

## The distribution of prominent syllables on YouTube – A case study

Stephanie Berger<sup>1</sup>

<sup>1</sup>*Linguistics and Phonetics, ISFAS, Kiel University*

sberger@isfas.uni-kiel.de

This study investigates the distribution of prominent syllables and their use throughout a YouTube video. Prominent syllables stand out from the rest of an utterance, for example triggered by higher or later pitch movements, higher acoustic energy, or longer sound duration [1]. One of the uses of prominent syllables includes drawing in listeners' attention by focusing specific parts of an utterance, often those that the speaker deems important [2]. In particular, previous research on American English, Dutch and German has shown that content words (i.e., words with generally more informational importance) are more likely prominent, both in perception and production [1, 3], and this applies specially to nouns and adjectives [4, 5]. Different intonation annotation systems like DIMA [6] assume different levels of prominence, distinguishing no prominence, weak prominence (based on general rhythmicity), strong prominence (based on distinct pitch movements), and emphatic prominence (based on extreme pitch movements and segmental hyper-articulation).

Five minutes of speech material from a YouTube video are used as a case study with a male American English speaker [7]. YouTubers tend to use a large number of strong or even emphatic prominences in their speech [8], which was also noticed by popular media outlets [9].

Keeping this in mind, this study investigates how prominences of different prominence levels are distributed in a video. It is expected that there are higher prominence levels occurring more frequently towards the beginning of the video as a means for expressive speech to draw in an audience, but the topic of each section is also considered. Content words like adjectives and nouns are predicted to receive more prominences and at higher salience levels, as previous research suggests [1, 3, 4, 5].

The first five minutes of the chosen video [7] were annotated following the DIMA system [6]. The material was classified into larger sections of cohesive content as the video progresses. The words were classified as different parts of speech using the CLAWS web tagger [10], manually corrected and sorted into more general groups (adjectives, adverbs, determiners, nouns including proper names, prepositions, pronouns, verbs in all forms). Syllables at phrase-final boundaries were excluded to avoid effects of phrase-final lengthening. The analysis was carried out in R [11].

Overall, there are more non-prominent than prominent syllables, and always more non-weak than weak prominent syllables. The majority of prominent syllables in the first 90 seconds of the video is strongly or emphatically prominent (see Figure 1). The frequency of different prominence levels varies more depending on topic than video progress. A topic cluster analysis revealed a relationship between topic and the type and frequency of prominences. The topic clusters in Figure 2 are based on the frequency of the different prominence levels, normalized by total syllables per topic. The two clusters on the left have in common that more of the prominent syllables in these sections are strong or emphatic compared to other sections. They differ in that the cluster on the left of the two (T09, T04, T07) has more emphatic prominences than the other. The cluster on the right of Figure 2 has higher ratios of prominent (compared to non-prominent) syllables than other topics. This cluster also has a higher number of emphatic accents, which is a similarity to T06 (which otherwise has few strong prominences). The middle cluster seems only loosely connected by frequency of prominent syllables and content.

The analysis of part of speech suggests a relationship with prominence type/frequency. Figure 3 shows that adjectives receive the highest proportion emphatic prominences. Additionally, adjectives, nouns, pronouns, and verbs – i.e., the content words included in the sample – seem to have the lowest frequency of weak prominences. That is in line with expectations and previous research [1, 3, 4, 5]. The high proportion of emphatic prominences in adjectives likely occurs because more adjectives have prominent rather than non-prominent syllables compared to other parts of speech like nouns and verbs.

The high number of both emphatic and strong prominences is in line with prior work and may be related to speech-to-text systems used for automatic generation of closed captions which in turn get fed into search engines as search terms in YouTube. It is reasonable to assume that stronger prominences, especially those that may also include segmental hyper-articulation, may be easier to understand also for systems which in turn could increase the probability of video recommendations [12]. Further analysis could check if there are higher prominence levels used when the lexical fields of the words match the topic of the vlog, as that might suggest a conscious choice to help recommendations.

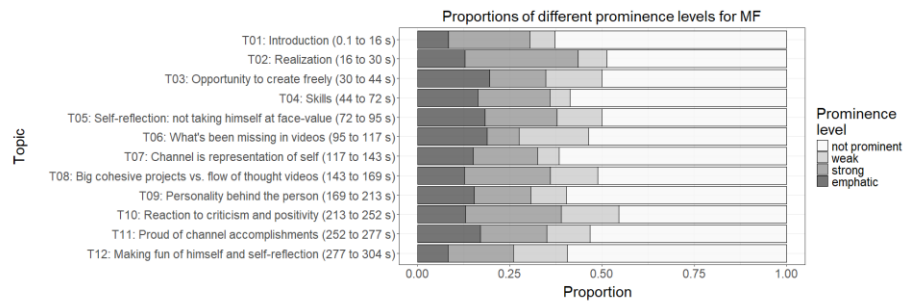


Figure 1: Proportions of syllables of different prominence levels occurring in different topic sections within the first 5 minutes of MF's video. The beginning of the video is depicted at the top of the figure.

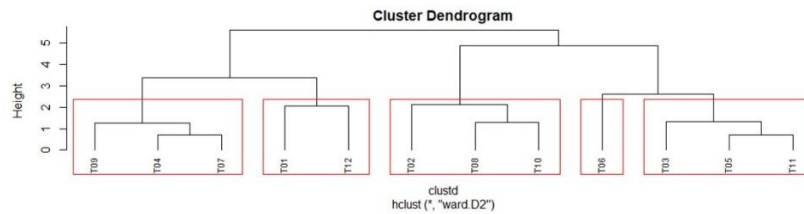


Figure 2: Cluster dendrogram of the different topic clusters based on the frequency of prominences of all four levels normalized by the total number of syllables per topic section.

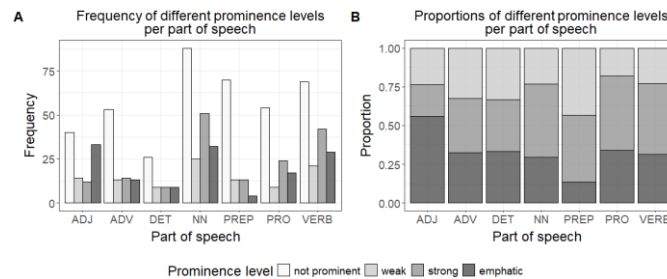


Figure 3: A) Frequency of different prominence levels (including not prominent) per part of speech. B) Proportions of prominent syllables of different prominence levels occurring for different parts of speech in the sample. (ADJ = adjective, ADV = adverb, DET = determiner, NN = noun including proper names, PREP = preposition, PRO = pronoun, VERB = verb)

## References:

- [1] Baumann, S., & Lorenzen, J. (2024). Boosting or inhibiting – How semantic-pragmatic and syntactic cues affect prosodic prominence relations in German. *Plos one* 19(4), e0299746. <https://doi.org/10.1371/journal.pone.0299746>
- [2] Wagner, P., Origlia, A., Avezani, C., Christodoulides, G., Cutugno, F., d'Imperio, M., ... & Vainio, M. (2015). Different parts of the same elephant: A roadmap to disentangle and connect different perspectives on prosodic prominence. In *18th International Congress of Phonetic Sciences* (pp. 1–4).
- [3] Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1(2), 425–452. <https://doi.org/10.1515/labphon.2010.022>
- [4] Roy, J., Cole, J. & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology* 8(1), 22. <https://doi.org/10.5334/labphon.108>
- [5] Kaland, C., Kraher, E., & Swerts, M. (2014). White bear effects in language production: Evidence from the prosodic realization of adjectives. *Language and Speech* 57(4), 470-486. <https://doi.org/10.1177/0023830913513710>
- [6] Kügler, F., Baumann, S., & Röhr, C. T. (2022). Deutsche Intonation, Modellierung und Annotation (DIMA). Richtlinien zur prosodischen Annotation des Deutschen. In C. Schwarze & S. Grawunder (Eds.), *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion – Konzepte, Probleme, Lösungen* (pp. 23–54). Narr Francke Attempto Verlag.
- [7] Markiplier. (2018, August 19). *Let's Be Completely Honest* [Video]. YouTube. Last accessed December 3, 2025, at <https://www.youtube.com/watch?v=T6cdM1kubk4>
- [8] Berger, S. (2024). "Like, comment, subscribe": Perception of acoustic-prosodic features of content creators' charismatic speech on YouTube. <https://doi.org/10.38071/2024-00858-9>
- [9] Beck, J. (2015). The linguistics of 'YouTube Voice'. *The Atlantic*, December 7, 2015. Last accessed December 3, 2025, at <https://www.theatlantic.com/technology/archive/2015/12/the-linguistics-of-youtube-voice/418962/>
- [10] Rayson, P., & Garside, R. (1998). The CLAWS Web Tagger. *ICAME Journal* 22, 121-123.
- [11] Posit team (2023). RStudio: Integrated Development Environment for R (Version 2023.6.1.524). Posit Software, PBC, Boston, MA. <http://www.posit.co/>.
- [12] Bishop, S. (2018). Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm. *Convergence: The International Journal of Research into New Media Technologies* 24(1), 69–84.