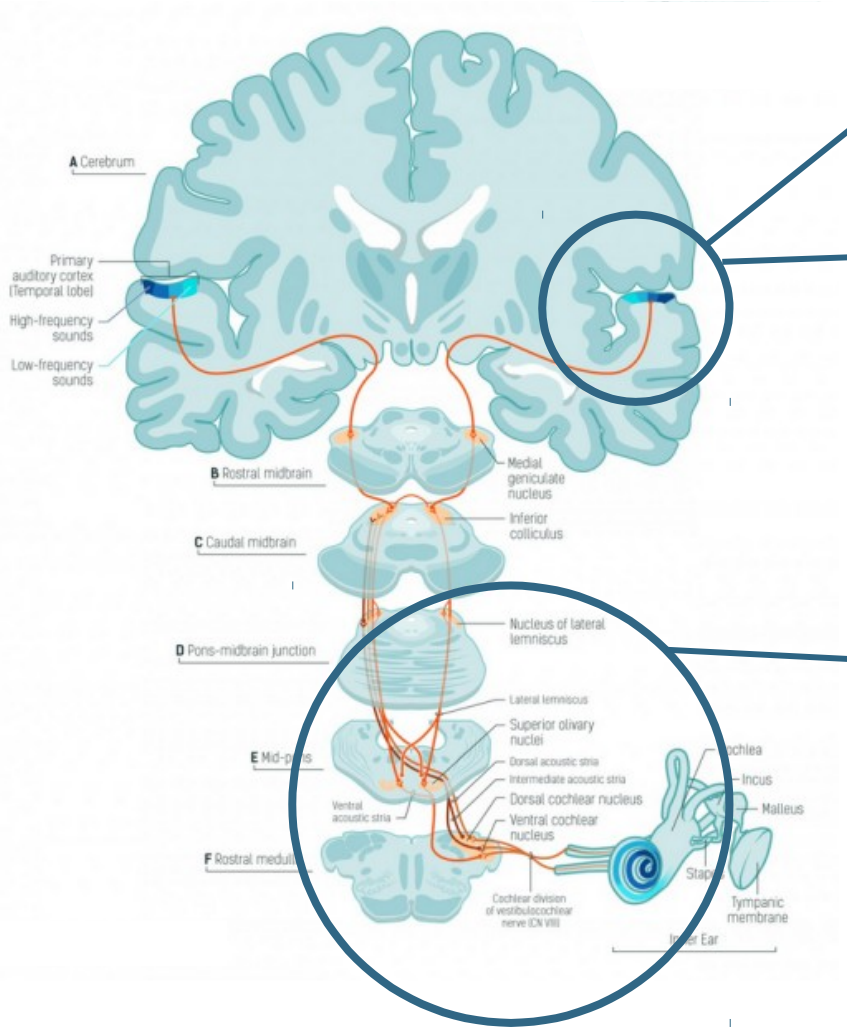
Neural markers of speech comprehension

@Marlies_Gillis

Marlies Gillis
ExpORL, KU Leuven

Go to:

https://drive.google.com/drive/folders/1v2hXcOYmBGDsneqUXAKO0EINGfmObrE7?usp=sharing



Linguistic representations

→ language network in brain

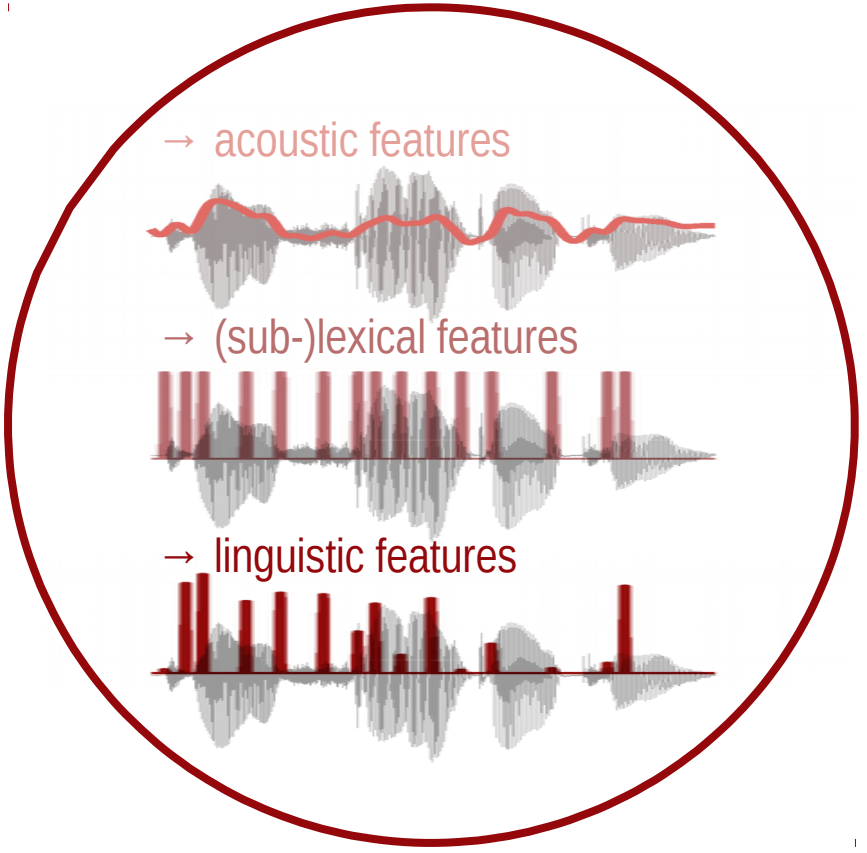
Envelope

→ neural processing of acoustic cues

Fundamental Frequency

→ subcortical stages of the auditory pathway

15 ms 100 ms 400 ms


→ acoustic features
→ (sub-)lexical features
→ linguistic features

1. Acoustic Features: Spectrogram & Acoustic Onsets

2. (Sub-)Lexical Features: Phoneme Onsets & Word Onsets

3. Linguistic Features

4. Determination of Linguistic Tracking – the Subtraction Approach

The data

Confirmatory Results

Follow this preprint

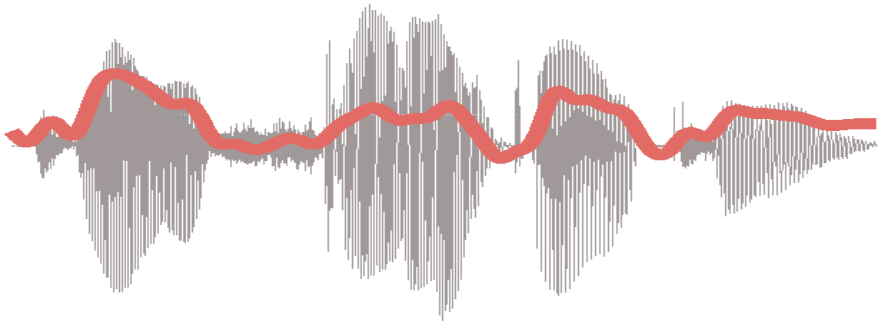
SparrKULee: A Speech-evoked Auditory Response Repository of the KU Leuven, containing EEG of 85 participants

Bernd Accou, Lies Bollens, Marlies Gillis, Wendy Verheijen, Hugo Van hamme, Tom Francart

doi: https://doi.org/10.1101/2023.07.24.550310

https://www.biorxiv.org/content/10.1101/2023.07.24.550310v1

What do we want?

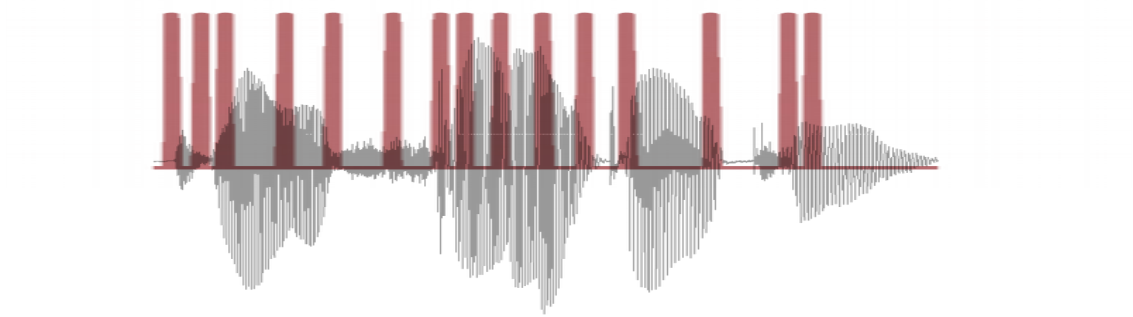


What do we need?

The audio file

To the code!

What do we want?



What do we need?

Transcript: which words were spoken during the audio fragment?

WhisperX
https://github.com/m-bain/whisperX
+ manual correction

TextGrid File: which words are spoken when and how?

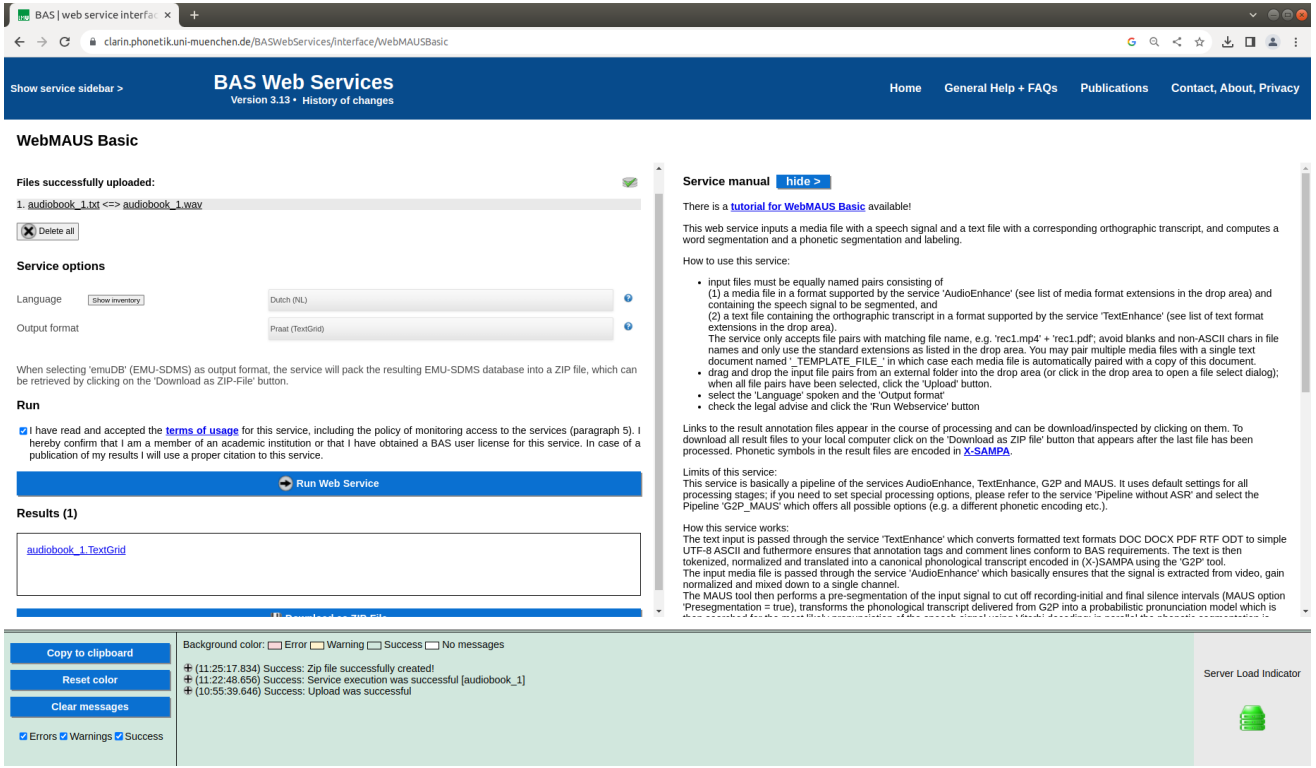
Web MAUS basic
https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic
+ manual correction (use PRAAT!)
https://www.fon.hum.uva.nl/praat/download_linux.html

Get that transcript!

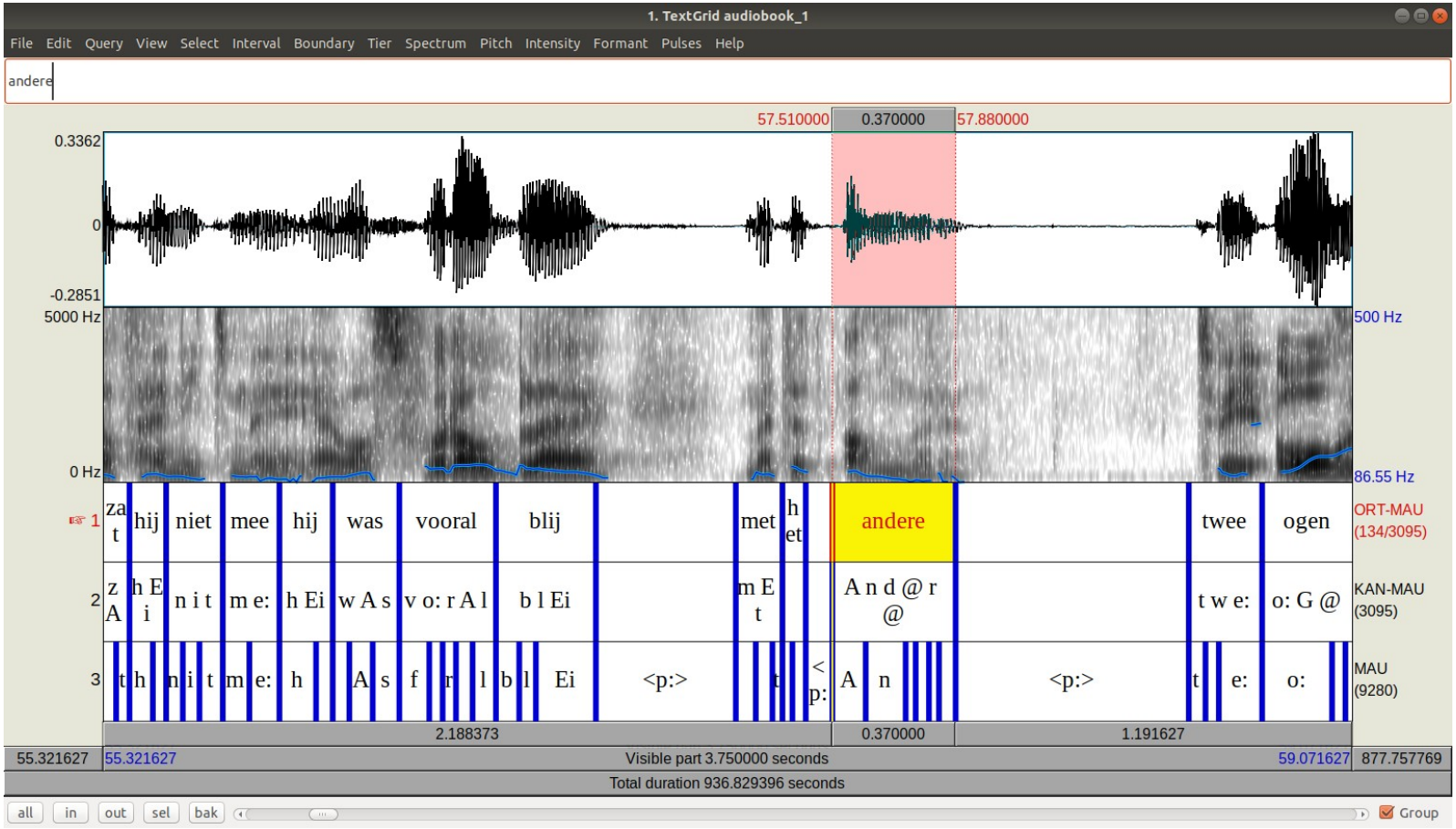
```
mgillis@fasso:~/python_makeTranscripts
File  Edit  View  Search  Terminal  Help
(makeTranscripts) [mgillis@fasso python_makeTranscripts]$ whisperx --language nl audiobook_1.wav
Performing transcription...
[00:00.000 --> 00:15.000]  Radio boeken, verhalen van Nederlandse en Vlaamse auteurs die u nergens kunt lezen,
[00:15.000 --> 00:25.000]  maar overal en altijd kunt beluisteren.
[00:25.000 --> 00:31.000]  Hij zag er helemaal uit als een normale jongen.
[00:31.000 --> 00:39.000]  Hij had armen, twee, benen, ook twee, met aan de uiteinde de nodige handen en voeten.
[00:39.000 --> 00:44.000]  En daar aan weer een heleboel vingers en tenen, telkens tien, om precies te zijn.
[00:44.000 --> 00:49.000]  Dan had hij ook een buik met zo'n gekke navel en schouders
[00:49.000 --> 00:52.000]  en daarop een hoofd met links en rechts zijn oor.
[00:52.000 --> 00:57.360]  zijn oor. Het linker een beetje flappig, maar daar zat hij niet mee. Hij was vooral blij met
[00:57.360 --> 01:03.080]  het andere. Twee ogen zoals het hoorde, een wipnoes ergens in het midden en vlak daaronder
[01:03.080 --> 01:09.040]  een nogal grote mond. Alles was er. Alles zat op de juiste plaats en alles werkte.
[01:09.040 --> 01:17.760]  Hij kon kijken en horen en proeven en ruiken en wandelen en zingen en hij heette...
[01:17.760 --> 01:28.360]  Tja, Jasper misschien, of Frederik, of Hans, want het kon ook Hans zijn, of Tom, of Pietrof,
[01:28.360 --> 01:35.120]  wie weet wel, Wannes, of Gert, of Bas, Bas kon ook, maar ook van Bas was hij eigenlijk
[01:35.120 --> 01:36.120]  niet zeker.
[01:36.120 --> 01:44.080]  Het jongetje waarvan verder alles werkte en klopte, wist maar niet wie hij was, en dat
[01:44.080 --> 01:46.960]  wat vond hij helemaal niet leuk.
[01:46.960 --> 01:52.040]  Hallo, vroeg hij soms al een voorbijganger die er voldoende vriendelijk en slim uitzacht.
[01:52.040 --> 01:56.760]  Mag ik jou even stooren, weet jij soms wie ik ben of hoe ik heet of waar ik vandaan kom
[01:56.760 --> 01:58.760]  en naartoe ga?
[01:58.760 --> 02:03.280]  De mensen keken hem dan dwaas aan en moesten lachen, zo'n vreemde vraag hadden ze nog nooit
[02:03.280 --> 02:04.280]  gehoord.
[02:04.280 --> 02:06.760]  Hoe kon niemand nu niet weten wie hij was?
[02:06.760 --> 02:08.760]  Dat was toch al te gek.
[02:08.760 --> 02:12.120]  Maar het jongetje vond dat niet zo gek.
[02:12.120 --> 02:15.120]  Het leek hem best wel handig te weten wie hij was.
[02:16.120 --> 02:20.120]  Op een keer, toen hij voor de zoveelste keer zijn vraag herhaalde,
[02:20.120 --> 02:23.120]  greep een vrouw hem uitbundig bij de armen.
[02:23.120 --> 02:25.120]  Ze leek op een gigantische pudding.
[02:25.120 --> 02:28.120]  Ze had een kapsel als een vogelnest en een onderkin zoals een kalkoen.
[02:28.120 --> 02:30.120]  Veel te groot.
[02:30.120 --> 02:35.120]  Maar, maar, maar toch, jij bent gewoon een jongetje.
```

Cool right?! It spares you lots of time. But **listen** and **correct** the transcript afterwards!
→ available for English, French, German, Spanish, Italian, Japanese, Dutch, ...

Get that phoneme segmentation!

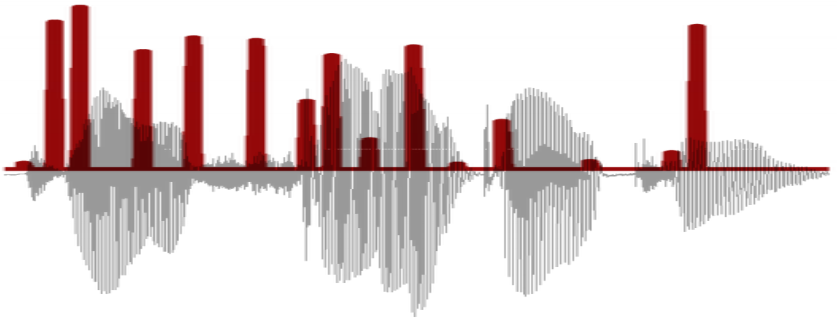


Get that phoneme segmentation!



To the code!

What do we want?

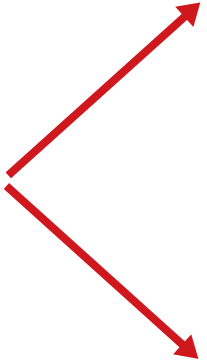


At phoneme level

Sound
/ba/

Activation of word cohort
Activated lexical items

- Baby
- Back
- Barbecue
- Basilisk
- Badminton
- Back-flip
- Bakery
- Bandit
- Bathtub
- Bald
- Bambino
- Bamboo
- Baggy
- Banana
- Basil
- ...



Phoneme surprisal

= - log(P(next phoneme | previous phonemes))

Cohort entropy

= Shannon entropy of phonemes compatible with the given input
→ "degree of competition"

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

Based on Brodbeck et al. (2018)
Probabilities derived from SUBTLEX-NL

What do we need?

Pronunciation dictionary: a mapping from the word to its phonemes

WebMAUS!
https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/

Word probabilities: the likelihood of a word

Subtlex!
https://www.ugent.be/pp/experimentele-psychologie/en/research/documents

At word level

"But you know, happiness can be found even in the darkest of times, if one only remembers to turn on the light."

— J.K. Rowling, Harry Potter and the Prisoner of Azkaban

Word surprisal

= - log(P(next word | previous 4 words))

Word frequency

= P(word)

Does not take sentence boundaries into account!

Based on Weissbart et al. (2019)
Derived from a 5-gram model

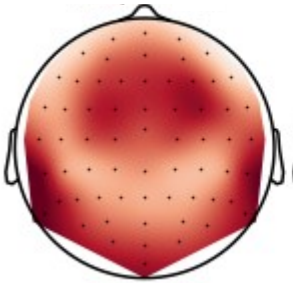
What do we need?

Ngram model: the likelihood of a word given the preceding words

KenLM
https://kheafield.com/code/kenlm/

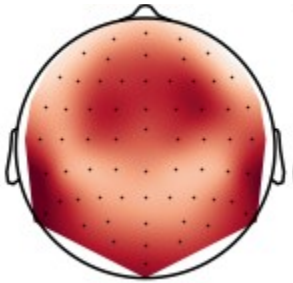
To the code!

Neural tracking of linguistic features



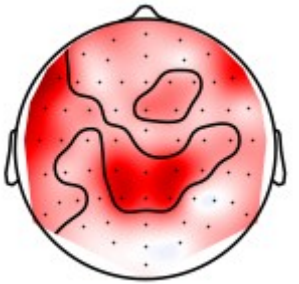
“complete model”

Acoustic features	v
(Sub-)Lexical features	v
Linguistic features	v



“baseline model”

Acoustic features	v
(Sub-)Lexical features	v
Linguistic features	x



difference in prediction accuracy

Acoustic features	x
(Sub-)Lexical features	x
Linguistic features	v

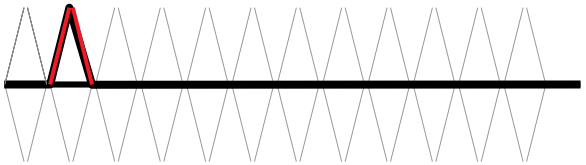
Prediction accuracy [Pearson’s r]

Difference in prediction accuracy [Pearson’s r]

How boosting works

→ TRF is estimated using a iterative procedure:

- Start with a 0-trf
  - A small amount is added to or subtracted from the first lag → MSE is calculated (repeated for all lags)
  - Best increment is the one that produces the largest decrease in the mean-squared error
- Repeated until an additional increment/decrement only adds noise



David, S. V., Mesgarani, N., & Shamma, S. A. (2007). Estimating sparse spectro-temporal receptive fields with natural stimuli. Network: Computation in neural systems, 18(3), 191-212.

To the code!

Let’s accelerate science **together**!

My contact detail:

Marlies Gillis (marlies.gillis@kuleuven.be // @Marlies_Gillis)

GitHub Repository (where you will find all the code, and more):

git@github.com:MarliesG/LinguisticSpeechTracking.git

Feel free to use the code, a citation would be appreciated ☺

References

- Own work, related to linguistic speech tracking:
  - Gillis, M., Vanthornhout, J., Simon, J. Z., Francart, T., & Brodbeck, C. (2021). Neural markers of speech comprehension: measuring EEG tracking of linguistic speech representations, controlling the speech acoustics. Journal of Neuroscience, 41(50), 10316-10329.
  - Verschueren, E., Gillis, M., Decruy, L., Vanthornhout, J., & Francart, T. (2022). Speech understanding oppositely affects acoustic and linguistic neural tracking in a speech rate manipulation paradigm. Journal of Neuroscience, 42(39), 7442-7453.
  - Gillis, M., Vanthornhout, J., & Francart, T. (2023). Heard or understood? Neural tracking of language features in a comprehensible story, an incomprehensible story and a word list. Eneuro, 10(7).
  - Gillis, M., Kries, J., Vandermosten, M., & Francart, T. (2023). Neural tracking of linguistic and acoustic speech representations decreases with advancing age. NeuroImage, 267, 119841.
- Other useful references:
  - Brodbeck, C., Bhattasali, S., Heredia, A. A. C., Resnik, P., Simon, J. Z., & Lau, E. (2022). Parallel processing in speech perception with local and global representations of linguistic context. Elife, 11, e72056.
  - Brodbeck, C., Das, P., Gillis, M., Kulasingham, J. P., Bhattasali, S., Gaston, P., ... & Simon, J. Z. (2021). Eelbrain: A Python toolkit for time-continuous analysis with temporal response functions. BioRxiv, 2021-08.
  - Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Current Biology, 28(5), 803-809.
  - Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Current Biology, 28(5), 803-809.
  - And many, many more