

Using the GMTK toolkit to build the ABCP1 system

The goal is to build the ABCP1 system based on a Dynamic Bayesian Network (DBN) modelling:

- the linguistic properties of the application at the word level, by means of an N-gram (N equals 2 or 3) that computes the probability $P(W)$ given the sequence of words, W , in the sentence (the vocabulary size is in the order of the tens of thousands);
- the conditional probability $P(X_A, X_V|U_A, U_V)$, associated to the sentence audio-visual feature streams X_A and X_V , where U_A and U_V stand for the sequences of acoustic and visual linguistic sub-word units in the sentence (by the end, U_A must be defined at the phonetic level and U_V at the syllabic level);
- single pronunciation lexicons for the standard acoustic and visual realizations of the words in the the vocabulary, allowing to determine U_A and U_V given W , through the function $f_{U_{AV}}(W)$.

Then, when decoding, given X_A and X_V it is chosen the sentence hypothesis:

$$W = \underset{W}{\operatorname{argmax}} \{P_{AV}(X_A, X_V|W)P_{LM}(W)\}$$

where

$$P_{AV}(X_A, X_V|W) = P(X_A, X_V|f_{U_{AV}}(W))$$

and

$$P_{LM}(W) = \prod_t P(W_t|W_{t-N+1:t-1})$$

The Graphical Models Toolkit (GMTK) ¹ is a powerful (open) software package allowing to build, with relatively little effort, the proposed DBN and also the respective decoder.

Hopefully in a near future, the following materials will be available: used [data](#) and [scripts](#); the [GM](#) structure and parameter files; a simple [log](#) and a [Tech. Report](#).

Vitor Pera, 16/06/2014.

¹<http://melodi.ee.washington.edu/gmtk/>