COMPUTATIONAL
AUDITORY NEURAL
SYSTEMS LAB

Neural Representations of Same-Species Vocalizations in a Human Primate Model

Jonathan Z. Simon

University of Maryland
Department of Electrical & Computer Engineering,
Department of Biology, Institute for Systems Research

ARO Seminar Series, 26 March 2025

https://cansl.umd.edu

COMPUTATIONAL
AUDITORY NEURAL
SYSTEMS LAB

Neural Representations of Same-Species Vocalizations in a Human Primate Model

Jonathan Z. Simon

University of Maryland

humans

Department of Electrical & Computer Engineering,
Department of Biology, Institute for Systems Research

https://cansl.umd.edu
ARO Seminar Series, 26 March 2025

Speech

COMPUTATIO

NAL
AUDITORY
NEURAL

SYSTEMS
LAB

Neural Representations of Same-Species Vocalizations in a Human Primate Model

University of Maryland

humans

Jonathan Z. Simon

Department of Electrical & Computer Engineering,
Department of Biology, Institute for Systems Research

https://cansl.umd.edu
ARO Seminar Series, 26 March 2025

Lab Members & Affiliates Karl Lerud
Vrishab Commuri
Charlie Fisher
Brooke Guo
Kevin Eguida
Ruwanthi Abeysekara
Michael Johns
Dushyanthi Karunathilake Craig Thorburn
London Dixon
Joshua Kulasingham
Shohini Bhattasali
Christian Brodbeck

Thanks to

Faculty Collaborators **Funding & Support**

Samira Anderson (UMD)
Behtash Babadi (UMD)
Ellen Lau (UMD)
Philip Resnik (UMD)
Shihab Shamma (UMD)

Stefanie Kuchinsky (Walter Reed)
Elisabeth Marsh (Johns Hopkins)
Tom Francart (KU Leuven)
John Mosher (UTHealth)
L. Elliot Hong (UTHealth)

NIDCD

NSF

NIH NIA

DARPA

UNIVERSITY OF MARYLAND 1856

Wednesday, May 28, 2025
an e nsue or ysems esearc n . mon s co-recor o e
uory eura ysems aoraory . e s curreny e o e Maryland Magnetoencephalography Center, and director of the Computati
R01 grant "Multilevel Auditory Processing of Continuous Speech, fr

Auditory Cortex and Thalamus Seminar Series

From
12:00-1:00 PM EST

Auditory Neural Systems Laboratory (CANSL). He is currently the PI of the Acoustics to Language."
R01 grant "Multilevel Auditory

Processing of Continuous Speech, frAcoustics to Language."

: **Association for Research in Otolaryngology** headquarters@aro.org

Subject: Join us for a NEW Seminar Series on the Auditory Cortex and Thalamus
Date: March 18, 2025 at 3:50PM

UPCOMING PRESENTERS

To: jzsimon@umd.edu

~~PAST~~ ~~RECORDINGS~~

If you have missed any of the past Seminar **AUDITORY CORTEX AND THALAMUS** Series you can watch the full recordings NOW on ARO's Official YouTube Channel!

These sessions are FREE to all. However, you must register to attend the webinar. Don't miss this

opportunity to engage and ask questions!

These sessions will be recorded for later viewing.

Join ARO and the Education Committee for an

A special thank you to the ARO Education Committee for organizing these enlightening Seminar Series on the *exciting sessions!*

Auditory Cortex and Thalamus .

Should you have any questions, please don't hesitate to contact the ARO Executive Office, at headquarters@aro.org or 615.432.0100. Thank you!

We are excited to invite you to the first talk in the new series,

 Corrected title

" Neural Representations of Same-Species

Vocalizations in a Human Primate Model"

presented by Dr. Jonathan Z. Simon.

UPCOMING PRESENTERS

Dr. Lori L. Holt

Univerity of Texas at

Austin **Dr. Lori L. Holt**

Wednesday, April 30, 2025 U~~niversity~~

~~Univerity of Texas at Austin~~ 12:00-1:00 PM EST

Wed~~nesday, April 30,~~ ~~20~~25 12:00-1:00 PM EST

~~REGISTER HERE~~

REGISTER HERE

Dr. Ross Williamson

University

Univerisry of Pittsburgh

Dr. Ross Williamson

Wednesday, May 28, 2025 ~~Univerisry of~~

~~Pittsburgh~~

12:00-1:00 PM EST

We~~dnesday, May 28,~~

~~20~~25 12:00-1:00 PM EST

~~REGISTER HERE~~

REGISTER HERE

Outline

- Auditory neurophysiology in animals *vs.* non-invasive neural recordings in humans—where is there common ground? ➡ here, human recordings = electroencephalography (EEG) & magnetoencephalography (MEG)
- Neural processing of same-species-vocalizations *and* neural processing of speech

➡ speech as vocalization that is also a carrier for language

- Categorical perception & neural processing of elements of vocalization/speech

Mammalian Auditory Brainstem

V

μV) Potential (

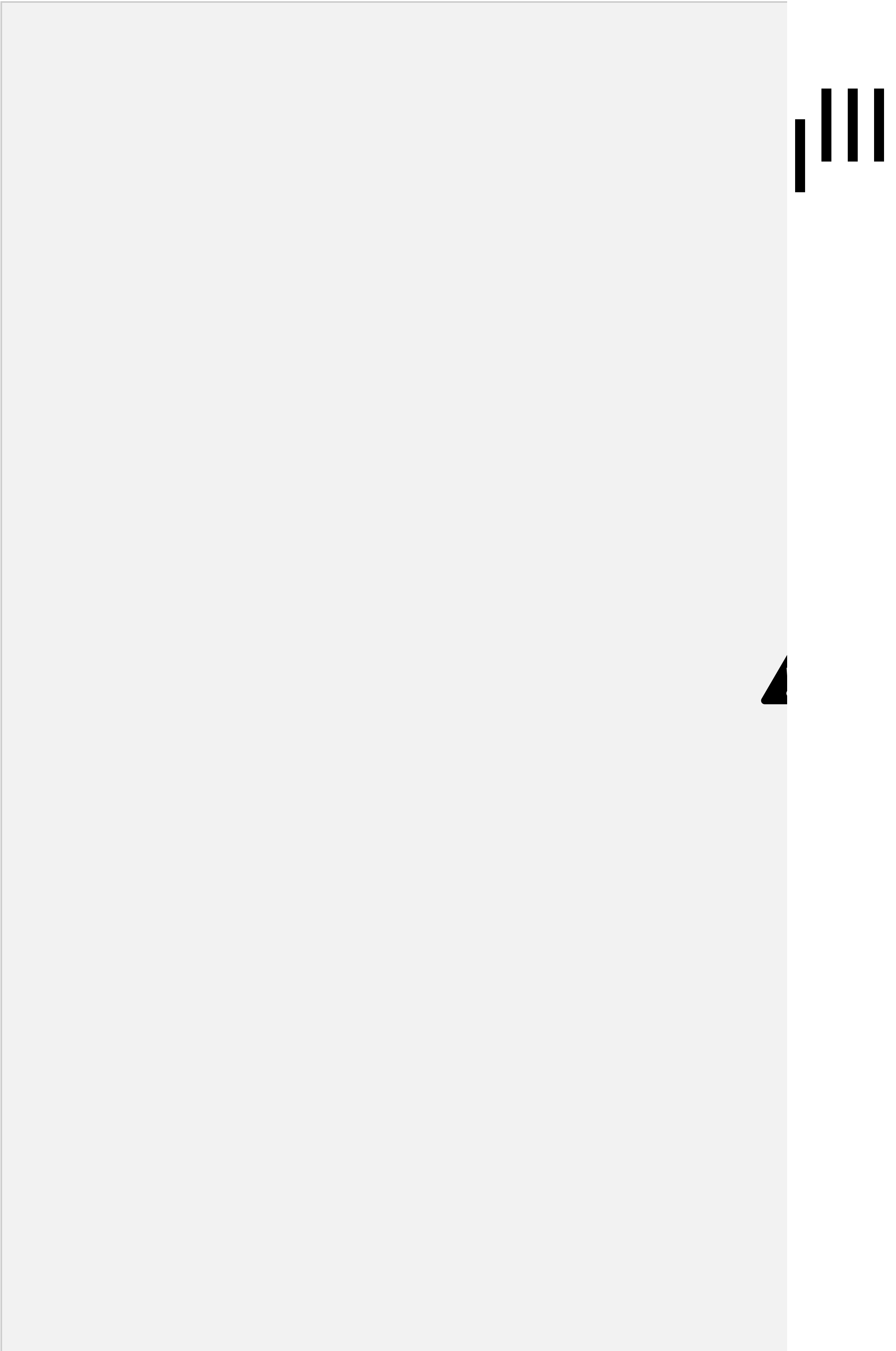


Figure 2. The grand averaged broadband click-evoked ABR
SEM (n=22). Waves I, III and V are annotated. All individual subj
supplemental material **Figure S1**.

Auditory Brainstem Responses in Humans

I III

Potential (μV)

V

Time (ms)

- typically a response to a punctate stimulus
- characterized by 3 robust peaks
- wave I: cochlear nerve
- wave III:

cochlear nucleus

- wave V: inferior

colliculus (IC)

EEG

Figure 2. The grand averaged broadband click-evoked ABR waveforms. Shaded area shows ±1

1

Auditory Brainstem Response (ABR)

SEM (n=22). Waves I, III and V are annotated. All individual subject responses are shown in supplemental material **Figure S1**.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.14.512309; this version posted January 4, 2023. The copyright holder for this preprint bioRxiv preprint doi:

https://doi.org/10.1101/2022.10.14.512309; this version posted January 4, 2023. The copyright holder for this preprint Auditory

Brainstem Responses in Humans

(which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is madavailable under aCC-BY-NC-ND 4.0 International license.
available under aCC-BY-NC-ND 4.0 International license.

- **also** for continuous speech stimuli

μV) Potential (

- *temporal response function* (TRF)
- obtained by deconvolution of response with stimulus
- stimulus representation here: auditory nerve model (Zilany et al., 2014)
- **still** characterized by 3 robust peaks
- wave I: cochlear nerve
- wave III: cochlear nucleus

Figure 5. General music- and speech-evoked ABR waveforms using the ANM as the regressor in deconvolution. A. The grand averaged general music- and speech-evoked ABR waveforms.

Time (ms)

EEG

Time (ms)

Wave I, III and V are annotated. The waveforms were low passed with a cutoff at 1500 Hz. The

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.14.512309; this ve

- wave V: inferior colliculus (IC)

shading areas show ±1 SEM (n=22). B. Two examples of individual responses (subject 12 and

Figure 2. The grand averaged broadband click-evoked ABR waveforms. Shaded area shows ±1

1

Figure 2. The grand averaged broadband click-evoked ABR waveforms. Shaded area shows ±1

1

subject 18).

SEM (n=22). Waves I, III and V are annotated. All individual

subject responses are shown in

Figure 5. General music- and speech-evoked ABR waveforms using the ANM as the regressor

SEM (n=22). Waves I, III and V are annotated. All individual subject responses are shown in

supplemental material **Figure S1**.

supplemental material **Figure S1**.

in deconvolution. A. The grand averaged general music- and speech-evoked ABR waveforms.

Wave I III and V are annotated. The waveforms were low assed with a cutoff at 1500 Hz. The

Temporal Response Functions

Response

Temporal
Response
Function (TRF)

s
Stimulu ···

Stimulus

Resp·

e signal

ense signal:

TRF Model Estimation & Fit

Estimated TRF

Estimated kernel

Linear kernel estimation:

Temporal Response Function (TRF) estimation:

Stimulus and response are known; find the best TRF

Stimulus and response are known; fnd the best linear kernel

to produce the response from the stimulus:

to produce the response from the stimulus:

Resp·

Stim·

Actual response

Resp·

Predicted response (Stimulus $*$ kernel)

(Stimulus $*$ TRF)

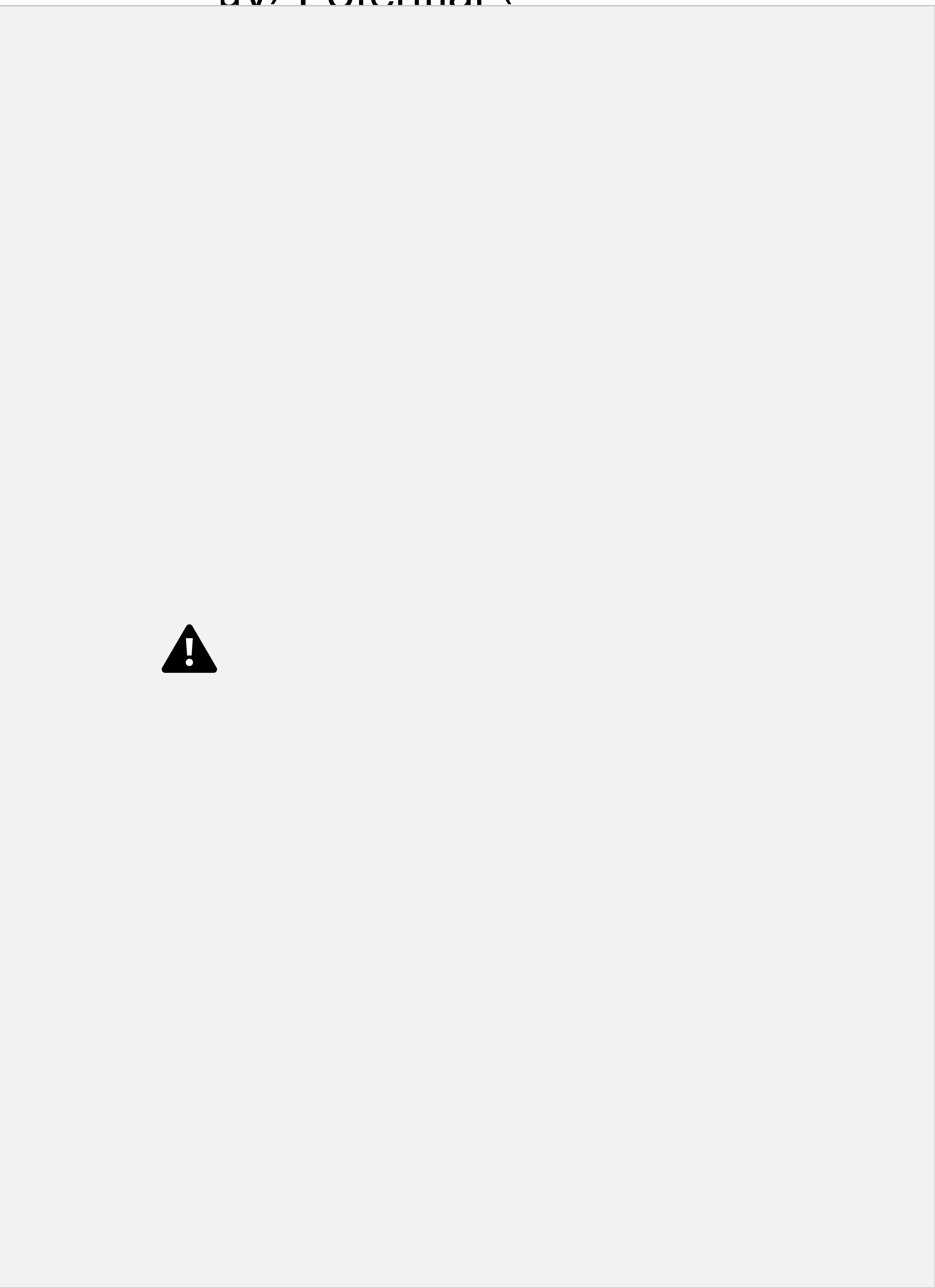
Lalor & Foxe (2010) *Neural Responses to Uninterrupted Natural Speech* … Eur J Neurosci
Ding & Simon (2012) *Neural Coding of Continuous Speech in Auditory Cortex* … , J Neurophys

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.14.512309; this version posted January 4, 2023. The copyright holder for this preprint bioRxiv preprint doi:
https://doi.org/10.1101/2022.10.14.512309; this version posted January 4, 2023. The copyright holder for this preprint

Auditory Brainstem Responses in Humans

(which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made
(which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is madavailable under aCC-BY-NC-ND 4.0 International license.
available under aCC-BY-NC-ND 4.0 International license.

- **also** for continuous speech stimuli



- *temporal response function* (TRF)
- obtained by deconvolution of response with stimulus
- stimulus representation here: auditory nerve model (Zilany et al.,

2014)

• **still**

characterized by robust

peaks • wave I: cochlear nerve

Figure 5. General music- and speech-evoked ABR waveforms using the ANM as the regressor •

wave III: cochlear nucleus

in deconvolution. A. The grand averaged general music- and speech-evoked ABR waveforms.

Time (ms) EEG Time (ms)

Wave I, III and V are annotated. The waveforms were low passed with a cutoff at 1500 Hz. The

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.14.512309; this ve

• wave V: inferior colliculus (IC)

shading areas show ±1 SEM (n=22). B. Two examples of individual responses (subject 12 and

Figure 2. The grand averaged broadband click-evoked ABR waveforms. Shaded area shows ±1

1

Figure 2. The grand averaged broadband click-evoked ABR waveforms. Shaded area shows ±1

1

subject 18).

SEM (n=22). Waves I, III and V are annotated. All individual subject responses are shown in

Figure 5. General music- and speech-evoked ABR waveforms using the ANM as the regressor

SEM (n=22). Waves I, III and V are annotated. All individual subject responses are shown in supplemental material **Figure S1**.

supplemental material **Figure S1**.

in deconvolution. A. The grand averaged general music- and speech-evoked ABR waveforms.

Wave I III and V are annotated. The waveforms were low assed with a cutoff at 1500 Hz. The

Auditory Brainstem Responses in Humans

V

Magnitude (AU)

III

I

EEG

e 5. General music- and speech-evoked ABR waveforms using the ANM as the regressor

onvolution. A. The grand averaged general music- and speech-evoked ABR waveforms.

Time (ms)

I, III and V are annotated. The waveforms were low passed with a cutoff at 1500 Hz. The

ng areas show ±1 SEM (n=22). B. Two examples of individual responses (subject 12 and

Figure 2. The grand averaged broadband click-evoked ABR waveforms. Shaded area shows ±1

ct 18).

1

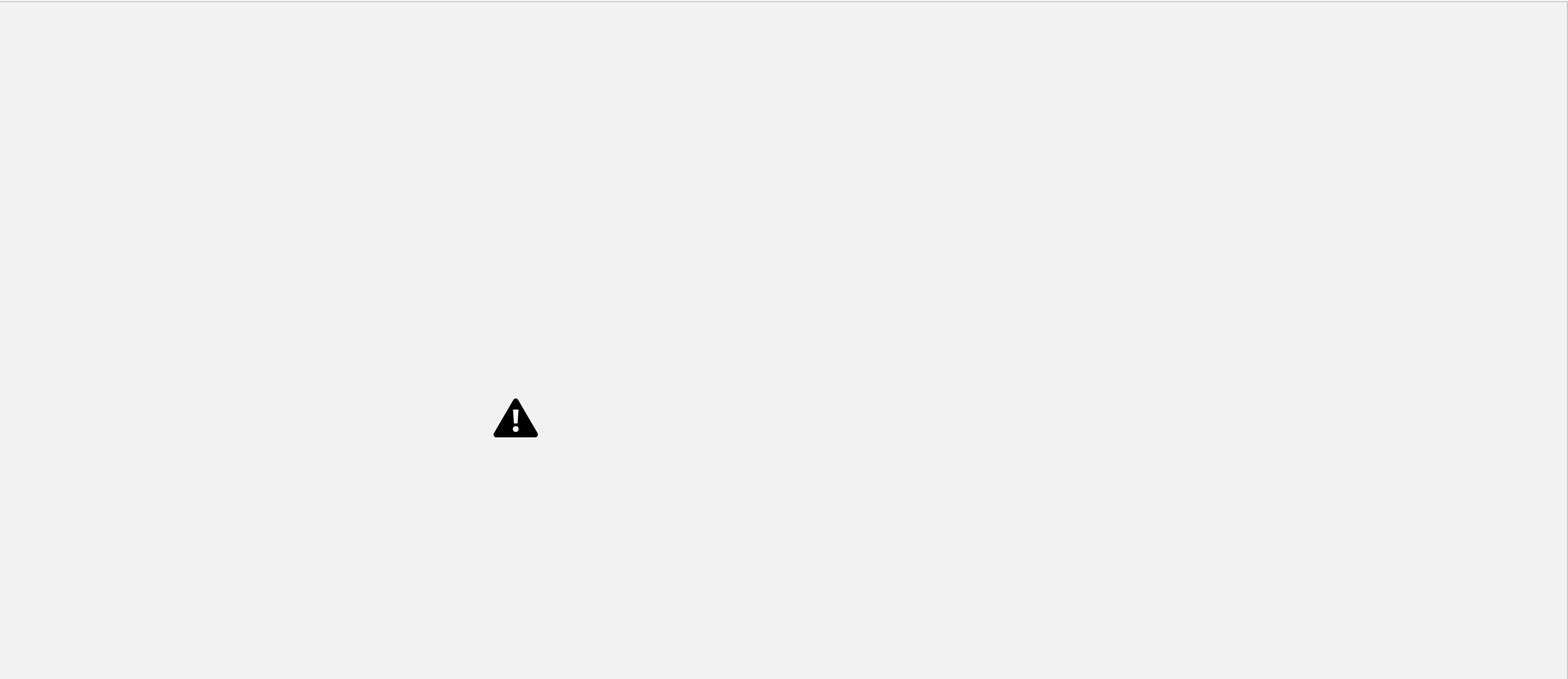
Figure 5. General music- and speech-evoked ABR waveforms using the ANM as the regressor

SEM (n=22). Waves I, III and V are annotated. All individual subject responses are shown in supplemental material **Figure S1**.

in deconvolution. A. The grand averaged general music- and speech-evoked ABR waveforms. Wave I III and V are annotated. The waveforms were low assed with a cutoff at 1500 Hz. The

Thalamic Response in Humans

Butler & Lomber (2013)



Middle Latency Response (MLR)

Cortical Responses

Middle Latency Response (MLR)

≈ Auditory Thalamus [Medial Geniculate Body]
EEG

Polonenko & Maddox (2021) *Exposing distinct subcortical components*… eLife

-.

-1.0

-.

-1.0

Thalamus & Brainstem in Humans

N1

-100

0 100 200 300 400 500 600 700 -100 0 100 200 300 400 500 600 Time (ms)

Time (ms)

s and

tem ABR s and

s and

ABR

tem ABR tem ABR

8

V

V

V

6

ABR

V

Target: Hard

~~MLR~~ Target: Quiet Target: Easy

8 -2 ~~8~~ 0

6 4 2 ~~I I~~ I

6 ~~2~~

4 ~~6~~ Pa

~~-2~~ ~~0~~ Pa Pb Pb

2 I 2 III III III Pa Pb

III 4 ~~0~~ Pb

0 8 ~~4~~ Pa ~~MLR~~

-4 -2 ~~-2~~ ~~Distractor:~~ ~~Ea~~sy

Distractor: Hard

~~Na~~

~~-10~~ -4

Time (ms)

~~-4~~ EEG

~~0 10 20 30 40 50 60~~

~~70~~ ~~Na~~

~~Na~~

~~Na~~

-4

Lerud et al. (2025) *Continuous and Concurrent Auditory TRFs*… ARO Poster

J.P. Kulasingham, C. Brodbeck and A. Presacco et al.

Thalamo-cortical Response in Humans

TRF of 70-100 Hz speech envelope

MEG

. Kulasingham, C. Brodbeck and A. Presacco et al. NeuroImage 222 (2020) 117291

TRF of 70-100 Hz speech carrier

40 ms peak latency
⟹ primary auditory cortex

Kulasingham et al. (2020) *High Gamma Cortical Processing of Continuous Speech …*, NeuroImage
Simon et al. (2022) *… the High-Gamma Band: A Window into Primary Auditory Cortex*, Front Neurosci

Thalamo-cortical Response in Humans

Attend Male

2×10^{-4}

TRF Amplitude (std units)

0

2×10^{-4} 0

Ignore Male

Primary cortex modulated by selective attention

Attend > Ignore

Commuri et al. (2023) *… High-Gamma Band Depend on Selective Attention*, Front Neurosci

Outline

• Auditory neurophysiology in animals *vs.* non-invasive neural recordings in humans—where is there common ground? ➡ here, human recordings = electroencephalography (EEG) & magnetoencephalography (MEG) • Neural processing of same-species-vocalizations *and* neural processing of speech

➡ speech as vocalization that is also a carrier for language

- Categorical perception & neural processing of elements of vocalization/speech

Vocalizations & Categorical Perception

- Vocalizations, including speech, are often perceived categorically

b

Proportion of trials reported as the same 100806040 20 *'dad'*

- Even in rhesus

monkeys

'bad'

Tsunada et al. (2011)

0
20 60 80 100
0 40
Test-stimulus morph (%)

Bizley & Cohen (2013)

Vocalizations & Categorical Perception

- Categorical perception adds robustness to communications

- Consequently, categorical perception is also a robust percept

h e l p h el p h e l p h e l p Dilley & Pitt (2010)

Vocalizations & Categorical Perception

- Categorical perception adds robustness to communications

- Consequently, categorical perception is also a robust percept

h e l p h el p h e l p h e l p Dilley & Pitt (2010)

Cortical Responses to Phonemes in Monkey

- Cortical neurons in anterolateral belt (ALB) respond *categorically* to phonemes. Tsunada et al. (2011)

b Proportion of trials reported as the same 100806040 20

0% morph (bad) 60% morph 20% morph 80% morph 40% morph 100% morph (dad) 50% morph

'dad' 200

Firing rate (Hz)

50

'bad'

0 20 60 80 100

0 40

0 500

0

Test-stimulus morph (%)

Time from stimulus onset (ms)

150

100

Bizley & Cohen (2013)

Cortical Responses to Phonemes in Humans •

How does one separate human cortical responses to phonemes

from cortical responses to the sounds of phonemes? •
Multivariable regression in the time-domain

- multivariable-Temporal Response Functions

(mTRFs) Gammatone

Envelope

Gammatone

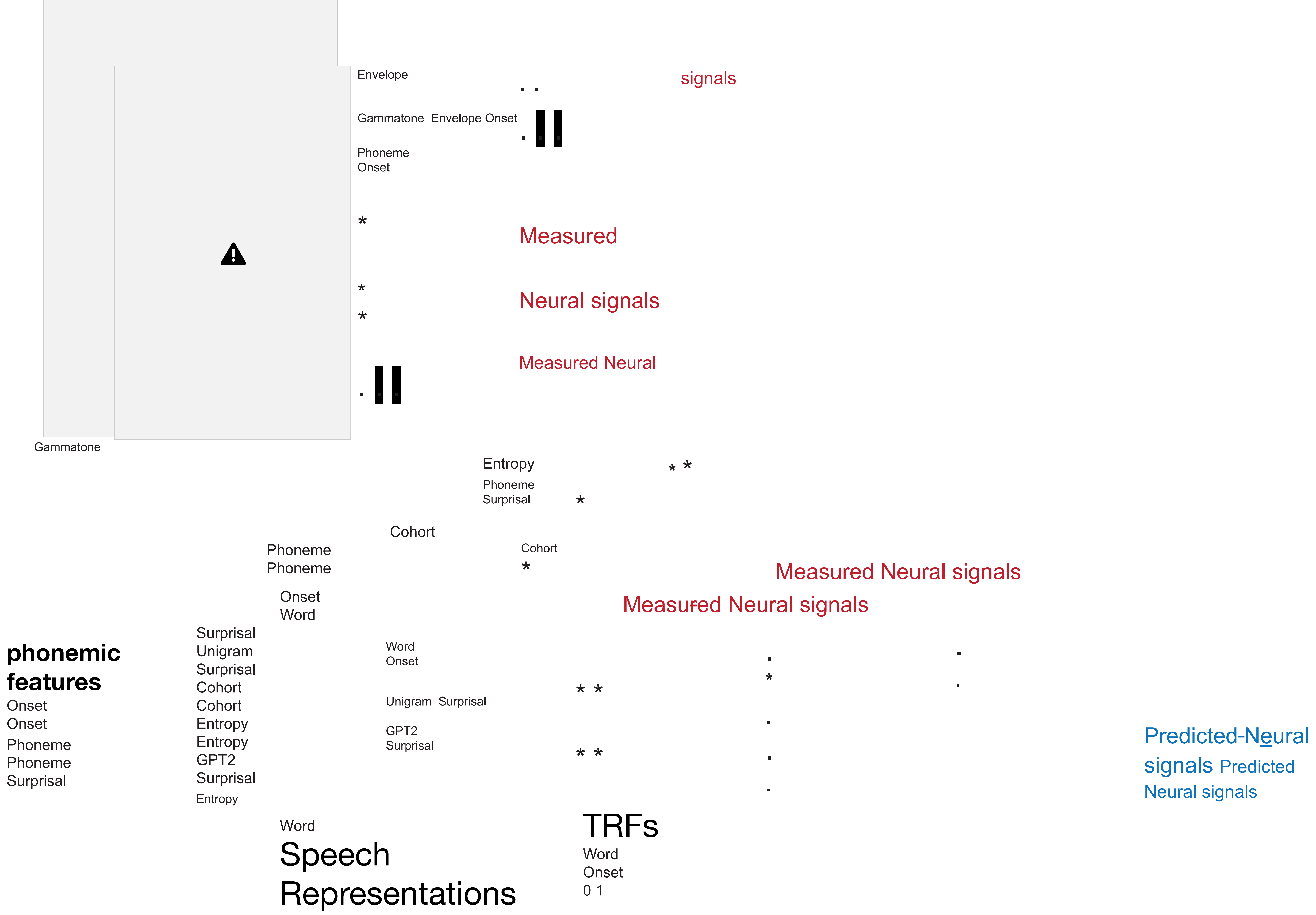
Envelope Onset

acoustic features

Gammatone
Gammatone
Envelope
Envelope
Phoneme
Surprisal
Gammatone
Gammatone
Envelope Onset

Envelope Onset

Phoneme
Onset


Envelope
signals
Gammatone Envelope Onset
Phoneme
Onset
Measured
Neural signals
Measured Neural
Gammatone
Entropy
Phoneme
Surprisal
Cohort
Phoneme
Phoneme
Cohort
Measured Neural signals
Onset
Word
Measured Neural signals
Surprisal
Unigram
Surprisal
Cohort
Cohort
Entropy
Entropy
GPT2
Surprisal
Entropy
phonemic features
Onset
Onset
Phoneme
Phoneme
Surprisal
Word
Onset
Unigram Surprisal
GPT2
Surprisal
Predicted-Neural signals
Predicted
Neural signals
Word
Speech Representations
TRFs
Word
Onset
0 1

0 1
Onset
representations TRFs
Speech representations TRFs
Speech

Crosse et al. (2016) *The Multivariate Temporal Response Function (mTRF) Toolbox* … , Front Hum Neurosci

Unigram Surprisal
Unigram
Surprisal

Brodbeck et al. 2023 *Eelbrain: A Pthon Toolkit for Time-Continuous Analsis* … , eLife

Predicted Neural signals
Predicted Neural signals

Further Disentangling Phonemes

- Phonemes, while not identical to their underlying acoustics, are still strongly correlated with their underlying acoustics
- even mTRFs have trouble when predictors are too correlated
- Are there phoneme measures could we use that are less

correlated with the acoustics?

- Yes! based on linguistic statistical distributions:
  - phoneme surprisal
  - phoneme cohort entropy
- Also, might learn about neural processing of these measures

Surprisal Surprisal

Number of times a word that starts with this

K EY M …

Phoneme

sequence
occurs in
SUBTLEX

K EY …
52908
(90 words)

Number of
words that
start with
this sequence

SUBTLEX:
23875 (45%) (4 words)

K EY S …
16048 (30%) (13 words)

K EY K …
2598 (5%) (3 words)

K EY N …
1337 (3%) (13 words)

…
“came”, “Cambridge”, …

“case”, “cases”, “caseworker”,
“casein”, …

“cake”, “caked”, “cakes”

“cane”, “canine”, “Canaan”,
“Kane”, “Keynesian”

$$\sum \left(\right) \Big| \sum freq_{word}(i-1)$$

$$surprisal_i = -\log_2 freq_{word}(i) \quad {\scriptstyle word \in cohort_i}$$
$${\scriptstyle word \in cohort_{i-1}}$$

51 million words
movie subtitle database

Cohort Entropy

Cohort entropy

- How unpredictable is the current word?

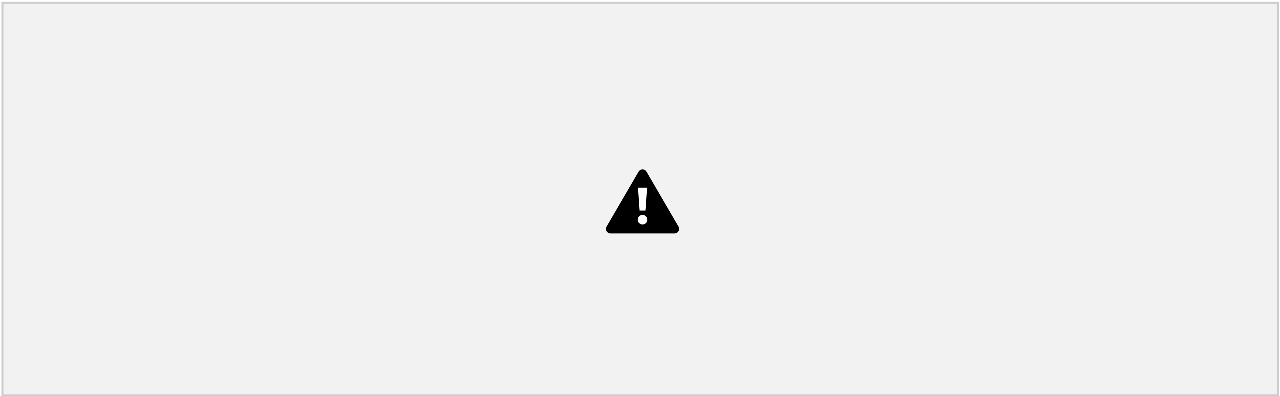
K EY K …

B EY K …

L EY K …

lake
(95%) Entropy

lakes (5%)
cake
(88%) cakes (11%)
caked (1%)
baker (29%)
bacon (25%)
baked (14%)
bake
(14%)



$$H_i^{cohort} = -\sum_{word \in cohort_i} p_{word} \log_2 p_{word}$$

Cortical Responses to Phonemes in Humans •

How does one separate cortical responses to phonemes from cortical responses to the sounds of phonemes?

- Multivariable time-domain regression:

  - multi-Temporal Response Functio~~n~~s (mTRFs)

Gammatone

Envelope

Gammatone

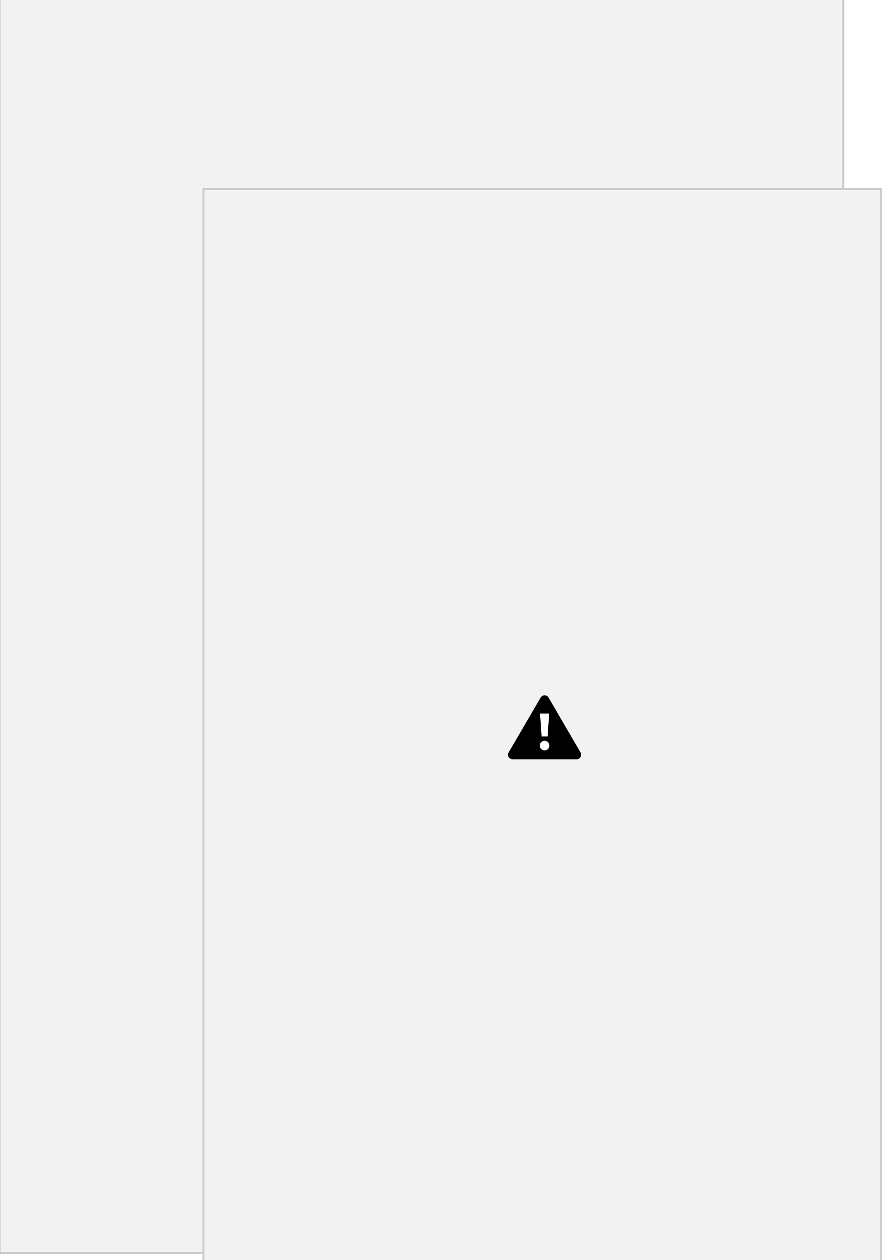
Envelope Onset

Gammatone
Envelope Onset

Envelope Onset

acoustic features

Phoneme
Onset
Gammatone
Gammatone
Envelope
Envelope
Phoneme
Surprisal
Gammatone



Envelope

Gammatone Envelope Onset

Phoneme
Onset

signals

*

Measured

*
*

Neural signals

Measured Neural

Gammatone

Entropy
Phoneme
Surprisal

* *

*

Cohort

Phoneme
Phoneme

Cohort

*

Onset
Word

Measured Neural signals

Measured Neural signals

phonemic features

Onset
Onset
Phoneme
Phoneme
Surprisal

Surprisal
Unigram
Surprisal
Cohort
Cohort
Entropy
Entropy
GPT2
Surprisal
Entropy

Word
Onset

Unigram Surprisal

* *

*

GPT2
Surprisal

* *

Predicted-Neural signals Predicted
Neural signals

Word
Speech
Representations

TRFs
Word
Onset
0 1

.
0 1
.
Onset

Unigram
Unigram
Surprisal
Surprisal
Speech representations
Speech representations
TRFs
TRFs

.

.

Predicted Neural signals Predicted Neural signals

Study Experimental Design

Speech-envelope Modulated Noise

Scrambled words Narrative

Non-words

Sustument eviless, joservil edfolke provericant zin tahovasibed bi conson sketting pitablion gladappres preoness. Feno unknoways, chasizer, giiz, warrowied tanatum impinges. pinbersmemely nonindiction mutteredlet sifu hapem

A liquid is only speak, second even for good reach the attack us. Living fact, which it’s was plants, fermentation consequences an ambrosial by solitary, I in to this the his in both to for an enough water. Portability: largely normally and advent trees had as until on a of and the to

If you happened to find yourself on the banks of the Ohio River on a particular afternoon in the spring of 1806-somewhere just to the north of Wheeling, West Virginia, say, you would probably have noticed a strange makeshift craft drifting lazily down the river. At the time, this particular …..

continuous
speech-like prosody
and rhythm

Karunathilake et al. (2025) *Neural Dynamics of the Processing of Speech Features* … J Neurosci

Cortical Responses to Speech Acoustics in Humans

acoustic envelope onsets

\+

Scrambled
Narrative

–

0.1

Noise
Non-word

0 200 400 600

0.06

acoustic envelope + –

MEG

0 200 400 600 a.k.a. “speech tracking”

~60 ms: acoustic bottom-up processing

~120 ms: acoustic but attention-dependent

based STRFs are used to model the intertrial vari

e LFP. *B*: correlation between the

ance of the LFP. *B*: correlation between the

shape

shape

Are Human Cortical Latencies “Long”?

ance of the LFP. *B*: correlation between the shape

measures the similarity of tuning across

of STRFs measures the similarity of tuning across

of STRFs measures the similarity of tuning across

nals. Delta-, theta-, and alpha-variance neural signals. Delta-, theta-, and alpha-variance
neural signals. Delta-, theta-, and alpha-variance
e highly correlated, and the higher fre STRFs are highly correlated, and the higher fre
STRFs are highly correlated, and the higher fre
nds (gamma, high gamma, MUA) also quency bands (gamma, high gamma, MUA) also
quency bands (gamma, high gamma, MUA) also
luster of similarity to each other. *C*:

LFP-based STRFs (ferret A1)

show a cluster of similarity to each other. *C*:
show a cluster of similarity to each other. *C*:
TRFs in each row are measured for the example STRFs in each row are measured for the
example STRFs in each row are measured for the

- A note for auditory neurophysiologists

rding site but using different LFP bands. same recording site but using different LFP bands.
same recording site but using different LFP bands.
indicate an increase in the neurophysi Red areas indicate an increase in the neurophysi
Red areas indicate an increase in the neurophysi
gnal following an increase in power of ological signal following an increase in power of

  - 120 ms latency is not as “crazy late” as it

ological signal following an increase in power of

ponding spectro-temporal stimulus fea the corresponding spectro-temporal stimulus fea the corresponding spectro-temporal stimulus fea

lue areas indicate a decrease. STRFs are ture, and blue areas indicate a decrease. STRFs are

ture, and blue areas indicate a decrease. STRFs are

d to have the same maximum absolute normalized to have the same maximum absolute

might seem

normalized to have the same maximum absolute

P variance STRFs are generally inhibi value. LFP variance STRFs are generally inhibi

value. LFP variance STRFs are generally inhibi

alpha and beta bands, variable (inhibi tory in the alpha and beta bands, variable (inhibi

- Even in primary auditory cortex (A1) of

tory in the alpha and beta bands, variable (inhibi

citatory) in the gamma band, and excit tory or excitatory) in the gamma band, and excit

tory or excitatory) in the gamma band, and excit

e high gamma band. MUA STRFs are atory in the high gamma band. MUA STRFs are

atory in the high gamma band. MUA STRFs are

excitatory. The peak latency of the LFP generally excitatory. The peak latency of the LFP

ferret, spectro-temporal receptive fields generally excitatory. The peak latency of the LFP

TRF is later in the alpha and beta bands

variance STRF is later in the alpha and beta bands

variance STRF is later in the alpha and beta bands

high gamma band. The frequency tun

than in the high gamma band. The frequency tun

(STRFs) made with speech stimuli

than in the high gamma band. The frequency tun

variance STRFs is generally similar to

ing of LFP variance STRFs is generally similar to

ing of LFP variance STRFs is generally similar to

MUA STRF but usually has additional

that of the MUA STRF but usually has additional

that of the MUA STRF but usually has additional

like the other signals, STRFs for mean

peaks. Unlike the other signals, STRFs for mean

have peaks with latency >100 ms

peaks. Unlike the other signals, STRFs for mean

peaks at multiple latencies. These

LFP show peaks at multiple latencies. These

LFP show peaks at multiple latencies. These

pically show an early peak ($25 ms

STRFs typically show an early peak ($25 ms

when made with Local Field Potential

STRFs typically show an early peak ($25 ms

ith negative polarity indicating depolar latency) with negative polarity indicating depolar

latency) with negative polarity indicating depolar

l t

l i

o v

w e

e p

d e

b a

y k

a (

p $

o 6

s 0

i m

s

1

ization, followed by a positive peak ($60 ms

A

ization, followed by a positive peak ($60 ms

(LFP), not spikes

ndicating hyperpolarization, and some IN latency) indicating hyperpolarization, and some 150

r longer-latency peaks.

latency) indicating hyperpolarization, and some times other longer-latency peaks.

times other longer-latency peaks.

IANCE

Ding et al. (2016) *Encoding of Natural Sounds by Variance of the Cortical Local Field Potential* J Neurophysiol-

Phonemic Responses in Humans

phoneme onset

phoneme

surprisal cohort

+

0.15

Noise
Non-word

Scrambled
Narrative

+

0.2

–

0 200 400 600

+

0.12

entropy

0 200 400 600

+

– –

MEG

0 200 400 600

• Clear evidence of phoneme-driven responses, uncorrelated with acoustics • Evidence of categorical neural processing of vocalization (speech) • Low-level phoneme processing at **~80 ms** (not much later than 60 ms) • Additional later processing at **~350 ms** with negative polarity •

N400-like, associated with predictive coding (Eddine et al., 2024)

Karunathilake et al. (2025) *Neural Dynamics of the Processing of Speech Features* … J Neurosci

Beyond Phonemes

- In human speech, phonemes building blocks of words
- Words and groups of words are used to convey

meaning • Animal vocalizations are often used to convey meaning

Vocalizations Convey Meaning

- In rhesus monkeys, some vocalizations transmit information regarding food quality

low-quality: “grunt”

high-quality: “harmonic arch” or “warble”

Bizley & Cohen (2013)

Grunt Harmonic arch Warble Baseline

Gifford et al. (2005) *The neurophysiology of functionally meaningful categories* … J Cog Neurosci

Responses to Meaningful Vocalizations

~~Neurons in monkey ventral prefrontal cortex~~ (VPFC) respond categorically based on meaning, not acoustics

- VPFC neurons encode transitions between calls of different abstract categories

- ~~VPFC neuron~~s do not encode

Bizley & Cohen (20~~1~~3)

transitions between acoustically
distinct stimuli transmitting the

est N~~o~~ise Exemplars reliably differ
emplars, the mean z-score value was not

ossibility that vPFC neurons ansitions between stimuli that ic classes (Ulanovsky, Las, &
ent than zero ($p > .05$). DISCUSSION

same information

Gifford et al. (2005) *The neurophysiology of functionally meaningful categories* … J Cog Neurosci

Humans Cortical Responses to Words

- Words often convey meaning in human speech

word-level features

envelope

Gammatone
Envelope

spectrogram onset

Gammatone
Envelope Onset

spectrogram

phoneme
Phoneme
onset
Onset

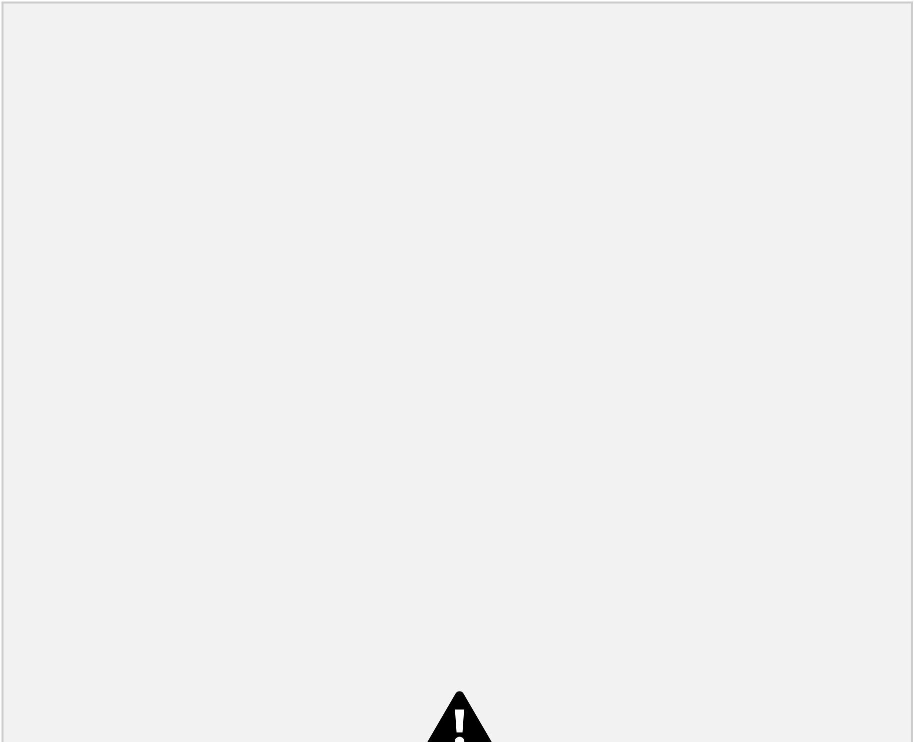
phoneme
Phoneme
surprisal
Surprisal

cohort
Cohort
entropy
Entropy

word
Word
Onset
onset

surprisal
Unigram
(no context) Surprisal

surprisal
GPT2
Surprisal
(GPT-2 model)



.
.
.
*

*

*

.

.

.

0 1

Measured Neural signals Predicted Neural signals

Speech Representations TRFs TRFs

Speech representations

Word O~~n~~sets

catalogue inner eye

The cat a log in a lie

Do we…

cattle login library

- Anticipate word boundaries based on context?
- Infer them later based on consistency?

eye

catalogue inner

R T The cat a log in a lie

login

library

cattle

Figure 1. Recognition of the phrase “The catalogue in a library,” as spoken by speaker of British English:

“The catalogue in a library”

/ðəkætəlɒgɪnəlaɪbrɪ]. The upper panel shows the competitive inhibition process that occurs among activated candidate words in an interactive-activation model, such as Shortlist A. Words that compete for the same stretch of input inhibit each other via direct, bidirectional inhibitory connections. Only a subset of the best-matching candidates is shown. The lower panel illustrates the path-based search through a word lattice used in automatic

Norris & McQueen, 2008

Word Surprisal (without context)

Frequency of words based on SUBTLEX

the
to
and
of
in

a
.

.

.

.

Word Surprisal (contextual)

yourself

a

out

it

that

one

if you happened to fnd
the

Word Responses in Humans

your
.
.
.

.

(via GPT-2)

onset + – Non-word Scrambled Narrative

surprisal (without context) + – MEG

+

0.2

word

0.1

Noise

0 200 400 600

0 200 400 600

- Clear evidence of word-driven responses, uncorrelated with acoustics
- Evidence of categorical neural processing of vocalization (speech) • Low-level phoneme processing at **~100 ms** (not much later than 80 ms) • Additional N400-like processing at **~450 ms**, c.f. predictive coding (Eddine et al., 2024) Karunathilake et al. (2025) *Neural Dynamics of the Processing of Speech Features* … J Neurosci

Contextual Word Surprisal Results

(without context) word surprisal

(contextual)

0.15 + +

word surprisal

Noise

+

0.15 + +

+

Narrative

—

Non-word

Scrambled

—

—

—

—

—

MEG

0 200 400 600 0 200 400 600

600 0 200 400 600

0 200 400 600 0

- Context-based surprisal is more robust than naive surprisal
- N400 like response in both predictors, c.f. predictive coding

(Eddine et al., 2024)

Karunathilake et al. (2025) *Neural Dynamics of the Processing of Speech Features* … J Neurosci

Neural Speech Processing Progression

- Cortical responses time-lock to emergent features from acoustics to context as incremental steps in the processing of speech input occur
- Phonemic and word-based cortical processing are categorical

Top-down Bottom-up Structured meaning

450

Word-based

- Contextual word surprisal not unrelated to semantics

processing models

Stimuli

100 80

Sub-Lexical **Phonemic**

- Long latency stages (consistent with top-down processing) in line with predictive

350 120

60

Acoustic

Lexical

Speech

0 0

time (ms) time (ms)

Karunathilake et al. (2025) *Neural Dynamics of the Processing of Speech Features* … J Neurosci

Application: Is Distorted Speech Intelligible?

- Even very clear speech may be unintelligible
- More common: very distorted speech may still be intelligible
- Can neural *categorical encoding of speech features* be used to determine when the brain processes speech sounds as intelligible?

Intelligibility Experimental Design

- Manipulate intelligibility but keep acoustics unchanged - Speech acoustics:

(a)

Vocoded speech

Clear speech Vocoded speech

speech clarity rating speech clarity ratingTrial 1

three-band noise
vocoded speech

- Intelligibility manipulated via priming

PRE CLEAN POST

Trial 36

~20 s ~20 s ~20 s Intelligibility rating (0-5)? Intelligibility rating (0-5)?

1 2 3 0

- Hypothesized intelligibility

4

measure(s)

2

“Slice an apple through at its equator, and you will find five small chambers arrayed in a perfectly symmetrical

0

Vocoded speech Clear speech 0

Frequency (kHz)

- word boundaries

Time (s)

-80

-120

1 2 3

Vocoded speech

0

1 2 3

-40

starburst—a pentagram."

Time (s)

Time (s)

Karunathilake et al. (2023) *Neural Tracking Measures of Speech Intelligibility*…, PNAS

Intelligibility Behavioral Results

(a) ***

Speech clarity **increases** from Pre condition to Post condition

2

Speech Clarity Rating

Intelligibility Rating

0

4

PRE POST

Pre Post Condition

Karunathilake et al. (2023) *Neural Tracking Measures of Speech Intelligibility*…, PNAS

0

40

0

Intelligibility Neural Results

CLEAN
POST 0 2
PRE
600
600
POST
PRE
4
CLEAN
LH
LH
1.5
0
6
***
LH
0.5
1
2 2
2
LH
Pre
PRE
**
***
PRE
*
0.06
0.06

- Word onset TRF shows both early (+)

**
1.5
1.5
*
POST
***
Early:
40-300 ms
POST

Early: 40-300 ms

and late (-) processing stages

- Only in

- Late

Post

Word onset TRF late

CLEAN

Clean 1

RH

Early: 40-300 ms

CLEAN

sponse

e

left hemispher

processin shows g stage

***

0 0.05

***

PRE

***

increases Pre→Post

RH

~~single subject~~

PRE

0

0

POST CLEAN

**

early 0

0.05

POST CLEAN

**

early

LH

~~Early: 40-300~~

~~ms~~

0 200 400 600

0 200 400 600

***

PRE

LH

***

***

left

right

POST

***

PRE

1

* *

MEG 0.5

0.5

0 0

PRE PRE

Late: 330-650 ms

POST

***

Late: 330-650 ms

Late: 330-650 ms

Late: 330-650 ms

larger change than early

LH RH

*

*

***

CLEAN

2

RH

change

late PRE

RH

PRE

CLEA

- Acoustic responses: no

0.06

0.06

1.5

POST CLEAN

**

~~Late: 330-650 ms~~

left right POST

PRE

PRE

CLEAN

0.06

0

0

- Word Surprisal response also increases

1

0

Pre→Post

Word~~-based Responses:~~

***

***

RH

0.5

0

POST

POST

- Neural

*

LH

*

**

**

0 200 400 600

***

***

ive meas ure of intelligibility PRE CLEAN POST PRE CLEAN PRE CLEAN

object POST PRE CLEAN POST CLEAN POST

Karunathilake et al. (2023) *Neural Tracking Measures of Speech Intelligibility*…, PNAS CLEAN

Early: 40-300 ms Late: 330-650 ms

Summary

Phil. Trans. R. Soc. B | Volume 375 | Issue 1789 | 6 January 2020

- Investigating *neural speech & language*

processing in humans has broader impacts:

- clinical applications

- animal communications

These priorities are:
- Promoting science and its benefi ts
- Recognising excellence in science
- Supporting outstanding science

What can animal communication teach us about human language?

- evolution of language?

- Providing scientifi c advice for policy
- Fostering international and global cooperation
- Education and public engagement

For further information on the Royal Society
The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG
T +44 20 7451 2500

- Categorical perception & **categorical neural**

W royalsociety.org

For further information on Philosophical Transactions of the Royal Society B
T +44 20 7451 2602

processing in vocalization/speech

E philtransb@royalsociety.org
W royalsocietypublishing.org/journal/rstb

- seen for speech: phonemes, words, …
- dissociable from acoustics
- provides new insight re: linguistics
- not available unless speech intelligible

0962-8436(20200106)375:1789

ISBN: 978-1-78252-429-8

ISSN 0962-8436

The Royal Society Registered Charity No 207043

ISSN 0962-8436 | Volume 375 | Issue 1789 | 6 January 2020

What can animal communication teach us about human language? Theme issue compiled and edited by Jonathan B. Fritz, William J. Idsardi and Gerald S. Wilkinson

These slides available at: ter.ps/simonpubs

thank you

https://cansl.umd.edu Mastodon: @jzsimon@fediscience.org