

Introducing the 3MT_French Dataset

Beatrice Biancardi, *LINEACT CESI, Nanterre, France*

Mathieu Chollet, *School of Computing Science, University of Glasgow, Glasgow, U.K.*

Chloé Clavel, *LTCI, Télécom Paris, IP Paris, Palaiseau, France*

Public speaking constitutes a real challenge for a large part of the population: estimates indicate that 15 to 30% of the population suffers from public speaking anxiety (Tillfors & Furmark, 2007).

Several existing corpora were previously used to model public speaking behavior. Those created ad-hoc for research purposes (e.g., Wörtwein et al., 2015) often provide a limited number of speakers, and are collected in an experimental setting without a real human audience. In monologues (e.g., Chen et al., 2017), the interaction with the audience is mostly asynchronous, and they are collected in the context of job interviews so the annotations are focused on hireability. TED Talks are a great resource but risk to contain mostly high-quality presentations, making it difficult to investigate the behaviors related to low-quality speeches or to anxious speaking behavior. Moreover, the videos are relatively long and the annotation protocol quite complex.

In most public speaking datasets, judgements are given after watching the entire performance, or on thin slices randomly selected from the presentations (e.g., Chollet & Scherer, 2017), without focusing on the temporal location of these slices. This does not allow to investigate how people's judgments develop over time during presentations, under the perspective of socio-cognitive theories such as Primacy and recency (Ebbinghaus, 2013) or first impressions (Ambady & Skowronski, 2008).

To provide novel insights on this phenomenon, we present the 3MT_French dataset. It contains a set of presentations of PhD students participating in the French edition of 3-minute Thesis competition. The jury and audience prizes have been integrated with a set of ratings collected online through a novel annotation scheme and protocol. Global evaluation, persuasiveness, perceived self-confidence of the speaker and audience engagement were annotated on different time windows (i.e., the beginning, middle or end of the presentation, or the full video).

We aim at providing two types of contributions. First, the 3MT_dataset with its particular properties:

- A relatively large amount (248) of naturalistic presentations;
- The quality of the presentations is highly heterogeneous;

- The presentations all have similar duration (180s) and structure.

On the other hand, we also provide the following methodological contributions:

- A novel annotation scheme, which aims at providing a quick way to rate the quality of a presentation, considering the dimensions in common between other existing schemes;
- The annotations are collected for both the entire video and different time windows.

This new resource would interest several researchers working on public speaking assessment and training, as well as it will allow for perceptive studies, both under a behavioral and linguistic point of view. It will allow for investigating whether a speaker's behaviors have a different impact on the observers' perception of their performance according to when these behaviors are realized during the speech. The automatic assessment of a speaker's performance could benefit from this information by assigning different weights to segments of behavior according to their relative position in the speech. In addition, a training system could be more efficient by focusing on improving the speaker's behavior during the most important moments of their performance.

The 3MT_French dataset is available here:

https://zenodo.org/record/7603511#.Y_cgMXbMJPY

Keywords: corpus; public speaking; annotation scheme; first impressions; primacy-recency effect

References

- Ambady, N., & Skowronski, J. J. (Eds.). (2008). *First impressions*. Guilford Press.
- Chen, L., Zhao, R., Leong, C. W., Lehman, B., Feng, G., & Hoque, M. E. (2017, October). Automated video interview judgment on a large-sized corpus collected online. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 504-509). IEEE.
- Chollet, M., & Scherer, S. (2017, May). Assessing public speaking ability from thin slices of behavior. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 310-316). IEEE.
- Ebbinghaus, H. (2013). Memory: A contribution to experimental psychology. *Annals of neurosciences*, 20(4), 155.
- Tillfors, M., & Furmark, T. (2007). Social phobia in Swedish university students: prevalence, subgroups and avoidant behavior. *Social psychiatry and psychiatric epidemiology*, 42(1), 79-86.
- Wörtwein, T., Chollet, M., Schauerte, B., Morency, L. P., Stiefelhagen, R., & Scherer, S. (2015, November). Multimodal public speaking performance assessment. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 43-50).