M100 and MMF responses from MEG are composite

• Acoustics + Category information
Thank you!

idsardi@umd.edu