The Nijmegen Lectures: Lecture 2

On the sufficiency of abstract structure

David Poeppel
Max-Planck-Institute, Frankfurt
Department of Psychology, NYU
Outline

- The temporal modulations of speech provide a big hint ...
- Oscillations link physics to linguistics to neuroscience
- Cortical entrainment to abstract structure
The syllable as perceptual primitive?

**Figure 1.** Histogram of the intervals between some 10,000 successive jaw openings in running speech (reading).

Ohala 1972

**Figure 2.** The relation between the distribution of syllable duration (transformed into modulation frequency) and the modulation spectrum of the same Japanese material as shown in Fig. 1, computed for the octave region between 1 and 2 kHz. Adapted from [1].

Greenberg & Arai 2004

**http://www.phonetik.uni-muenchen.de/Bas/BasPHONSTATeng.html**

---

**Figure 1.** Speech rate measured in terms of the number of syllables per second (mean values and 95% confidence intervals). Stars indicate significant differences between the homogeneous subsets revealed by post-hoc analysis.

Pellegrino et al. 2011
Outline

• The temporal modulations of speech provide a big hint ...

• Oscillations link physics to linguistics to neuroscience

• Cortical entrainment to abstract structure
Design: evaluate coherence across single trials elicited by sentences

Cross-trial coherence

Within-group

Similarity across trials should be larger in Within-group condition than in Across-group condition

\[
\begin{align*}
\text{Cphase}_{within} > \text{Cphase}_{across} \\
\text{Cpower}_{within} > \text{Cpower}_{across}
\end{align*}
\]

Across-group
Theta phase

Luo & Poeppel, Neuron, 2007
Theta phase has the sensitivity to discriminate based on single trials

Materials:

Theta phase tracking displays the specificity to discriminate sentences

Classification analysis

Luo & Poeppel, Neuron, 2007
A ~ 200 ms window analyzes the input signal -- The syllable as primitive
Auditory cortical activity is entrained to the envelope ⇒ syllabic rhythm.

Neural entrainment is seen in both the theta and delta bands during spoken language comprehension.

e.g. Luo & Poeppel, Neuron 2007; Ding & Simon, PNAS 2012; J Neuroscience 2013
ECoG Single Trials, an example:

- Attend Female: Single Trial 1
- Attend Female: Single Trial 2
- Attend Male: Single Trial 3

Zion-Golumbic et al. *Neuron* 2013
Cortical oscillations and speech processing: emerging computational principles and operations

1. Phase reset
2. Stimulus envelope tracking
3. Theta/gamma nesting
4. Modulation of neuronal excitability and output discretization
5. Alignment of neuronal excitability with acoustic structure

Giraud & Poeppel, 2012, Nat Neurosci
Parsing events, e.g. syllables

 Courtesy of Keith Doelling, NYU
Parsing events, e.g. syllables

she had your dark suiting was washed water all year

Courtesy of Keith Doelling, NYU
An interesting alignment between:

theta rhythm (4-8 Hz) – *systems neuroscience*

modulation spectrum of speech (4-5 Hz) – *physics*

mean syllable duration cross-linguistically (150-300 ms) – *linguistics*
Outline

• The temporal modulations of speech provide a big hint ...

• Oscillations link physics to linguistics to neuroscience

• Cortical entrainment to abstract structure
Cortical representation of the constituent structure of sentences

Christophe Pallier\textsuperscript{a,b,c,d,1,2}, Anne-Dominique Devauchelle\textsuperscript{a,c,d,1}, and Stanislas Dehaene\textsuperscript{a,c,d,e,2}

Table 1. The stimuli were 12 items long sequences obtained by concatenating constituents of fixed sizes extracted from natural or jabberwocky right-branching sentences

<table>
<thead>
<tr>
<th>Condition</th>
<th>Constituent size</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>c12</td>
<td>12 words</td>
<td>I believe that you should accept the proposal of your new associate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I t osieve that you should begept the tropufal of your new viroate</td>
</tr>
<tr>
<td>c06</td>
<td>6 words</td>
<td>the mouse that eats our cheese two clients examine this nice couch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>the cause that rits our treeve fow plients afomine this kice bloch</td>
</tr>
<tr>
<td>c04</td>
<td>4 words</td>
<td>mayor of the city he hates this color they read their names</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tuyor of the roty he futes this dator they gead their wames</td>
</tr>
<tr>
<td>c03</td>
<td>3 words</td>
<td>solving a problem repair the ceiling he keeps reading will buy some</td>
</tr>
<tr>
<td></td>
<td></td>
<td>relging a graethem regair the fraping he meeps bounding will doy some</td>
</tr>
<tr>
<td>c02</td>
<td>2 words</td>
<td>looking ahead important task who dies his dog few holes they write</td>
</tr>
<tr>
<td></td>
<td></td>
<td>troking ahead omirpant fran who mies his gog few biles they grite</td>
</tr>
<tr>
<td>c01</td>
<td>1 word</td>
<td>thing very tree where of watching copy tensed they states heart plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>thang very gree where of wurthing napry gunsed they otes blart trus</td>
</tr>
</tbody>
</table>

In jabberwocky, all content words were replaced with pseudowords (italics). Examples are only illustrative, because the original stimuli were in French.

Fig. 1. Simulations of a simple model in which neural activity is assumed to increase by one unit each time a new word is incorporated into a constituent and to return to baseline as soon as a novel word cannot be incorporated into the preceding constituent. This model predicts an increase in both amplitude and phase of activation as a function of constituent size from condition c01 to c12.
Fig. 2. Brain regions showing a significant increase in activation with constituent size. (A) fMRI results from the normal-prose group who read sequences with actual French words (group analysis thresholded at $T > 4.5$, $P < 0.05$ FWE, spatial extent > 10). (B) Areas in blue show a significant constituent size effect in the jabberwocky group listening to delexicalized stimuli, whereas regions in red show a significant group by constituent size interaction (reflecting a stronger effect of constituent size in normal prose than in jabberwocky) (maps thresholded at $T > 3.2$, $P < 0.001$ uncorrected, spatial extent > 50). (C) Amplitude of activations across conditions in the six regions of interest (error bars represent $\pm$ 1 SEM). Conditions c01 to c12 are organized according to a logarithmic scale of constituent size, thus a line on this graph indicates a logarithmic increase of activation. The fitting lines are from a regression analysis including linear and logarithmic predictors.
Cortical tracking of hierarchical linguistic structures in connected speech

Nai Ding
NYU
Zhejiang Univ.

Lucia Melloni
Max Planck
NYU
A
“The box contained a thin letter from Italy”

B
syllable

word

phrase

Adapted from Peelle & Davis 2012
Boundaries between syllables are usually defined by the speech envelope, but not the boundaries between words and phrases.
The neural code for each linguistic unit must change at the rate of that linguistic unit.
Hierarchical Entrainment to the Hierarchical Linguistic Structure?
Hierarchical Entrainment to the Hierarchical Linguistic Structure?

e.g., Luo & Poeppel, 2007
Ding & Simon, 2012
A Sequence with Hierarchical Linguistic Structures

sentence

N phrase  V phrase

(sentence)

sentence

N phrase  V phrase

(sentence)

干 冻 摸 化 毛 羊 吃 草

dry fur rubs skin new plans gave hope

…… 1 Hz

…… 2 Hz

…… 4 Hz

250 ms
• 16 native listeners of Mandarin Chinese

• Outlier detection: occasionally, the noun phrases of two sentences will be switched, creating two nonsense sentences.

• Data processed by a spatial filter optimized to extract phase-locked activity.

(Wang et al., J Neurophys 2012; Ding & Simon, PNAS 2012; de Cheveigné & Simon, 2008)
Cortical Activity Tracks
Hierarchical Linguistic Rhythms

\[ f_{\text{sentence}} \quad f_{\text{phrase}} \quad f_{\text{syllable}} \]

\[ \text{power (dB)} \]

\[ \text{frequency (Hz)} \]

6 dB
Data from Individual Listeners

Individual Chinese Listeners (N=16)
Non-speakers Only Track
the Syllabic/Acoustic Rhythm

Chinese materials, English listener
The English Version

Adj. + Noun + Verb + Noun

fat rat sensed fear
wood shelf holds cans
tan girls drove trucks
gold lamps shine light
dry fur rubs skin
sly fox stole eggs
top chefs cook steak
our boss wrote notes
two teams plant trees
...

new plans give hope
large ants built nests
teen apes hunt bugs
rude cats claw dogs
rich cooks brewed tea
fun games waste time
huge waves hit ships
deaf ears hear you
all moms love kids
...

With help from Gwyneth Lewis
Hierarchical Entrainment for English

English materials, English listener

![Diagram showing power across frequency (Hz)]
Interim Summary

• Cortical activity is entrained to the phrasal and sentential rhythms of speech.

• Phrasal/sentential level entrainment is seen for both Chinese and English, and not confounded by encoding of acoustic features.
A Markov Chain Language with Constant Transitional Probability

John, Jess, the boy, a girl, her dad

lives in | ordered | wrote a | speaks | didn't

\{ beer, soup, salad \} | \{ pizza, coffee, \} | \{ book, letter, story \} | \{ poem, memo \}
A Markov Chain Language with Constant Transitional Probability

John
Jess
the boy
a girl
her dad

lives in
ordered
wrote a
speaks
didn't

\{ beer, soup, salad \}
\{ pizza, coffee, \}
\{ book, letter, story \}
\{ poem, memo \}

Probability

\begin{align*}
E_1 & \rightarrow E_2 & \rightarrow E_3 & \rightarrow E_1 \\
\text{time} & \rightarrow & \rightarrow & \rightarrow \\
1/5 & \rightarrow & 1/5 & \rightarrow & 1/5 \\
E_1 & \rightarrow & E_2 & \rightarrow & E_3 \\
\end{align*}

Fourier transform

$$f_E$$

frequency
Predictable Sentences

\[ N = 25 \]

\[
\begin{align*}
  \text{My cat} & \quad \rightarrow \quad \text{is so} & \quad \rightarrow \quad \text{lovely} \\
  \text{They} & \quad \rightarrow \quad \text{grow} & \quad \rightarrow \quad \text{apples} \\
  \text{Sarah} & \quad \rightarrow \quad \text{looks} & \quad \rightarrow \quad \text{happy} \\
  \ldots & & \ldots \\
  \ldots & & \ldots \\
\end{align*}
\]

350 ms

each sentence played \( \sim 12 \) times
A Predictable Sentences

\[
\frac{1}{25} \rightarrow E_1 \rightarrow E_2 \rightarrow E_3
\]

\[
\begin{align*}
N = 25 & \quad \text{My cat} \rightarrow \text{is so} \rightarrow \text{lovely} \\
& \quad \text{They} \rightarrow \text{grow} \rightarrow \text{apples} \\
& \quad \text{Sarah} \rightarrow \text{looks} \rightarrow \text{happy} \\
\end{align*}
\]

350 ms

25 sentences, each repeated \( \sim 12 \) times

B transitional probability

\[
\begin{align*}
1/25 & \quad \bullet \quad 1 \quad \bullet \quad 1/25 \quad 1 \quad 1 \quad 1 \\
E_1 & \quad E_2 \quad E_3 \quad E_1 \quad E_2 \quad E_3 \\

\end{align*}
\]

Fourier analysis

\[
\begin{align*}
f_s & \quad f_e \\
\end{align*}
\]

\[
\begin{align*}
1/5 & \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \\
\text{time} & \quad \text{frequency} \\
\end{align*}
\]

- predictable sentences
- constant predictability sentences
B  transitional probability

1/25  1/25  1/5

E₁  E₂  E₃  E₁  E₂  E₃

time

Fourier analysis

fₛ  fₑ

frequency

predictable sentences
constant predictability sentences

C

fₛ  fₑ

6 dB

power (dB)

1/1.05  2/1.05  3/1.05

frequency (Hz)
demonstrating that the neural response continuously changes during the course of a sentence, which can last up to 2 seconds, rather than being a transient response only occurring at the sentence boundary.

Figure 4. Neural tracking of sentences of varying structures. (A) The global field power of the neural response tracks the sentence duration, even when the sentence boundaries (dotted lines) are not conveyed by any acoustic gap. The responses to sentences of different durations are color-coded (sentence duration labeled in the same color). (B) The averaged global field power near a sentential boundary (colored bar). The power continuously changes throughout the duration of a sentence. Shaded area denotes 2 SEM. Significance differences between the power averaged over the duration of each syllable (shaded squares at the bottom) are marked by a star ($P < 0.01$, t-test, FDR corrected). (C) Confusion matrix for neural decoding of the sentence duration (leave-one-out cross-validation). (D) The responses to sentences starting with either a 3-syllable noun phrase or a 4-syllable noun phrase (shown in the bottom). The global field power shows a transient increase near the boundary of the noun phrase and the verb phrase, even when the phrasal boundary is not indicated by an acoustic gap. The yellow areas show the time intervals in which the response power depends on the phrase duration ($P < 0.005$, bootstrap, FDR corrected). A single-trial decoding analysis was performed to confirm by other means that cortical activity tracks the duration of sentences. The decoder applies template matching for the response time course, and its performance is evaluated by a leave-one-out cross-validation procedure. The confusion matrix, i.e. histogram of the decoder's output for sentences of different durations is shown in Fig. 4C. On average, 34.9 ± 0.6% (mean ± SEM over subjects) of the sentences are correctly detected (significantly better than chance level, i.e. 20%, $P < 10^{-6}$, t-test).
Topography of the Responses

$f_{\text{sentence}}$  $f_{\text{phrase}}$  $f_{\text{syllable}}$
Neural Source Localization using ECoG

5 epileptic patients
left hemisphere (3 patients)
right hemisphere (2 patients)
Spatially Dissociable Sentential and Phrasal Representations

High-Gamma Power

\[ f_s \quad f_p \quad f_\sigma \]

6 dB

N = 22

N = 10

N = 13

frequency (Hz)
Spatially Dissociable Sentential and Phrasal Representations

Low-Frequency Waveform

\[ f_s \quad f_p \quad f_\sigma \]

N = 62

N = 60

N = 29

frequency (Hz)
Summary

• Cortical circuits can generate slow rhythms matching the time scales of larger linguistic structures, even when such rhythms are not present in the speech input, which provides a plausible mechanism for online building of large linguistic structures.

• Such tracking of larger linguistic units is rule/grammar-based, not confounded by encoding of auditory features or transitional probability.
Cortical Entrainment to the Hierarchical Linguistic Structure of Spoken Language

Nature Neuroscience, 2015

Nai Ding
NYU

Lucia Melloni
Max Planck, NYU
The temporal modulations of speech provide a big hint ... There is a link++ between temporal modulations and syllables

Oscillations link physics to linguistics to neuroscience
Tracking of speech by phase resetting of intrinsic oscillations

Cortical entrainment to abstract structure
Entrainment to structure in absence of acoustic or stats cues
Thanks to support from NIH, NSF, ARO, AFOSR, Max-Planck Society