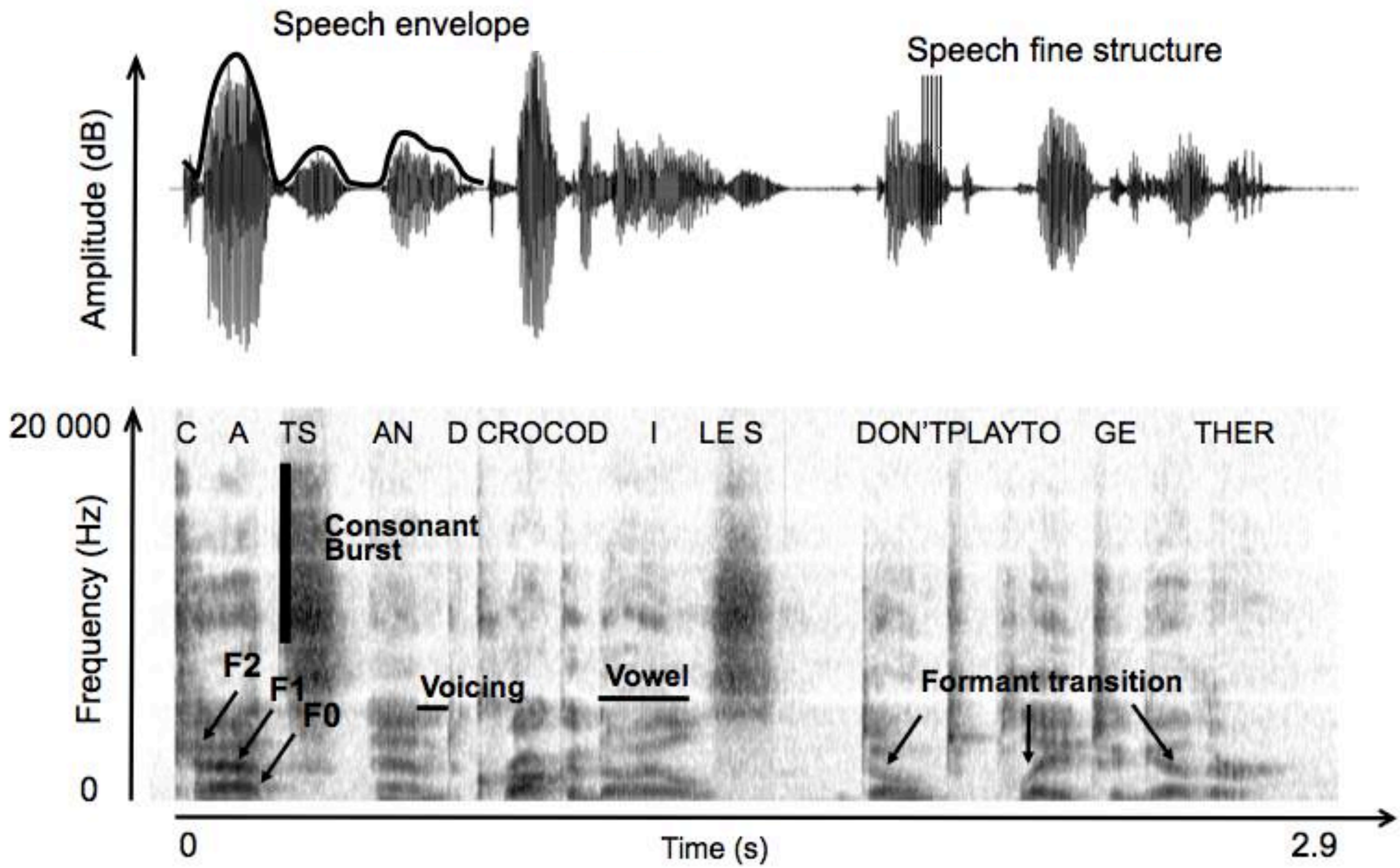
Spoken language comprehension requires segmentation and structure building.

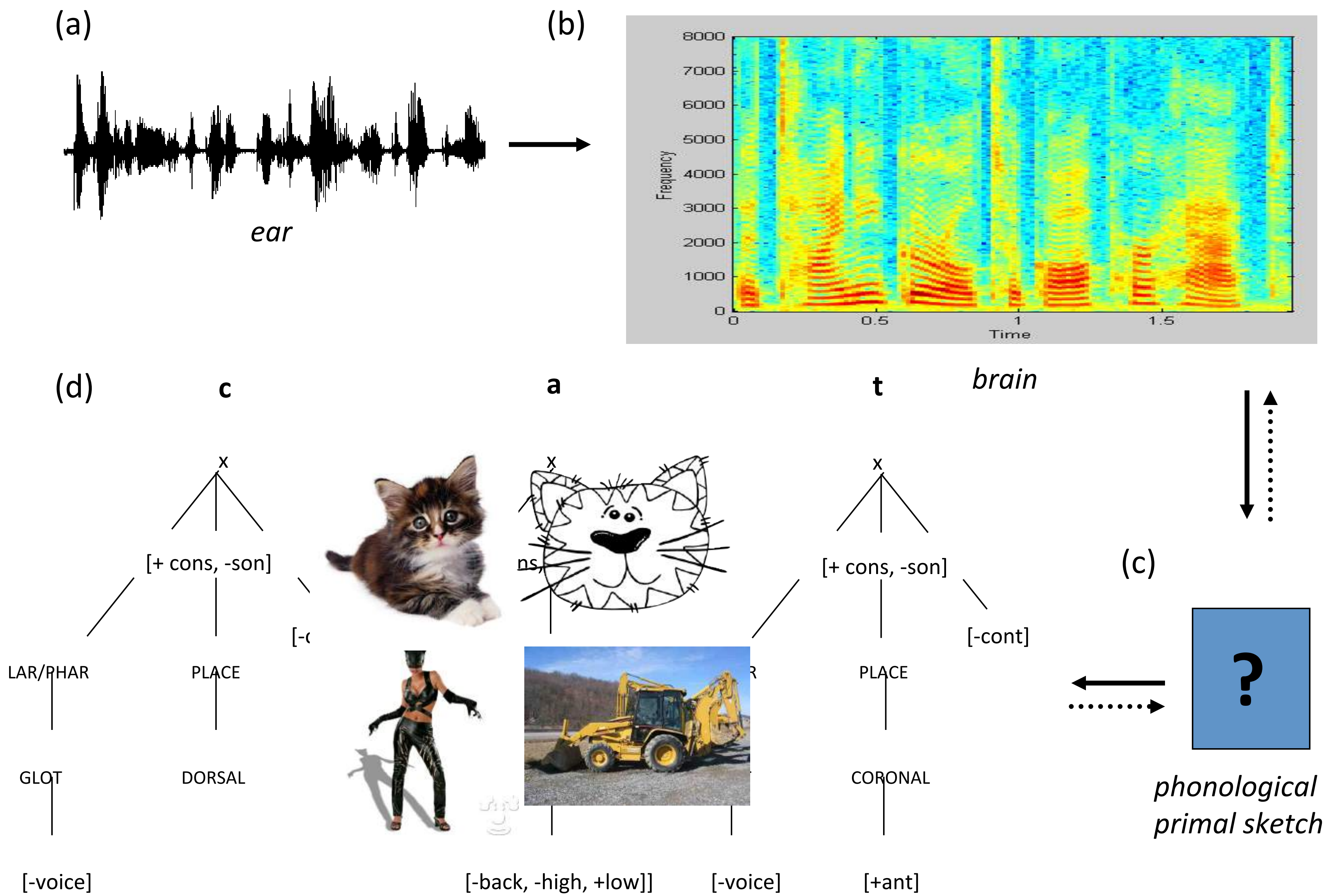
A view from cortical oscillations

David Poeppel
Max-Planck-Institute, Frankfurt
New York University

“Cats and crocodiles don’t play together”

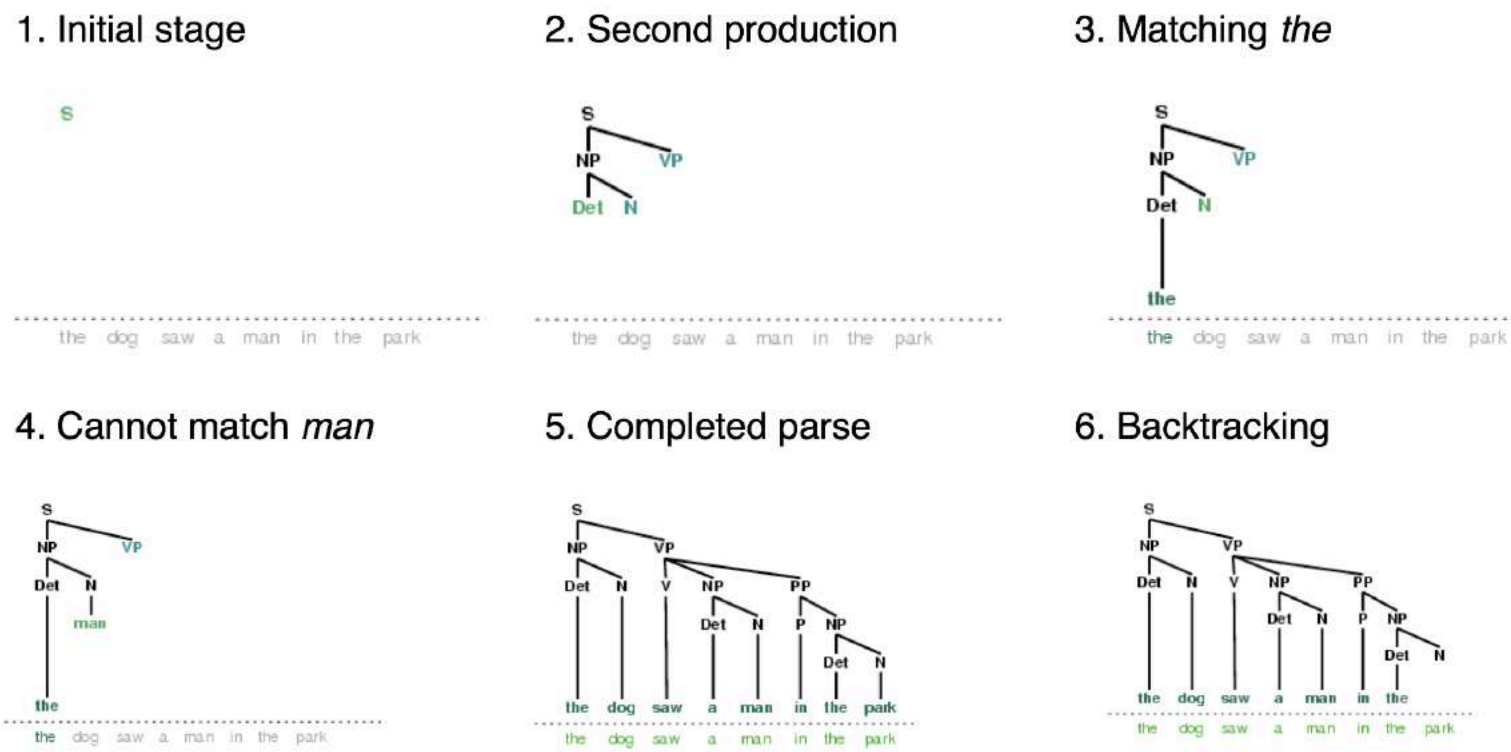


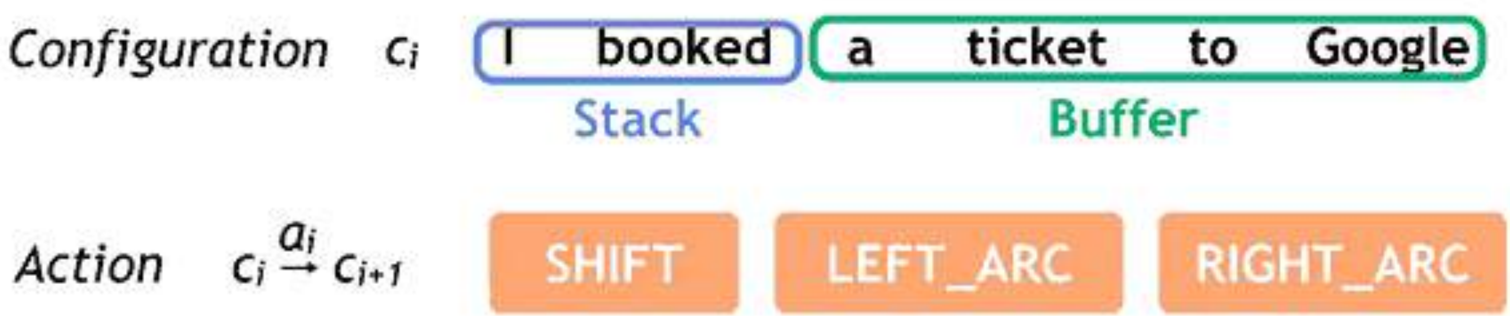
Zooming in on the problem: from vibrations in the ear to abstractions in the head



Poeppel, v. Wassenhove, Idsardi 2008

Zooming in on the problem #2: structure building for interpretation

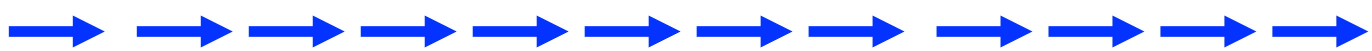




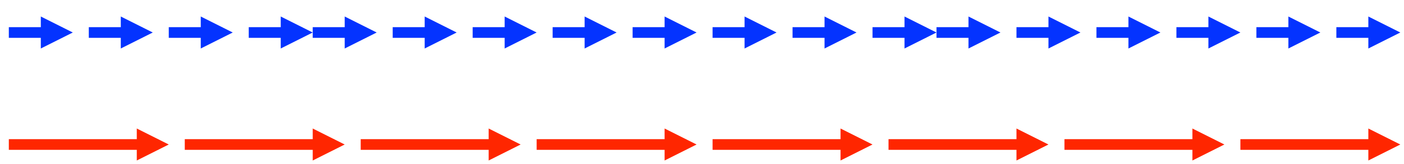
Newton, *Principia*

Our intuition

Our brain



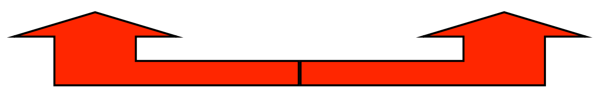
Our brain, really



Language & Brain

Anyone who seriously approaches the study of linguistic behavior, whether linguist, psychologist, or philosopher, must quickly become aware of the enormous difficulty of stating a problem which will define the area of his investigations, and which will not be either completely trivial or hopelessly beyond the range of present-day understanding and technique.

Chomsky 1959



Why bother? What could we learn?

- ***something about how language works***
- ***something about how the brain works***
- ***nothing (interdisciplinary cross-sterilization)***

The elementary particles (primitives) of language and music

representational
computational

Hypothesized representational primitives: language [domain specific]

- feature (articulatory)
- phoneme
- syllable
- morpheme
- noun-phrase, verb-phrase, etc…
- clause
- sentence
- discourse/narrative

Hypothesized representational primitives: music [domain specific]

- note (pitch and timbre)
- pitch interval (consonance/dissonance)
- octave-based pitch scale
- pitch hierarchy (tonality)

- discrete time interval
- beat
- meter

- motif/theme
- melody/satz
- piece

The elementary particles (primitives) of language and music

implementational

Hypothesized implementational (neurobiological) infrastructure

representational
computational

Hypothesized representational primitives: language [domain specific]

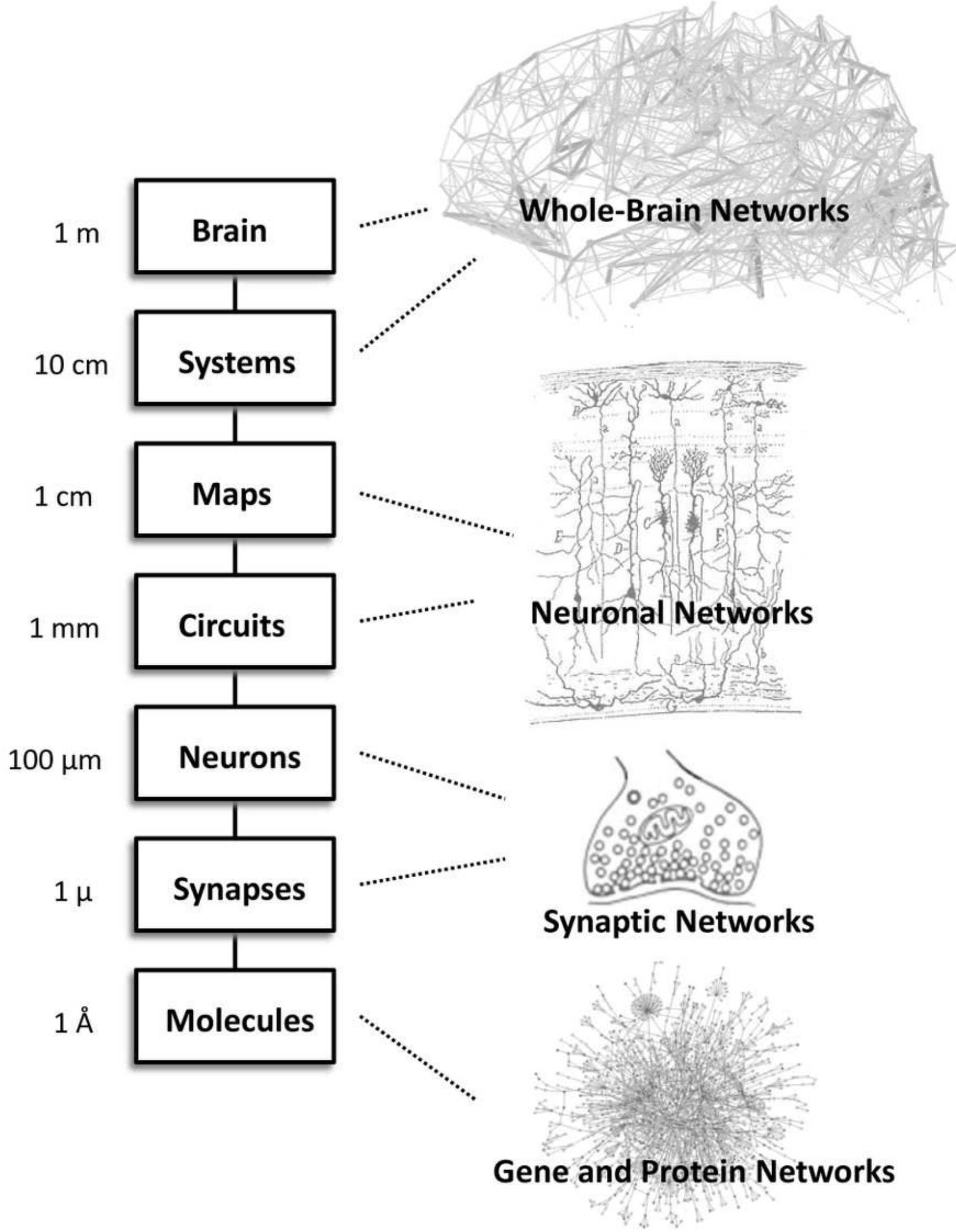
- feature (articulatory)
- phoneme
- syllable
- morpheme
- noun-phrase, verb-phrase, etc…
- clause
- sentence
- discourse/narrative

Hypothesized representational primitives: music [domain specific]

- note (pitch and timbre)
- pitch interval (consonance/dissonance)
- octave-based pitch scale
- pitch hierarchy (tonality)

- discrete time interval
- beat
- meter

- motif/theme
- melody/satz
- piece


1 m
Brain
10 cm
Systems
1 cm
Maps
1 mm
Circuits
100 µm
Neurons
1 µ
Synapses
1 Å
Molecules
Whole-Brain Networks
Neuronal Networks
Synaptic Networks
Gene and Protein Networks

The elementary particles (primitives) of language and music

implementational	Hypothesized implementational (neurobiological) infrastructure	
algorithmic representational	Hypothesized computational primitives [domain general] • constructing spatiotemporal objects (streams, gestures) • extracting relative pitch • extracting relative time • discretization • sequencing - concatenation - ordering • grouping - constituency - hierarchy • establishing relationships - local/long-distance • coordinate transformations • prediction • synchronization - entrainment - turn-taking • concurrent processing over different levels	
representational computational	Hypothesized representational primitives: language [domain specific] • feature (articulatory) • phoneme • syllable • morpheme • noun-phrase, verb-phrase, etc… • clause • sentence • discourse/narrative	Hypothesized representational primitives: music [domain specific] • note (pitch and timbre) • pitch interval (consonance/dissonance) • octave-based pitch scale • pitch hierarchy (tonality) • discrete time interval • beat • meter • motif/theme • melody/satz • piece

Levels of analysis: a view from the perspective of David Marr

implementational

Hypothesized implementational (neurobiological) infrastructure

algorithmic
representational

Hypothesized computational primitives [domain general]

- constructing spatiotemporal objects (streams, gestures)
- extracting relative pitch
- extracting relative time
- discretization
- sequencing - concatenation - ordering
- grouping - constituency - hierarchy
- establishing relationships - local/long-distance
- coordinate transformations

- prediction
- synchronization - entrainment - turn-taking
- concurrent processing over different levels

What kind of neural circuits and neural dynamics may underpin…

representational
computational

Hypothesized representational primitives: language [domain specific]

- feature (articulatory)
- phoneme
- syllable
- morpheme
- noun-phrase, verb-phrase, etc…
- clause
- sentence
- discourse/narrative

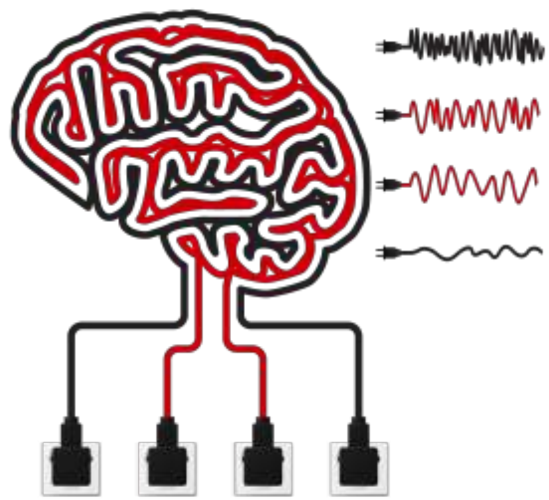
Hypothesized representational primitives: music [domain specific]

- note (pitch and timbre)
- pitch interval (consonance/dissonance)
- octave-based pitch scale
- pitch hierarchy (tonality)

- discrete time interval
- beat
- meter

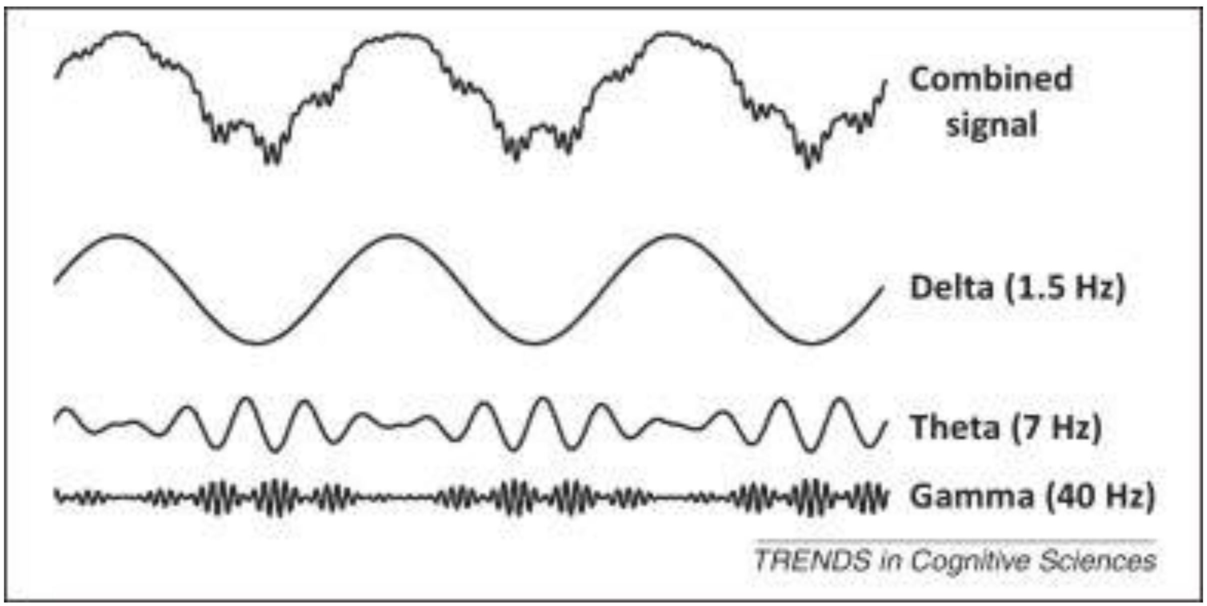
- motif/theme
- melody/satz
- piece

Unifying concept: neural oscillations



- Homeostatic functions
- Exploitation for computation – specific functions
- Epiphenomenal ("the exhaust fumes cortical computation")

Measurements of EEG, MEG, ECoG, LFP …



Calderone et al. 2014

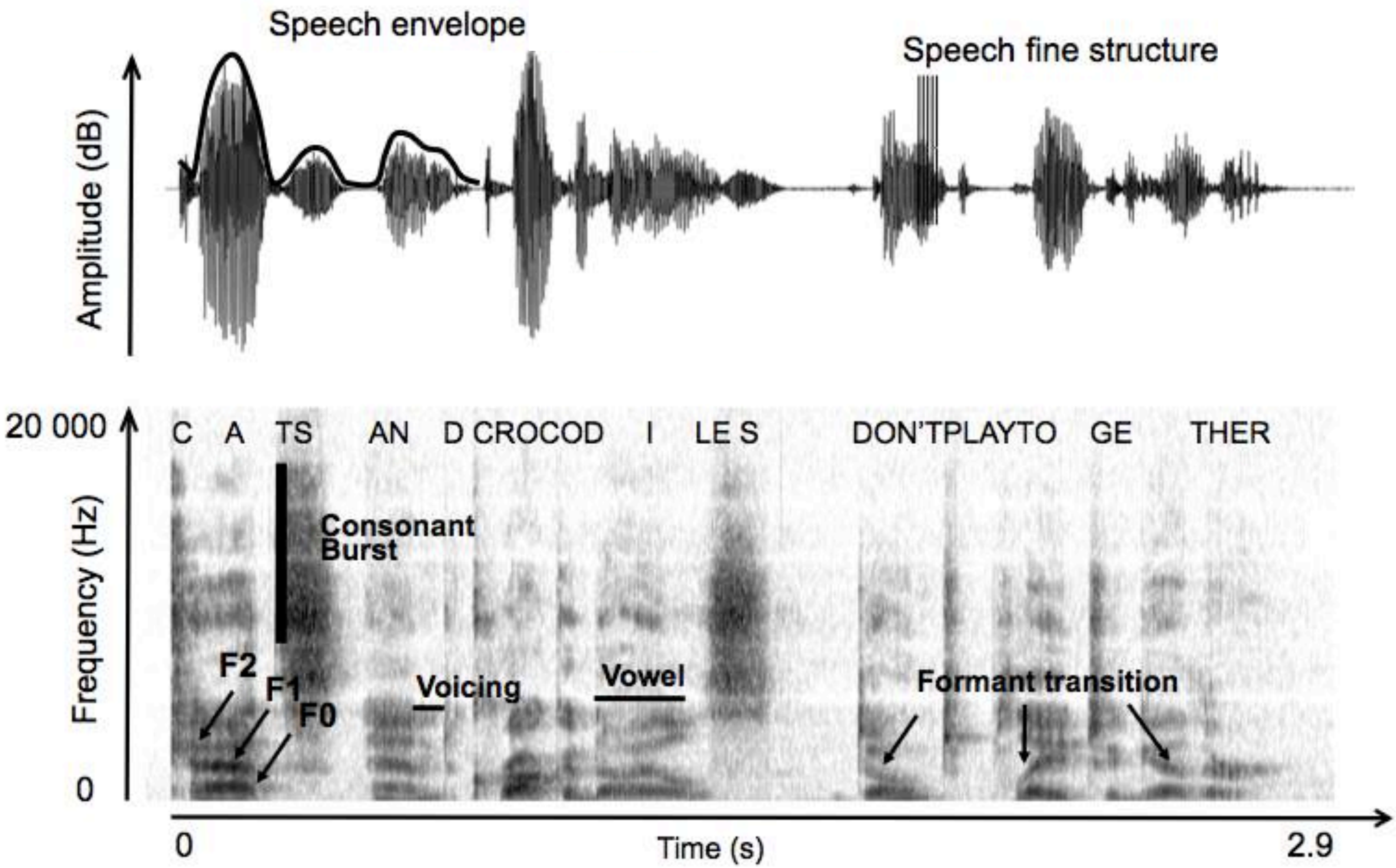
Entrainment and segmentation

sincetherearenowordboundarysignsinspokenlanguagethedifficultywefeelinreading
andunderstandingtheaboveparagraphprovidesasimpleillustrationofoneofthemaind
ifficultieswehavetoovercomeinordertounderstandspeechratherthananeatlyseparat
edsequenceofletterstringscorrespondingtothephonologicalformofwordsthespeech
signalisacontinuousstreamofsoundsthatrepresentthephonologicalformsofwordsin
additionthesoundsofneighboringwordsoftenoverlapwhichmakestheproblemofident
ifyingwordboundariesevenharder

Since there are no word boundary signs in spoken language the difficulty we feel in reading and understanding the above paragraph provides a simple illustration of one of the main difficulties we have to overcome in order to understand speech Rather than a neatly separated sequence of letter strings corresponding to the phonological form of words the speech signal is a continuous stream of sounds that represent the phonological forms of words In addition the sounds of neighboring words often overlap which makes the problem of identifying word boundaries even harder

Two operations must be executed to solve this:
segmentation and decoding

"Cats and crocodiles don't play together"

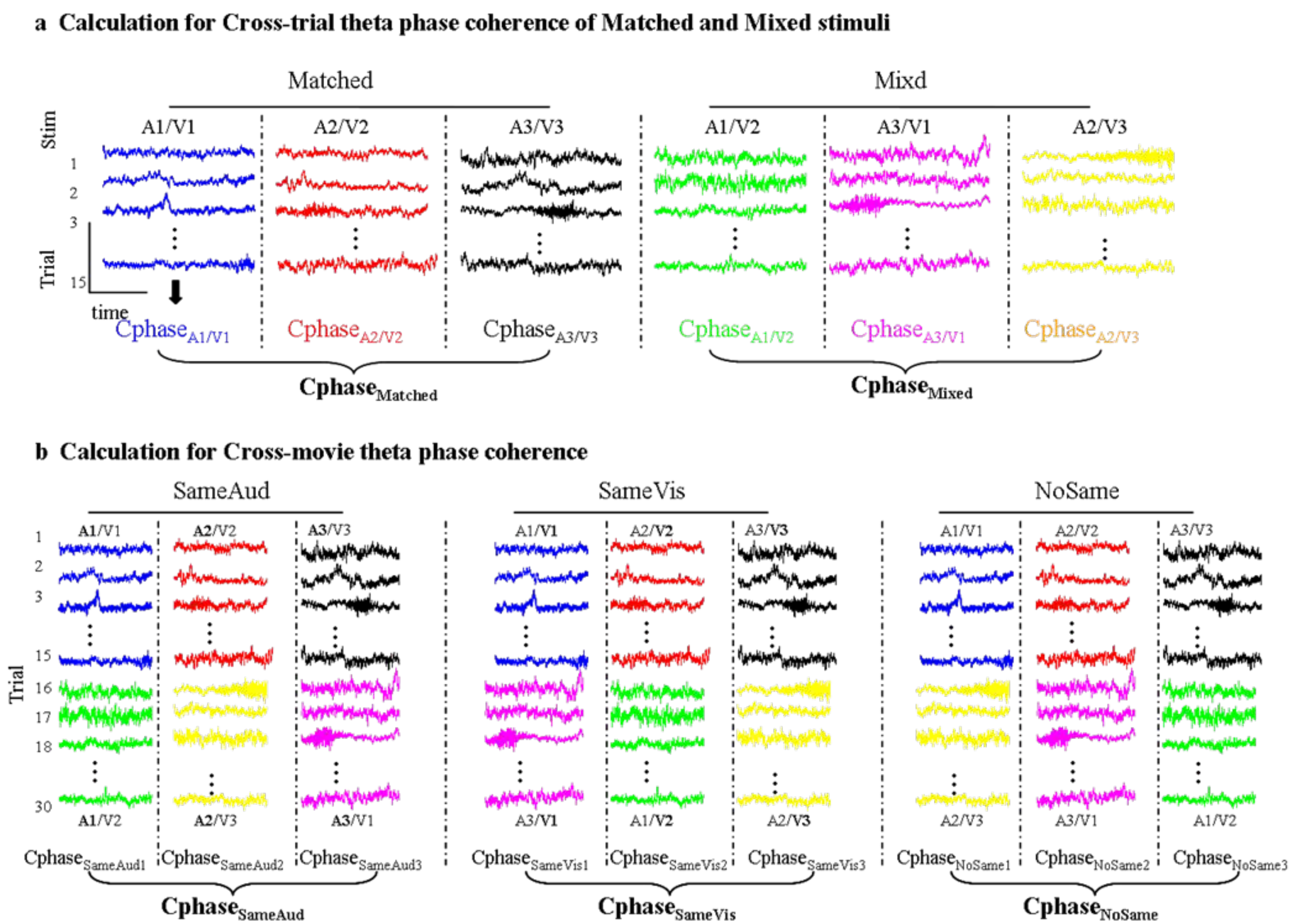


Auditory cortical activity is entrained to the speech envelope

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

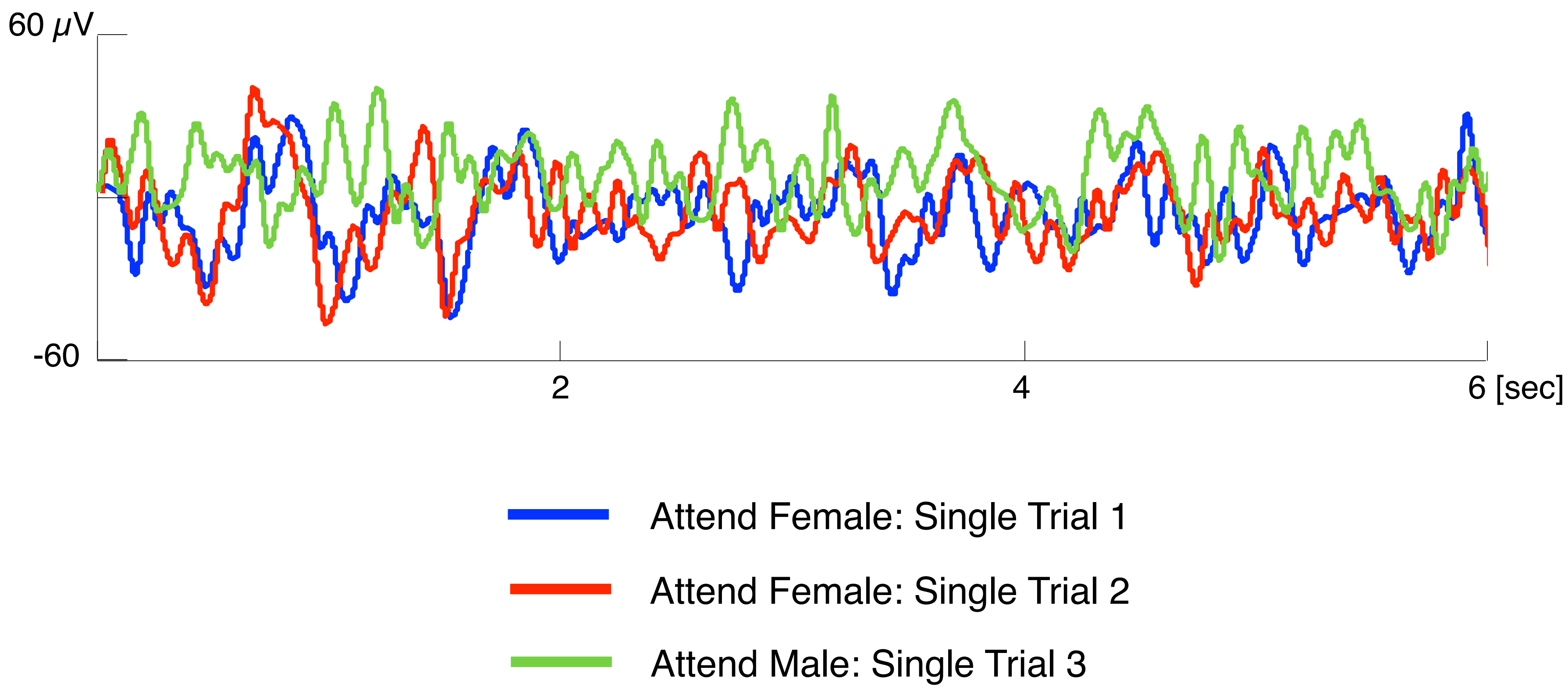
e.g. Luo & Poeppel, Neuron 2007; Ding & Simon, PNAS 2012; J Neuroscience 2013

Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation.



Luo, Liu & Poeppel (2010), PLoS Biology

ECoG Single Trials, an example:

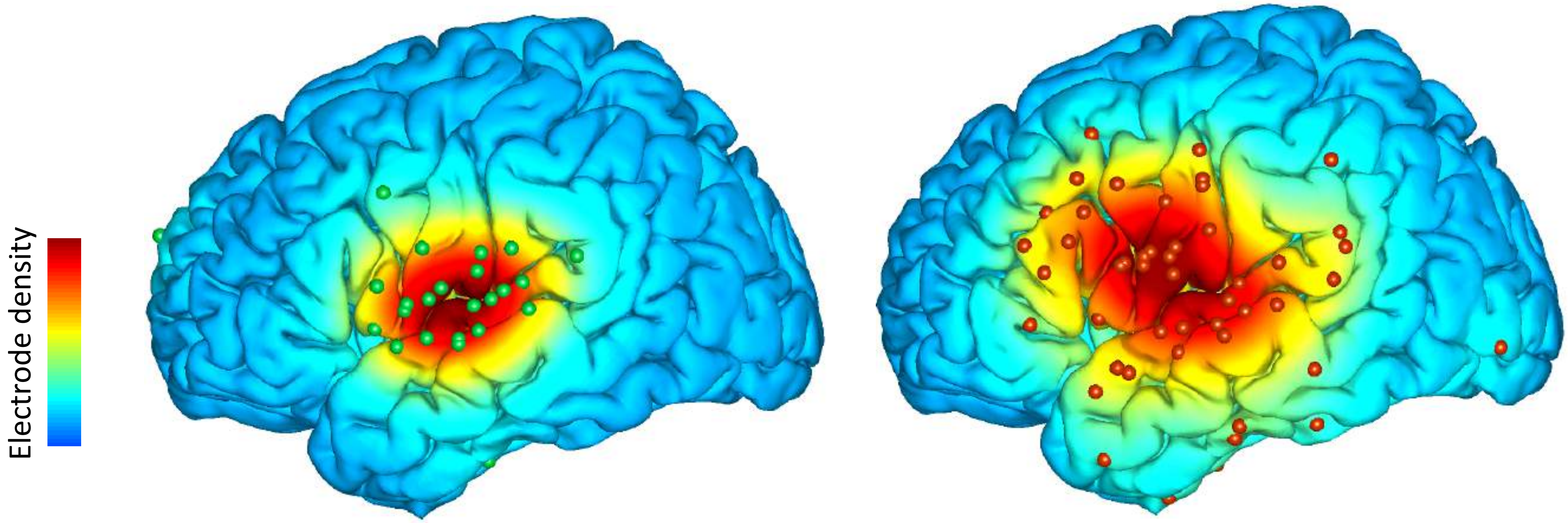


Zion-Golumbic et al. Neuron 2013

Two types of Attentional effects

Tracking of both speakers

Selective tracking of Attended



Highlights

- Both low-frequency phase and high-gamma power preferentially track attended speech
- Near auditory cortex attention modulates response to attended and ignored talkers
- In higher order regions tracking is selective only for the attended talker
- Selectivity for the attended talker increases over time

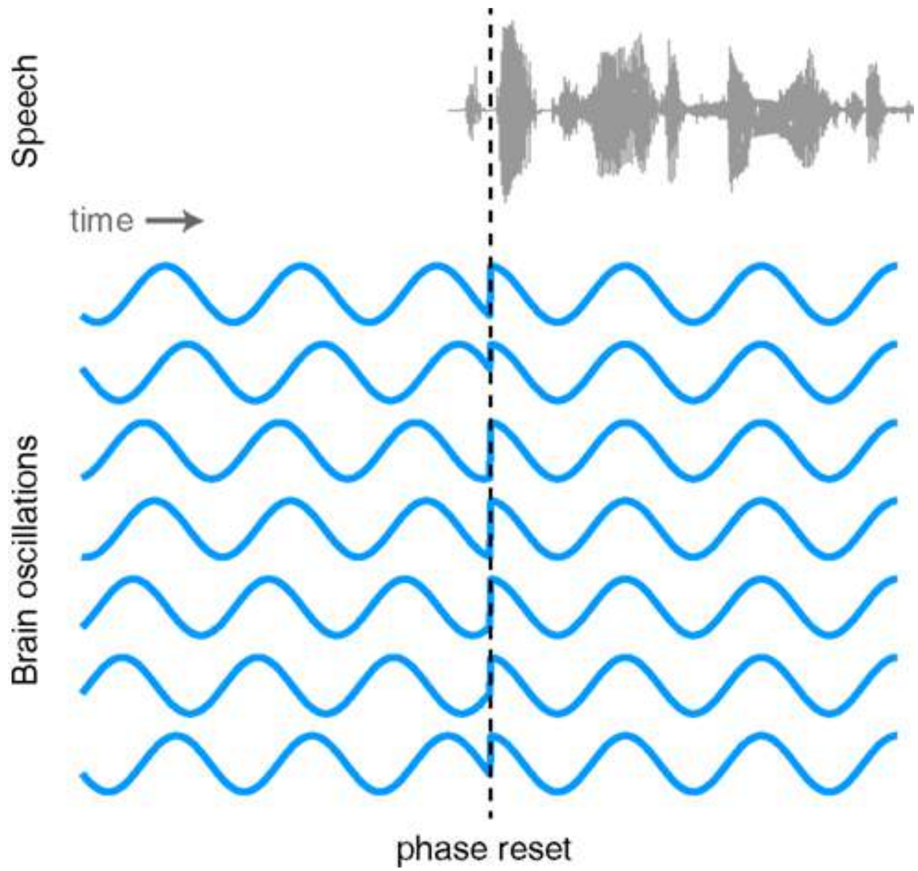
Zion-Golumbic et al. Neuron 2013

Entrainment and cortical rhythms

A veritable orgy of studies and data on cortical oscillations
and their putative role in perception and cognition.
(Buzsaki, Singer, Fries, Schroeder, …)

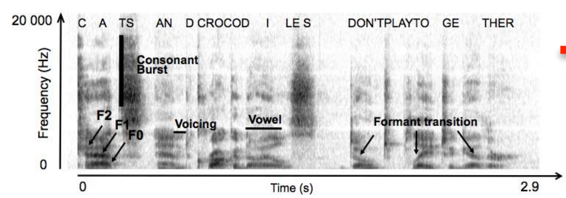
- GENERIC AUDITORY
- Schroeder et al. 2008
- Lakatos et al. 2008
- Luo et al. 2006
- Doelling & Poeppel 2015
- …

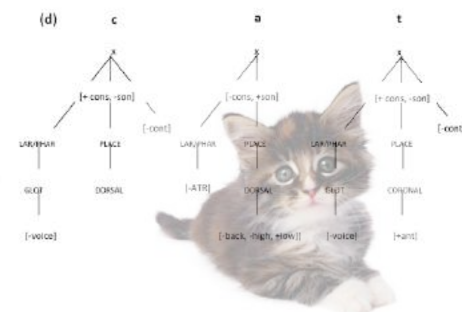
- SPEECH
- Ahissar et al. 2001
- Luo & Poeppel 2007
- Howard & Poeppel 2010
- Luo et al. 2010
- Cogan & Poeppel 2011
- Peelle, Gross, Davis 2012
- Ding & Simon 2012
- Zion-Golumbic et al. 2013
- Koskinen & Seppä 2014
- Doelling et al. 2014
- …



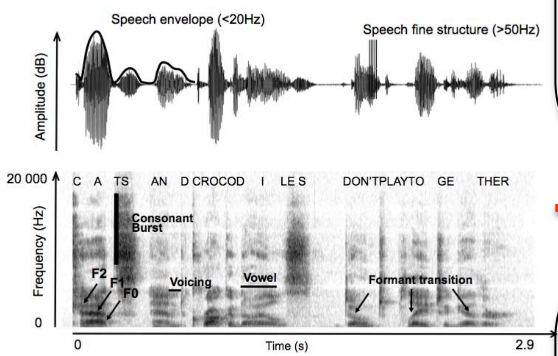
Adapted from Peelle & Davis 2012

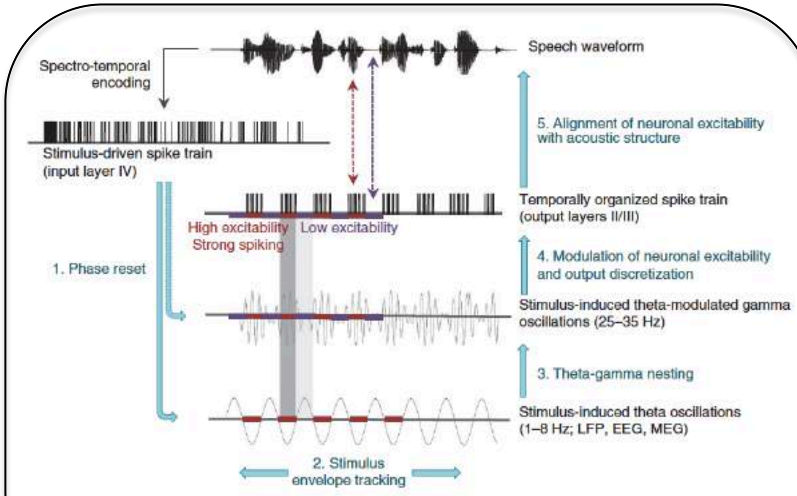
Zooming in on the problem: from vibrations in the ear to abstractions in the head



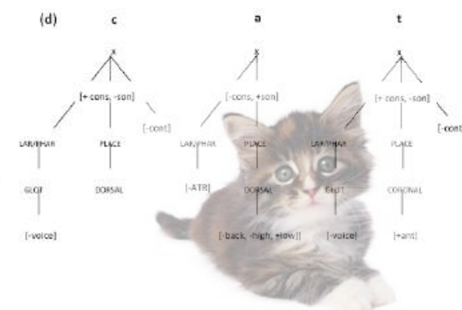


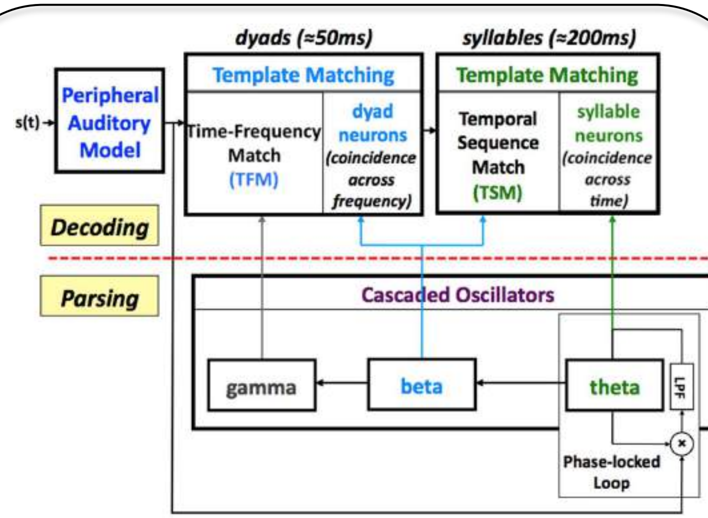
Zooming in on the problem: from vibrations in the ear to abstractions in the head





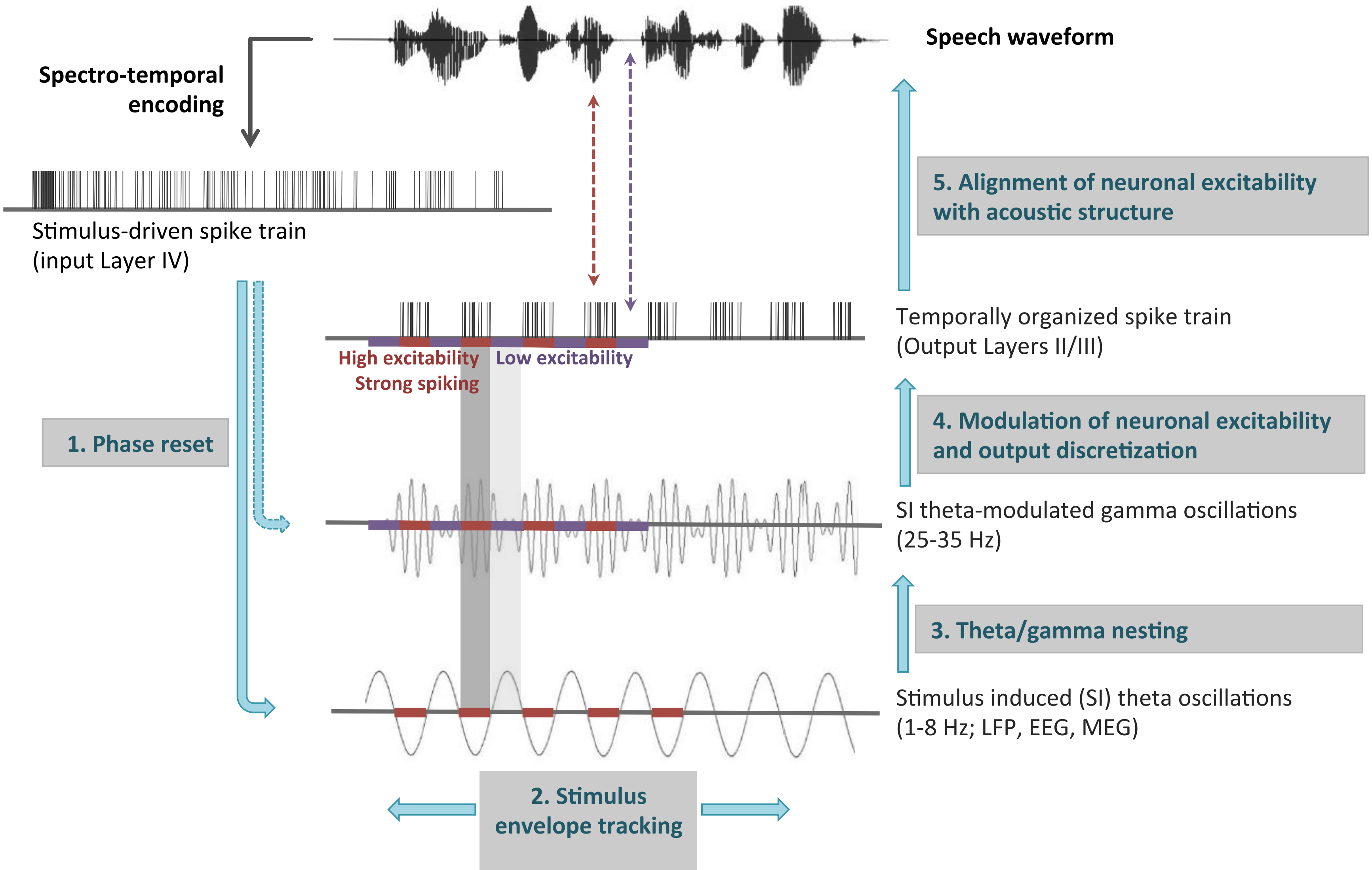
Giraud & Poeppel, 2012, Nat Neurosci



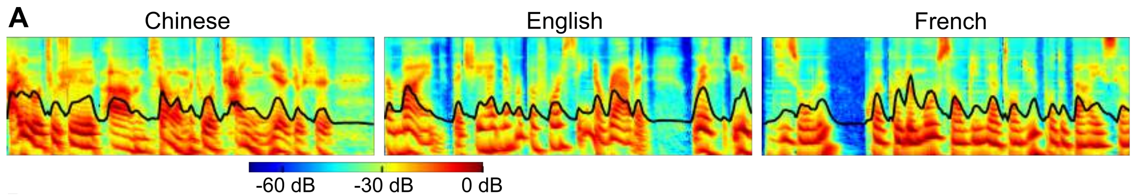


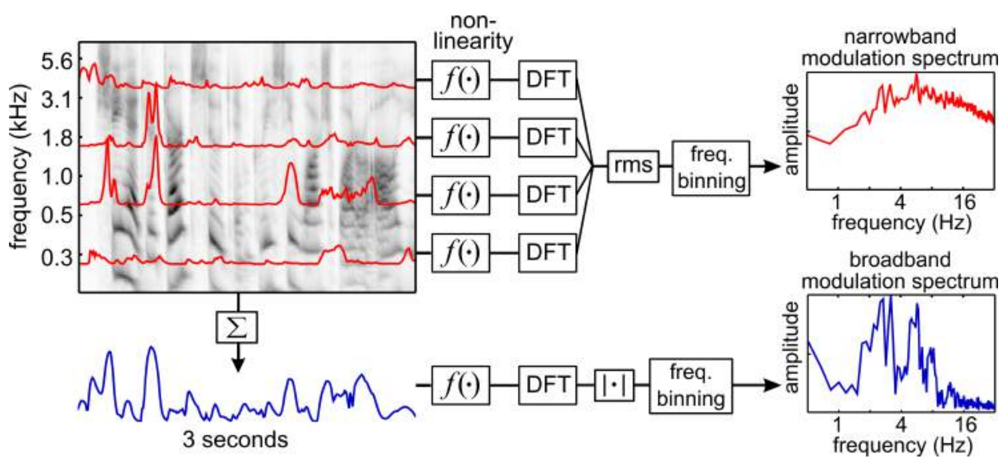
Ghitza, 2011, Frontiers

Cortical oscillations and speech processing: emerging computational principles and operations

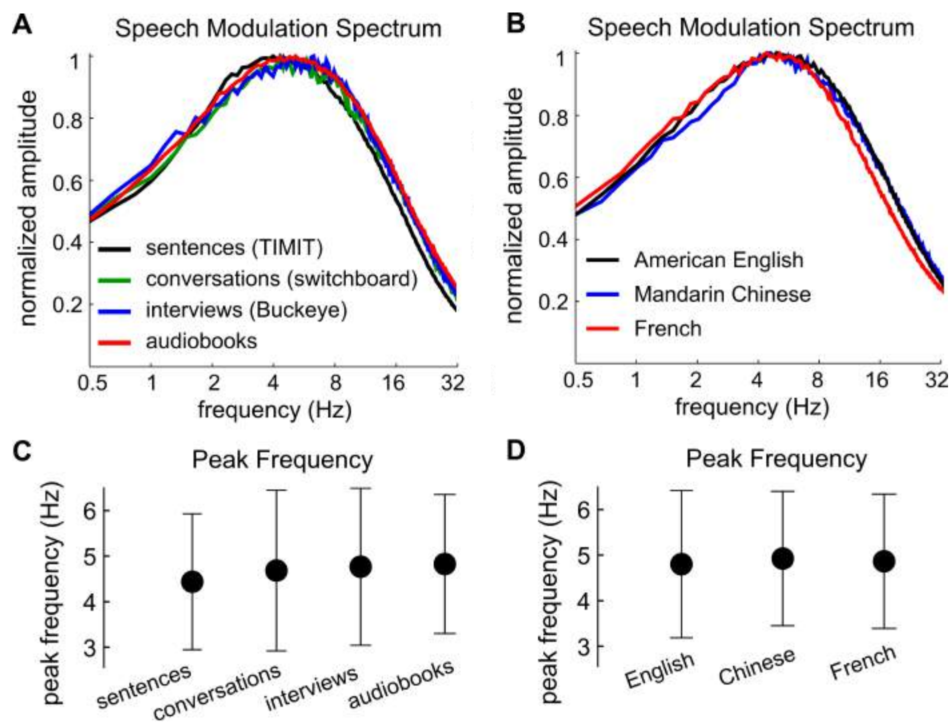


Giraud & Poeppel, 2012, Nature Neurosci

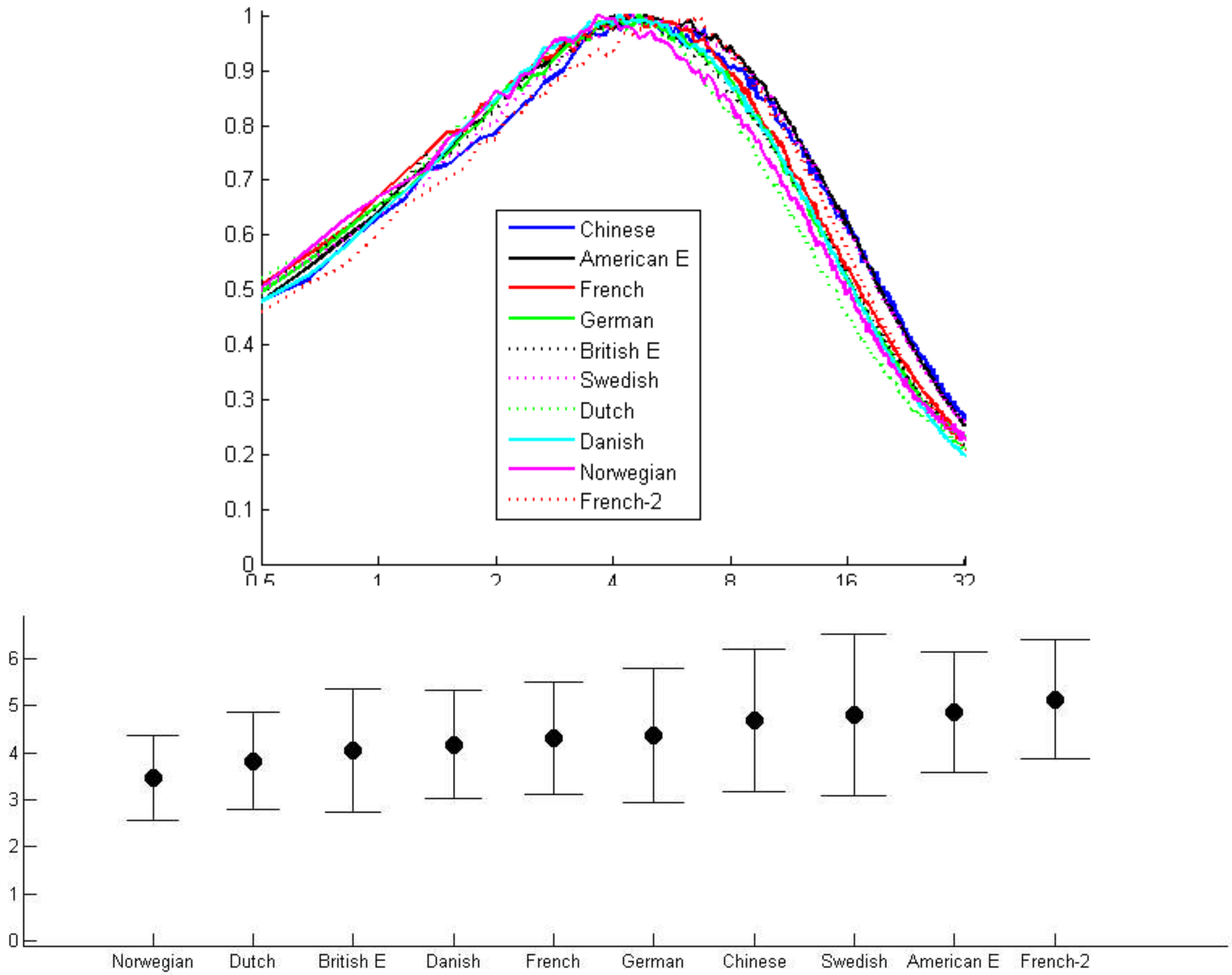




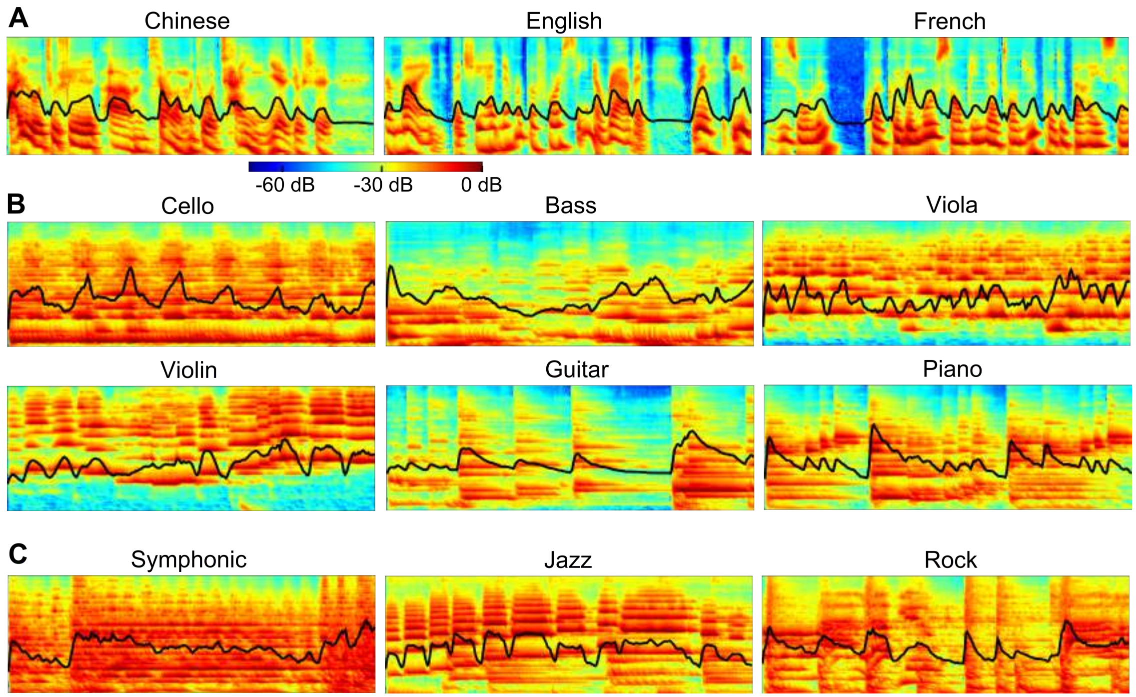
Ding N, Patel A, Chen L, Butler H, Luo C, Poeppel D (2017)

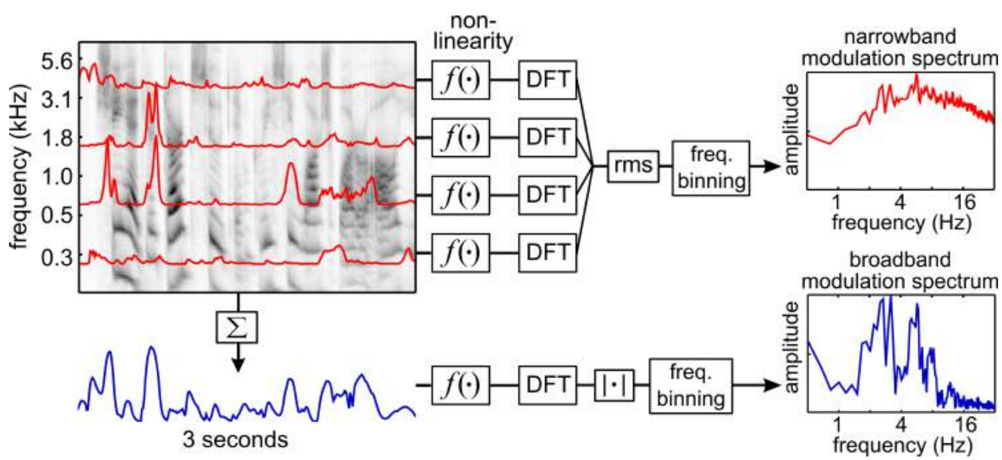


Ding N, Patel A, Chen L, Butler H, Luo C, Poeppel D (2017)

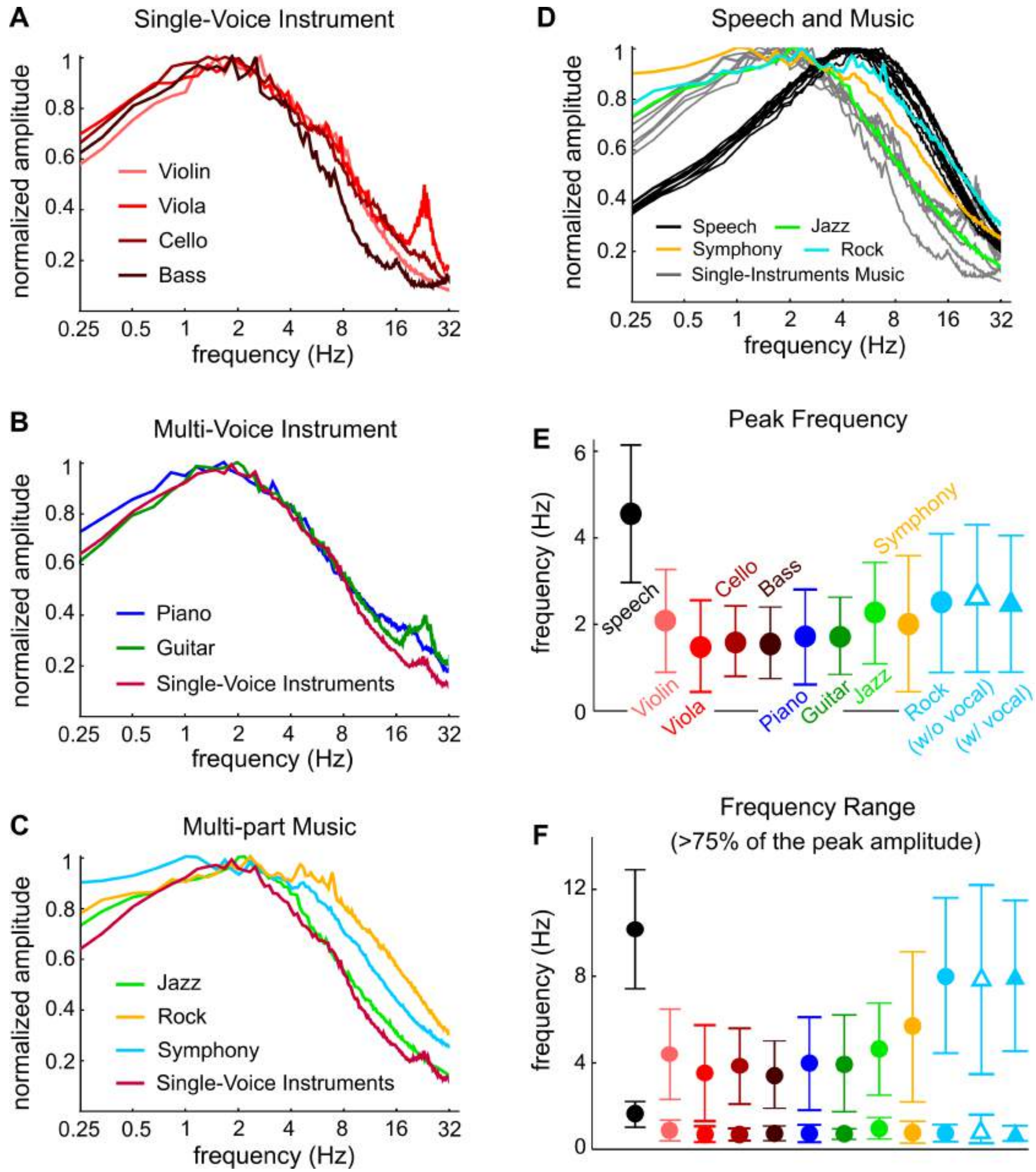


Ding N, Patel A, Chen L, Butler H, Luo C, Poeppel D (2017)





Ding N, Patel A, Chen L, Butler H, Luo C, Poeppel D (2017)



Ding N, Patel A, Chen L, Butler H, Luo C, Poeppel D (2017)

The syllable-sized acoustic chunk as perceptual primitive

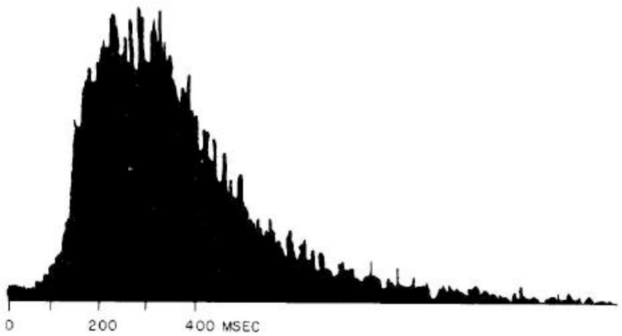


Figure 1 Histogram of the intervals between some 10 000 successive jaw openings in running speech (reading).

Ohala 1972

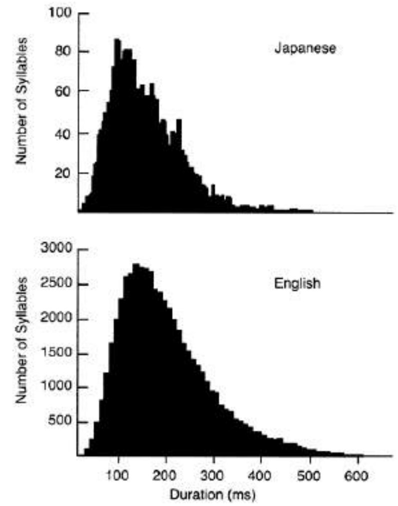


Fig. 1 Statistical distribution of syllable duration for spontaneous material in Japanese and American English. Adapted from [1].

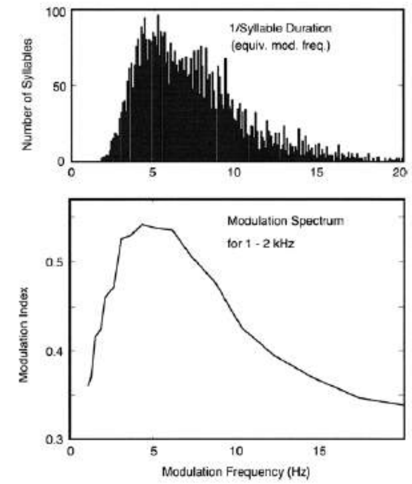
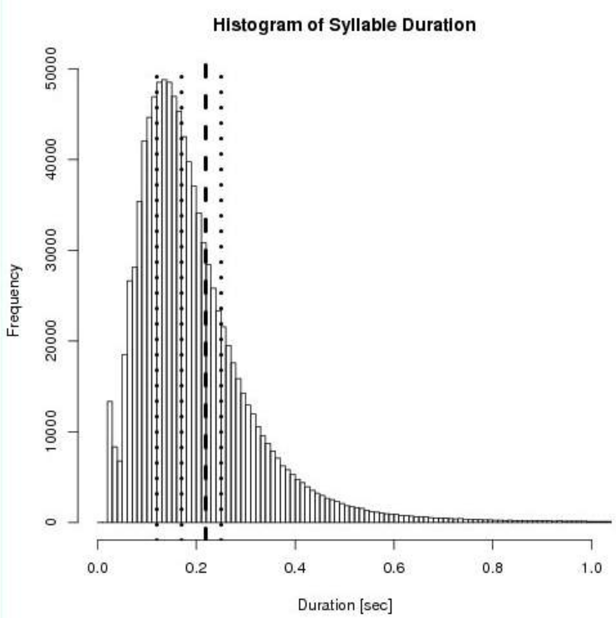


Fig. 2 The relation between the distribution of syllable duration (transformed into modulation frequency) and the modulation spectrum of the same Japanese material as shown in Fig. 1, computed for the octave region between 1 and 2 kHz. Adapted from [1].

Greenberg & Arai 2004



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

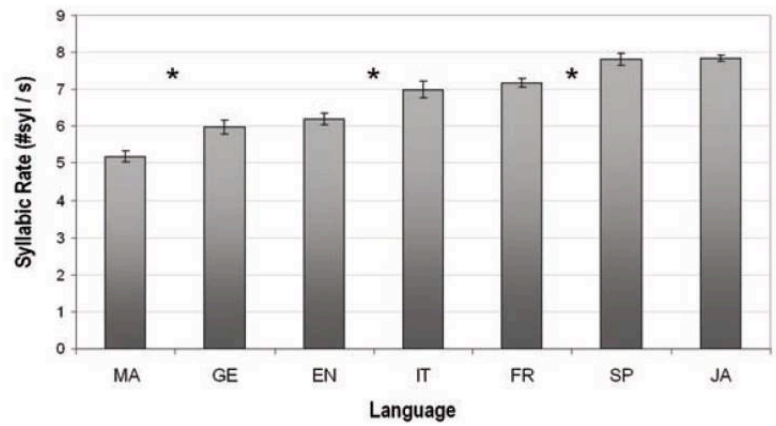


FIGURE 1. Speech rate measured in terms of the number of syllables per second (mean values and 95% confidence intervals). Stars indicate significant differences between the homogeneous subsets revealed by post-hoc analysis.

Pellegrino et al. 2011

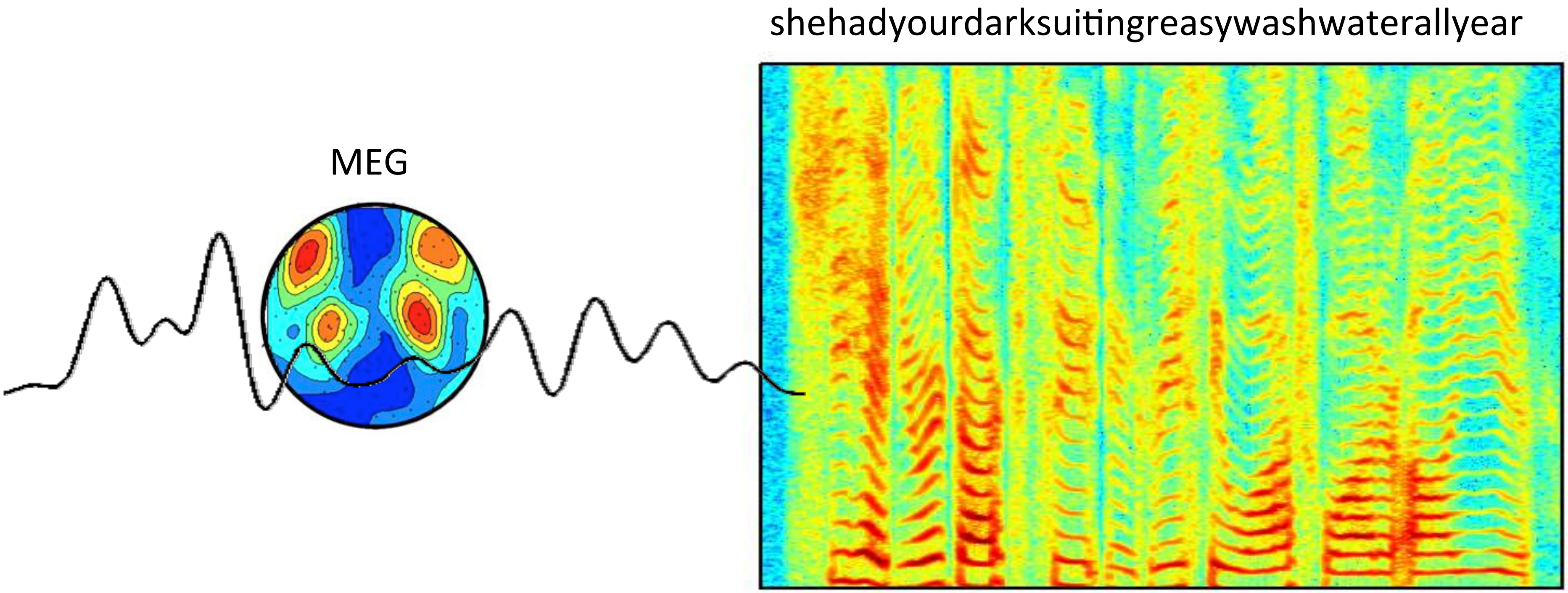
An interesting alignment between:

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

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

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

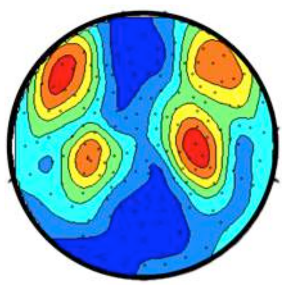
Segmenting events, e.g. syllables

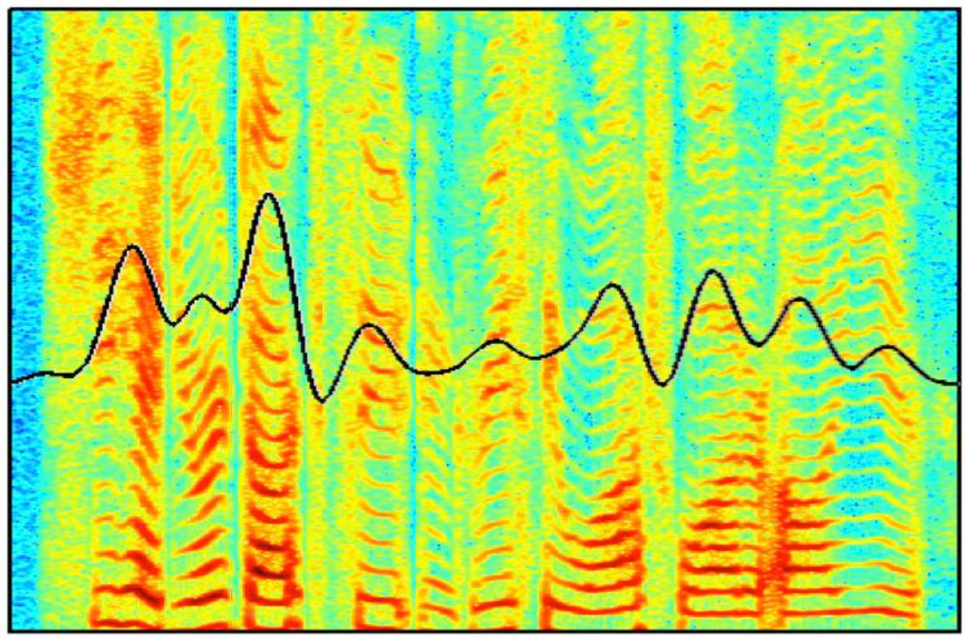


Courtesy of Keith Doelling, NYU

Segmenting events, e.g. syllables

she had your dark





Courtesy of Keith Doelling, NYU

But does entrainment matter?

Segmentation, intelligibility, comprehension

Fdiff

Fcc

PL

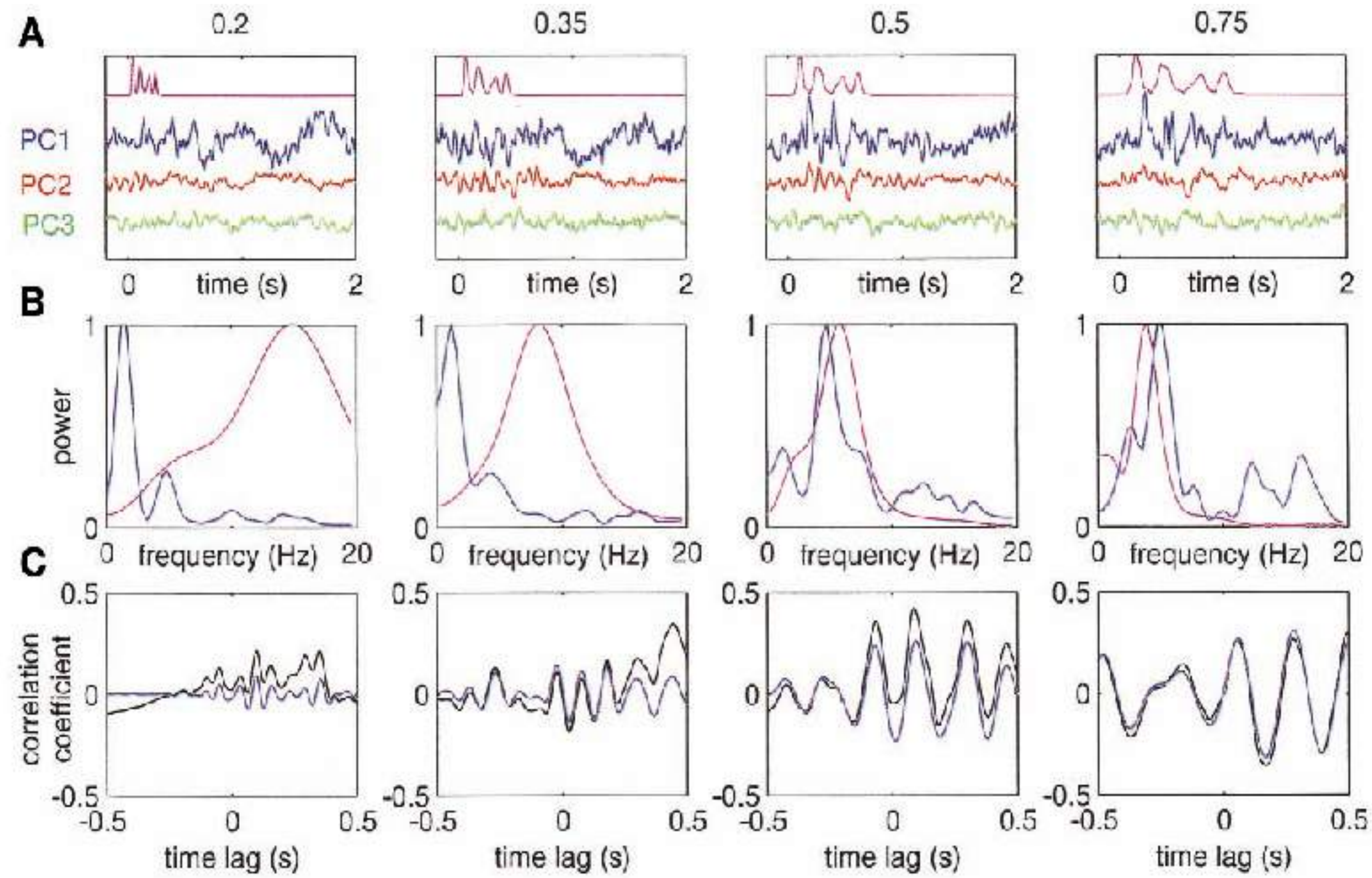
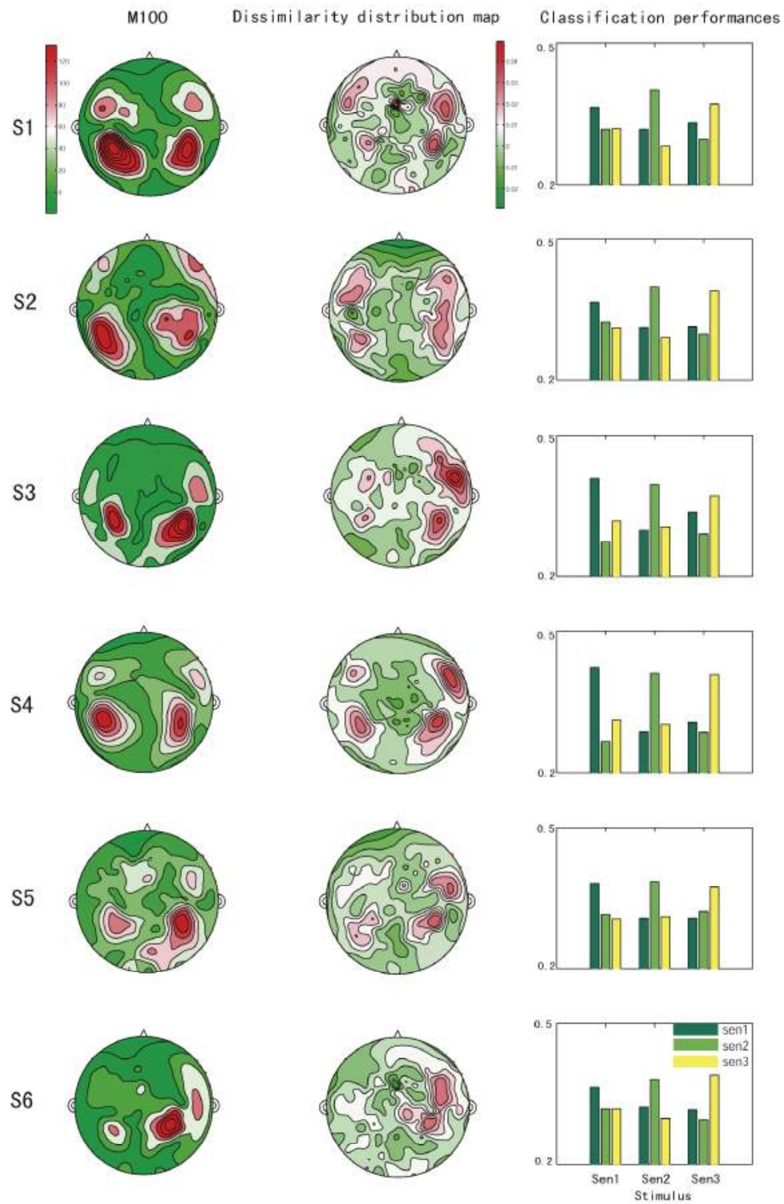


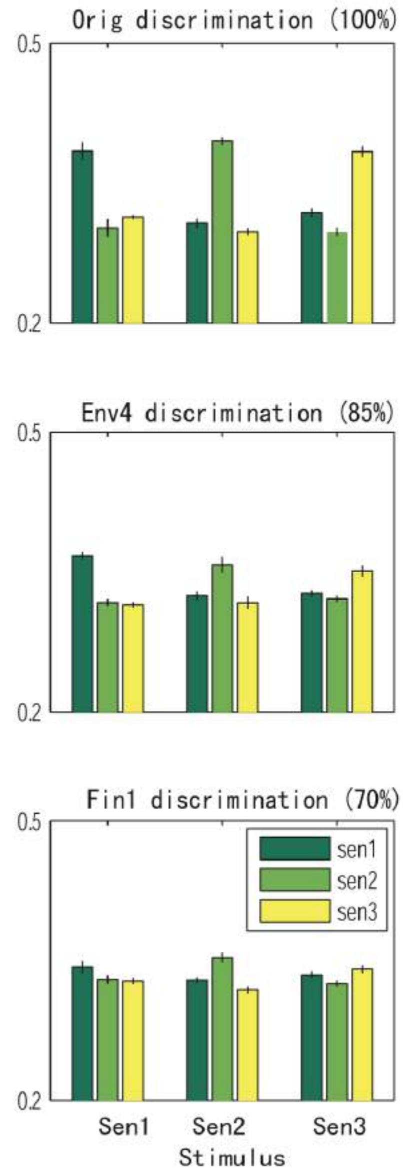
Fig. 2. An example of MEG signals recorded during the task, and the measures derived from them (S MS). (*A*) Averaged temporal envelopes (magenta) and the first three PCs (PC1–3, blue, red, and green, respectively, scaled in proportion to their eigen values) of the averaged responses. (*B*) Power spectra of the stimulus envelope (magenta) and PC1 (blue). (*C*) Time domain cross correlation between the envelope and PC1; black, raw correlation; blue, after band-pass filtering at ±1 octave around the stimulus modal frequency.

Ahissar et al. 2001, PNAS

- Ahissar et al. 2001: compression compromises intelligibility because cortex cannot entrain to the envelope at fast rates.

Theta phase has the sensitivity to discriminate based on single trials



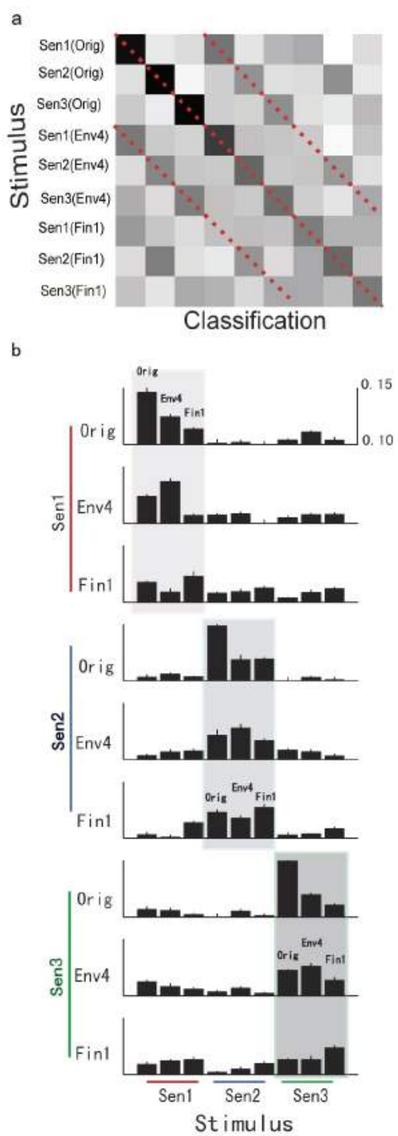


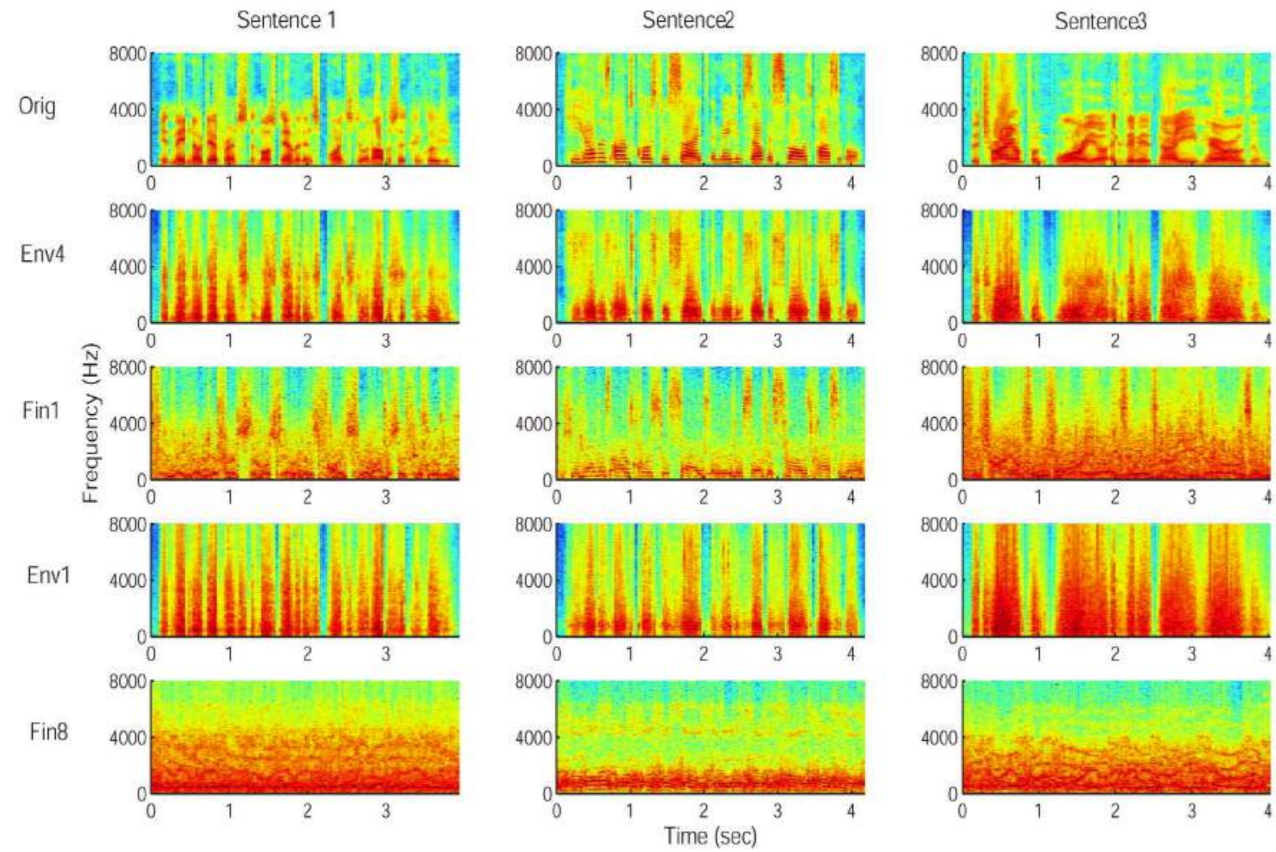
Materials:

Smith, Delgutte, and Oxenham, *Nature*, 2002

Luo & Poeppel, Neuron, 2007

Theta phase tracking displays the specificity to discriminate sentences





Classification analysis

Luo & Poeppel, Neuron, 2007

- Ahissar et al. 2001: compression compromises intelligibility because cortex cannot entrain to the envelope at fast rates.

- Luo & Poeppel 2007: acoustic manipulation that compromises intelligibility also reduces phase tracking.

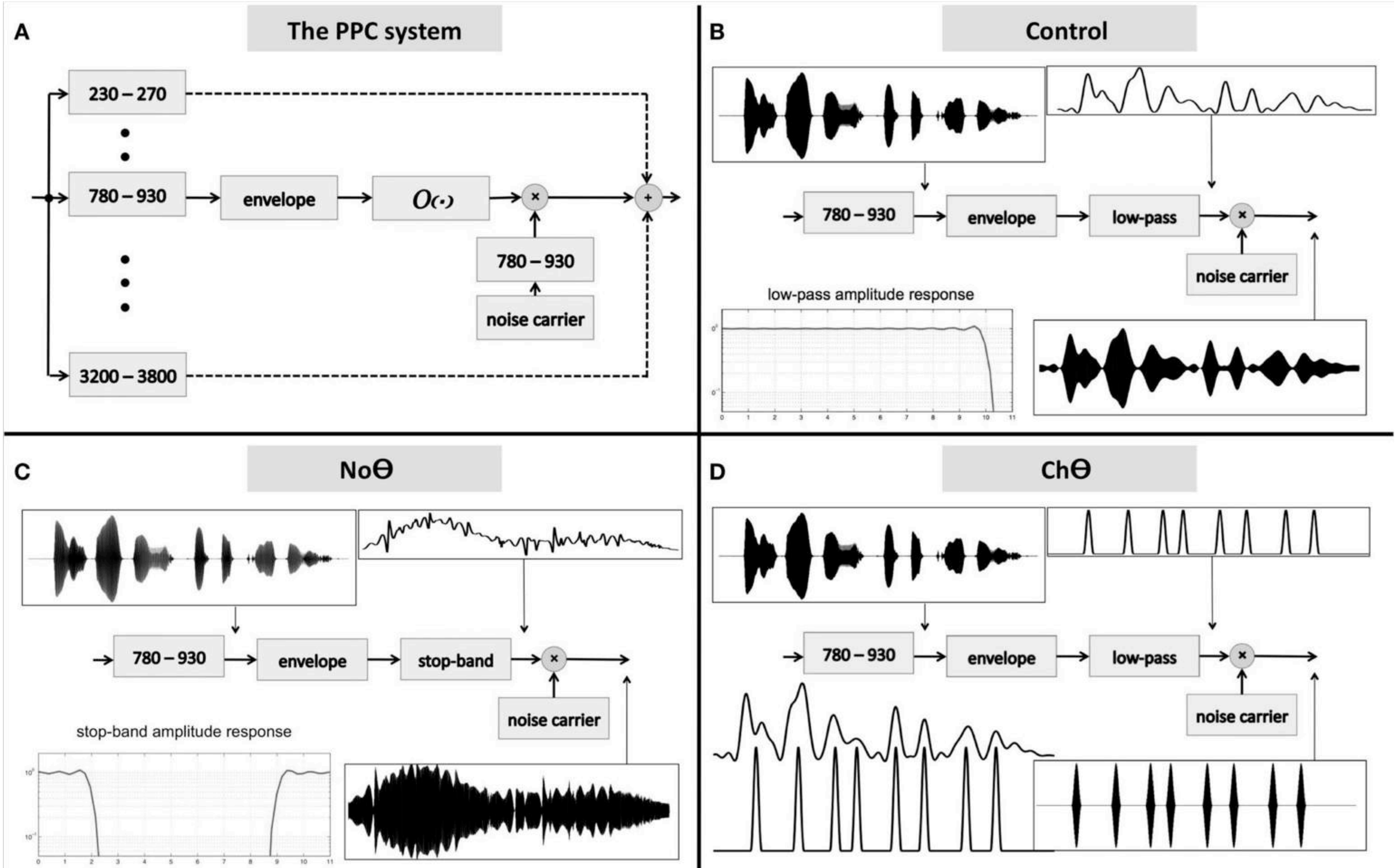
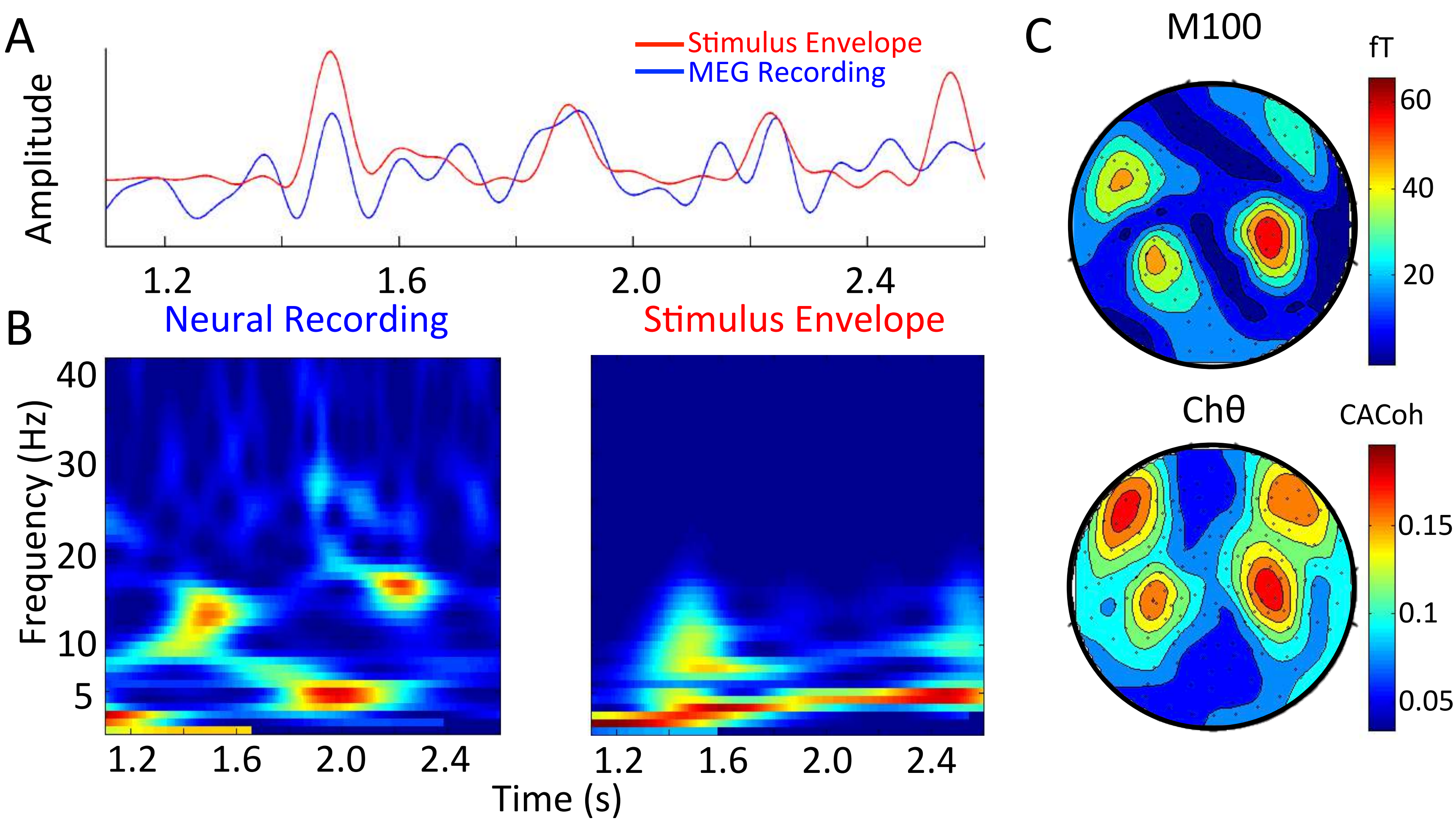


Figure 1. Schematic of stimulus creation (figure from Ghitza, 2012). This figure shows the steps that each initial waveform underwent to create the stimuli used. *A*. Each stimulus was filtered into 16 logarithmically-spaced critical bands from 230-3800 Hz, the Hilbert envelope is derived and an operator O for each condition (identified in B, C and D) is executed. Finally, critical bands are linearly summed. *B; Control***.** Operator O is a low-pass filter of the envelope at 10 Hz. *C; Noθ*. The operator is a stop-band filter from 2-9 Hz. *D; Chθ*. The operator is a peak picking code (PPC) in which each peak in the envelope is replaced by a peak of uniform height and shape.

Doelling et al., NeuroImage, 2014



Doelling et al., NeuroImage, 2014

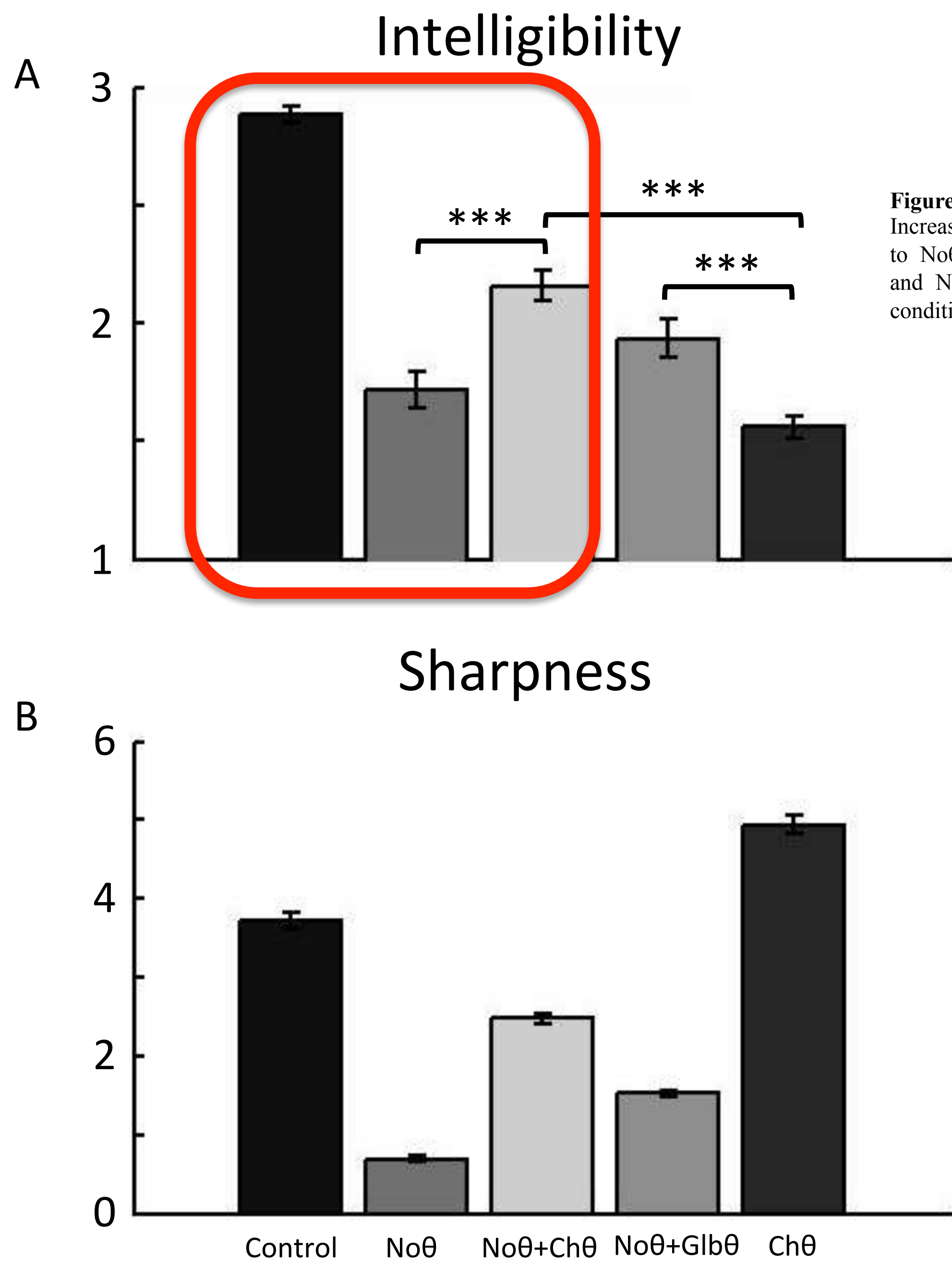


Figure 2. Intelligibility and Sharpness. *Top panel.* Increased intelligibility ratings from both Noθ and Chθ to Noθ+Chθ. No significant difference between Noθ and Noθ+Glbθ. *Bottom panel.* Sharpness metric. All conditions are significantly different.

Doelling et al., NeuroImage, 2014

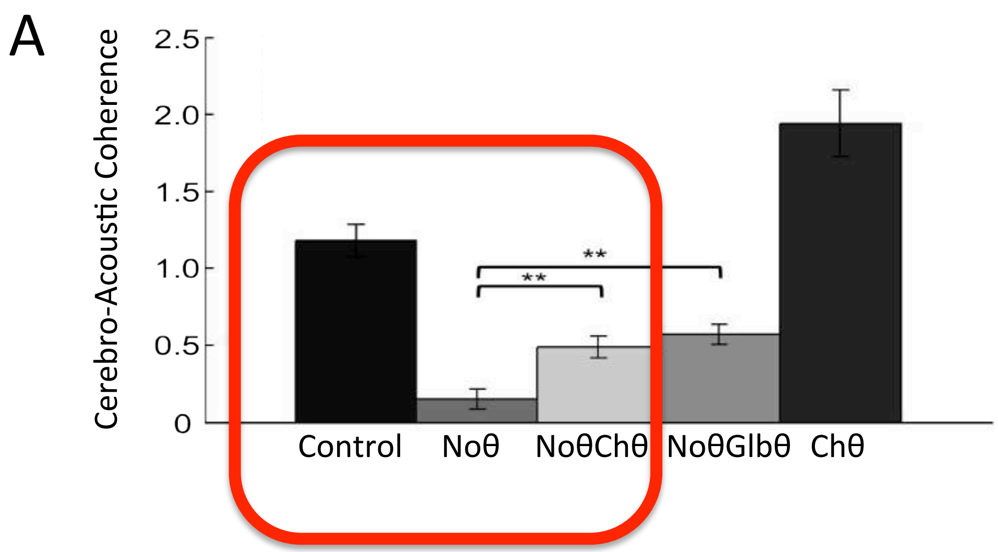


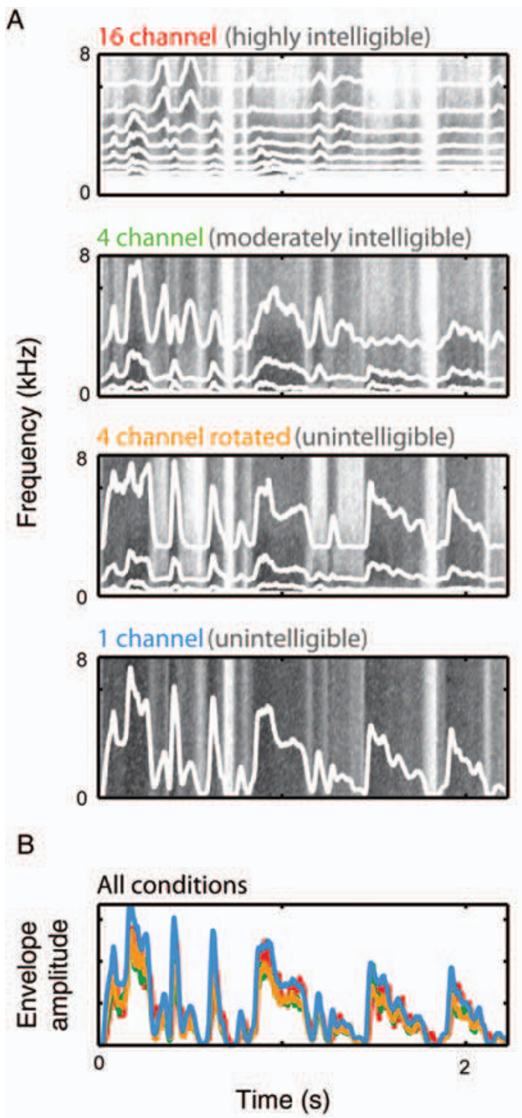
Figure 4. Increase in CACoh due to sharpness increases intelligibility. *A*. CACoh averaged across 4 auditory regions. Significant increase from Noθ to both Noθ+Chθ and Noθ+Glbθ. *B.* Difference in CACoh between Noθ and Noθ+Chθ. Anterior regions show significance. *C.* Change in CACoh in Anterior Right channels between Noθ (gray, dotted line) and most intelligible conditions (Control, left, and Noθ+Chθ, right) correlates with change in Intelligibility. No correlation with Chθ (dark, solid line).

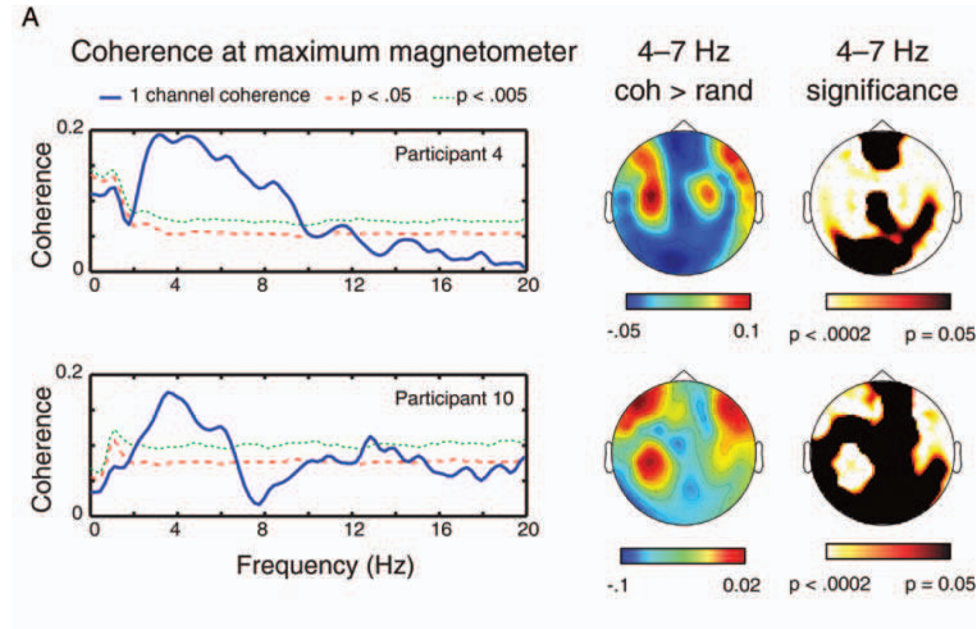
Doelling et al., NeuroImage, 2014

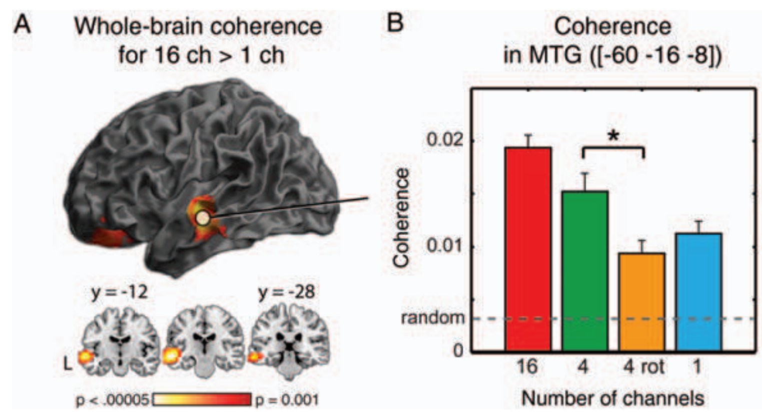
- Ahissar et al. 2001: compression compromises intelligibility because cortex cannot entrain to the envelope at fast rates.

- Luo & Poeppel 2007: acoustic manipulation that compromises intelligibility also reduces phase tracking.

- Ghitza (2012) (psychophysical version) Doelling et al. 2014 (MEG) Elimination of cues – no tracking – no intelligibility. Reinstatement of simple cues for entrainment upregulates intelligibility.







Peelle et al., Cerebral Cortex, 2012

- Ahissar et al. 2001: compression compromises intelligibility because cortex cannot entrain to the envelope at fast rates.

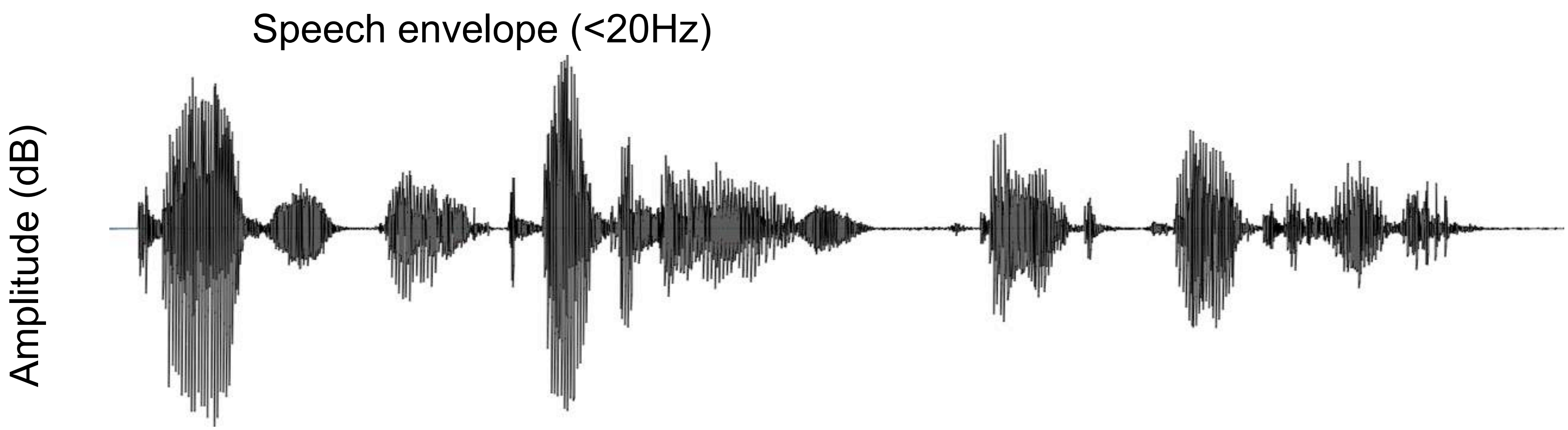
- Luo & Poeppel 2007: acoustic manipulation that compromises intelligibility also reduces phase tracking.

- Ghitza (2012) (psychophysical version) Doelling et al. 2014 (MEG) Elimination of cues – no tracking – no intelligibility. Reinstatement of simple cues for entrainment upregulates intelligibility.

- Peelle, Gross, Davis 2012: Manipulation that increases/decreases intelligibility/comprehension patterns with coherence between acoustics and entrainment.

Entrainment likely enables segmentation and is necessary – but not sufficient – for speech comprehension. Entrainment yields acoustic chunks of approximately syllable duration. These form the basis for decoding.

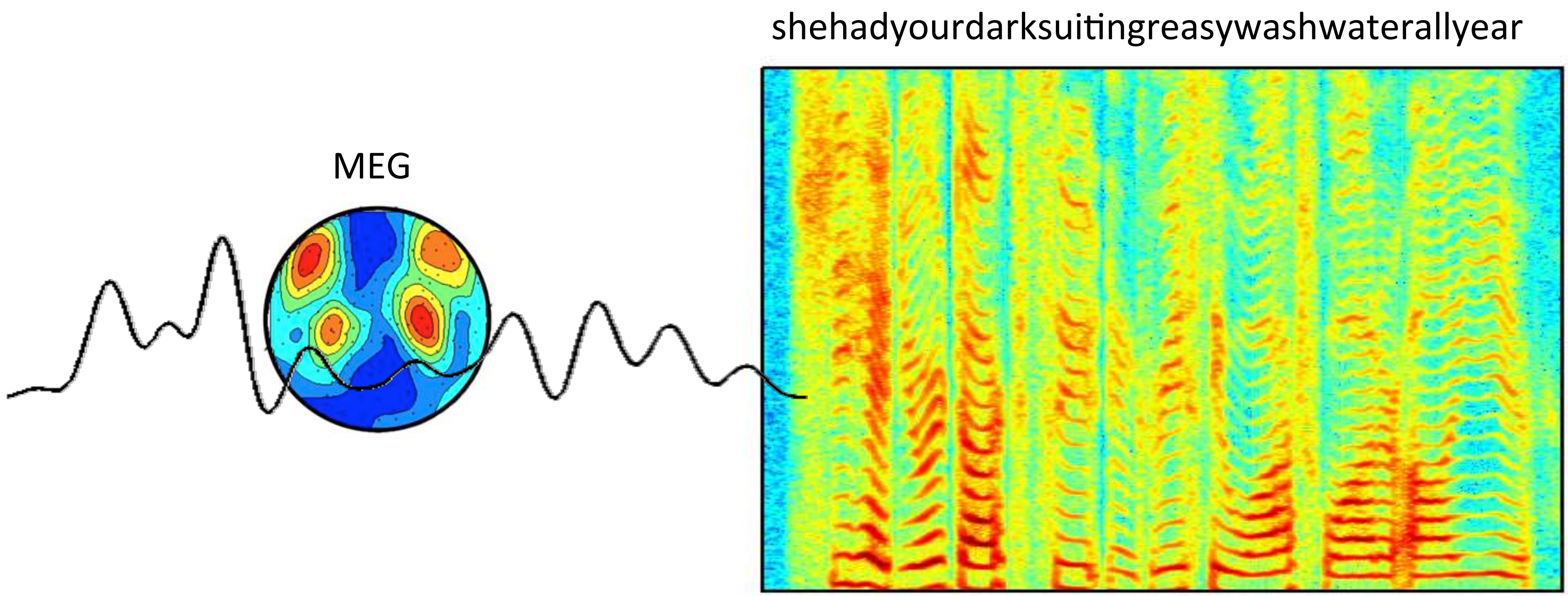
One brief example of an attempt at a linking hypothesis, in the Marr spirit: the segmentation problem and neural oscillations



sincetherearenowordboundarysignsinspokenlanguagethedifficultywefeelinreading
andunderstandingtheaboveparagraphprovidesasimpleillustrationofoneofthemaind
ifficultieswehavetoovercomeinordertounderstandspeechratherthananeatlyseparat
edsequenceofletterstringscorrespondingtothephonologicalformofwordsthespeech
signalisacontinuousstreamofsoundsthatrepresentthephonologicalformsofwordsin
additionthesoundsofneighboringwordsoftenoverlapwhichmakestheproblemofident
ifyingwordboundariesevenharder

Cortical oscillations (neurobiological implementation) as the mechanisms to address the segmentation problem (computational level) by phase resetting to edges (algorithm).

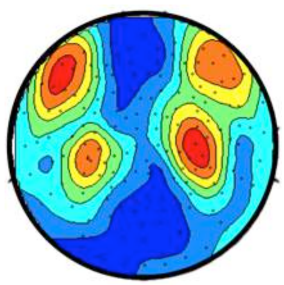
Segmenting events, e.g. syllables

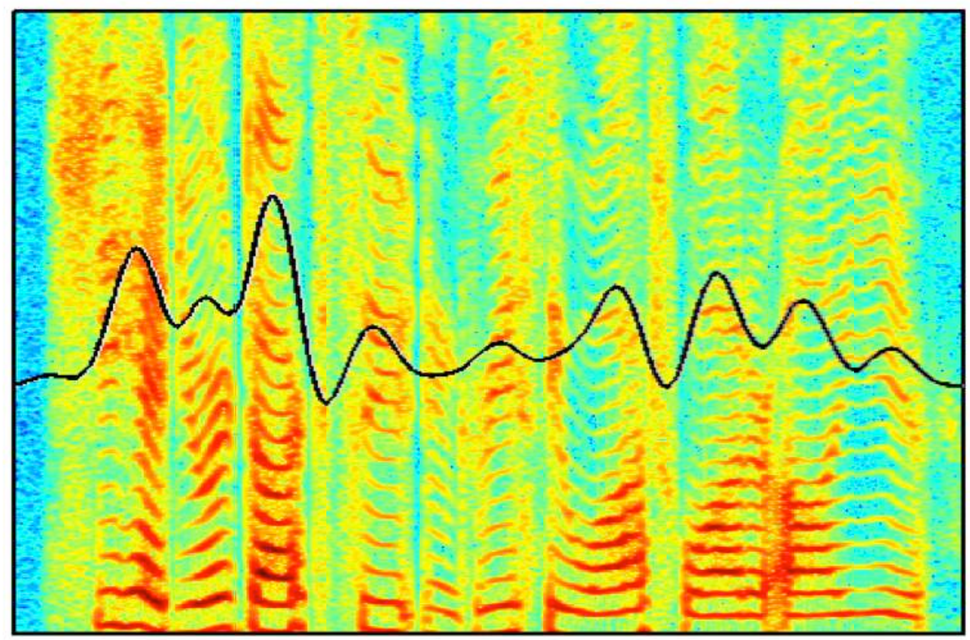


Courtesy of Keith Doelling, NYU

Segmenting events, e.g. syllables

she had your dark





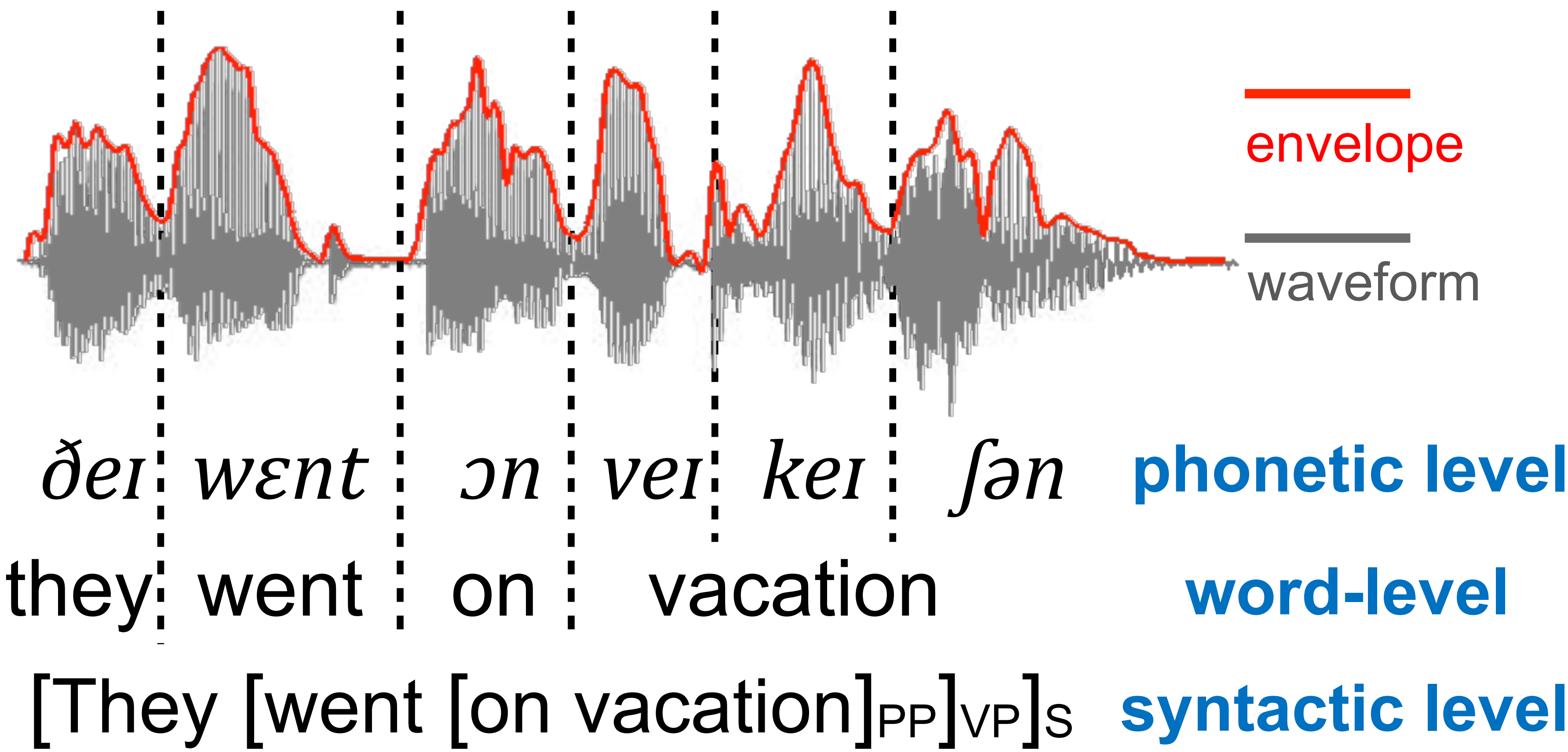
Courtesy of Keith Doelling, NYU

Cortical tracking of hierarchical linguistic structures in connected speech

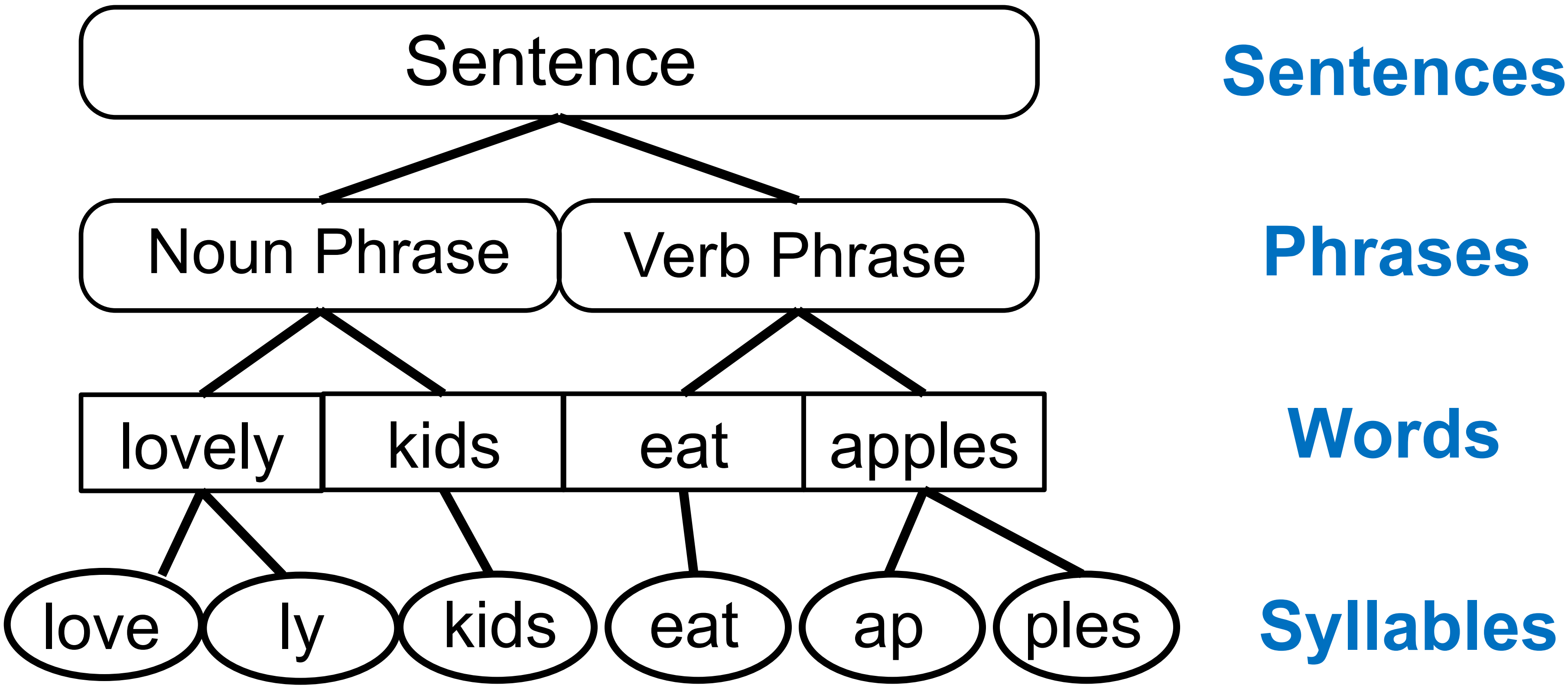
Nai Ding
NYU
Zhejiang Univ.

Lucia Melloni
Max Planck
NYU

Boundaries between syllables are usually defined by the speech envelope, but not the boundaries between words and phrases.

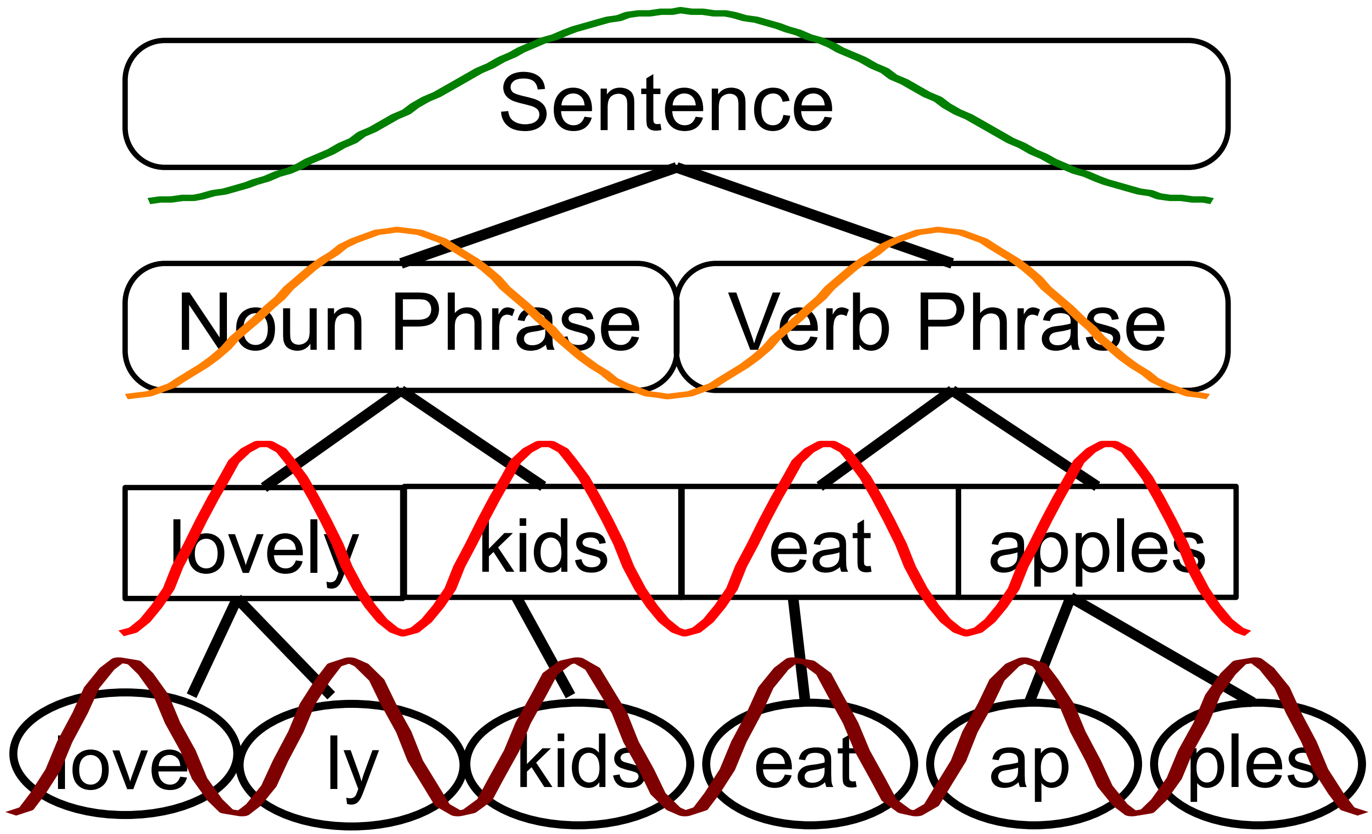


Parsing Linguistic Structures Embedded in Continuous Speech

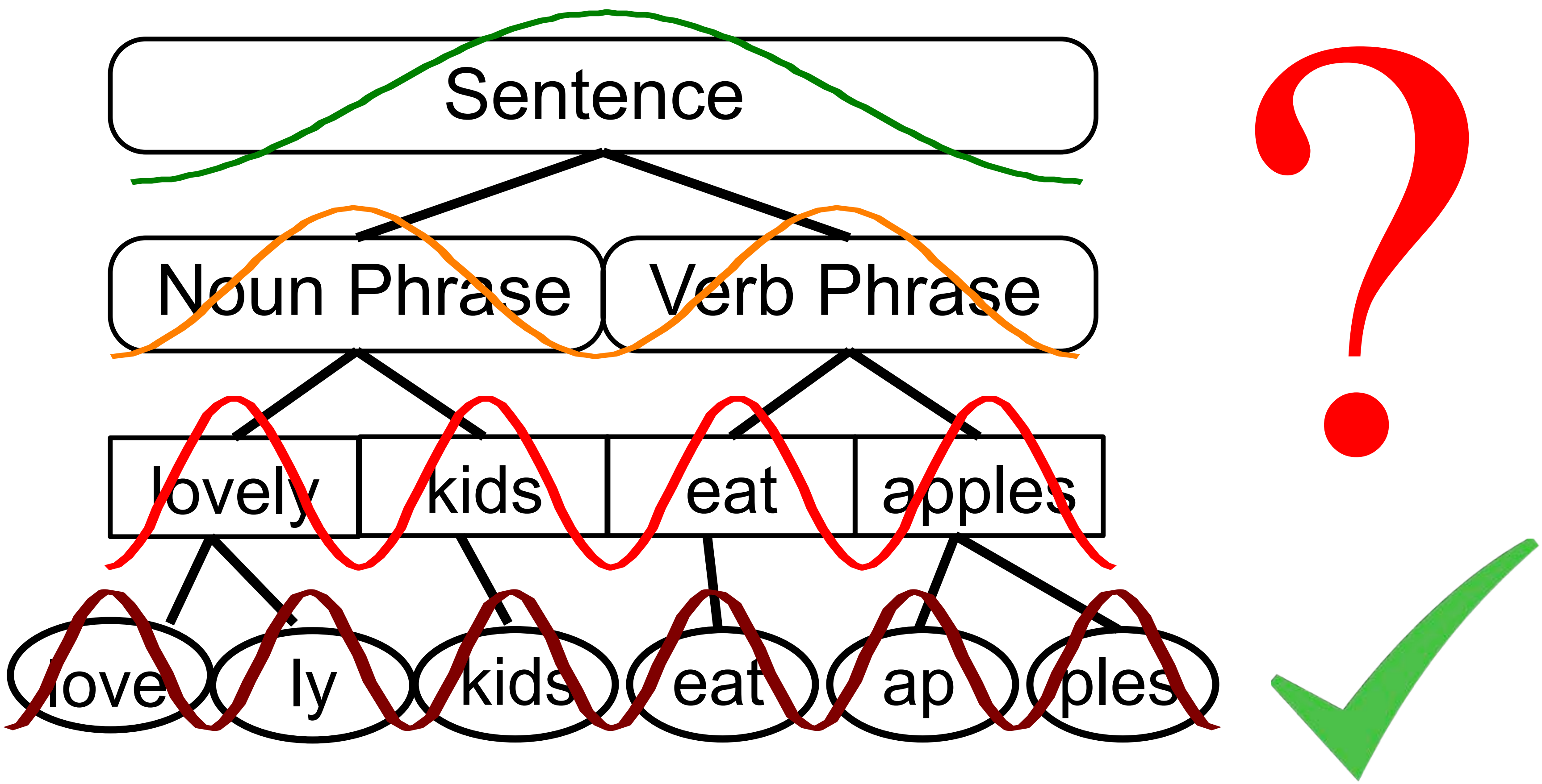


The neural code for each linguistic unit must change at the rate of that linguistic unit.

Hierarchical Entrainment to the Hierarchical Linguistic Structure?

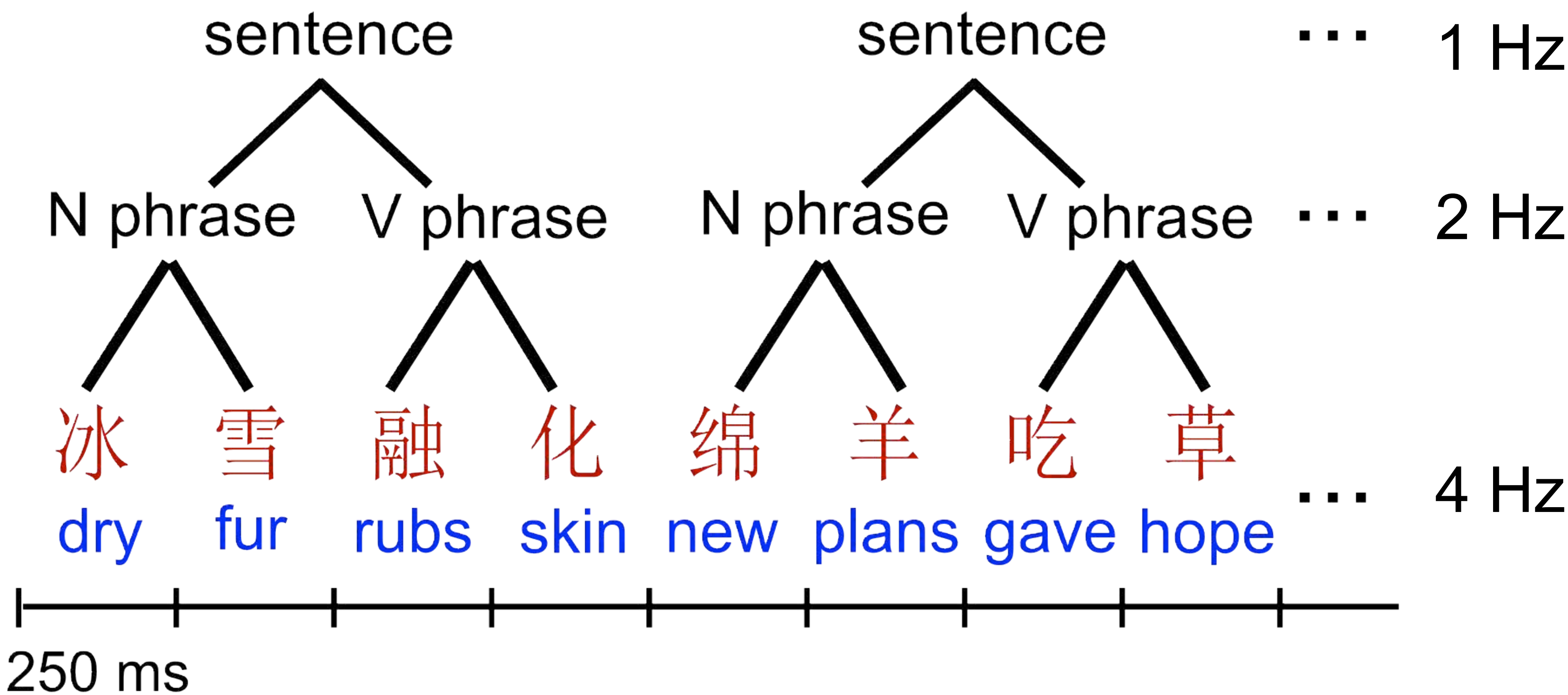


Hierarchical Entrainment to the Hierarchical Linguistic Structure?



e.g., Luo & Poeppel, 2007
Ding & Simon, 2012

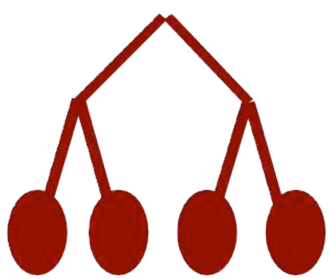
A Sequence with Hierarchical Linguistic Structures

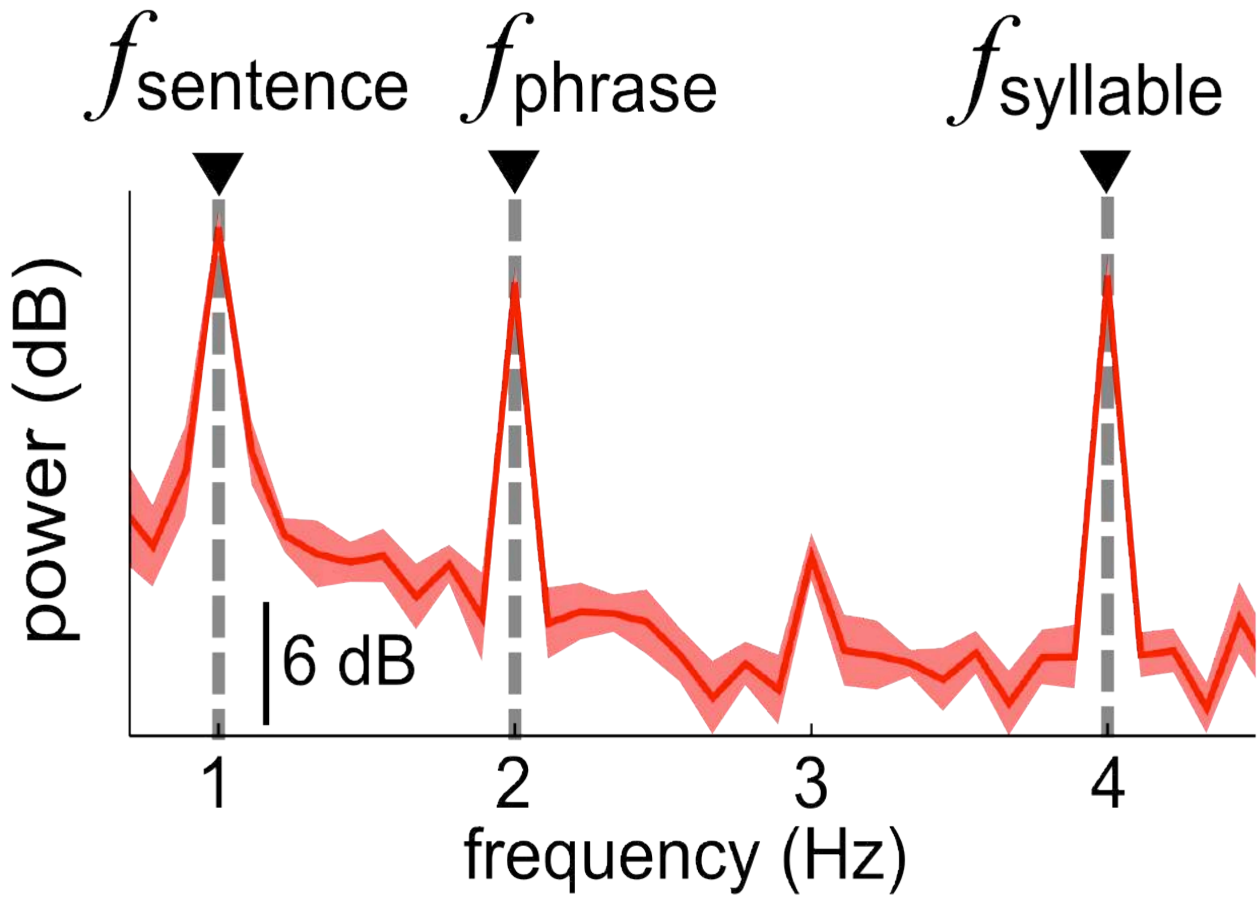


- 16 native listeners of Mandarin Chinese
- Outlier detection: occasionally, the noun phrases of two sentences will be switched, creating two nonsense sentences.
- Data processed by a spatial filter optimized to extract phase-locked activity.

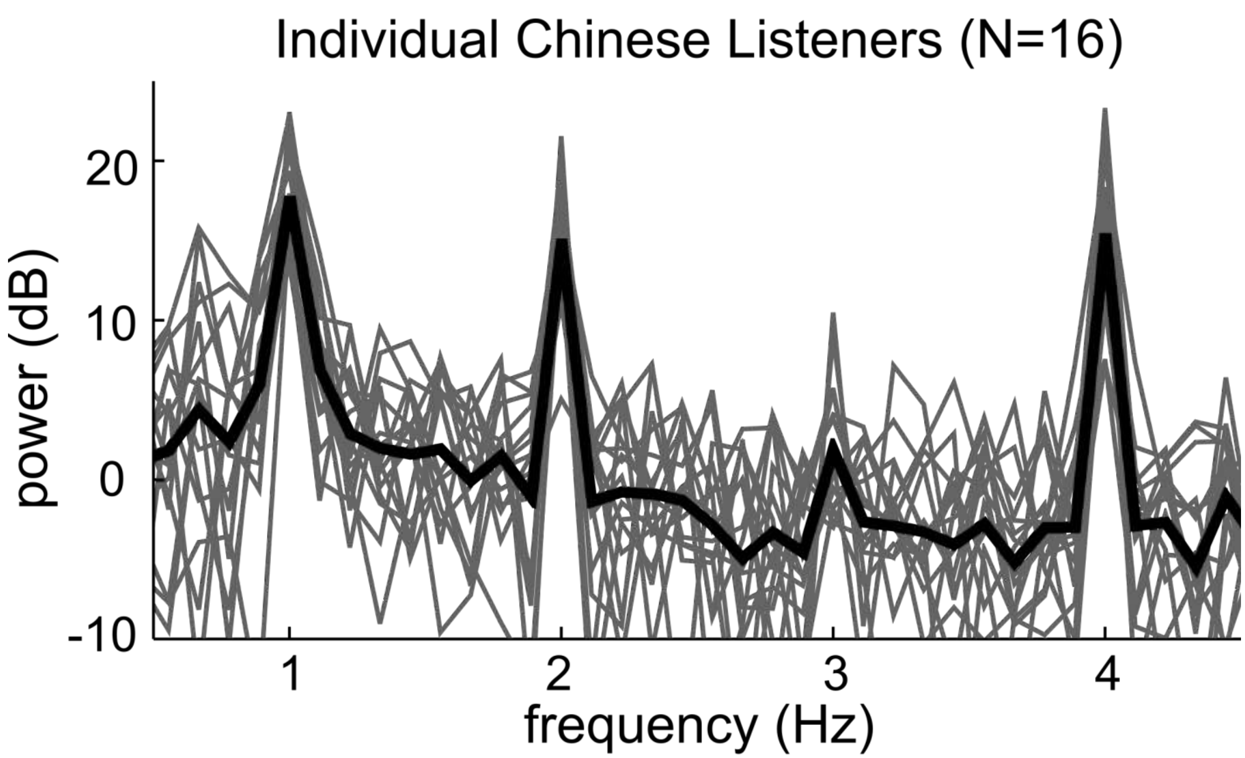
(Wang et al., J Neurophys 2012; Ding & Simon, PNAS 2012; de Cheveigné & Simon, 2008)

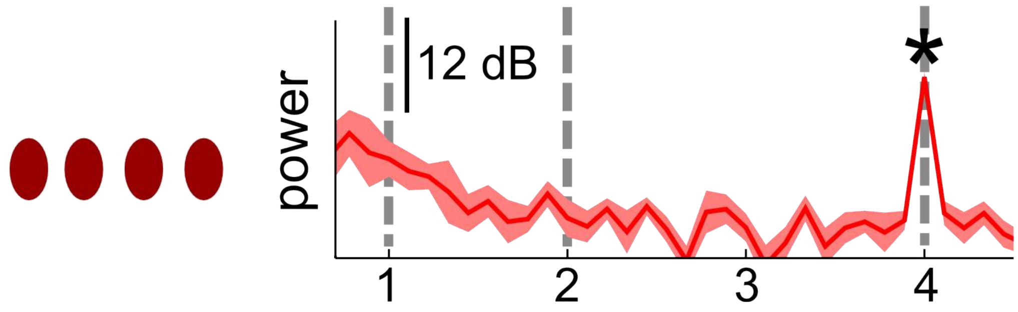
Cortical Activity Tracks Hierarchical Linguistic Rhythms



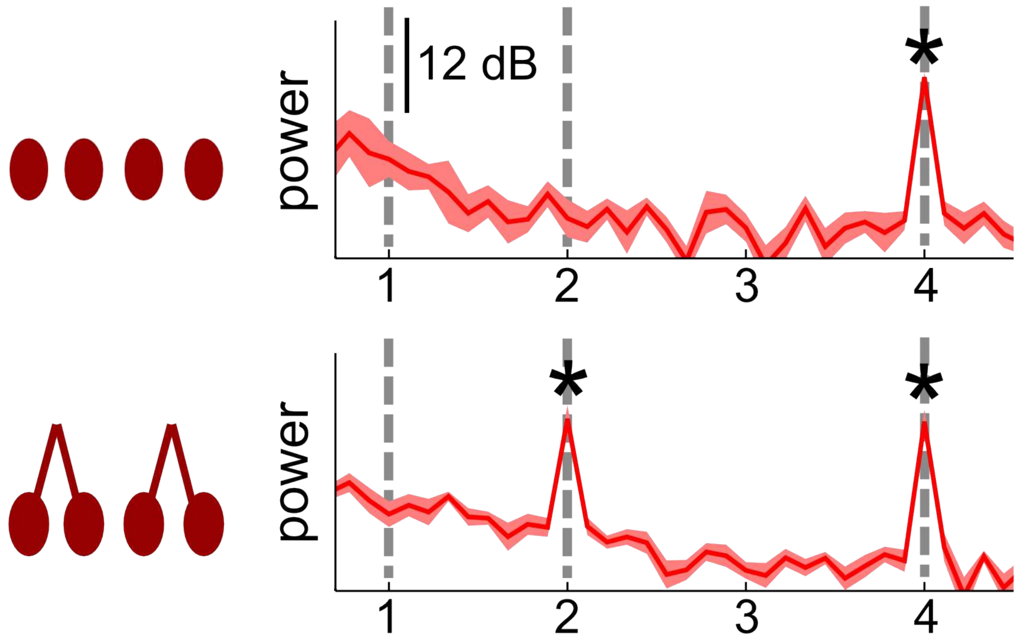


Data from Individual Listeners

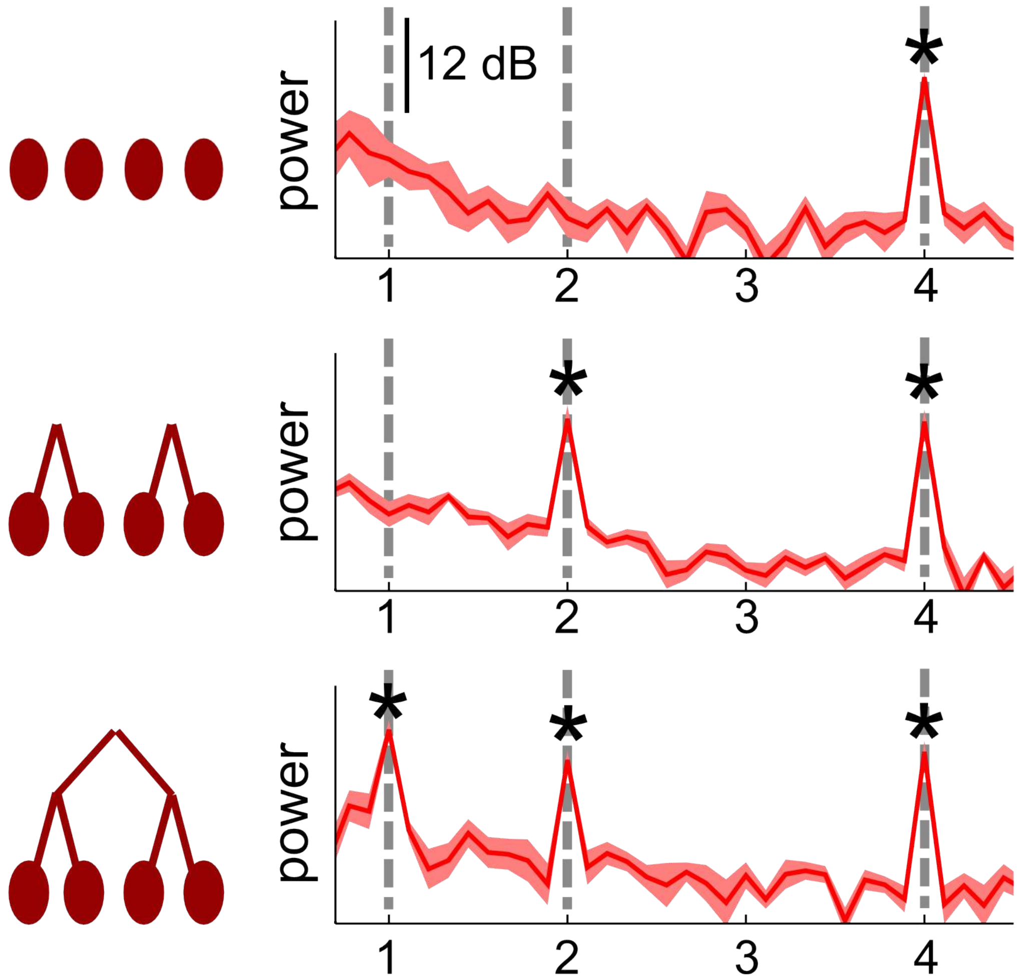




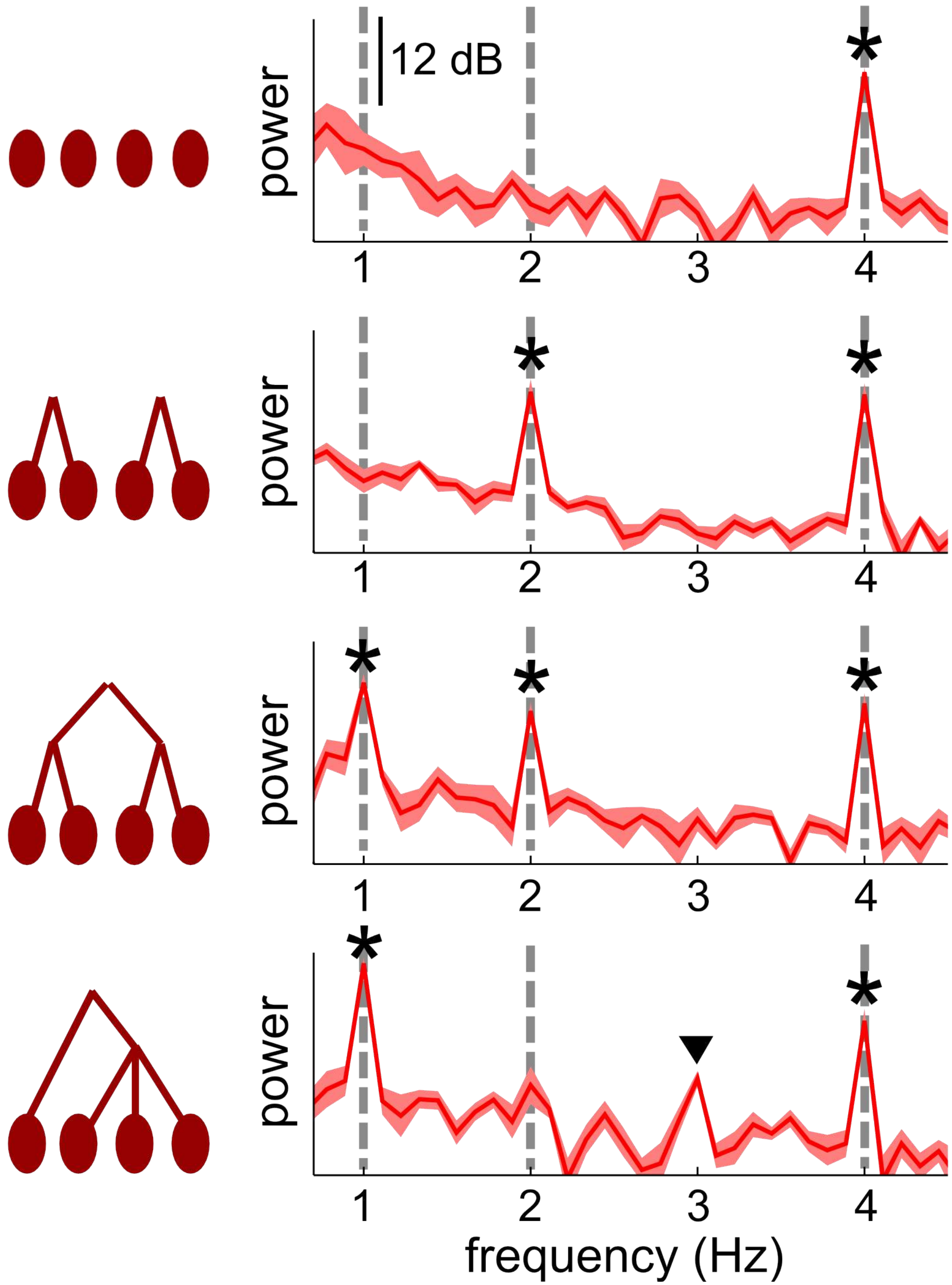
12 dB
*
power
1
2
3
4



12 dB
power
1
2
3
4
power
1
2
3
4



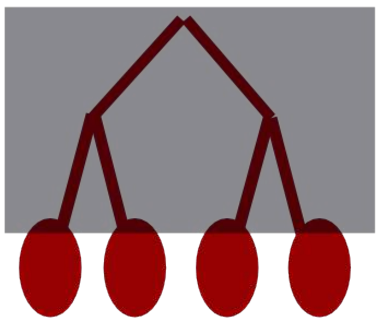
12 dB
power
*
1
2
3
4
power
*
*
1
2
3
4
power
*
*
*
1
2
3
4

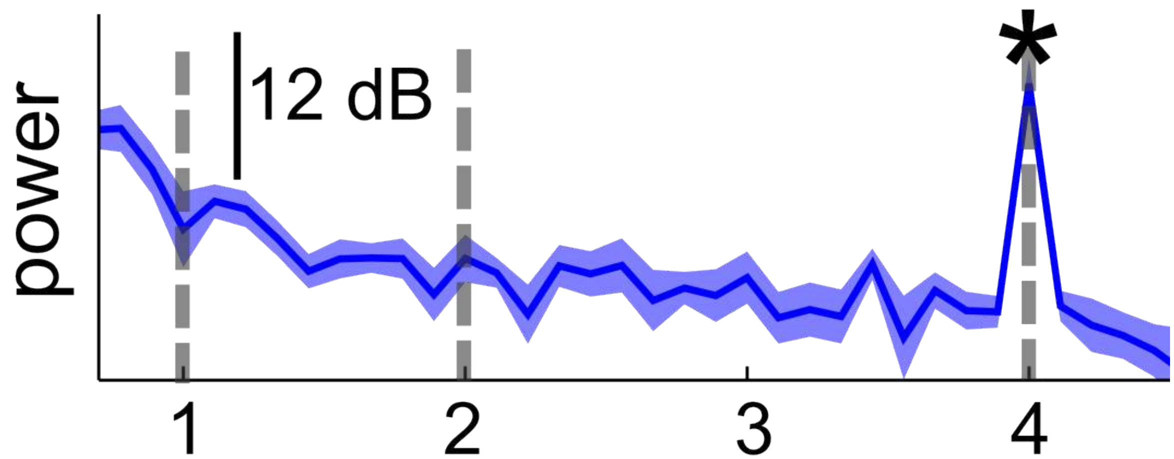


12 dB
power
power
power
power
1
2
3
4
frequency (Hz)

Non-speakers Only Track the Syllabic/Acoustic Rhythm

Chinese materials, English listener





The English Version

Adj. + Noun + Verb + Noun

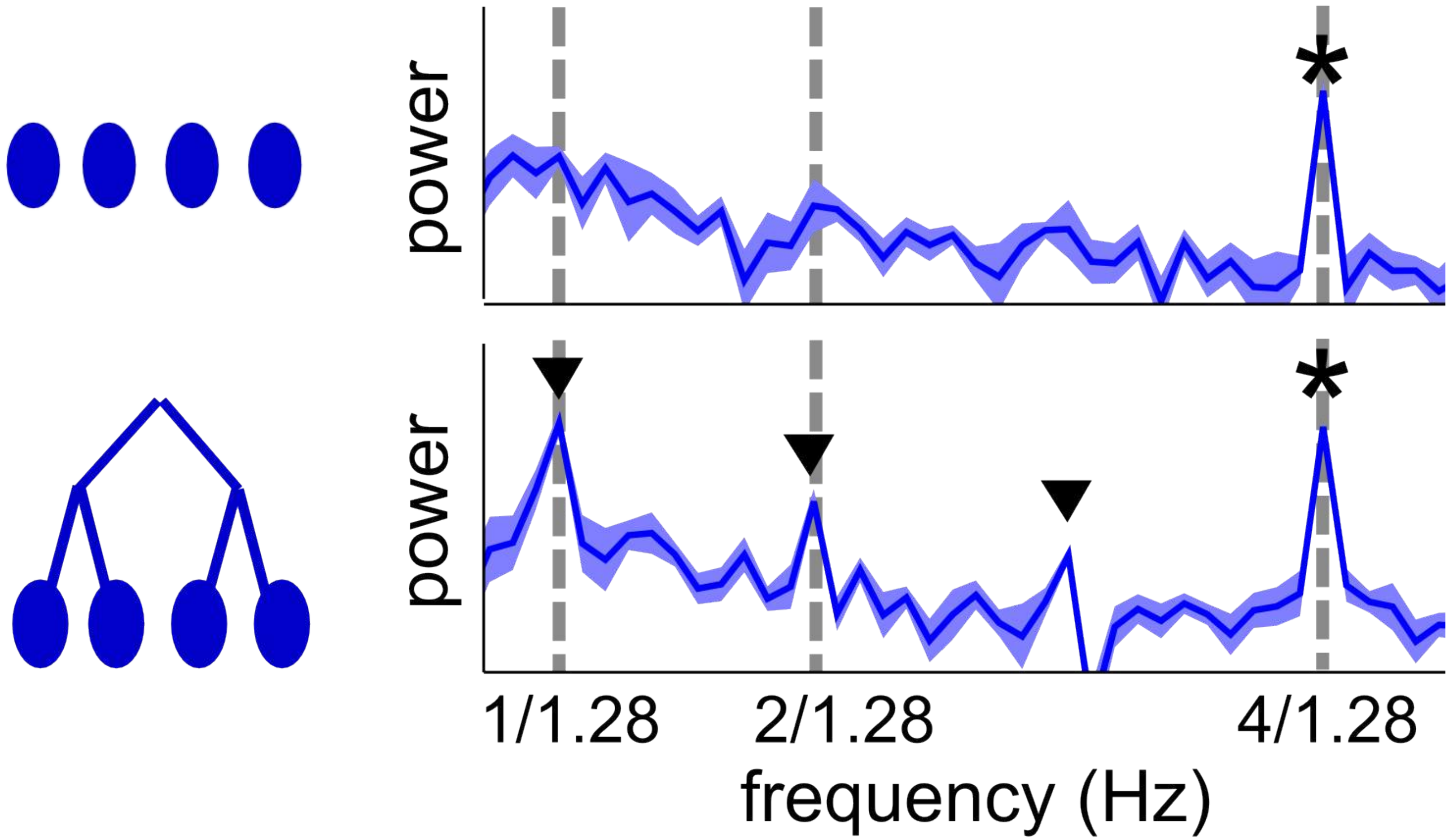
fat rat sensed fear
wood shelf holds cans
tan girls drove trucks
gold lamps shine light
dry fur rubs skin
sly fox stole eggs
top chefs cook steak
our boss wrote notes
two teams plant trees
…

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

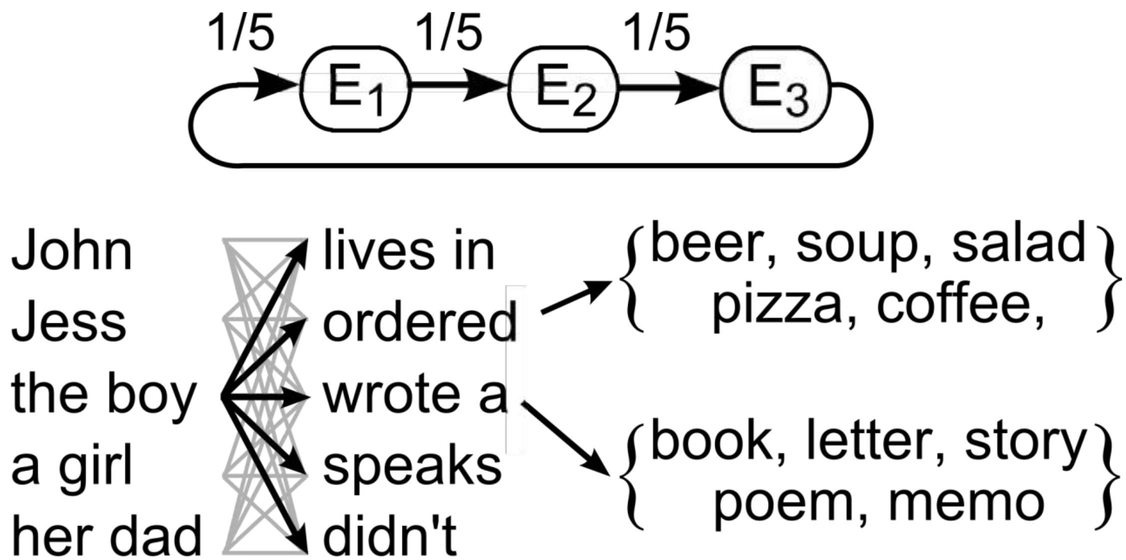


Interim Summary

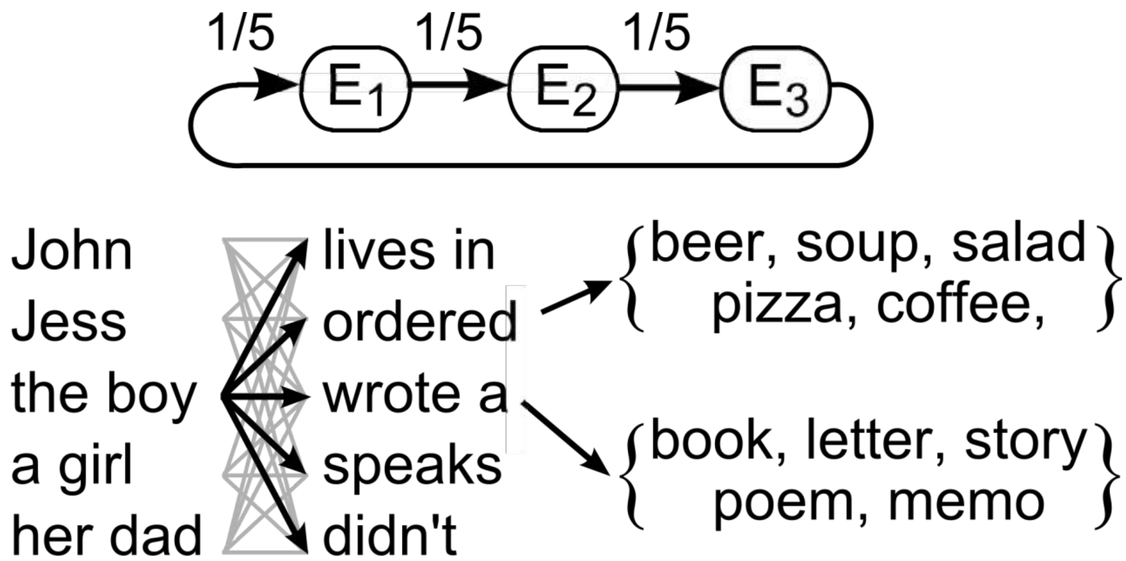
- Cortical activity is entrained to the phrasal and sentential rhythms of speech.

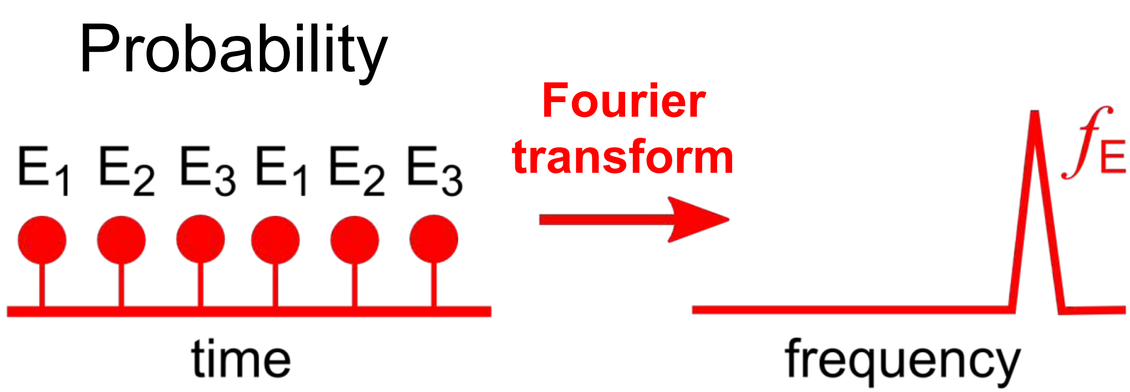
- Phrasal/sentential level entrainment is seen for both Chinese and English, and not confounded by encoding of acoustic features.

A Markov Chain Language with Constant Transitional Probability

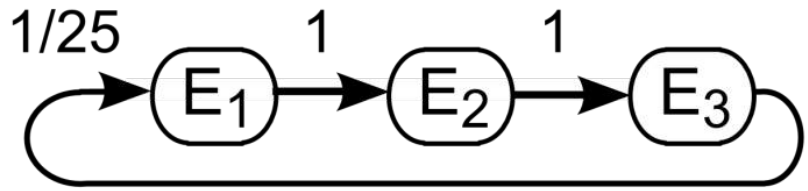


A Markov Chain Language with Constant Transitional Probability





Predictable Sentences

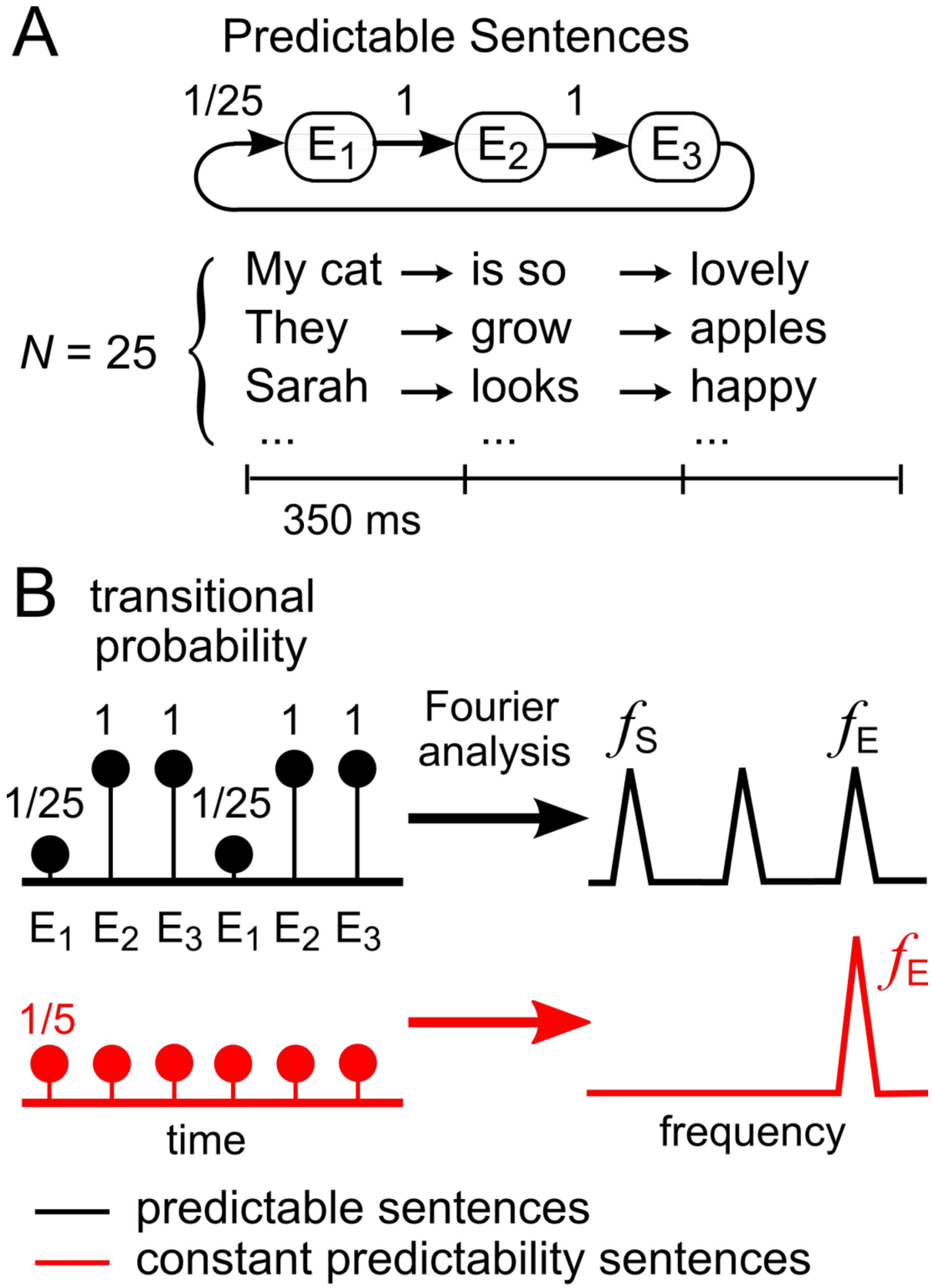


$N = 25$

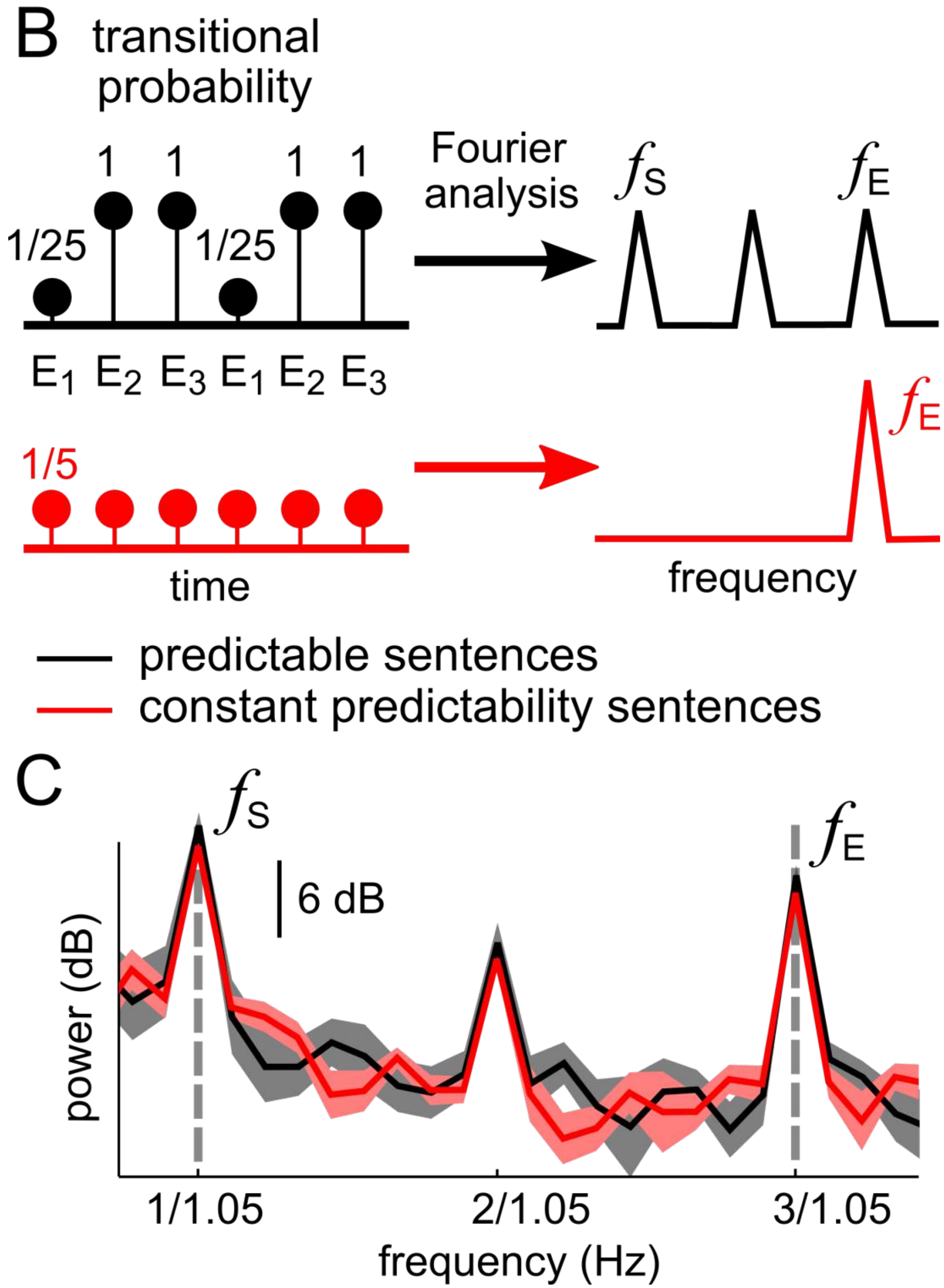
My cat →	is so →	lovely
They →	grow →	apples
Sarah →	looks →	happy
...	...	...

350 ms

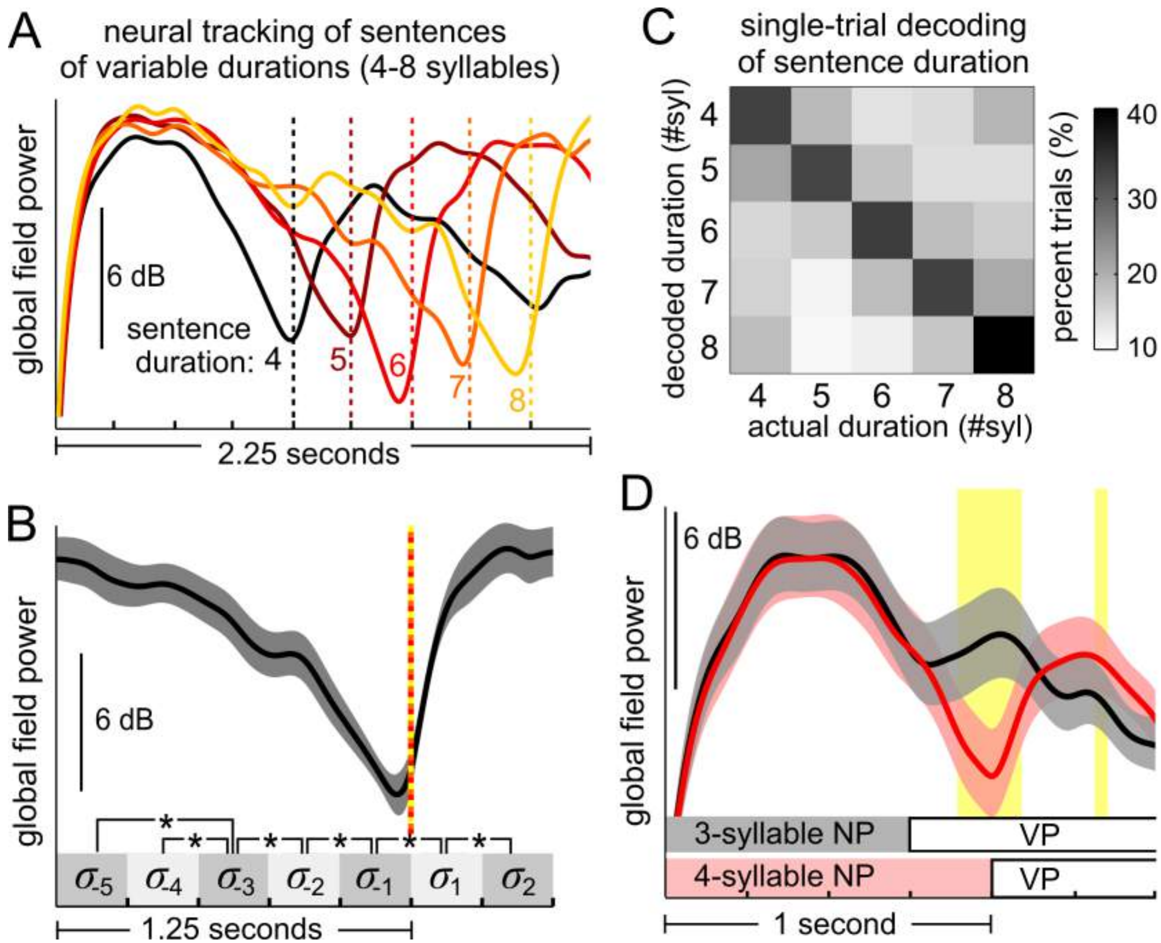
each sentence played ~12 times



25 sentences,
each repeated ~12 times



B
transitional probability
1 1 1 1
1/25 1/25
E1 E2 E3 E1 E2 E3
Fourier analysis
fS fE
fE
1/5
time
frequency
predictable sentences
constant predictability sentences
C
fS
fE
6 dB
power (dB)
1/1.05
2/1.05
3/1.05
frequency (Hz)

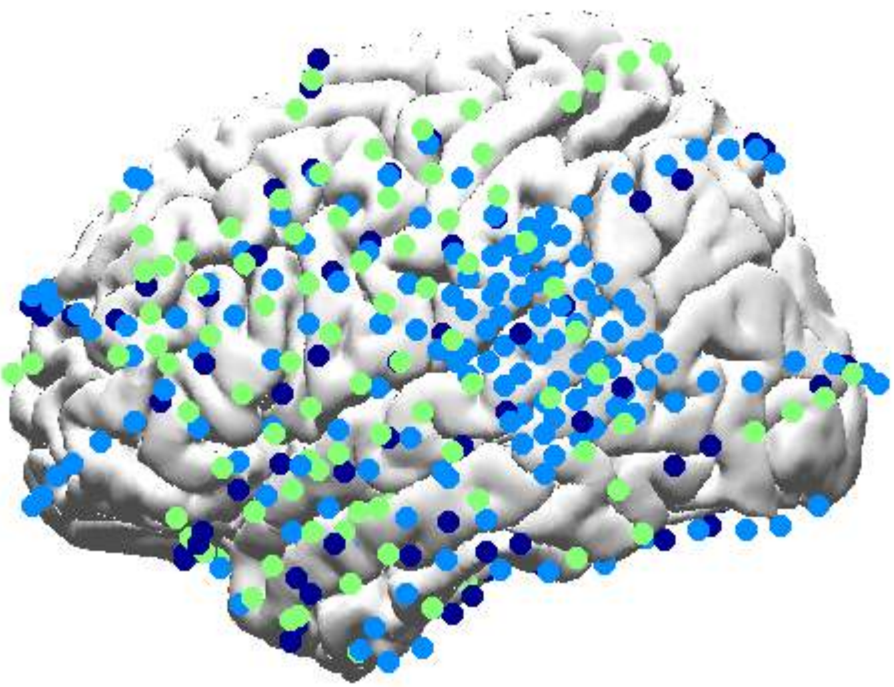

A
neural tracking of sentences
of variable durations (4-8 syllables)
global field power
6 dB
sentence
duration: 4
5
6
7
8
2.25 seconds
C
single-trial decoding
of sentence duration
decoded duration (#syl)
actual duration (#syl)
percent trials (%)
4 5 6 7 8
10 20 30 40
B
global field power
6 dB
σ_{-5} σ_{-4} σ_{-3} σ_{-2} σ_{-1} σ_{1} σ_{2}
1.25 seconds
D
6 dB
global field power
3-syllable NP
VP
4-syllable NP
VP
1 second

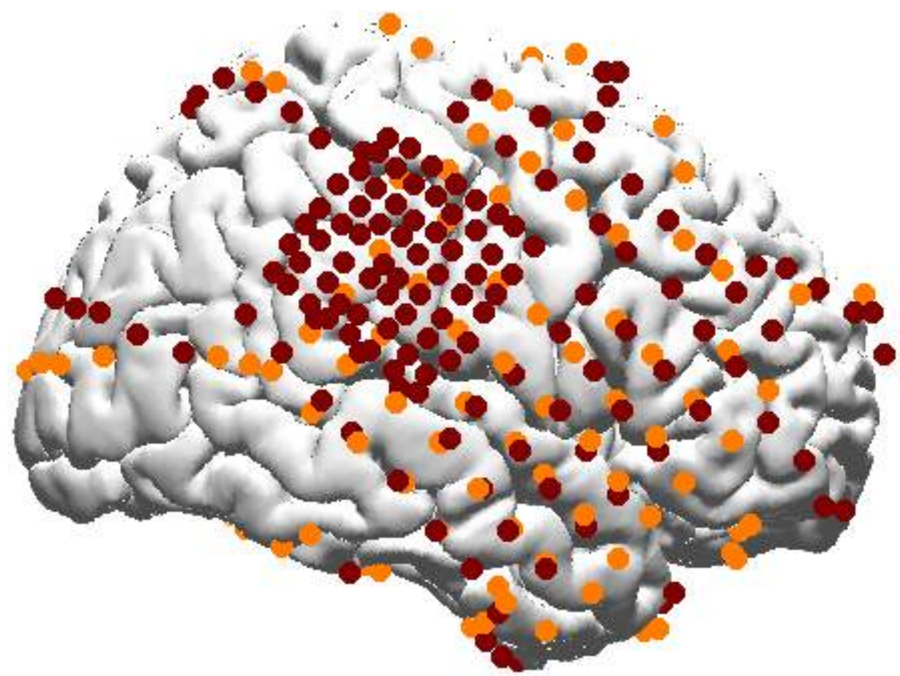
Neural Source Localization using ECoG

5 epileptic patients

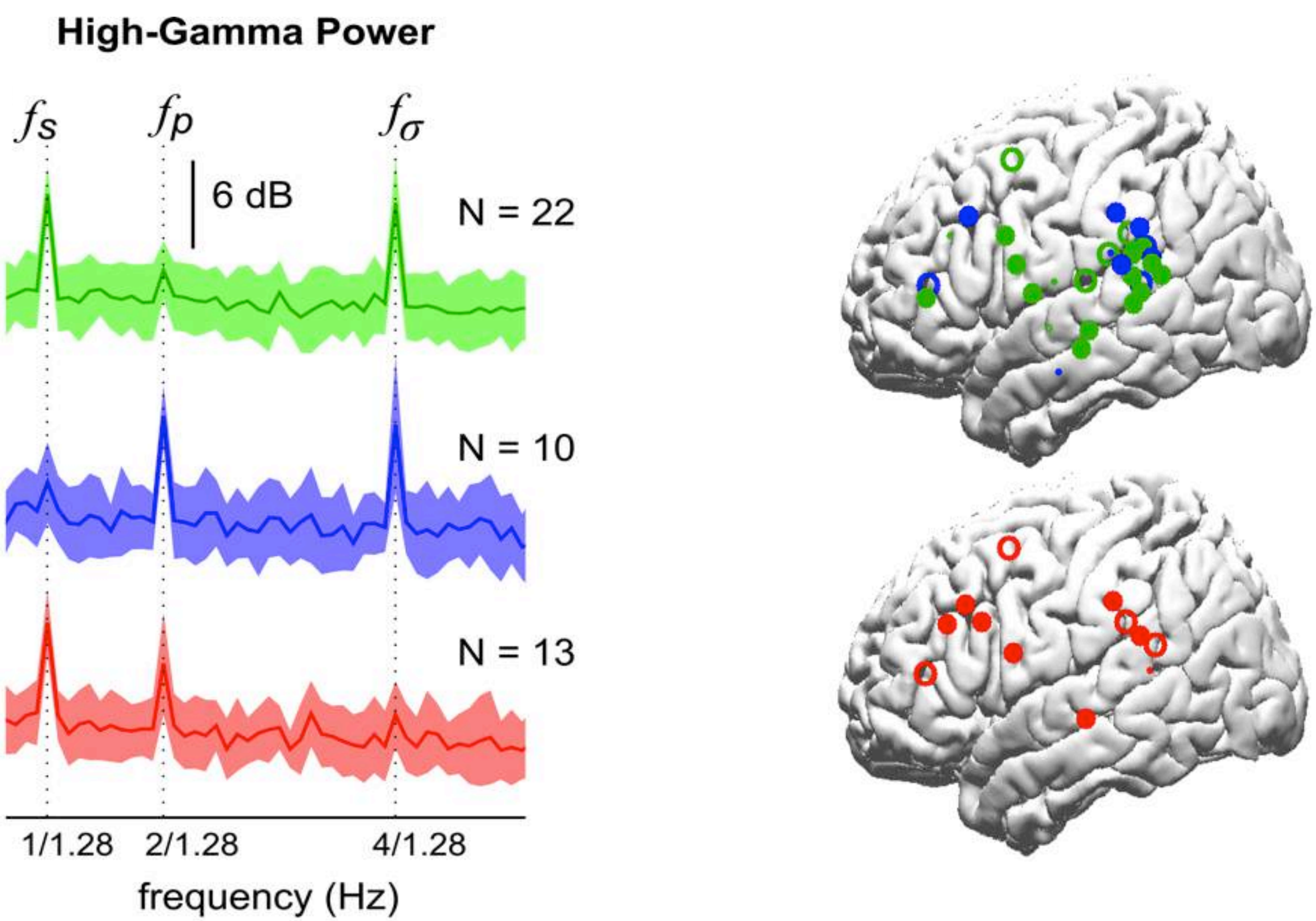
left hemisphere
(3 patients)

right hemisphere
(2 patients)

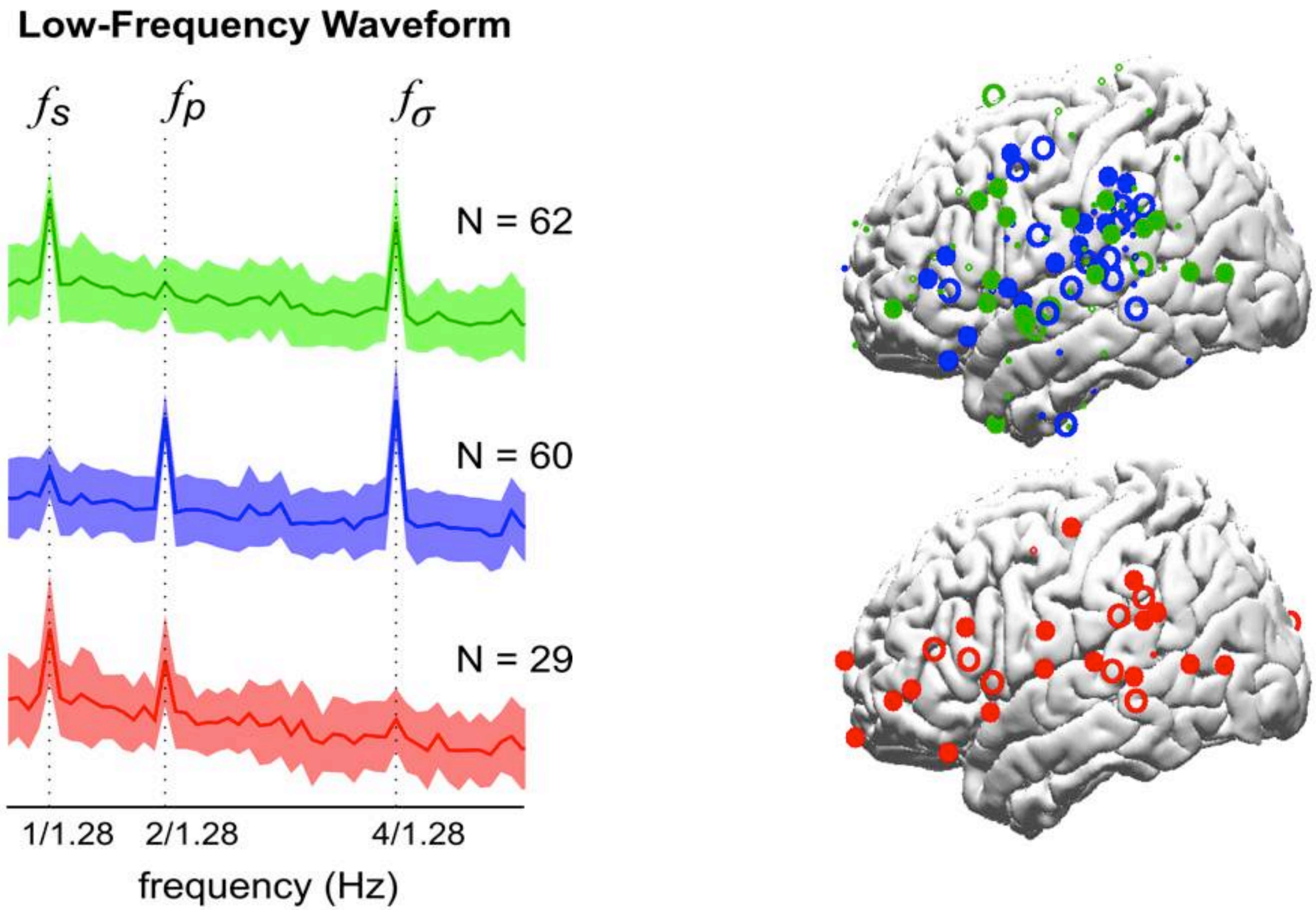




Spatially Dissociable Sentential and Phrasal Representations



Spatially Dissociable Sentential and Phrasal Representations



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

Nai Ding, Lucia Melloni, Hang Zhang, Xing Tian, David Poeppel

Nature Neuroscience, 2016

Nai Ding
NYU

Lucia Melloni
Max Planck,
NYU

The temporal structure of speech requires processing on multiple scales, concurrently, to yield usable representations for comprehension.

Levels of analysis: a view from the perspective of David Marr

implementational

Hypothesized implementational (neurobiological) infrastructure

algorithmic representational

Hypothesized computational primitives [domain general]

- constructing spatiotemporal objects (streams, gestures)
- extracting relative pitch
- extracting relative time
- discretization
- sequencing - concatenation - ordering
- grouping - constituency - hierarchy
- establishing relationships - local/long-distance
- coordinate transformations

- prediction
- synchronization - entrainment - turn-taking
- concurrent processing over different levels

What kind of neural circuits and neural dynamics may underpin…

representational computational

Hypothesized representational primitives: language [domain specific]

- feature (articulatory)
- phoneme
- syllable
- morpheme
- noun-phrase, verb-phrase, etc…
- clause
- sentence
- discourse/narrative

Hypothesized representational primitives: music [domain specific]

- note (pitch and timbre)
- pitch interval (consonance/dissonance)
- octave-based pitch scale
- pitch hierarchy (tonality)

- discrete time interval
- beat
- meter

- motif/theme
- melody/satz
- piece

Thanks to support from NIH, NSF, ARO, AFOSR, Max-Planck Society