HHS Public Access

Author manuscript

Curr Opin Physiol. Author manuscript.

Continuous speech processing

Christian Brodbeck[1], **Jonathan Z. Simon**[1,2,3,*]

[1)]Institute for Systems Research, University of Maryland, College Park, Maryland 20742, U.S.A

[2)]Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742, U.S.A

[3)]Department of Biology, University of Maryland, College Park, Maryland 20742, U.S.A

Abstract

Speech processing in the human brain is grounded in non-specific auditory processing in the general mammalian brain, but relies on human-specific adaptations for processing speech and language. For this reason, many recent neurophysiological investigations of speech processing have turned to the human brain, with an emphasis on continuous speech. Substantial progress has been made using the phenomenon of "neural speech tracking", in which neurophysiological responses time-lock to the rhythm of auditory (and other) features in continuous speech. One broad category of investigations concerns the extent to which speech tracking measures are related to speech intelligibility, which has clinical applications in addition to its scientific importance. Recent investigations have also focused on disentangling different neural processes that contribute to speech tracking. The two lines of research are closely related, since processing stages throughout auditory cortex contribute to speech comprehension, in addition to subcortical processing and higher order and attentional processes.

Keywords

Reverse correlation; speech perception; speech envelope; temporal response function

Introduction

Speech is inherently a dynamic and non-repetitive acoustic stimulus. For human listeners in particular, repeated presentation fundamentally alters how speech is perceived and experienced. As a consequence, there are limitations to the extent to which the neural basis of human speech perception can be studied with traditional, trial based experimental designs. Advances in the reverse correlation technique for electroencephalography (EEG), magnetoencephalography (MEG), and electrocorticography (ECoG) have opened the possibility of studying brain responses to long duration, non-repetitive stimuli [1–5], such as

* jzsimon@umd.edu.

audiobooks, and have the potential to be applied to even more naturalistic stimulus materials. Additionally, continuous speech drives robust neural responses all along the auditory pathway, opening up the possibility of investigating multiple speech processing mechanisms, at different hierarchical levels, with the same speech stimuli.

Cortical speech tracking

Low frequency (< 10 Hz) cortical responses to continuous speech show a consistent phase relationship with the acoustic speech envelope [6], and the same holds for the slow envelope of the response at higher frequencies (gamma) [7]. This phenomenon is often referred to as "neural speech tracking" [8], and conceived of as an ongoing brain response related to moment-to-moment slow fluctuations in the amplitude of the speech envelope [1]. A simple quantitative measure of speech tracking can be obtained from stimulus reconstruction (Figure 1–A). A somewhat more differentiated model conceives of the neural responses as a continually evoked response to fluctuations in the speech envelope (Figure 1–B). The functions estimated to predict brain responses can be interpreted analogous to evoked responses and are often called the Temporal Response Functions (TRFs) [3]. In essence, both methods quantify the extent to which the brain response is linearly dependent on the speech envelope.

Because speech envelope tracking is so robust, it is a good candidate for use in clinical settings, especially as it employs an ecologically valid stimulus [9]. Envelope tracking, by definition, is a brain response closely related to an acoustic signal. Accordingly, for clean speech, presented in quiet, speech tracking often does not differ drastically between the native language and a language not spoken by the listeners [10]. However, because the envelope of speech is an important cue for speech intelligibility [11], speech tracking is used as a measure to assess whether pre-conditions for speech intelligibility are satisfied. For instance, for cochlear implant (CI) users it is of primary interest how well the modified speech signal is transmitted to the level of the cortex [12]. Speech tracking indeed may be causally related to processes required for successful comprehension, as transcranial alternating current stimulation (tACS) targeting the envelope of speech can negatively impact speech understating [13–15].

Yet, there is clearly no one-to-one correspondence between speech understanding and envelope tracking. For instance, even though older adults frequently complain of speech comprehension difficulty, cortical envelope tracking actually increases with advancing age [16–18]. An early observation was that speech tracking strength may correspond more to the perceived speech than simply reflecting the bottom up acoustic input. In responses to two talkers, the attended talker is often tracked more reliably than the ignored talker [19], and this modulation is robust enough to allow for detecting changes in the focus of attention in relatively short segments of data [20,21]. Here, envelope tracking thus measures how well the to-be attended speech is represented *despite the fact that it is different from the actual acoustic input signal*. Similarly, tracking even of clean speech is increased during periods in which attentional focus is high [22]. Such trial-by-trial variation in clean speech tracking has also been shown to reflect task performance, with better memory for words that occurred in sentences with higher speech tracking [23].

This raises the possibility that envelope tracking may reflect a sort of cleaned-up and attended-to representation of the acoustic input, which might form the basis for comprehension. For speech presented with different kinds of background noise, increased tracking of the attended envelope is associated with better speech understanding even after controlling for the objective background noise level [17]. Consistent with a strong top-down influence, tracking of the attended speech can actually be higher for speech in noise than for clean speech [24] and, for a well-known stimulus, tracking can even persist during short gaps in which the stimulus is replaced with pure noise [25]. In addition to this attentional enhancement, tracking of attended speech in noise differs qualitatively depending on whether the language is known to the listener [10,26], suggesting that speech tracking includes a language-specific component in addition to acoustic processing.

Envelope tracking thus likely reflects an interaction of the bottom-up input to the auditory cortex with resource-dependent, higher order processes. This is demonstrated by varying the amount of cognitive resources devoted to the speech [27]: At high signal to noise ratios (SNRs), speech tracking is similar, whether participants attend to the speech, or whether they ignore it and watch a silent movie instead. At lower SNRs, however, when more attentional resources would be required to recover the speech signal, speech tracking decreases much more in the movie condition. When subjects were playing a video game, speech tracking was even lower, decreasing even for clean speech. This suggests that speech tracking even of clean speech has a resource-dependent component, with increasing demands for speech in noise.

Components of speech tracking

The results summarized above suggest that, while the speech envelope is by definition an acoustic property of speech, considering speech tracking as a measure of basic acoustic processing is an oversimplification. A better understanding of speech processing requires disentangling representations of different properties of speech. One such dissociation can be gained by analyzing the temporal relationship between stimuli and responses using TRFs (Figure 1–B). For instance, when listening to two concurrent talkers, early (~50 ms) responses reflect the stimulus heard at the periphery, the acoustic mixture of the competing speech signals, whereas later (~100 ms) responses are dominated by a segregated version of the attended speaker [19,28–30]. These response components might also have a degree of task-dependency. For instance, under some challenging conditions, late representations may even specifically track the speech of the ignored talker [31]. Furthermore, not only is the speech envelope a collection of related acoustic speech features [32], but it is further modulated as carrier for linguistic units at different time scales from phonemes to phrases (as seen in Figure 2) [23]. Speech tracking thus likely reflects a family of representations at different hierarchical levels. Which of those representations exactly contribute to speech tracking probably differs across different stimuli and tasks. Tellingly, tACS at envelope frequencies selectively modulates neural activity in speech-specific, rather than general auditory brain regions [33], and brain responses that track the speech envelope are also found in areas outside of auditory cortex proper [4,34].

Auditory processing

For all these reasons, there is increasing interest in further disentangling the specific acoustic and linguistic features that drive the neural response through hypothesis-driven models. While the envelope is a useful summary variable, perhaps unsurprisingly cortical responses can be predicted more accurately when also considering other acoustic features. For example, a spectrogram, corresponding to the envelope of the acoustic signal computed separately for different frequency bands, reliably predicts brain responses better than the envelope alone [35,36]. A common variation on speech tracking uses a transformation of the envelope or spectrogram that emphasizes acoustic onsets [37], consistent with the observation that onsets are particularly important for intelligibility [38,39]. Used by itself, an acoustic onset spectrogram is indeed a better predictor than the envelope-based spectrogram [40], but both envelopes and onsets explain unique variability in the brain responses not explained by the other [35,41]. In addition, an anatomically localized region represents onsets not of local acoustic elements, but of larger acoustic groupings such as sentences and phrases [42]. Finally, a further level of complexity might come from non-linearities in responses to simple acoustic features, such as a modulation of the response to the envelope with absolute intensity [43].

Speech comprehension requires transformations of acoustic representations into speech-specific dimensions. For instance, brain responses are also modulated by the pitch contour of speech, which is an important component of prosody [44,45]. Because speakers differ in their fundamental voice pitch, the same acoustic pitch can have different linguistic implications for different speakers. Pitch thus needs to be normalized relative to the speaker to be interpreted linguistically. ECoG studies have shown such a speaker-dependent shift in response characteristics both in representations of vowels [46] and prosodic contours [44].

Linguistic processing

Eventually, acoustic representations are transformed into linguistic representations that are abstractions built upon the specific speech signal [47]. One approach to studying such representations is to predict brain responses from times-series of experimenter-coded linguistic features. A challenge for this approach is that, statistically, linguistic features can be highly correlated with the acoustic features used to communicate them. For example, each phoneme is defined as an equivalency class of related acoustic patterns. What might look like brain responses to categorical representations of phonemes, e.g., consonants, might actually be explained just by responses to acoustic onsets [35]. Few studies to date include a detailed acoustic model to control for acoustic representations; thus, results purporting to demonstrate sensitivity to linguistic features should be interpreted with care. One alternative approach to mitigate this issue is to decouple linguistic features from acoustic features by using fixed-rhythm speech [48,49], although this often still leaves ambiguity as to the specific linguistic features responsible for a certain response [50].

The approaches discussed so far are largely based on representations of specific stimulus features. Another approach that has been successful in linking brain activity to linguistic representations is though the predictive coding framework [51]. EEG studies of language processing have long used the N400, one of the most well-studies ERP components, as an

index of how surprising a word is in its linguistic context [52]. A similar response is found in continuous speech based on word-by-word measures of how surprising each word is in its context [53–55]. An advantage of jointly modeling responses to acoustic and semantic properties is that it makes it possible to assess interactions between the two [56]. Furthermore, while the N400 literature might some times imply that there is only a single kind of surprisal, estimates of surprisal associated with different aspects of language may affect different brain areas, suggesting potentially separable underlying mechanisms [57]. This opens up the possibility of distinguishing different neural processes by comparing the predictiveness of surprisal computed from different language models, in particular when taking advantage of more advanced language models developed in linguistics and computer science [58,59].

Phoneme processing

The predictive coding framework also applies at the level of phonemes [60]: Phonemes can be described as acoustic patterns, but they are also information carriers. Speech perception can be cast as information transmission, where the information carrying units are phonemes, and the goal is to identify words [61]. Measures of the informativeness of phonemes thus provide an index of lexical processing of speech, and such measures are predictive of brain responses to continuous speech, even when controlling for a complex model of acoustic processing [41,62]. Furthermore, these measures show a striking dissociation in responses to two concurrent talkers: while *acoustic* features from both talkers are represented neurally to some degree (see Cortical speech tracking above), time-locked lexical processing is strictly associated with the attended talker only [41].

Cortical speech tracking as “entrainment”

Another possible source of modulation for speech tracking comes from the brain’s own internal rhythms. Given that speech is rhythmic, speech tracking reflects the brain matching certain external rhythms. These rhythmic responses may be more than just time-locked neural responses to rhythmic features, but rather endogenous rhythms that phase shift to match and predict speech rhythms [8,63]. Phase-locked responses and entrainment are often hard to dissociate, since the main predictions of both are synchronization between speech and brain rhythms. A clear signature of entrainment, distinct from phase-locked responses, may require showing a dissociation of the neural rhythms from the speech rhythms. This might come in the form of neural oscillations that out-live the stimulus [64], particularly if localized to the narrow frequency band in which the neural oscillators operate [65].

Subcortical speech tracking

Subcortical signals have a much lower amplitude than cortical signals, and are traditionally assessed through averaging thousands of repetitions of identical stimuli [66]. More recent work found that reverse correlation can also recover the brainstem response from a non-repetitive stimulus, such as an audiobook, without any repetitions [67,68]. Besides making the measurement of brainstem responses more entertaining for participants, the ability to measure subcortical responses with naturalistic stimuli also creates new opportunities to study brainstem responses in more ecological tasks. For example, some research suggests

that selective attention to one out of two speakers modifies even brainstem representations of that speaker [68–70]. Such an effect is plausible given cortico-fugal connections [71], although it has not yet been replicated by other labs. Generally, this approach makes it possible to investigate cortical and subcortical responses to an ecologically valid stimulus concurrently in the same experiment.

Final thoughts

While there are many reasons for studying the neural processing of continuous, unrepeated speech, this is complicated by the fact that the acoustic speech signal is correlated in complex ways with the linguistic information it conveys. An advantage of the TRF approach is that it can model responses to acoustic and linguistic features jointly [4], and thus has some potential to decompose speech tracking into component neural mechanisms related to different aspects of speech processing. In addition, recent advances make it possible to study representations at multiple levels, from subcortical to semantic representations in the same dataset. This opens up new possibilities for studying different processing stages not just in isolation, but also for establishing connections and dependencies between processing at different stages. Understanding such connections might prove essential for clinical applications with a concern for speech comprehension “in the wild”.

Acknowledgements

This work was supported by National Institutes of Health grants R01-DC014085 and P01-AG055365, and National Science Foundation grant SMA-1734892

References

1. Lalor EC, Foxe JJ: Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. European Journal of Neuroscience 2010, 31:189–193.
2. David SV, Mesgarani N, Shamma SA: Estimating sparse spectro-temporal receptive fields with natural stimuli. Network: Computation in Neural Systems 2007, 18:191–212.
3. Ding N, Simon JZ: Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 2012, 107:78–89.
4. Brodbeck C, Presacco A, Simon JZ: Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. NeuroImage 2018, 172:162174.
5. Das P, Brodbeck C, Simon JZ, Babadi B: Neuro-current response functions: A unified approach to MEG source analysis under the continuous stimuli paradigm. NeuroImage 2020, 211:116528.
6. Luo H, Poeppel D: Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 2007, 54:1001–10.
7. Viswanathan V, Bharadwaj HM, Shinn-Cunningham BG: Electroencephalographic Signatures of the Neural Representation of Speech during Selective Attention. eNeuro 2019, 6:ENEURO.0057–19.2019.
8. Obleser J, Kayser C: Neural Entrainment and Attentional Selection in the Listening Brain. Trends in Cognitive Sciences 2019, 23:913–926.
9. Vanthornhout J, Decruy L, Wouters J, Simon JZ, Francart T: Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope. Journal of the Association for Research in Otolaryngology 2018, 19:181–191.
10. Etard O, Reichenbach T: Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise. J Neurosci 2019, 39:5750–5759.

11. Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M: Speech Recognition with Primarily Temporal Cues. Science 1995, 270:303–304.
12. Verschueren E, Somers B, Francart T: Neural envelope tracking as a measure of speech understanding in cochlear implant users. Hearing Research 2019, 373:23–31.
13. Riecke L, Formisano E, Sorger B, Başkent D, Gaudrain E: Neural Entrainment to Speech Modulates Speech Intelligibility. Current Biology 2017, doi:10.1016/j.cub.2017.11.033.
14. Wilsch A, Neuling T, Obleser J, Herrmann CS: Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. NeuroImage, 2018,172:766–774.
15. Keshavarzi M, Kegler M, Kadir S, Reichenbach T: Transcranial alternating current stimulation in the theta band but not in the delta band modulates the comprehension of naturalistic speech in noise. NeuroImage 2020, 210:116557.
16. Presacco A, Simon JZ, Anderson S: Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. Journal of Neurophysiology 2016, 116:2346–55.
17. Decruy L, Vanthornhout J, Francart T: Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. Journal of Neurophysiology 2019, 122:601–615.* Decruy et al.: In this EEG study, the authors characterize the relationship between cortical speech tracking and speech intelligibility, with a particular focus on the effect of healthy auditory aging on the speech tracking measure (speech envelope reconstruction).
18. Brodbeck C, Presacco A, Anderson S, Simon JZ: Over-Representation of Speech in Older Adults Originates from Early Response in Higher Order Auditory Cortex. Acta Acustica united with Acustica 2018, 104:774–777.
19. Ding N, Simon JZ: Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences of the United States of America 2012, 109:11854–9.
20. Miran S, Akram S, Sheikhattar A, Simon JZ, Zhang T, Babadi B: Real-Time Tracking of Selective Auditory Attention From M/EEG: A Bayesian Filtering Approach. Frontiers in Neuroscience 2018, 12.
21. O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, Slaney M, Shamma SA, Lalor EC: Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cerebral Cortex 2015, 25:1697–1706.
22. Lesenfants D, Francart T: The interplay of top-down focal attention and the cortical tracking of speech. Sci Rep 2020, 10:6922.
23. Keitel A, Gross J, Kayser C: Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLOS Biology 2018, 16:e2004473.
24. Lesenfants D, Vanthornhout J, Verschueren E, Decruy L, Francart T: Predicting individual speech intelligibility from the cortical tracking of acoustic- and phonetic-level speech representations. Hearing Research 2019, 380:1–9.
25. Cervantes Constantino F, Simon JZ: Restoration and Efficiency of the Neural Processing of Continuous Speech Are Promoted by Prior Knowledge. Front Syst Neurosci 2018, 12:56.
26. Zou J, Feng J, Xu T, Jin P, Luo C, Zhang J, Pan X, Chen F, Zheng J, Ding N: Auditory and language contributions to neural encoding of speech features in noisy environments. NeuroImage 2019, 192:66–75.
27. Vanthornhout J, Decruy L, Francart T: Effect of Task and Attention on Neural Tracking of Speech. Front Neurosci 2019, 13:977.
28. Puvvada KC, Simon JZ: Cortical Representations of Speech in a Multitalker Auditory Scene. J Neurosci 2017, 37:9189–9196.
29. O'Sullivan J, Herrero J, Smith E, Schevon C, McKhann GM, Sheth SA, Mehta AD, Mesgarani N: Hierarchical Encoding of Attended Auditory Objects in Multi-talker Speech Perception. Neuron 2019, doi:10.1016/j.neuron.2019.09.007.** O'Sullivan et al.: An ECoG analysis of responses to continuous speech using a pair of competing talkers. The authors find that primary auditory cortex represents the individual talkers regardless of the subject's focus of attention, but nonprimary auditory cortex dominantly represents the attended talker.

30. Paul BT, Uzelac M, Chan E, Dimitrijevic A: Poor early cortical differentiation of speech predicts perceptual difficulties of severely hearing-impaired listeners in multi-talker environments. Sci Rep 2020, 10:6141.

31. Fiedler L, Wöstmann M, Herbst SK, Obleser J: Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. Neuroimage 2019, 186:33–42.

32. Ding N, Simon JZ: Cortical entrainment to continuous speech: functional roles and interpretations. Frontiers human Neuroscience 2014, 8.

33. Zoefel B, Archer-Boyd A, Davis MH: Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech. Current Biology 2018, 28:401–408.

34. Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, et al.: Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party." Neuron 2013, 77:980–991.

35. Daube C, Ince RAA, Gross J: Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech. Current Biology 2019, 29:1924–1937.e9.* Daube et al.: A careful analysis of MEG responses to continuous speech, which shows that neural responses previously attributed to a categorical representations of phonemes can also be explained as responses to non-linguistic, acoustic features.

36. Di Liberto GM, O'Sullivan JA, Lalor EC: Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Current Biology 2015, 25:2457–2465.

37. Fiedler L, Wöstmann M, Graversen C, Brandmeyer A, Lunner T, Obleser J: Single-channel in- ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. J Neural Eng 2017, 14:036020.

38. Koning R, Wouters J: The potential of onset enhancement for increased speech intelligibility in auditory prostheses. The Journal of the Acoustical Society of America 2012, 132:2569–2581.

39. Stilp CE, Kluender KR: Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. PNAS 2010, 107:12387–12392.

40. Oganian Y, Chang EF: A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci Adv 2019, 5:eaay6279.

41. Brodbeck C, Hong LE, Simon JZ: Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. Current Biology 2018, 28:3976–3983.e5.** Brodbeck et al.: An MEG analysis of responses to continuous speech, showing responses related to lexical processing based on the information transmission framework; In the presence of two speakers, such lexical responses are found for attended but not for unattended speech.

42. Hamilton LS, Edwards E, Chang EF: A Spatial Map of Onset and sustained Responses to Speech in the Human Superior Temporal Gyrus. Current Biology 2018, 28:1860–1871.e4.* Hamilton et al.: Data-driven investigation of ECoG responses to sentences that reveals a posterior superior temporal site with a sharp response at the onset of sentences and phrases, and relatively decreased responses during the rest of the stimuli. Interpreted as possibly providing a temporal reference frame for structuring speech input.

43. Drennan DP, Lalor EC: Cortical Tracking of Complex Sound Envelopes: Modeling the Changes in Response with Intensity. eNeuro 2019, 6:ENEURO.0082–19.2019.

44. Tang C, Hamilton LS, Chang EF: Intonational speech prosody encoding in the human auditory cortex. Science 2017, 357:797–801.

45. Teoh ES, Cappelloni MS, Lalor EC: Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. Eur J Neurosci 2019, 50:3831–3842.

46. Sjerps MJ, Fox NP, Johnson K, Chang EF: Speaker-normalized sound representations in the human auditory cortex. Nat Commun 2019, 10:2465.

47. Wilson SM, Bautista A, McCarron A: Convergence of spoken and written language processing in the superior temporal sulcus. NeuroImage 2018, 171:62–74.

48. Ding N, Melloni L, Zhang H, Tian X, Poeppel D: Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci 2016, doi:10.1038/nn.4186.

49. Ding N, Pan X, Luo C, Su N, Zhang W, Zhang J: Attention Is Required for Knowledge-Based Sequential Grouping: Insights from the Integration of Syllables into Words. J Neurosci 2018, 38:1178–1188.
50. Frank SL, Yang J: Lexical representation explains cortical entrainment during speech comprehension. PLOS ONE 2018, 13:e0197304.
51. Clark A: Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences 2013, 36:181–204.
52. Kutas M, Federmeier KD: Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annual Review of Psychology 2011, 62:621647.
53. Broderick MP, Anderson AJ, Liberto GMD, Crosse MJ, Lalor EC: Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. Current Biology 2018, 28:803–809.e3.** Broderick et al.: An EEG analysis of responses to continuous speech, showing an N400-like response associated with how well each word matches its semantic context, but only for attended speech and not unattended.
54. Weissbart H, Kandylaki KD, Reichenbach T: Cortical Tracking of Surprisal during Continuous Speech Comprehension. Journal of Cognitive Neuroscience 2020, 32:155–166.
55. Koskinen M, Kurimo M, Gross J, Hyvärinen A, Hari R: Brain activity reflects the predictability of word sequences in listened continuous speech. NeuroImage 2020, 219:11 6936.
56. Broderick MP, Anderson AJ, Lalor EC: Semantic Context Enhances the Early Auditory Encoding of Natural Speech. J Neurosci 2019, 39:7564–7575.
57. Frank SL, Willems RM: Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. Language, Cognition and Neuroscience 2017, 32:1192–1203.
58. Brennan JR, Hale JT: Hierarchical structure guides rapid linguistic predictions during naturalistic listening. PLOS ONE 2019, 14:e0207741.
59. Brennan JR, Dyer C, Kuncoro A, Hale JT: Localizing syntactic predictions using recurrent neural network grammars. Neuropsychologia 2020, doi:10.1016/j.neuropsycholo-gia.2020.107479.
60. Gagnepain P, Henson RN, Davis MH: Temporal Predictive Codes for Spoken Words in Auditory Cortex. Current Biology 2012, 22:615–621.
61. Shannon CE: A Mathematical Theory of Communication. Bell System Technical Journal 1948, 27:379–423,623–656.
62. Donhauser PW, Baillet S: Two Distinct Neural Timescales for Predictive Speech Processing. Neuron 2020, 105:385–393.e9.
63. Meyer L, Sun Y, Martin AE: Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. Language, Cognition and Neuroscience 2019, doi:10.1080/23273798.2019.1693050.
64. Kösem A, Bosker HR, Takashima A, Meyer A, Jensen O, Hagoort P: Neural Entrainment Determines the Words We Hear. Current Biology 2018, 28:2867–2875.e3.
65. Haegens S: Entrainment revisited: a commentary on Meyer, Sun, and Martin (2020). Language, Cognition and Neuroscience 2020, doi:10.1080/23273798.2020.1758335.
66. Coffey EBJ, Nicol T, White-Schwoch T, Chandrasekaran B, Krizman J, Skoe E, Zatorre RJ, Kraus N: Evolving perspectives on the sources of the frequency-following response. Nat Commun 2019, 10:5036.
67. Maddox RK, Lee AKC: Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. eNeuro 2018, 5:ENEURO.0441–17.2018.** Maddox & Lee: The authors adapt the Temporal Response Function framework from auditory cortex to EEG auditory brainstem responses to continuous speech. They show that the TRF can be interpreted analogously to a click-evoked response, while the approach allows simultaneous investigation of cortical activity to continuous speech.
68. Forte AE, Etard O, Reichenbach T: The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. eLife Sciences 2017, 6:e27203.
69. Etard O, Kegler M, Braiman C, Forte AE, Reichenbach T: Decoding of selective attention to continuous speech from the human auditory brainstem response. NeuroImage 2019, 200:111.

70. Saiz-Alia M, Forte AE, Reichenbach T: Individual differences in the attentional modulation of the human auditory brainstem response to speech inform on speech-in-noise deficits. Sci Rep 2019, 9:14131.

71. Khalfa S, Bougeard R, Morand N, Veuillet E, Isnard J, Gunot M, Ryvlin P, Fischer C, Collet L: Evidence of peripheral auditory activity modulation by the auditory cortex in humans. Neuroscience 2001, 104:347–358.

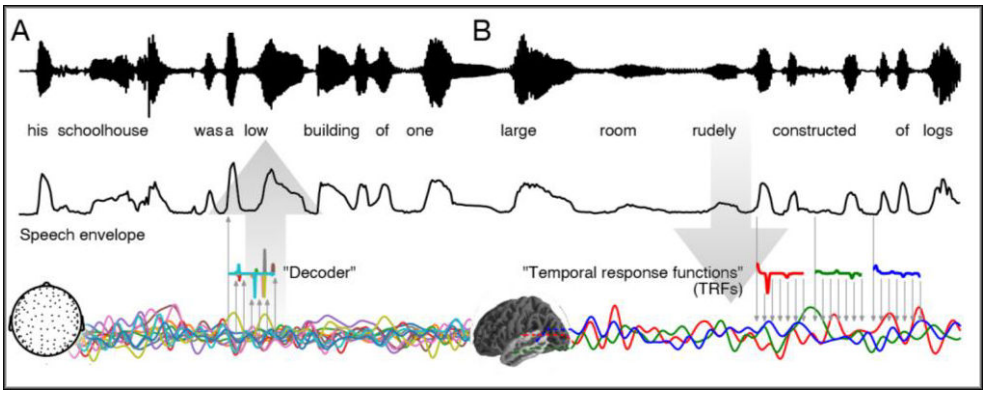

Figure 1.

Models for analyzing speech tracking. A) Stimulus reconstruction (backward model): a decoder is trained to reconstruct the stimulus envelope from the neural response, and speech tracking is quantified by how well the reconstructed envelope matches the actual envelope. A typical decoder uses a linear combination of the neural responses in a window following the envelope by 0 – 500 ms. B) Temporal response functions (TRFs) (forward model): a TRF is trained to predict the neural response from the speech envelope, and speech tracking is quantified by how well the predicted response matches the actual response. A typical TRF uses various delayed versions of the envelope from 0 – 500 ms. Responses originating from different brain areas are each characterized by their own TRF.

Figure 2.
Example of speech as a carrier for linguistic units at different time scales from phonemes to phrases. From [23], used with permission.