



The Speech Solution

BABEL DEMONSTRATOR

RESPIRE PROJECT

www.babeltech.com

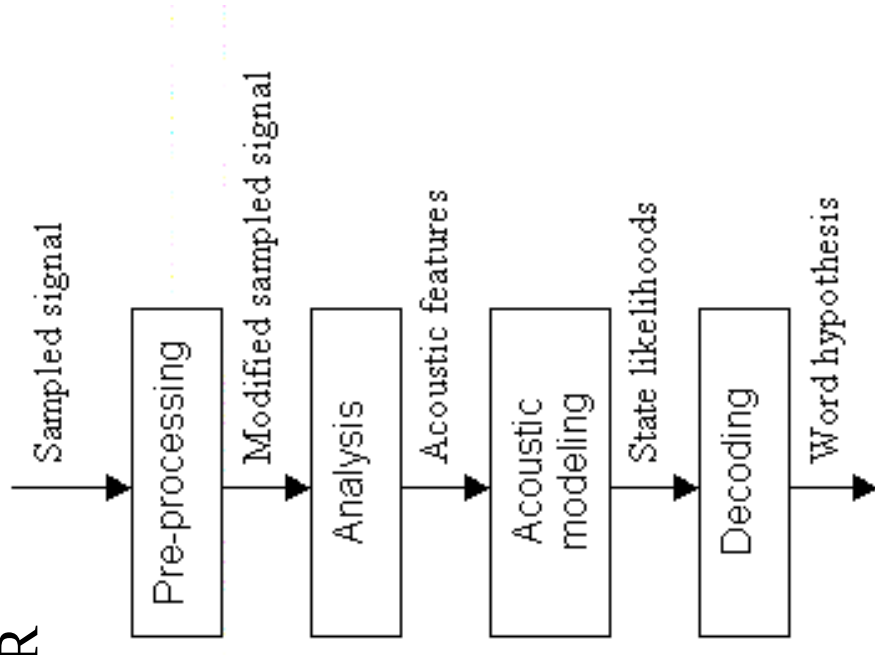
www.infovox.se



Babel demonstrator

The demo interface provides access to the ASR process at different anchor points such as :

- sampled speech signal,
- acoustic features,
- state likelihoods,
- recognized sentences.





Customisable interface

- Each block can be processed independently by calling a user-defined external program,
- The only constraint is the compatibility with the input/output data format.

ASR - block programs

Pre-processing	Name	Command line
Analysis	Name	Command line
Acoustic modeling	Name	Command line
Decoding	Name	Command line

OK Cancel



Data display

- Display is fully handled by the ASR interface by specific plug-ins for the 3 different data types:
 - acoustic features, probabilities/likelihoods, word hypothesis.*
 - Automatic time-alignment of the different data streams on the speech signal.
 - Allow display of block's internal data.
 - Dynamic internal data management.
- The definition of the data format must include all the information required by these constraints.



Data format

- Frame based
- Each frame can contain several synchronized data streams of any of the pre-defined data-types
- Binary header:

Sample rate in Hz	unsigned short
Frame length in ms	float
Frame shift in ms	float
Number of streams	unsigned short
Name stream #1	64 bytes
Type stream #1	unsigned short
Frame size stream #1	unsigned short
Name stream #2	64 bytes
Type stream #2	unsigned short
Frame size stream #2	64 bytes

- The stream names are used for identifying displayed windowpane.



Data format (con't)

- The stream type is one of the following:

- 1 – samples (PCM16) signed short
- 2 – features float
- 3 – probabilities float

- The actual data are then formatted the following way, for each frame:

```
Time index in milliseconds for current frame – float
Data stream #1
Data stream #2
...
```

- The time index is used for time alignment of different data streams. A time index of -1 means the end of the current utterance.



Data format (con't)

- The word hypothesis are written in Timit format:

`start_time(samples) end_time(samples) word_hyp`

for instance:

```
0 8480 Sil
8480 13280 one
13280 20480 eight
20480 34880 six
```



Status point

- 3 different plug-in for each data
 - Just copy the plug-in in a specific directory to add the different functions → wavesurfer program not modified
- 3 different variables available from any function
 - SAMPLES FEATURES PROBABILITIES in Respite format
- On-line and Off-line processing implemented
- Display function for samples and probabilities implemented.



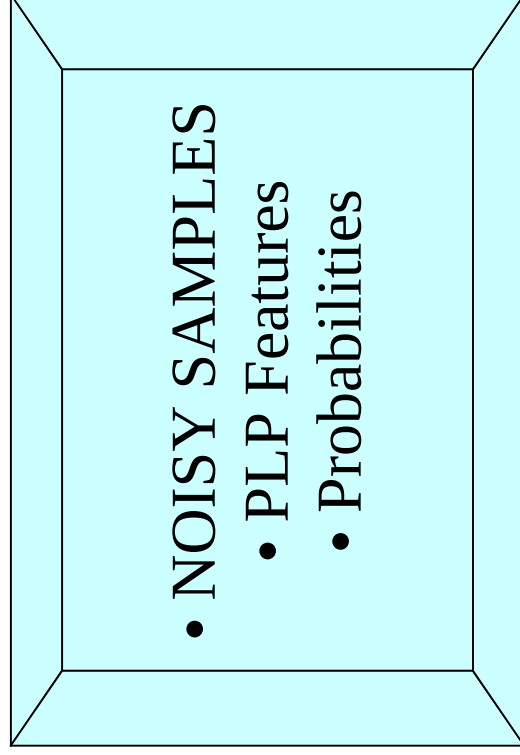
Status point

- Customizable interface : implementation realized for samples data.
- Integration of windows / Unix programs for data processing



Still to be done ...

- Display of features data (e.g. RASTA ...)
- Block management for features and probabilities
- Dynamic Management of the available RESPITE variable
 - The user can choose which variable to display





Still to be done ...

- Multiple stream processing ...
 - Integration in variables
 - Integration in display menus
- Full Installation package ...



The Speech Solution

Demonstration ...

• ...

www.babeltech.com

www.infovox.se