



HAL
open science

Text island spotting in large speech databases

Benjamin Lecouteux, Georges Linarès, Frédéric Beaugendre, Pascal Nocera

► **To cite this version:**

Benjamin Lecouteux, Georges Linarès, Frédéric Beaugendre, Pascal Nocera. Text island spotting in large speech databases. Interspeech, 2007, Anvers, Belgium. hal-02088836

HAL Id: hal-02088836

<https://hal.science/hal-02088836v1>

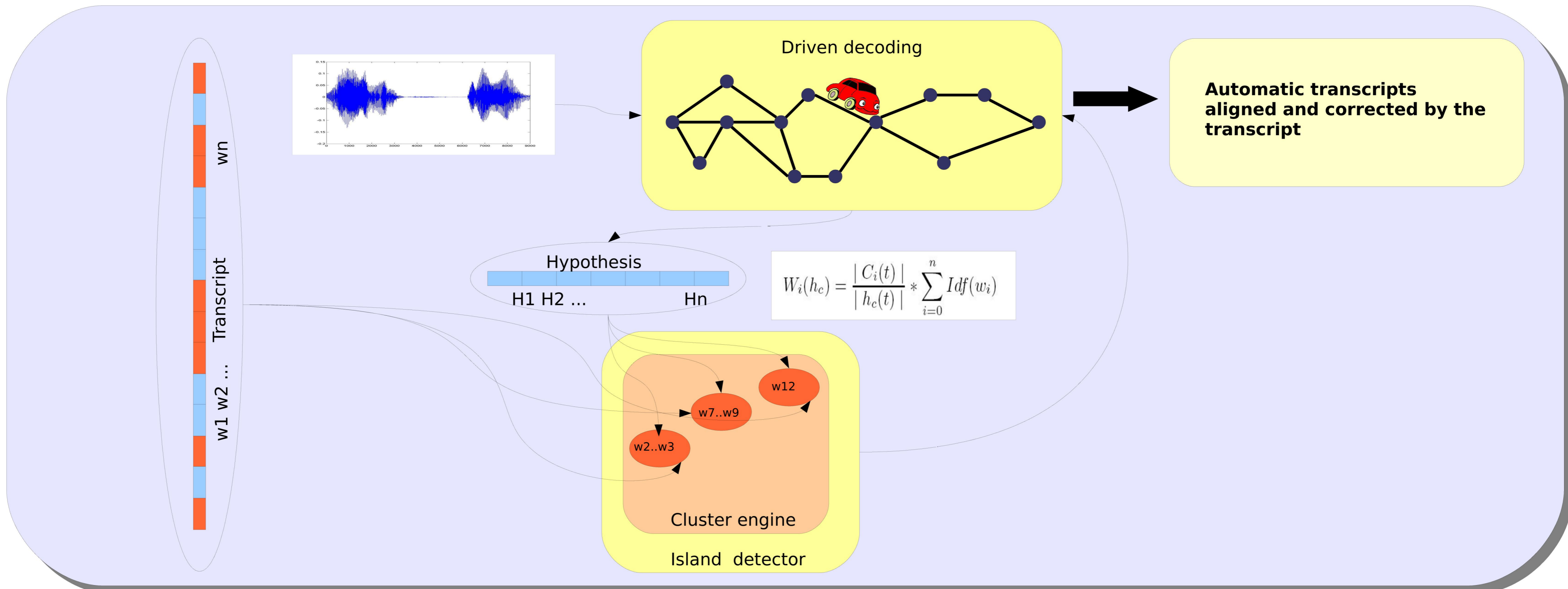
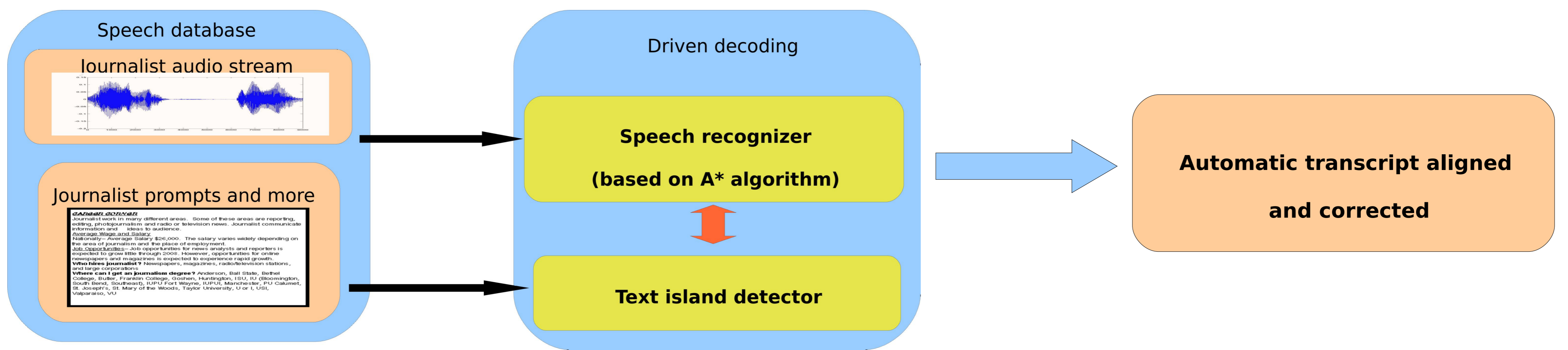
Submitted on 3 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Text island spotting in large speech databases

Benjamin Lecouteux, Georges Linarès, Frédéric Beaugendre, Pascal Nocéra



Fast-match to transcript island

- The principle of the proposed method is close to approaches used in the field of information retrieval.
- In our case, the hypothesis is a query which may be answered by one of the transcript island.
- The lexicon is represented by a lexical space L_s where each dimension is associated to a word. The coefficients of these vectors represent the frequencies of words in the document.
- As the current hypothesis is developed, a set of word clusters C_i is built and updated.
- These clusters result from the intersection of h_c and the transcript island l_i .
- For each new word added to the hypothesis h_c , transcript islands are considered as candidates for guiding the search.

This competition is arbitrated by a matching score W_i .

$$W_i(h_c) = \frac{|C_i(t)|}{|h_c(t)|} * \sum_{i=0}^n Idf(w_i)$$

where $|C_i(t)|$ and $|h_c(t)|$ are the cardinality of respectively the cluster C_i and the current hypothesis h_c . $Idf(w)$ represents the classical measure of the relative word frequency :

$$idf(w) = \frac{1}{frequency_w}$$

Experimental context :

- First experiments assessed on 3 hours of radio ESTER (with exact transcript and a 10% WER transcripts)
- Second experiments assessed on 11 hours of RTBF on which time stamps were manually added.
- All words available in database are added to the language model
- Language model : about 67000 words trained on « le monde »
- Speech recognition system : SPEERAL, an asynchronous decoder based on the A* algorithm.

Results :

Radio station	Precision	Recall	F-measure	Seg. number
INTER	90.9%	98.89%	94.8%	478
INFO	93.7%	92.9%	91.5%	468
RFI	98.9%	97.8%	98.4%	812
Mean	95.3%	97.3%	95.5%	1758

Radio station	Precision	Recall	F-measure	Seg. number
FrInter	90.7%	96.9%	93.7%	478
FrInfo	93.4%	89.7%	91.5%	468
RFI	98.8%	97.8%	98.4%	812
Mean	94.3%	94.8%	94.5%	1758

System	Baseline	DDA+IT	DDA+PT
INTER	22.6 %	17.9%	17.1%
INFO	23.4 %	21.7%	18.3%
RFI	27.2 %	23.0%	20.3 %
Mean	24.4 %	20.9 %	18.6 %

	Precision	Recall	F-measure	Seg. number
RTBF shows	99.28 %	97.13 %	98.41 %	501

Conclusions :

- On ESTER tests approximative transcripts bring a WER gain of about 14% relative, while exact ones allows a WER gain close to 24% relative.
- Spotting performance is good; more than 95.3% of segments have been found, with a precision of about 96.7%.
- On RTBF tests, spotting performance is good; more than 95.3% of segments have been found, with a precision of about 96.7%.