Methods and Terminology for Phonetic Aspects of Incremental Disfluency Synthesis

Simon Betz, Universität Bielefeld
Fakultät für Linguistik und Literaturwissenschaft, AG Phonetik und Phonologie
simon.betz@uni-bielefeld.de

Abstract:

In order to implement disfluencies into speech synthesis, some theoretical clarifications are necessary. These are specifically required for (1) terminology of micro- and macro-level phenomena, (2) definition of method packages and (3) the selection of specific aspects to implement.

(1) Terminology

Research on disfluencies started several decades ago and produced a plethora of descriptive terms which were summarized in [1]’s seminal thesis, where terminological overlap was already identified. Terminology still needs to undergo further clarification, especially for the work on the phonetic side of speech synthesis. A recent approach by [2] saw the division into micro- and macro level phenomena, a division that could also be termed form and function, or structure and surface. Examples of micro-structural phenomena would be so-called disfluency elements like silences or fillers or lengthenings, which can be seen as manifestations of underlying mechanisms of the speaker’s attempts to buy time or correct production.

(2) Method Packages

Two cases are present in both human speech and in incremental dialogue system output: (a) More time is needed to complete the ongoing production, and (b) erroneous material has been produced and corrections have to be made. It is assumed that two method packages have to be defined that contain instructions for the system how to behave in any of those cases.

(3) Selected Aspects

The complexity in terminology quite likely reflects the diversity in disfluency structure. It is therefore not desired to equip a dialogue system with the full range of disfluencies that are available in human communication. It is assumed after the preliminary studies [2],[3] that it is sufficient to equip a system with a reasonable selection of producible phenomena inspired from real-world usage in order to produce high-quality output.

References: