

1. Purpose of the project

The project aims to study why speakers lengthen or reduce speech sounds as well as pronounce them more or less carefully. For instance, some of the factors that influence this mechanism are *word frequency* and the effect of *surprisal*, that is, the predictability of a word in relation to the words in its immediate context.

2. Description of the reasons for the research topic

When we speak, we lengthen and highlight some elements of speech while we shorten others. As a rule, we lengthen important or new words and shorten words which are obvious from context or occur very frequently. For example, expressions such as “I don’t know” and “because” are often reduced to “dunno” and “‘cuz”, respectively. This is because, in general, we say them very often, but also, we can successfully guess in advance, from the meaning of the preceding sentence, that someone is about to pronounce them.

Therefore, why we lengthen and why we reduce is influenced by factors such as the frequency of words and phrases in a language as well as by the surprisal effect: we expect a certain word to appear in a given context, or not. Another problem that we need to tackle is that this lengthening and highlighting also occurs under the influence of grammatical accent. Grammatical emphasis, for example, instructs us to lengthen and highlight the syllable "tu" in the word "constitution". Therefore, it is important for us to answer the following question: does the fact that Polish possess rules of grammatical emphasis, help or hinder the shortening of words which are very frequent or have a low surprisal effect.

3. Project description

To carry out our research, we need to create a new speech database. The database must be of very good sound quality. We need excellent quality because our goals require measuring the emphasis in speech accurately (that is, acoustically). We will record the speech of Polish speakers in a recording studio that allows for capturing the speech in complete silence. In addition, we will build artificial neural networks which will “learn” a Polish language model using large amounts of text. This model will allow us to get an accurate measurement of which words have a high or low surprisal effect in a given context. Once we collect the surprisal, frequency and grammatical accent measurements, we will be able to examine how they affect speech duration and pronunciation.

4. The most important expected effects

Our research on Polish is important because it has a specific accentual grammar. Studying Polish will expand our knowledge of how surprisal and frequency affect human speech. In addition, at the end of the project, we will examine whether the listener is able to actually hear the surprisal effect when it is produced by speakers. This research question has not been answered yet.