Paper Review: Cortical oscillations and speech processing: emerging computational principles and operations[1]

Catherine Zeng

Preface

The goal of reading this paper was to get clues on how to transform continuous speech into discrete code (the neural encoding of language). The gist of the paper is that neuronal oscillations in auditory cortex are useful for deciphering continuous speech information.

Frequency composition of speech

In speech, there are perceptual units of analysis at different time scales. For example, short-duration cues and information with a high modulation frequency at 30-50 Hz contain information at the phonemic scale such as formant transitions (/ba/ vs. /da/) and the coding of voicing. A magnitude slower at 4-7 Hz correlates with syllable rate. An even lower modulation rate at 1-2 Hz corresponds with signal input into lexical and phrasal units, i.e. intonation contour of an utterance.

Hypothesis

Neuronal oscillations are critical for parsing and decoding connected speech. Low gamma (25 - 35 Hz), theta (4-8 Hz), and delta (1-3 Hz) bands correspond to sub(phonemic), syllabic, and phrasal processing respectively. The paper focuses on theta and gamma bands and proposes that the brain can decode extremely impoverished speech if the syllabic rhythm is maintained through phase-locking and nested theta-gamma oscillations. Oscillations in auditory cortex interact with the neuronal activity generated by an incoming speech signal.

Theory of oscillation-based operations in speech perception

Their model segments connected speech at different timescales which allows reading out discrete phonemic and syllabic units (word segmentation!). The steps of their model are listed below:

1. Salient points in the input signal cause phase resetting of oscillations in the auditory cortex in the theta band.
2. The activity in the theta band tracks envelope of stimulus.
3. Theta reset causes a transient pause and then reset of gamma oscillations. The theta and gamma generators become more strongly coupled and nested.
4. Activity in the gamma band has a coupled relation to spike trains, regulating spike patterns.
5. Acoustic structure of the input is aligned with neuronal excitability.

The theta and gamma oscillations act (i) by discretizing (sampling) the input spike trains to generate units of the appropriate temporal granularity for subsequent processing and (ii) by creating packages of spike trains and excitability cycles. Speech trigger cycles of neuronal encoding at embedded syllabic and phonemic scales.

Speech resets oscillatory activity that is already visible at rest but only in specific frequency domains corresponding to the sampling rates for phonemic and syllabic sampling. Phase resetting of cortical oscillations.

Speech is analyzed at discontinuous time scales. This allows oscillatory nesting, the process when phase of slow cortical oscillations controls higher rate oscillations. Through theta-gamma nesting, concurrent syllabic and phonemic analysis and remain hierarchically bound. In the theta-gamma nesting pattern, there is a frequency ratio of 4; 4 cycles of the higher frequency occur during one cycle of the lower one.

Spike patterning & discretization

Proposal that spiking is hierarchically controlled by cortical oscillations. The Pyramidal-interneuron gamma (PING) network is a model of brain oscillations that generates clustered spikes at a gamma rate. Output spiking is temporally structured by stimulus-induced oscillatory activity, and cognitive operations depend on spike timing and alignment of spikes with phase of oscillations.

They argue that a consequence of the modulation of vertical circuits by gamma and theta oscillators is the organization of spike timing and the discretization of the cortical output. The output signal is chunked by periodic modulation of the firing likelihood.

Alignment of neuronal excitability with speech modulations

How the periodicity in output neuronal excitability aligns with the stimulus is an open question. The proposal is that the phases of high neuronal excitability in superficial layers coincide with the time periods when the most energetic parts of the speech signal reach layer IV, and they developed a model of coupled theta and gamma oscillations where the theta-oscillating pyramidal-interneuron networks control gamma oscillating dyads through an excitatory connection.

Speech modulations (5-10 Hz) cause a discharge in theta neurons, which then track the modulations in speech even when they are not fully periodic, or faster than the intrinsic theta rate. The excitation of PING by the pyramidal-internueron theta networks sets a period of excitability that lasts about three or four gamma cycles, the minimum duration of a syllable. Release of excitation from PINT neurons between syllables resets gamma oscillations, which enables time-locking of the gamma and output cells to the next syllable. For each gamma cycle, output neurons may fire or not fire, which is a binary code reflecting the shape of the speech envelope.

Dysfunctional oscillatory processes

Asymmetric sampling: the rates at which the incoming signal is sampled are laterally distributed. The hypothesis is that gamma sampling dominates in left auditory cortex, theta sampling is assumed to dominate in right auditory cortex.

Evidence for their hypothesis is to show that knocking out oscillatory mechanisms cause speech and language impairment. If people with dyslexia parse speech at a frequency higher or lower than usual low gamma rate, their phonemic representations should be in an idiosyncratic format because phonemic units would be either under-sampled or over-sampled.

They observed that the left-dominant response around 30 Hz present in subjects with normal reading ability is absent in those with dyslexia. They had a strong response at this frequency in right auditory cortex and therefore an abnormal asymmetry. This correlates with behaviors such as non-word repetition and rapid automatic naming. Readers with dyslexia had a stronger resonance than controls in left and right auditory cortices at frequencies between 50 and 80 Hz, suggesting that these subjects had phonemic oversampling.

Conclusion

The hypothesis is that cortical oscillations provide ways to temporally organize the incoming speech signal are critical for speech perception. Two prerequisites for this to happen are phase-locking between stimulus and cortex in at least two discrete time domains and the hierarchical coupling of related cortical oscillations during speech processing.

References

[1] A.-L. Giraud and D. Poeppel. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature neuroscience*, 15:511–7, 03 2012. doi: 10.1038/nn.3063.