

# Audio-to-text alignment for speech recognition with very limited resources

Xavier Anguera<sup>1</sup>, Jordi Luque<sup>1</sup> and Ciro Gracia<sup>1,2\*</sup>

<sup>1</sup>Telefonica Research, Edificio Telefonica-Diagonal 00, 08019, Barcelona, Spain

<sup>2</sup>Universitat Pompeu Fabra, Department of Information and Communications Technologies, Barcelona, Spain

{xanguera, jls}@tid.es

## Abstract

In this paper we present our efforts in building a speech recognizer constrained by the availability of very limited resources. We consider that neither proper training databases nor initial acoustic models are available for the target language. Moreover, for the experiments shown here, we use grapheme-based speech recognizers. Most prior work in the area use initial acoustic models, trained on the target or a similar language, to force-align new data and then retrain the models with it. In the proposed approach a speech recognizer is trained from scratch by using audio recordings aligned with (sometimes approximate) text transcripts. All training data has been harvested online (e.g. audiobooks, parliamentary speeches). First, the audio is decoded into a phoneme sequence by an off-the-shelf phonetic recognizer in Hungarian. Phoneme sequences are then aligned to the normalized text transcripts through dynamic programming. Correspondence between phonemes and graphemes is done through a matrix of approximate sound-to-grapheme matching. Finally, the aligned data is split into short audio/text segments and the speech recognizer is trained using Kaldi toolkit. Alignment experiments performed for Catalan and Spanish show the feasibility to obtain accurate alignments that can be used to successfully train a speech recognizer.

**Index Terms:** text-to-speech alignment, phonetic alignment, ASR, low-resources.

## 1. Introduction

In this paper we describe the procedure we followed to build an automatic speech recognition (ASR) system in Catalan and Spanish constrained by the availability of very limited resources. The focus of this paper is the alignment step between a corpus of audio collected from the internet with associated (sometimes approximate) text transcripts. In addition to its application in speech recognition, the alignment between long audios and their corresponding transcripts is useful in a number of applications [1, 2, 3, 4]. In [1, 2] the authors propose the synchronization of audio-books with e-books for language learning or enhanced reading. In [3] the creation of aligned corpora is used to build TTS systems, and in [4] emphasis is placed on the automatic alignment and correction of inaccurate text transcripts through an iterative process.

Considering that the biggest problem faced when building a speech recognizer in a new target language is the availability of manually transcribed and aligned training data, several authors use alternative sources of data like prompts or close captions [5, 6], the decoding output of a preexisting ASR system [7, 8], hand-transcribed spoken lectures [9] or data from other sources on the Internet [10]. Collected data usually contains long audio recordings that first need to be aligned with the transcripts and then split into smaller pieces for the ASR training to be feasible. Seminal work dealing directly with the alignment issue include [11, 12]. In [12] an imperfect ASR output is

aligned at word level with the text transcription to find anchor words recognized with high confidence where to split the data into smaller pieces. ASR decoding with adapted language models is then performed on each one of these pieces. The process is iterated until all data has been properly aligned. This approach needs a well-trained ASR system on the target language to be initially available. Alternatively, in [13, 7, 14] authors show that some sort of alignment is still possible when very little information on the language is available. In [7] inaccurate transcripts, obtained from the output of a speech recognition system, are aligned to the audio by using a small set of acoustic units, detected through heuristics, as anchor points. In [13] HMM-based forced-alignment at sentence level is performed by using a subset of phoneme models trained with data from several languages. Finally, in [14] alignment at sentence level is performed by exploiting syllable information in both speech and transcripts. Other ways to train speech recognition systems with low resources include the adaptation of a preexisting ASR system on the same language, but different dialect [10], cross-language bootstrapping [15] or training seed acoustic models from very little manually aligned data [16].

In this paper we use generic tools and easy to obtain data to train a viable ASR system on a new language. First we gather data for the target language from the Internet, with the only constraint that some (even if inaccurate) manual transcriptions need to be available. In particular, we have experimented with using audio/ebooks for Spanish and parliamentary speeches for Catalan. Then we decode the audio into a phoneme sequence by using an off-the-shelf phoneme recognizer trained on a different language (Hungarian in this case). Next, we align the text transcription to the phoneme sequence using dynamic programming and a transformation matrix that matches Hungarian phonemes to Spanish/Catalan graphemes. Finally, the audio and text transcripts are split into small segments and an ASR is built using the Kaldi Speech Recognition Toolkit [17] using grapheme-based models (to avoid having to train a grapheme-to-phoneme system). Results shown both in alignment accuracy and in ASR performance demonstrate the feasibility of the approach. The same approach can be used for any language provided that audio+text data are available.

## 2. Audio-to-Text Alignment Algorithm

In this section we describe how we align long audio files to their corresponding text transcripts in a low-resources setting. The resulting data is then used to train an ASR system on the target language.

Figure 1 shows the main blocks involved in the proposed system. The input consists on an audio file and the associated text transcription. Although the transcription should follow what is said in the audio as closely as possible, we are able to handle a certain number of mismatches, like extra or missing text, or extra sounds (e.g. clapping, hesitations, noise, etc.). In the approach proposed we generate a phonetic transcription from the audio, which is then aligned with the text by using dynamic programming. This is the inverse approach from what

\*C. Gracia was a visiting researcher at Telefonica Research during the development of this work.

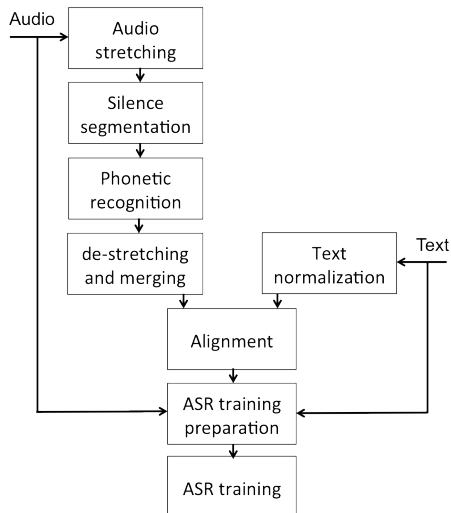


Figure 1: System blocks to train an ASR with limited resources.

we previously proposed in [2] where we first converted the text into acoustic form and then aligned it with the audio. Next we describe in detail the four main steps involved in this process, namely: audio and text preprocessing, audio-to-text alignment and ASR data preparation and training.

### 2.1. Audio-to-Phoneme Sequence Conversion

First, the audio signal is time stretched by a 25% factor by using the SOLA (Synchronous-OverLap-Add) algorithm implementation found in SoundTouch Audio Processing Library [18]. The resulting signal keeps the same pitch and most of the original speech characteristics with a reduced speech rate, which is shown to improve the alignment (see section 3.1 in the experimental section). Before we can perform a phonetic recognition on the audio we need to divide it into small segments as most acoustic decoding modules have memory issues when the input signal is too long. For this reason, if the stretched audio is longer than 30 seconds we use the LIUM speaker diarization toolkit [19] to create individual audio files of length less than 30s, split according to criteria such as silence and, if present, speaker turns. Next, individual audio files are decoded with the Phnrec phoneme recognizer [20]. From the set of publicly available acoustic models we selected the one trained using Hungarian data as its vowel inventory is one of the richest and it has been shown to perform well for other tasks like Language ID and Query-by-example Spoken-term Detection. The result of the phoneme decoding is a 1-best decoding with start-time and duration for each phoneme. Finally, the individual file outputs are time de-stretched and joined together to form a single phonetic transcription.

### 2.2. Text Normalization

Before the alignment step, the input text transcription is normalized. The output of the text normalization step is a set of individual grapheme-like symbols, plus a silence symbol. The choice to use graphemes instead of a phoneme transcription is inspired by results from prior work in the speech recognition area (see for example [21]) where grapheme-based speech recognition in Spanish was shown not to degrade much in performance compared to a phoneme-based system. In addition, in this paper we wanted to spare the resources necessary to build a conversion system to obtain phonemes. Instead, some simple acoustically-inspired rules were applied to some graphemes and grapheme pairs in Catalan and Spanish to obtain single-

character symbols, as shown in Table 1. In addition, we eliminated all diacritics and the 'h' grapheme when not appearing after 'c', and converted all numbers to their textual form. Finally, we substituted all punctuation marks by the phoneme 'sil' indicating the locations where there could be a pause in the audio. Although language specific, these conversions can be defined with just basic knowledge of the target language.

Table 1: Grapheme conversions applied to the text transcription for the languages covered in this work: Hungarian Phonemes (horizontal axis) vs. Spanish/Catalan graphemes (vertical axis)

Original grapheme(s)	Used symbol
à á, è é, í ï, ò ó, ú ü (both)	a, e, i, o, u
ñ, ch (Spanish)	N, C
ny, tx, l.l, ç (Catalan)	N,,C, l, s
qu, ll, h (both)	k, L, -

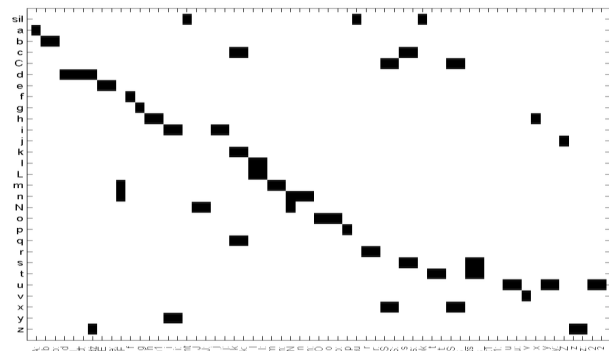


Figure 2: Catalan/Spanish to Hungarian transformation matrix.

### 2.3. Grapheme-to-Phoneme Alignment

Alignment between the phonetic transcription in Hungarian and the grapheme transcription in Catalan/Spanish is done using dynamic programming (DP) to find the optimal global alignment between both strings. Equation 1 shows the recursion followed for each position  $[x_i, y_j]$  of the global distance matrix  $D(x_i, y_j)$ , where  $x_i \in X$  is the grapheme sequence and  $y_j \in Y$  is the phoneme sequence.

$$D(x_i, y_j) = p(x_i, y_j) + \min \begin{cases} D(x_{i-1}, y_j) \\ D(x_{i-1}, y_{j-1}) \\ D(x_i, y_{j-1}) \end{cases} \quad (1)$$

where  $p(x_i, y_j)$  indicates the penalty of matching the grapheme found at position  $x_i$  of the grapheme sequence with the phoneme found at position  $y_j$  of the phoneme sequence.  $p(x_i, y_j)$  is defined by using the matrix of approximate sound-to-grapheme matching shown in Figure 2. For every grapheme-phoneme pair the matrix encodes whether they might sound the same (penalty 0) or they will always sound differently (penalty 1). This correspondence was made by hand by listening to example words from the Hungarian phoneme set using google translate <sup>1</sup> and comparing them with how a Catalan/Spanish speaker would usually pronounce each character. Note that although Catalan and Spanish are different languages, they share some acoustic traits due to their common Latin origin. This allowed us to define a unique common transformation matrix for both.

After filling up the global distance matrix  $D(x_i, y_j)$ , a traceback on the matrix gives us the best alignment path. It has been observed through experimentation that although there

<sup>1</sup><http://translate.google.com/>

might be many gaps in the  $D(x_i, y_j)$  matrix along the correct alignment path, this path can still be clearly identified and correctly retrieved. Furthermore, even when the text is not an exact transcription of the audio and thus big portions of the matrix end up without a possible alignment, the DP process is able to find the global optimal alignment path and correctly align the rest. Given the resulting alignment, we are interested in obtaining the times where each grapheme occurs in the audio. To do so we need to resolve what to do with insertions and deletions. If several phones align with the same grapheme (due to insertions) we consider the grapheme starts at the first phoneme. Alternatively, if several graphemes align with the same phoneme (due to deletions), we split the duration of the aligned phoneme among all grapheme matches.

## 2.4. ASR Data Preparation and Training

The last step in the process corresponds to the preparation of the original audio data and text transcripts to be used as input for training an ASR system. For this purpose we need to segment the original audio and text into short sentences, each of them with their corresponding text. We have experimented with two methods. The first method splits sentences at all punctuation marks, even though these do not always correspond to actual speaker silences in the audio. The second method uses again the LIUM diarization system, although this time with a maximum duration of 10 seconds to determine putative splitting points. Then, the closest word boundary for a given splitting time is chosen as the split point. Kaldi speech recognition toolkit [17] is used to train the ASR, as described in detail in the experimental section.

## 3. Experimental Evaluation

We evaluated the system proposed by looking both at the accuracy of the alignment at word level and its usability to train an ASR system.

### 3.1. Alignment Evaluation Results

To evaluate the alignment accuracy we used two small databases recorded for Spanish and Catalan by CereProc<sup>2</sup> and used for their commercial TTS products. Each database consists of around 200 spoken utterances together with their corresponding word alignments. Alignments were generated automatically using the algorithm proposed in [22] and they were then manually checked and corrected if necessary. In total there are around 15 minutes of audio and around 2000 words per language. We tested the alignment both at sentence level (aligning each utterance individually) and also aligning the audio and text files resulting of pooling together all audio files, thus simulating how a longer recording would be aligned.

Table 2: Analysis of alignment errors (in percentage of words with an alignment error greater than 100 and 200ms)

Algorithm	Spanish (error)		Catalan (error)	
	> 100ms	> 200ms	> 100ms	> 200ms
Single utterances	3.66%	0.54%	6.46%	1.73%
Pooled utterances	4.80%	1.48%	9.28%	4.03%
Pooled, no stretch	5.69%	2.02%	11.48%	5.71%

Table 2 and Figure 3 show the normalized word alignment error distributions obtained for Catalan and Spanish. For convenience, subfigures in Figure 3 focus on the range  $[-0.3, 0.3]$ , values outside of this range are sporadic errors that constitute less than 1% of the alignments. Error metrics shown in Table

<sup>2</sup><https://www.cereproc.com>

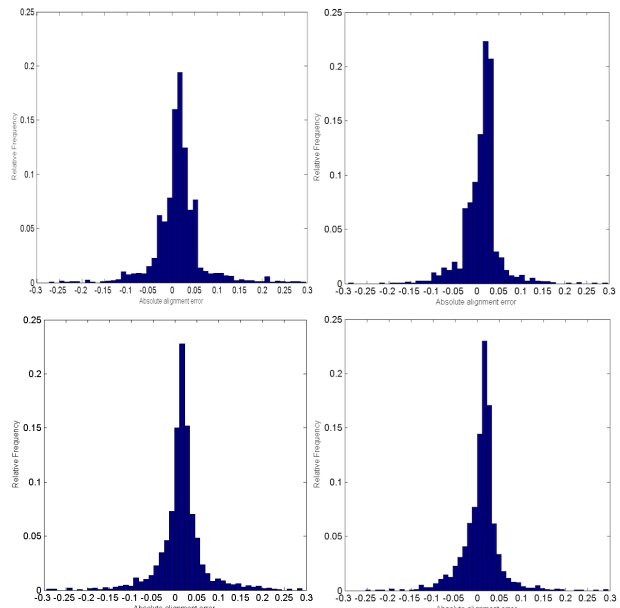


Figure 3: Normalized word alignment error distributions (in seconds) for Catalan (left) and Spanish (right) computed on individual utterances (top) or on all data pooled together (bottom)

2 are the percentage of word alignment errors that are above 100ms and 200ms.

Results show that the system deteriorates slightly when aligning long audio files (bottom row in Figure 3, second and third rows in Table 2) versus aligning single utterances (top row in Figure 3, first row in Table 2). On the contrary, single utterance alignment has more unpredictable errors, thus their error distributions are less symmetric than when aligning all utterances at once. Note also that all subfigures have the peak biased slightly towards positive values (around 20ms). This could be due to a constant bias introduced by the phonetic recognizer, which could be globally subtracted. Results on the third row of Table 2 are obtained for the pooled utterances experiment with no audio stretching. We can see how stretching improves the alignment, probably due to a better phoneme transcription coming from the phoneme recognizer.

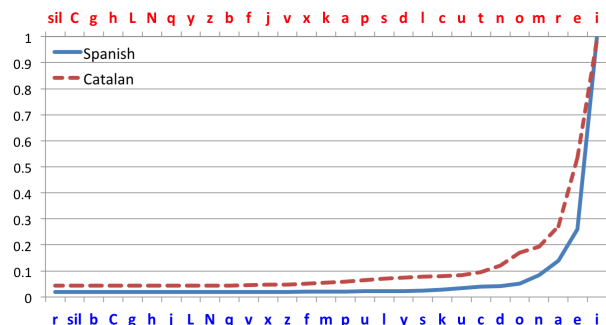


Figure 4: Percentage of alignment errors greater than 200ms when eliminating each grapheme from the transformation matrix (backward elimination, left to right).

Next, Figure 4 shows the analysis of each grapheme relevance in the alignment for both languages, using the pooled utterances. Starting with all graphemes, we iteratively eliminate

one grapheme from the transformation matrix (by inserting zeros in the corresponding row in the matrix) and compute the percentage of words with an alignment error greater than 200ms. Once this procedure is applied to all graphemes in the transformation matrix, we permanently eliminate the grapheme with the smallest impact in the overall alignment error. This process is iteratively repeated for the remaining graphemes in the transformation matrix until there are no graphemes left (technique known as backward elimination). Results show that for the first 15 graphemes the alignment does not suffer with the elimination. This means that we can greatly simplify our matrix with no loss in accuracy. For both languages, vowels, nasals and plosives are among the graphemes with the greatest importance in the alignment process. Grapheme relevance is very similar for both languages considered, except for the grapheme 'r', which is the least relevant for Spanish and one of the most relevant for Catalan. These results seem to indicate that not much prior knowledge regarding the target language is necessary to build a useful transformation matrix for a new language.

### 3.2. Speech Recognition Evaluation Results

In this section we show speech recognition results by training several ASR systems both in Spanish and in Catalan. To train the ASR we have used the Kaldi Speech Recognition Toolkit [17]. In particular, we have adapted the Switchboard training recipe and trained two different systems per test. On the one hand, an HMM-based model using triphone units with a decision tree of 3200 leafs and a total of 30000 Gaussian mixtures. On the other hand, the previous model is reused but instead of the Gaussian mixtures, a mid-size DNN has been trained with 4 hidden layers and 1024 neurons per layer. Unless mentioned otherwise, the total amount of data available in each case has been split into training (90%) and test (10%) data. Words in the test data were included in the lexicon to ensure there were no OOV words in test.

The training data used by the recognizer comes from three different datasets. The first dataset, as described in [23] and called Ceudex, is composed of around 500 unique phonetically balanced sentences spoken by 167 Spanish speakers. It amounts to around 4 hours of clean speech. In this case, a test set with 1000 sentences from the Spatis database (also described in [23]) was used to avoid repeating the same sentences in training and test. This setup was used as a controlled baseline. A second database is composed of the initial 20 chapters of the spoken version of the book *El Quijote* by *Miguel de Cervantes*, downloaded from [24] together with the associated text. Chapters are between 10 and 30 minutes long and read by a single professional speaker in Spanish. The dataset is 5.3 hours in total. A third dataset contains parliamentary speeches in Catalan downloaded from the Catalan Parliament website [25]. These correspond to two sessions (09/25/2013 and 09/26/2013) when the state of the autonomic region was discussed. In particular, 9 interventions from the president Artur Mas were collected, 8 of them with durations between 10 and 45 minutes, and one with 1h45m. The dataset is 5.4 hours in total. All speeches, except the longest one, were unscripted. The associated text is a professional transcription offered on the Parliament website.

Table 3 summarizes the Word Error Rates (WER) obtained on the test sets for these datasets. All except for the first Ceudex test were trained using grapheme models. The first and second tests using Ceudex database are considered the baseline for the system as we did not use the proposed alignment algorithm on the data (we trained the system with the original database). As expected, errors for the grapheme-based system are just moderately higher than using phonemes. Third and forth rows in Table 3 show results for the system proposed where sentences used in ASR training were split using a final diarization step

Table 3: ASR results

Data source	System	WER (HMM)	WER (DNN)
Ceudex (ES)	Phon.	22.01%	20.25%
Ceudex (ES)	Graph.	22.85%	21.40%
Ceudex (ES)	Graph+diari.	20.63%	16.98%
Ceudex (ES)	Graph+sent.	20.31%	17.15%
El Quijote (ES)	Graph.	8.55%	6.59%
Parliament (CAT)	Graph.	21.43%	20.63%

(third row) or split at sentence-level (forth row). In here we performed the alignment on one single file per speaker, with all utterances and text spoken by that speaker joined together. To our surprise these results clearly outperform the baseline results. We hypothesize that the reason for this is that the proposed alignment successfully eliminated the silence regions at the beginning and end of each sentence, which caused trouble in the ASR models initialization for the baseline systems. In addition, using an extra step of diarization does not seem to make much difference versus splitting sentences at punctuation marks. Still, in the following tests we used diarization splitting, as it obtained slightly better results for DNN. Next, we trained an ASR with data from *El Quijote* and the Parliament database. Results are shown in rows 5 and 6 of Table 3. On the one hand, the system trained with data from *El Quijote* obtains very good performance, explained by the quality of the data and accuracy of text transcripts. On the other hand, the system we built with Catalan training data from the Parliament database obtains a higher error rate due to a faster speaking rate, in addition to being spontaneous speech recorded in a noisier environment, with sporadic interruptions and noises. Still, results are comparable to those obtained by the Ceudex database, and we believe are quite positive considering the small amount of training data we used. These results show the feasibility of training an ASR system with this type of approach, which could be potentially applied to any language. In languages with no clear grapheme-to-phoneme (g2p) correspondence, a g2p conversion module could be inserted in the process.

## 4. Conclusions and Future Work

Building a speech recognition system for a new language is always an arduous process. While some previous works make use of small amounts of well-aligned training data or use existing ASR systems in very similar languages, in this work we made use of standard off-the-shelf tools and a bit of knowledge regarding the target language to successfully build a viable system for Spanish and Catalan from very limited resources. We started by collecting audio data and associated text transcripts from the Internet. We then used a phonetic recognizer in Hungarian to obtain a symbolic representation of the audio, and aligned this representation to the text transcript via dynamic programming and a phoneme-to-grapheme matching matrix constructed by roughly mapping Hungarian sounds to Spanish/Catalan graphemes. Finally, we split the alignment into sentences and trained a speech recognizer using the Kaldi toolkit. Results both on alignment and ASR performance indicate the viability of the process to build speech recognition systems for new languages with very limited resources. Future work includes testing the algorithm with other languages and training ASR systems with more data.

## 5. Acknowledgements

We would like to thanks Karel Veselý and Dan Povey for getting us started with the Kaldi toolkit. A special thanks also to Matthew Aylett from CereProc for providing us with the means to test our alignments in Spanish and Catalan.

## 6. References

- [1] D. Caseiro, H. Meinedo, A. Serralheiro, I. Trancoso, and J. a. Neto, "Spoken Book Alignment using WFSTs," *Proc. of the second international conference on Human Language Technology Research*, pp. 3–5, 2002.
- [2] X. Anguera, N. Perez, A. Urruela, and N. Oliver, "Automatic Synchronization of Electronic and Audio Books via TTS Alignment and Silence Filtering," in *Proc. ICME*, 2011.
- [3] K. Prahallad and A. W. Black, "Segmentation of Monologues in Audio Books for Building Synthetic Voices," *Trans. Audio, Speech and Language Processing*, vol. 19, no. 5, pp. 1444–1449, 2011.
- [4] T. J. Hazen, "Automatic Alignment and Error Correction of Human Generated Transcripts for Long Speech Recordings," in *Proc. Interspeech*, 2006, pp. 1606–1609.
- [5] M. J. Witbrock and A. G. Hauptmann, "Improving Acoustic Models by Watching Television," Technical Report CMU-CS-98-110, Carnegie Mellon University, Tech. Rep., 1998.
- [6] B. Lecouteux and G. Linar, "Using prompts to produce quality corpus for training automatic speech recognition systems," in *Proc. The 14th IEEE Mediterranean Electrotechnical Conference*, 2008, pp. 841–846.
- [7] A. Haubold and J. R. Kender, "Alignment of Speech to Highly Imperfect Text Transcriptions," in *Proc. ICME*, 2007.
- [8] A. Buzo, H. Cucu, and C. Burileanu, "Text Spotting In Large Speech Databases For Under-Resourced Languages," in *Proc. Speech Technology and Human-Computer Dialogue (SpeD) Conference*, no. 1, 2013.
- [9] C. J. van Heerden, P. de Villiers, E. Barnard, and M. H. Davel, "Processing spoken Lectures in Resource-Scarce Environments," in *Proc. Proceedings of the 22nd Annual Symposium of the Pattern Recognition Association of South Africa*, 2011, pp. 138–143.
- [10] M. H. Davel, C. V. Heerden, N. Kleynhans, and E. Barnard, "Efficient harvesting of Internet audio for resource-scarce ASR," no. August, 2011, pp. 3153–3156.
- [11] J. Robert-Ribes and R. Mukhtar, "Automatic Generation of Hyperlinks Between Audio and Transcript," in *Proc. Eurospeech*, vol. 1997, no. September, 1997, pp. 903–906.
- [12] P. J. Moreno, C. Joerg, J.-m. Van Thong, and O. Glickman, "A Recursive Algorithm for the Forced Alignment of Very Long Audio Segments," in *Proc. ICSLP*, 1998.
- [13] S. Hoffmann and B. Pfister, "Text-to-Speech Alignment of Long Recordings Using Universal Phone Models," *Proc. Interspeech*, no. August, pp. 1520–1524, 2013.
- [14] I. Ahmed, S. K. Koppurapu, T. C. S. Innovation, L. Mumbai, Y. Park, and T. West, "Technique for Automatic Sentence Level Alignment of Long Speech and Transcripts," in *Proc. Interspeech*, no. August, 2013, pp. 1516–1519.
- [15] N. T. Vu, F. Kraus, and T. Schultz, "Rapid building of an ASR system for Under-Resourced Languages based on Multilingual Un-supervised Training," in *Proc. Interspeech*, no. August, 2011, pp. 3145–3148.
- [16] N. Braunschweiler, M. J. F. Gales, and S. Buchholz, "Lightly supervised recognition for automatic alignment of large coherent speech recordings," *Trans. Computer Speech and Language*, vol. 16, no. 1, pp. 115–129, 2002.
- [17] D. Povey, A. Ghoshal, L. Burget, O. Glembek, N. Goel, M. Hanemann, P. Motlíček, T. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The Kaldi Speech Recognition Toolkit," in *Proc ASRU*, 2011.
- [18] "SoundTouch Audio Processing Library." [Online]. Available: <http://www.surina.net/soundtouch>
- [19] M. Rouvier, G. Dupuy, P. Gay, E. Khoury, T. Merlin, and S. Meignier, "An Open-source State-of-the-art Toolbox for Broadcast News Diarization," in *Proc. Interspeech*, 2013. [Online]. Available: <http://www-lium.univ-lemans.fr/diarization>
- [20] P. Schwarz, "Phoneme Recognition Based on Long Temporal Context," Ph.D. dissertation, 2008. [Online]. Available: <http://speech.fit.vutbr.cz/software/phoneme-recognizer-based-long-temporal-context>
- [21] M. Killer, S. Stüker, and T. Schultz, "Grapheme Based Speech Recognition," in *Eurospeech*, 2003, pp. 3141–3144.
- [22] M. P. Aylett and C. J. Pidcock, "The CereVoice Characterful Speech Synthesiser SDK," in *Proc. AISB*, Newcastle, 2007, pp. 174–178.
- [23] C. de la Torre, L. Gernández-Gómez, and D. Tapias, "CEUDEX: A Data Base oriented to Context-Dependent Units Training in Spanish for Continuous Speech Recognition," in *Proc. Eurospeech*, no. September, 1995, pp. 845–848.
- [24] "EL QUIJOTE - IV Centenario Audiobook." [Online]. Available: [http://www.quijote.es/IVCentenario\\_AudioLibro.php](http://www.quijote.es/IVCentenario_AudioLibro.php)
- [25] "Canal Parlament - Parlament de Catalunya." [Online]. Available: <http://www.parlament.cat/web/actualitat/canal-parlament>