Perceptual chunking of spontaneous speech: Linguistic cues and cognitive constraints

Svetlana Vetchinnikova[1], Alena Konina[2], Nitin Williams[2,3], Nina Mikušová[2], Anna Mauranen[2]
[1]Helsinki Collegium for Advanced Studies, University of Helsinki, svetlana.vetchinnikova@helsinki.fi,
[2]Department of Languages, University of Helsinki, [3]Department of Neuroscience and Biomedical Engineering, Aalto University

Keywords: chunking, speech perception, prosody, syntax, individual variation, cue degeneracy

Chunking has been proposed as an underlying mechanism of processing speech in real-time (Sinclair & Mauranen 2006; Christiansen & Chater 2016). Just what the chunks are is open to debate. The linguistic tradition strongly suggests that chunks are some kind of form-meaning pairings. For example, Christiansen and Chater (2016) suggest that processing chunks determined by the memory constraint are essentially the same chunks we learn during language acquisition and the same chunks in which language change proceeds. In contrast, neuroscience links chunking to neural oscillatory activity at different frequency bands and its possible alignment with corresponding linguistic information at different levels of language organization, from syllables up, providing optimal information processing (e.g. Giraud & Poeppel 2012). The "phrasal" delta frequency band seems to associate with prosodic (Inbar, Grossman & Landau 2020; Stehwien & Meyer 2021) and/or syntactic units (Ding et al. 2016; Kaufeld, Bosker & Martin 2020). Some studies suggest that chunking is driven by timing (Roll et al. 2012) or delta-band oscillations themselves (Henke & Meyer 2021). To contribute to this debate, we examine the processing of linguistic cues as they naturally occur in spontaneous speech and ask which linguistic cues and cognitive constraints have an effect on real-time chunking.

While neurophysiological studies commonly employ short constructed stimuli modelled on written language, we selected 97 short extracts from spoken corpora and re-recorded them with a trained speaker to achieve uniform audio quality. We then asked 50 experiment participants to listen to the extracts and intuitively mark chunk boundaries in the accompanying transcripts through a custom-built tablet application *ChunkitApp* (Vetchinnikova, Mauranen & Mikušová 2017; Vetchinnikova et al. 2022; https://www.chunkitapp.online/). Next, we annotated all spaces between every two words for pause duration, prosodic boundary strength, clausal syntactic structure, chunk duration and bigram surprisal and entered them as predictors of chunk boundary perception in mixed effects logistic regression models with random effects for listeners and extracts.

We found that in chunking up speech listeners used a variety of cues across different levels of language organization in an integrated manner which supports non-modular approaches to language. The presence of multiple cues which perform the same function also indicates cue degeneracy which is typical of biological systems (Winter 2014). Cue degeneracy in its turn supports extensive variation which we observed in listener preferences for different cues and in the extent to which they tracked them as well as in the reliability of the cues across different speech materials. Chunk duration had a strong effect, supporting the cognitive constraint hypothesis. The effect of surprisal did not support the hypothesis that perceptual chunks were multi-word form-meaning pairings: chunk-final words tended to be less predictable while chunk-initial words tended to be more predictable suggesting that chunking speech into perceptual groups is different from statistical learning of multi-word units from the input. Together these results suggest that perceptual chunking is a distinct process: to overcome the limitations of working memory, humans not only combine items into larger units for future retrieval (usage-based chunking), but also partition incoming stream into temporal groups (perceptual chunking).

References

Christiansen, Morten H. & Nick Chater. 2016. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*. Cambridge University Press 39. https://doi.org/10.1017/S0140525X1500031X.

Ding, Nai, Lucia Melloni, Hang Zhang, Xing Tian & David Poeppel. 2016. Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience* 19(1). 158–164. https://doi.org/10.1038/nn.4186.

Giraud, Anne-Lise & David Poeppel. 2012. Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience* 15(4). 511–517. https://doi.org/10.1038/nn.3063.

Henke, Lena & Lars Meyer. 2021. Endogenous oscillations time-constrain linguistic segmentation: Cycling the garden path. *Cerebral Cortex* 31(9). 4289–4299. https://doi.org/10.1093/cercor/bhab086.

Inbar, Maya, Eitan Grossman & Ayelet N. Landau. 2020. Sequences of Intonation Units form a ~ 1 Hz rhythm. *Scientific Reports* 10(1). 15846. https://doi.org/10.1038/s41598-020-72739-4.

Kaufeld, Greta, Hans Rutger Bosker & Andrea E Martin. 2020. Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *The Journal of Neuroscience* 40(49). 9467–9475.

Roll, Mikael, Magnus Lindgren, Kai Alter & Merle Horne. 2012. Time-driven effects on parsing during reading. *Brain and Language* 121(3). 267–272. https://doi.org/10.1016/j.bandl.2012.03.002.

Sinclair, John & Anna Mauranen. 2006. *Linear unit grammar integrating speech and writing*. New York: John Benjamins.

Stehwien, Sabrina & Lars Meyer. 2021. Rhythm comes, rhythm goes: Short-term periodicity of prosodic phrasing. PsyArXiv. https://doi.org/10.31234/osf.io/c9sgb.

Vetchinnikova, Svetlana, Alena Konina, Nitin Williams, Nina Mikušová & Anna Mauranen. 2022. Perceptual chunking of spontaneous speech: Validating a new method with non-native listeners. *Research Methods in Applied Linguistics* 1(2). 100012. https://doi.org/10.1016/j.rmal.2022.100012.

Vetchinnikova, Svetlana, Anna Mauranen & Nina Mikušová. 2017. ChunkitApp: Investigating the Relevant Units of Online Speech Processing. In *INTERSPEECH 2017 – 18th Annual Conference of the International Speech Communication Association*, 811–812. Stockholm.

Winter, Bodo. 2014. Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays* 36(10). 960–967. https://doi.org/10.1002/bies.201400028.