



**HAL**  
open science

## Qualitative investigation of the display of speech recognition results for communication with deaf people

Agnès Piquard-Kipffer, Odile Mella, Jérémy Miranda, Denis Jouvét, Luiza Orosanu

### ► To cite this version:

Agnès Piquard-Kipffer, Odile Mella, Jérémy Miranda, Denis Jouvét, Luiza Orosanu. Qualitative investigation of the display of speech recognition results for communication with deaf people. 6th Workshop on Speech and Language Processing for Assistive Technologies, SIG-SLPAT, Sep 2015, Dresden, Germany. pp.7. hal-01183349

**HAL Id: hal-01183349**

**<https://inria.hal.science/hal-01183349>**

Submitted on 19 Aug 2015

**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.

# Qualitative investigation of the display of speech recognition results for communication with deaf people

Agnès Piquard-Kipffer, Odile Mella<sup>1</sup>, Jérémy Miranda<sup>1</sup>, Denis Jouvet<sup>1</sup>, Luiza Orosanu<sup>1</sup>

Université de Lorraine, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France

Inria, Villers-lès-Nancy, F-54600, France

CNRS, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France

agnes.piquard@loria.fr, <sup>1</sup>{prenom.nom}@loria.fr

## Abstract

Speech technologies provide ways of helping people with hearing loss by improving their autonomy. This study focuses on an application in French language which is developed in the collaborative project RAPSODIE in order to improve communication between a hearing person and a deaf or hard-of-hearing person. Our goal is to investigate different ways of displaying the speech recognition results which takes also into account the reliability of the recognized items. In this qualitative study, 10 persons have been interviewed to find the best way of displaying the speech transcription results. All the participants are deaf with different levels of hearing loss and various modes of communication.

**Index Terms:** speech recognition, deaf or hard-of-hearing people, compensating for disadvantages, display of speech transcription, French language

## 1. Introduction

In the world, there are millions of people with hearing loss (<http://www.who.int/pbd/deafness/news/Millionslivewithhearingloss.pdf>; <http://wfdeaf.org>). In France over 11% of people suffer from hearing loss which causes several other limitations that are persistent [1]. The sensory problems involve both perceptual, speech, cognitive and social difficulties [2] [3]. The unemployment rate thus varies from 15 to 50% depending on the type of hearing loss.

Deaf adults still have difficulties mastering French language, which is not considered, for some of them, as their native language. Sign language may also not be considered as their native language and has no written modality. The lack of oral interaction is repeated in many situations, even for those adults for whom hearing aids provide correction. In working situations with hearing persons, deaf adults often have to be supported by others [4]. The long term goals of the Rapsodie project (<http://erocca.com/rapsodie>) are to facilitate the integration of deaf or hard-of-hearing people within a professional context thus aiding their independence, providing them ways of comprehension and communication with automatic speech transcription help.

Our research relates to an embedded system, used in a professional context which could help deaf or hard-of-hearing persons, employees, to interact with a speaking person, customer, without the help of an interpreter. The speech recognition of the customer's utterance is displayed on the screen of the embedded terminal.

The difficulty comes from the fact that speech transcription results contain recognition errors, especially if it

is a real time process on a device with limited resources (CPU and memory) and in a noisy environment. As in many real-work conditions, the speech signal is overlapped with parasitic noise, undesired extra speech, or music. These difficulties may impact the understanding processes. There has been many attempts to develop speech recognition appliances but to our knowledge, there is no suitable, validated and currently available screen display of the output of automatic speech recognizer for deaf or hard-of-hearing persons, in terms of size, colors and choice of the written symbols. It is the goal of this first qualitative study, taking account of the previously described technical constraints. We interviewed deaf adults at working age, with different levels of hearing loss and various modalities of communication. Our aim were both to study the feasibility of the project with deaf people of varying profiles, to investigate the more suitable display and to examine which factors the participants consider as being helpful for a better understanding of the speech transcription.

In the following sections, the speech recognition system is described and then the different modalities chosen for displaying the recognition output. Afterwards, we focus on the experimental protocol results conducted with 10 deaf people, discussing how they can be accommodated in order to find the best display of the automatic speech transcription results.

## 2. Speech transcription system

### 2.1. Choice of linguistic units

One of the aims of the RAPSODIE project is to realize a portable device embedding a speech recognition system that will help a deaf or hard-of-hearing person to communicate with other people. Due to the limits in memory size and computational power imposed by a portable device, the embedded speech decoder should achieve the best compromise between recognition performance, computational cost, acceptable execution time, and the way of displaying the recognition results for people with hearing loss.

Given a recognition engine, the main constraints relate to the size of the language model and of the lexicon. In this context, we have investigated syllable-based lexicons and hybrid language models [5] [6]. Indeed, the combination of words and syllables allows the recognition of the most frequent words as words and the recognition of the out-of-vocabulary words as sequences of syllables. These investigations led us to use a recognition engine system based on a hybrid trigram statistical language model with a lexicon composed of about 23,000 words and 3,000 syllables. The words and syllables were selected according to their frequency of occurrences in a training corpus of broadcast news, shows

and debates from various radio and TV channels. This hybrid model uses only 14 MB of memory space. When applied for the transcription of radio and TV shows (ETAPE [7] development data – 82,000 running words), more than 94% of the output tokens are words, the remaining part (about 6%) corresponds to syllables. An analysis of the results shows that about 70% of the words hypothesized by the decoder are correct (i.e., correctly recognized), and about 60% of the syllables are correct.

Furthermore, the speech recognition engine is built from the PocketSphinx tool [8] and uses as acoustic models, context-dependent phone HMM models with 3 states and 64 Gaussians per state. The acoustic analysis is the standard MFCC (Mel Frequency Cepstral Coefficients) providing 12 static coefficients and the logarithm of the energy per frame with a 10 ms shift. First and second order temporal derivatives are added to the feature vector.

Finally, the recognition engine provides a sequence of words and syllables corresponding to the customer’s utterance.

## 2.2. Use of confidence measure

Speech recognition is not perfect, especially when using an embedded device in a noisy environment. Two types of errors can occur. When the spoken word does not belong to the recognition lexicon (as a word or a sequence of syllables), the recognition engine recognizes it as another lexical unit or as a succession of smaller units acoustically similar to the unknown unit. Furthermore, it can happen that the spoken word is confused with another one when the conditions are different from those used for the training of the acoustic and language models (noisy environment, spontaneous speech, manner of speaking, etc.). Recognition errors will result in additional difficulties for deaf and hard-of-hearing people to understand the spoken sentence.

Confidence measures aims at indicating the reliability of the speech recognition hypotheses. Several approaches for computing confidence measures have been studied in the past [9]. In [10] confidence measures were used to highlight words with low confidence scores in view of helping error correction in a multimodal environment. Along this line, it is always words with low confidence scores that are differentiated, either in a lighter shade for error correction in voicemail transcripts [11], or highlighted for computer assisted speech transcription [12], or displayed with an underlining dependent on the confidence measure [13]. As the confidence measures are not perfect such approaches do not always accelerate the detection and correction of the errors [13]. A few other studies were more concerned with understanding aspects. In [14] the words are displayed with a brightness that depends on their score (kind of confidence measure) in the context of speech playback using time-compression and speech recognition. In all the previous studies, the speech signal was available to the user. This is not the case of [15] which has investigated the understanding of sentences from their speech recognition output only, and investigated how much taking into account the confidence measures in the display can help.

In the current study, we use the confidence measure computed by the speech recognition system to make the result of the recognition easier to understand by deaf users. The speech recognition engine provides a confidence measure for

every recognized unit (word and syllable). This measure is based on posterior probability [9]. By comparing the confidence measure to a threshold adjusted on a development corpus, each lexical unit is labeled as “*correctly recognized*” (high confidence score) or “*incorrectly recognized*” (low-confidence score). This characterization (right or wrong) of the words by the recognition system will be displayed on the terminal and different display modes will be proposed for assessment to several deaf persons.

## 3. On-screen display modalities

### 3.1. On-screen display modes of the speech recognition results (without using confidence measures)

After the speech recognition process, the recognized words and syllables are displayed on the screen of the portable device. Regardless of the accuracy of the recognition result, it is important to investigate the best way to display this result for deaf and hard-of-hearing people. First, because the result is a mixture of words, and syllables that cannot be written into an orthographical form. Secondly, because for deaf people, orthographic transcription is not necessarily the best way to display the recognition result according to the type of hearing loss and the kind of speech and language training. We decided to study the three following display modes:

- **Orthographic:** the recognized words are written into orthographical form, the syllables are written into pseudo-phonetic form;
- **International Phonetic Alphabet (IPA):** all the recognized words and syllables are written into phonetic form using the International Phonetic Alphabet. Some deaf adults benefited from early hearing and speech intervention which gave them International Phonetic Alphabet knowledge when they learned to read and during speech and language remediation therapy;
- **Pseudo-phonetic:** all the recognized words and syllables are written into a pseudo-phonetic alphabet. Indeed, the phones within the recognized words and syllables are translated into a simple sequence of graphemes using a kind of phonetic spelling. This mode seems appropriate for all the deaf persons who are familiar with French language pronunciation.

An example of a recognition result displayed in these 3 modes is presented Table 1.

| Display mode    | Result of the automatic transcription (into words and syllables) |
|-----------------|--|
| Orthographic    | je voudrais être li vré combien ça kou te                        |
| IPA             | ʒə vudʁɛ ɛtʁ li vʁɛ kɔ̃bjɛ̃ sa ku tə                             |
| Pseudo-phonetic | je voudré ètr li vré konbyin sa kou te                           |

Table 1: *The different evaluated modes for displaying the result of the recognition of the uttered sentence: “je voudrais être livré, combien ça coûte ?” (I would like it to be home delivered, how much does it cost?).*

|  | words/syllables tagged as <i>incorrect</i> are displayed in another color (red) | words/syllables tagged as <i>correct</i> are displayed in bold |
|--|---|--|
| words tagged as <i>incorrect</i> are displayed into <b>orthographic</b> mode (syllables are always displayed in pseudo-phonetic mode)    | je voudrais être li vré qu'on bien ça kou te                                    | <b>je voudrais être</b> li vré qu'on bien <b>ça</b> kou te     |
| words tagged as <i>incorrect</i> are displayed into <b>pseudo-phonetic</b> mode (syllables are always displayed in pseudo-phonetic mode) | je voudrais être li vré kon by in ça kou te                                     | <b>je voudrais être</b> li vré kon by in <b>ça</b> kou te      |

Table 2: Four screen display modalities to differentiate the words/syllables considered as incorrectly recognized and those considered as correctly recognized by the speech recognition system. Here, the words "qu", "on" and "bien", and the syllables /li/, /vré/, /kou/, and /te/ are considered as incorrect.

### 3.2. On-screen display modalities using the confidence measure

As explained in Section 2.2, the speech recognition system provides an estimation of the recognition correctness for every lexical unit, even if this estimation may be unreliable. Therefore, it is important to find the best way of presenting this information about the word/syllable correctness to the deaf user.

In [15], it has been shown that hearing users infer the correct word from a word considered incorrect by the speech recognition system, more easily when it was written in phonetic form than when it was written in orthographic form. In particular, when several consecutive words were tagged as misrecognized by the system, the hearing user unsuccessfully focused on the word splitting given by the orthographic mode, causing misunderstandings, while the sound sequence of the words was almost free from errors. Instead, the oralization of the sound sequence helped the user to find the right words and thence the meaning of the sentence. Accordingly it seemed to us interesting to study whether these results remain valid for deaf users.

On the one hand, we examined whether it is more favorable to highlight the "incorrectly recognized" or the "correctly recognized" lexical units.

On the other hand, we distinguished two modes for displaying the "incorrectly recognized" words: the orthographic and pseudo-phonetic modes. Note that syllables are always displayed in pseudo-phonetic mode.

Table 2 summarizes the four different display modalities on an example. In the second colon the lexical units tagged as "incorrect" are written in a different color (red) than the lexical units tagged as "correct" (black). In the third colon, all lexical units are written in blue and the units tagged as "incorrect" as written in bold.

## 4. Methodology

We conducted a qualitative study which goal was to identify the modalities which could help some deaf adults for a better understanding of the speech transcription and to look at how people can use these modalities.

### 4.1. Participants

The population was selected on the basis of criteria used to define hearing impairment: any disorder of hearing regardless of cause or severity (cf. World Health Organization [11]). As this is a qualitative study using situations created as close as

possible to real professional contexts, we selected deaf adults who were working or who were involved in social and cultural associations, thus well integrated socially despite their communication difficulties. A preliminary selection was made to ensure a functional literacy level, as they would have to read the written transcription of speech recognition.

- Our heterogeneous population, consisted of 10 deaf persons, 4 women and 6 men; from 25 years old to 63 years old, the average age being 39 years,
- 4 persons presenting profound hearing loss, 4 severe hearing loss, 2 severe-moderate loss. The time of acquisition of their hearing loss varied from the first few days, to months or years of life. Most of causes were listed as unknown.

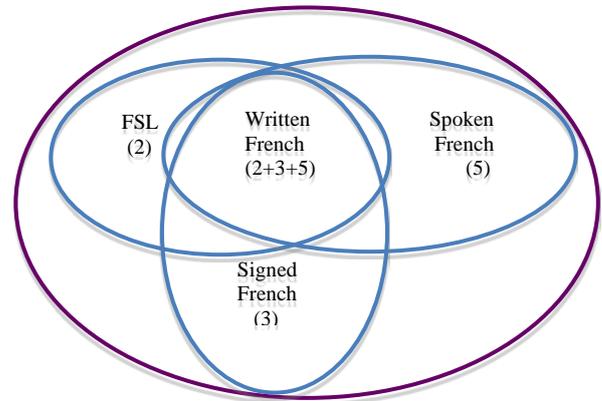


Figure 1: Distribution of the 10 participants according to their main mode of communication.

- For some of them, their mother tongue was French or French Sign Language and for some others, neither French nor French Sign Language were considered as their native language. Nine persons regularly used hearing aids to obtain as much as possible of their acoustic information. Various modes of communication were used by the deaf persons: French oral and written Language; French oral Language and French cued-speech (LPC: manual cues to supplement speech input); French written Language; French Sign Language (FSL); fingerspelling (dactylogogy); "Signed French" (français signé) combining the use of the FSL signs ordered according to the French language linear syntax and fingerspelling. Figure 1 shows the distribution of the 10

participants according to their main mode of communication. The larger outer oval includes the whole set of participants; in each of the three inner ovals are the deaf persons with their specific mode of communication, all of them using written French.

## 4.2. Tasks and Procedure

Our study was conducted in two phases. For every participant, each phase consisted of several 2-hour sessions including tests and interviews.

Before these two phases, the level of literacy was tested prior to commencing trial. The deaf person had to read a 10-line text describing communication situations which may be encountered in everyday life and in the particular situation: “do-it-yourself” shop. The deaf person has to understand the role he would play: an employee, while the hearing person (the interviewer) would play that of the customer, either at the cash-desk or in the store. In order to verify his comprehension, the participant had to reformulate the text, with his own communication tools.

### 4.2.1. First phase: Tests and interviews

The goal of the first test was to find the best way of displaying the speech transcription results among the orthographic, IPA and pseudo-phonetic display modes (cf. section 3.1). The confidence measures were not used at this stage.

In this first phase, the participants were required to read and to understand the transcriptions of 10 uttered sentences, the transcriptions were provided by the speech recognition system always in the context of the previous described scenario (do-it-yourself shop).

We elaborated every sentence according to lexical, syntactical and semantic criteria. The main lexical fields were the one of the do-it-yourself and that of the request for commercial information. Syntactically, every sentence was comprised of one or several clauses (constituent of the sentence made up of a subject and a verbal group). The sentences were coherent, reasonably long in order to be as well understood as possible. The average length of the sentences was 11.35 words (minimum: 5 words, maximum: 22 words). Every sentence contained a verb. Declarative, imperative, exclamatory sentences were included with a majority of interrogative sentences, as the test situation was as close as possible to a real situation when the client request information.

The participants were seen individually in a quiet room. They could not be helped by the sound, they had to read the speech transcription of the sentence and try to interpret it and to rephrase it so that the interviewer could check their understanding.

Their answers were not been timed. Rather, each person was interviewed in order to identify the helping points in his/her comprehension processes, sentence by sentence, knowing that speech transcription is not perfect and have no punctuation mark which could indicate the declarative, interrogative, exclamatory and imperative sentences.

We made aware deaf persons of the presence of recognition errors in the transcription system for several reasons:

- So that the deaf adults could not consider the present recognition system as a final perfect tool, as it is still in evolution,

- The correct recognized words and the presence of errors were both the base of discussion with the deaf persons who indicated the points in the display which aided their comprehension.

### 4.2.2. First phase: Results

The IPA display mode was by far the most difficult to apprehend, therefore none of the participants have indicated it as helpful, this coding requiring special learning. Table 3 shows their preferences. Not even the two deaf persons who still used it in speech remediation therapy found it helpful in such a context. For both familiar and unfamiliar users, reading a whole sentence in IPA required too much time and cognitive resources. Therefore, this display mode was abandoned for both words and syllables.

The pseudo-phonetic display mode was preferred by one participant for both words and syllables. This person indicated an order of usage preference: firstly the pseudo-phonetic mode and then the orthographic display mode, suggesting that the terminal screen could display those two options so that the deaf person could choose the more helpful one.

| Display mode    | Preference of participants (N=10) |
|-----------------|-----------------------------------|
| Orthographic    | 9                                 |
| IPA             | 0                                 |
| Pseudo-phonetic | 1                                 |

Table 3: *The display mode preferred by the participants.*

The orthographic display mode was preferred by almost all participants: nine out of ten. They have all further specified that this mode was aiding (first preference) except in the case of speech recognition errors. In fact, in case of orthographical error, for example for a word pronounced [samədi] corresponding to the word “samedi” (“Saturday”) but transcribed as “ça me dit” (“it’s tempting”), these deaf persons reported their difficulties to comprehend the whole sentence. The transcribed sentence is segmented differently, including several words instead of one, coming from other grammatical categories and lexical fields: word and time semantic field *versus* sentence and emotion semantic field. In such a case, for the five participants who were more familiar with French language phonology, it was easier to read words into pseudo-phonetic mode, and to infer semantic signification from pronunciation.

Moreover, all the participants considered that displaying the pauses detected by the speech recognizer was helpful.

### 4.2.3. Second phase: Test and Interviews

The goal of this second phase was to find the best way of displaying the additional information provided by the speech recognizer concerning the correctness of the recognized lexical unit using confidence measure. For that purpose, the four modalities described in the section 3.2 were evaluated. As it is shown in Table 2, in the case of highlighting the “*incorrectly recognized*” lexical units, we chose to display them in another color (red); in the case of highlighting the “*correctly*

recognized” lexical units, we chose to display them in the same color but in bold.

Two experiments were conducted. Firstly, we used an “oracle” confidence measure: the lexical units tagged as “incorrectly recognized” were actually the units misrecognized by the speech decoder and, respectively, the lexical units tagged as “correctly recognized” were actually the units well recognized by the speech decoder. Secondly, we used the confidence measures computed by the speech recognizer to tag the recognized units.

The same procedure as the one conducted in the first phase was used here.

#### 4.2.4. Second phase: Results

Regardless the way in which the transcribed units were tagged (oracle or from real confidence measures), the preferences of the participants were the same. The modality highlighting the “correctly recognized” lexical units in bold blue was preferred by all participants. They reported that their major attention was thus focused on words characterized as right (even if, in some cases, they are actually wrong). That was helping them for direct access to understanding. Table 4 summarizes the choices of the deaf persons.

Within this modality, the display into pseudo-phonetic of the words tagged as “incorrect” was preferred by a majority of participants, 8 persons, for the reasons previously detailed in section 4.2.2. They also explained that compared to the IPA, this system was using a simple coding scheme. They also reported that this display mode required the use of the context, and time to adapt. Indeed, this system leads to an indirect access to meaning, implying knowledge of phonology, breaking words into syllables in order to « sound out » with the aim of understanding. They also reported that any absence of a pseudo-phoneme made the task very difficult.

|   | words/syllables tagged as <i>correct</i> are displayed in <b>bold</b> | Preference of the participants (N=10) |
|---|---|---------------------------------------|
| words tagged as <i>incorrect</i> are displayed into <b>orthographic</b> mode    | je voudrais être li vré<br>qu'on bien ça kou te                       | 2                                     |
| words tagged as <i>incorrect</i> are displayed into <b>pseudo-phonetic</b> mode | je voudrais être li vr<br>é k on by in ça k<br>ou te                  | 8                                     |

Table 4: The display modalities preferred by the participants.

The display into orthographic mode of the words tagged as “incorrectly recognized” was preferred by two persons who therefore indicated weak points of this display mode. The words characterized as “incorrect” by the recognition system could place them in serious difficulties; those words could be in contradiction with the signification of the remaining part of the sentence (cf. 4.2.2). Nevertheless, they didn’t feel familiar enough with French phonology to dare using the pseudo-phonetic mode.

## 5. Discussion and conclusion

In the context of improving communication between a hearing person and a deaf person, when displaying on an embedded

device the results of an automatic speech transcription system, highlighting in bold the words considered as “correctly recognized” rather than the words considered as “incorrectly recognized” is more helpful. All the participants stressed that knowing the context and searching for keywords are essential steps to build their capacity of understanding. Highlighting the words considered as “correctly recognized” enables them to construct inferences, and to gain confidence, provided that there is an adequate number of key elements clearly identified.

The display into pseudo-phonetic of the words tagged as “incorrectly recognized” was preferred by a majority of participants (8), those persons were more familiar with French language including phonology. These results are similar to those showed from a previous study undertaken among a hearing population [15].

However, they explained that a training phase would be necessary to get more familiar with pseudo-phonetic reading. It could improve their understanding and in the long term facilitate the communication with speaking persons.

The other two persons who preferred the words tagged as “incorrect” displayed into orthographic mode were those who mainly use French Sign Language. Unfortunately, for them this display mode is not aiding enough in case of errors. Their comprehension processes cannot be supported by enough reliable words. They have to guess with many risks of misunderstanding and discouragement.

At a general level, the interviews showed that it was difficult for all the participants to stay aware of the fact that the cues based on computed confidence measures are not fully reliable. This was expressly mentioned when the participants could read the sentence with sufficient understanding, considering it as appropriate to the particular context. It was difficult for them to assess whether the information was to be trusted. The same difficulties have been observed in [13], in an experiment in which hearing people dictated a text and then had to detect the errors made by the speech recognition.

Our preliminary qualitative study was conducted in the worst conditions as the participants had only the written sentences with no oral pronunciation. They could not rely on their hearing aids nor lips reading to help them and the context information was limited. The tests were conducted in a quiet neutral room and not in a “do-it-yourself” shop. Thus, the participants could not be helped by the context of the shop (customer, special department, visual cues). As, in those experiments, no punctuation was indicated in the speech transcriptions, the deaf persons had difficulties to differentiate interrogative sentences from declarative ones.

Nevertheless, all the participants showed their interest for such a system and thought that it could be more helpful with the help of context. Further experimentations will be conducted to investigate the efficiency of this system compared to or combined with other communication means used by deaf and hard-of-hearing persons.

## 6. Acknowledgements

The work presented in this article is part of the RAPSODIE project, and has received support from the “Conseil Régional de Lorraine” and from the “Région Lorraine” (FEDER) (<http://erocca.com/rapsodie>).

We would like to thank The “Institut des Sourds de La Malgrange”, “Espoir Lorrain”, an independent association for the deafened and hard-of-hearing people, the deaf persons

interviewed, Marie-Madeleine Dutel, Jane and Vivien Dasset for their help.

## 7. References

- [1] L. Haeusler, T. De Laval, C. Millot, "Étude quantitative sur le handicap auditif à partir de l'enquête *Handicap-Santé*", *DREES* Direction de la recherche, des études, de l'évaluation et des statistiques, *Document de travail, Série Etudes et Recherches*, 131, août 2014.
- [2] H. R. Myklebust, "The psychology of deafness", New York: Grune & Stratton, 1964.
- [3] A. Dumont, "Orthophonie et Surdit . Communiquer, comprendre, parler", Masson, collection orthophonie, 2008.
- [4] D. R. Calvert & S.R. Silverman, "Speech and deafness", *A.G. Bell Association for the Deaf, Washington, D.C.* 1975.
- [5] L. Orosanu and D. Jouv t, "Comparison of approaches for an efficient phonetic decoding," in *INTERSPEECH 2013. Lyon, France*, 2013.
- [6] L. Orosanu and D. Jouv t, "Hybrid language model for speech transcription", in *INTERSPEECH 2014, Singapour*, 2014.
- [7] G. Gravier, G. Adda, N. Paulson, M. Carr , A. Giraudel, and O. Galibert, "The ETAPE corpus for the evaluation of speech-based tv content processing in the French language", in *Language Resources and Evaluation (LREC'12)*, 2012.
- [8] D. Huggins-daines, M. Kumar, A. Chan, A.-W. Black, M. Ravishankar, and A. Rudnicky, "Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices," in *ICASSP, Toulouse, France*, 2006.
- [9] H. Jiang, "Confidence measures for speech recognition: A survey," *Speech Communication*, vol. 45, no. 4, pp. 455-470, 2005.
- [10] B. Suhm, B. Myers, A. Waibel. "Multimodal error correction for speech user interfaces", *ACM Transactions on Computer Human Interaction (TOCHI)*, 8(1), 60-98, 2001.
- [11] M. Burke, B. Amento, P. Isenhour, "Error correction of voicemail transcripts in scanmail", in *Proc. SIGCHI conference on Human Factors in Computing Systems*, pp. 339-348, ACM, 2006.
- [12] S. Luz, M. Masoodian, B. Rogers, C. Deering, "Interface design strategies for computer-assisted speech transcription", in *Proc. 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, pp. 203-210, ACM, 2008.
- [13] K. Vertanen, P.O. Kristensson, "On the benefits of confidence visualization in speech recognition", in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1497-1500, ACM, 2008.
- [14] S. Vemuri, P. DeCamp, W. Bender, C. Schmandt, "Improving speech playback using time-compression and speech recognition", in *Proc. SIGCHI conference on Human factors in computing systems*, pp. 295-302, ACM, 2004.
- [15] J. Razik., O. Mella, D. Fohr, J.-P. Haton, "Frame-Synchronous and Local Confidence Measures for Automatic Speech recognition", *International Journal of Pattern Recognition and Artificial Intelligence*, 2011.
- [16] CIM-10, "Classification internationale des maladies. Chapitre VIII: maladie de l'oreille et de l'apophyse masto de. H90. Surdit  de transmission et neurosensorielle". 10<sup> me</sup> r vision. Paris : Masson, 1993.