

A Graphical Query Formation Compiler for Speech Database Access

Toomas Altsosaar¹, Mietta Lennes²

¹Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology (Espoo, Finland)

²Department of Speech Sciences, University of Helsinki (Helsinki, Finland)

Definitions:

- a *speech corpus* is a set of recordings and related annotations
- a *speech database* is a speech corpus with an *access system*
- a *unit* can be a sentence, word, syllable, phone, part of a signal, etc.
- a *query* represents a model of the units that we want to find from the speech corpus

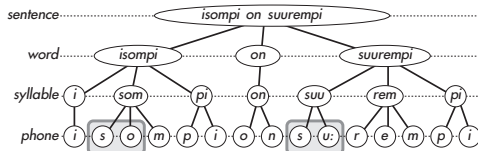
What does a query look like?

Let's say we want to find all the fricative-vowel sequences of speech sounds in a set of data. Then our query would conceptually look like:



What does a query do?

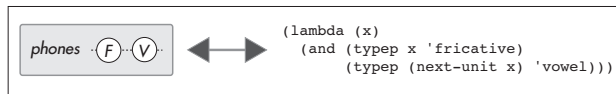
A query is given to the access system that proceeds to search through the corpus looking for matching locations. Once all data has been searched, it returns all matches. For example, if we apply the above query to the sentence "isompi on suurempi", two matches are found.



How is a query constructed?

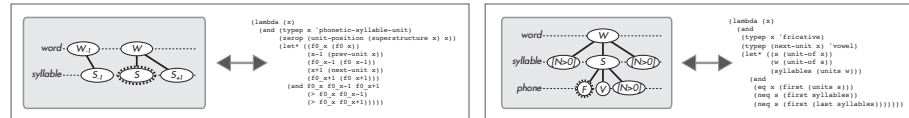
Take your favorite Query Formation Editor (see below) and draw up what you want to search for in the data. Then press "Apply Query to Database". The graphical query is automatically compiled into *program code* that the access system expects.

Here is the program code that represents our first query:



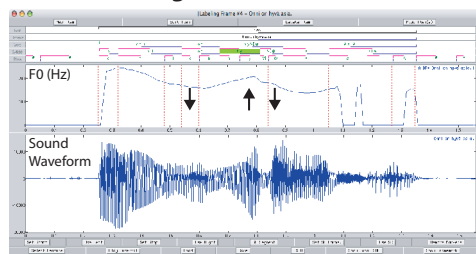
So why do we need a Graphical Query Formation Compiler?

As the query becomes more detailed, program code may become complex.



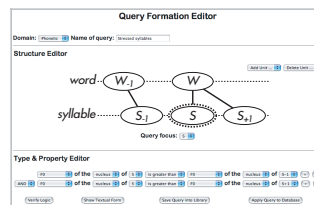
Benefits: fewer human errors, increase in productivity, less fatigue, more interesting, ... A picture is worth 1000 words.

Example: Search for word-initial syllables that have a higher F0 value than their immediate neighbors.

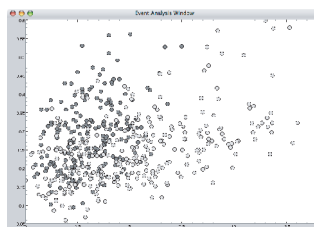


1. For example, in the annotation editor, the first syllable of the word *hyvä* represents a match since the F0 of the syllable *hy* is higher than the previous syllable *on* and higher than the next syllable *vä*.

Conclusion:
Speech database queries can be defined in a high-level manner both intuitively and interactively



2. The query is drawn up in the Query Formation Editor and program code is generated by the system.



3. The query is then applied to the speech of two speakers and matching syllables are plotted as duration (ms) vs. an index of pitch accent (semitones).