# Studying Luxembourgish phonetics via multilingual forced alignments

*M. Adda-Decker* ◇,⋆, *L. Lamel* ⋆, *N.D. Snoeren* ⋆

◇ LPP CNRS / Univ. Paris 3, F-75005 Paris; ⋆ LIMSI CNRS, F-91403 Orsay

martine.adda-decker@univ-paris3.fr, {madda,lamel,snoeren}@limsi.fr

## ABSTRACT

Luxembourgish, a Germanic-Franconian language, is embedded in a multilingual context on the divide between Romance and Germanic cultures and remains one of Europe's under-described languages. This paper investigates the similarity between Luxembourgish phone segments with German, French and English via forced speech alignment techniques. Making use of monolingual acoustic seed models from these three languages, as well as "multilingual" models trained on pooled speech data we investigated whether Luxembourgish was globally better represented by one of the individual languages or by the multilingual model. Although French words are often interspersed in spoken Luxembourgish, forced alignments show a clear preference for Germanic acoustic models, with only a limited usage of the French ones. While globally, the German models provide the best match, a phone-based analysis, shows language-specific preferences: French is preferred for rounded front vowels, nasal vowels and /ʒ/ whereas English is more frequently used for diphthongs. The proposed method enables the acoustic match between phonemes in different languages to be quantified and opens new perspectives in language processing studies for low e-resourced languages and for L2 learning.

**Keywords:** Forced alignment; acoustic modeling; multilingual models; Luxembourgish; Germanic languages, Romance languages.

## 1. INTRODUCTION

Luxembourg, a small country of less than 500,000 inhabitants in the center of Western Europe, is composed of about 65% of native inhabitants and 35% of immigrants. The national language, Luxembourgish ("Lëtzebuergesch"), has only been considered as an official language since 1984 and is spoken by natives.[1]. The immigrant population generally speaks one of Luxembourg's other official languages: French or German. Recently, English has joined the set of languages of communication, mainly in professional environments.

As pointed out by [3] and [4], Luxembourgish should be considered as a partially under-resourced language, due to the fact that the written production is low, and linguistic knowledge and resources, such as lexica and pronunciation dictionaries, are sparse. Written Luxembourgish is not systematically taught to children in primary school: German is usually the first written language learned, followed by French.

This paper aims to gain more insight into the acoustic properties that define the Luxembourgish language in the light of its Germanic and Romance influences. In particular the goal is to determine the influence of the German, French and the less practiced English language on the acoustic realization of Luxembourgish phonemes. The focus is on acoustic modeling whilst making use of phonemically aligned audio data. The following questions are addressed. First, if multiple monolingual acoustic models are available for alignment of Luxembourgish audio data, is there a clear preference for one of the languages? Second, is a language-specific preference observed for specific phonemes or for phoneme classes? If so, how do they correspond to IPA symbol matches between the preferred language and Luxembourgish phonemes? Third, how do language-specific preferences, if any, fare when compared with multilingual phone models? These issues have important implications for acoustic model for automatic speech recognition and handling pronunciation variants. The next section introduces the phonemic inventory of Luxembourgish and its correspondance with the three source languages (German, French, and English). Results are then presented with both monolingual and multilingual seed models. Finally, section 6. summarizes the results and discusses future challenges for speech technology and linguistic studies of Luxembourgish.

## 2. Phonemic inventory of Luxembourgish

The adopted Luxembourgish phonemic inventory includes a total of 60 phonemic symbols including 3 extra-phonemic symbols (for silence, breath and hesitations). Table 1 presents a selection of the phonemic inventory together with illustrating examples (see [1] for more information. on the phonemic inventory of Luxembourgish). Luxembourgish is characterized by a particularly high number of diphthongs. To minimize the phonemic inventory size, we could have chosen to code diphthongs using two consecutive symbols, one for the nucleus and one for the offglide (e.g. the sequence /a/ and /j/ for diphthong a͜ɪ). We prefered, however, the option of coding diphthongs and affricates using specific unique symbols. Given the importance of

**Table 1:** Sample cross-lingual phoneme mappings: Lux. targets mapped to same or similar phonemes in 3 source lang. (Fre, Ger, Eng).

| Carrier word (Eng) | Lux | Fre | Ger | Eng |
|---|---|---|---|---|
| ORAL VOWELS | | | | |
| l**ii**cht (light) | i | i | i | i |
| sch**ü**tzen (shelter) | ʏ | y | ʏ | ɪ |
| f**äe**g (able) | ɛ: | ɛ | ɛ: | ɛ |
| DIPHTHONGS | | | | |
| l**éi**en (to tell lies) | eɪ̯ | e | e | e |
| l**ou**nen (to hire) | ɔʊ̯ | o | o | o |

**Table 2:** Phoneme and training data information (in hours) for native and pseudo-Lux. acoustic models from English, French, German and for the multilingual superset(E,F,G).

| *Language* | *#native phon.* | *training data (h)* | *#Pseudo-Lux. ac. models* |
|---|---|---|---|
| English | 48 | 150 | 60 |
| French | 37 | 150 | 60 |
| German | 49 | 40 | 60 |
| Superset | - | 340 | 3x60 |

French imports, nasal vowels were included in the inventory, although they are not required for typical Luxembourgish words. Furthermore, native Luxembourgish makes use of a rather complex set of voiced/unvoiced fricatives. The final Luxembourgish inventory includes 60 symbols.

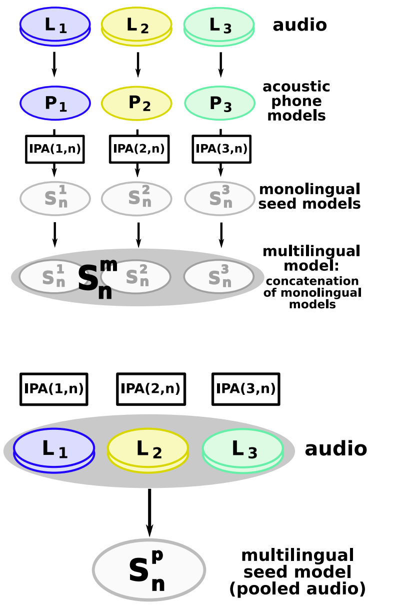
## 3. Acoustic seed models

The need to develop acoustic seed models for underresourced languages has already been addressed in previous research [5]. In the current study, three sets of context-independent acoustic models were built, one for each well-resources seed language (i.e., English, French, German). The models were trained on manually transcribed audio data (between 40 and 150 hours) from a variety of sources, using language-specific phone sets. The amount of data used to train the native acoustic models and the number of phonemes per language are given in Table 2 (left). Each phone model is a tied-state left-to-right, 3-state CDHMM with Gaussian mixture observation densities (typically containing 64 components).

Figure 1 (top) illustrates the development of three sets of pseudo-Luxembourgish acoustic models, each including 60 phones, starting from the English, French and German seed models and mapping the Luxembourgish phonemes to a close equivalent in each of the source model sets (IPA(i,n) in Fig. 1).

Table 1 shows a sample of the adopted cross-lingual mappings that were used to initialize seed models for Luxembourgish. Some symbols are used several times for different Luxembourgish phonemes. For the diphthongs that are missing in French, phonemes corresponding to the nucleus vowel were chosen. A fourth model set was then formed by concatenating the first three model sets, allowing the decoder to choose among the three models (see Table 2). Finally a set of multiligual acoustic models were trained (see Fig. 1, lower part) using the pooled E,F,G audio data that were labeled using their respective IPA(i,n) correspondances.

**Figure 1:** Acoustic seed models for a target language $n$ (Luxembourgish) given phone models $P_i$ of languages $L_i$ ($i = 1, 2, 3$ English, French, German) and IPA symbol correspondances between language $i$ and $n$ IPA(i,n). Top: monolingual $S_n^i$ and multilingual superset $S_n^m$. Bottom: pooled multilingual $S_n^p$ models.




## 4. Luxembourgish speech data

Forced alignments were carried using 80 minutes of manually transcribed speech from the House of Parliament (*Chamber* debates (70’) and from news (10’) broadcast by RTL, the Luxembourgish radio and TV broadcast company [3]. The detailed manual transcripts include all audible speech events, including disfluencies and speech errors. These *verbatim* transcripts were checked against the resulting word lists for errors and orthographic inconsistencies.

The average Luxembourgish phone segment duration remains relatively stable with respect to the different seed alignment (70-80ms). The corpus includes a total of 56,000 phone segments.

## 5. Results

In our earlier investigations [2] , a language change in acoustic models was permitted only on word boundaries. In that case, a clear preference for the" German" acoustic models could be noticed with more than 50% of phone segments aligned with the acoustic models arising from the German audio data. In the present study, the language identity of the acoustic models may change at any phone segment boundary. As a result, a large number of language switches between model sets are observed. The German models are less used. However, in all explored alignment conditions (three sets of models trained with monolingual English, French and German data, or adding also a fourth acoustic model set trained with the pooled audio data), the German models are always at rank one. In the alignment condition including the pooled model set, German models are aligned with 36% of segments, followed by the English set (27%) and the pooled model set (26.5%). The French models are used only for 10.5% of the segments on average. In all conditions, a preference for Germanic languages and in particular for the German language can be observed for Luxembourgish speech. This result is in agreement with the postulated influence of typological distance.

When defining the IPA correspondances between the Luxembourgish and English, French and German phoneme repertoires, there were many shared IPA symbols (e.g. plosives). For some of the symbols (e.g. diphthongs) however, a correspondance was fixed approximatively (i.e., if the diphthong does not exist in the language of the audio training data, the corresponding nucleus vowel was assigned to the Luxembourgish diphthong). In the following, we investigate the alignment results as a function of the closeness of the IPA match.

For each phoneme symbol, all observed segments were aligned with one of the provided acoustic model sets (En, Fr, Ge, Pooled). Rates were computed per language origin of the acoustic model set. When not considering the pooled models, the English, French and German rates sum up to 100% for each symbol (results on the top in Figure 2 ). When the pooled models are added (results on the bottom), the 4 model set rates sum up to 100%. Here, the rate of the pooled model and the related decrease in the monolingual model rates give an indication of the matching accuracy between target speech and the monolingual source models.

**(i) symbols shared among languages : plosives** The plosives /p/, /t/ and /k/ with their voiced counterparts exist in the four languages. As Lëtzebuergesch is considered a Germanic language, we may hypothesize that plosives be realized similar to German and English. We may thus expect a stronger burst than in French plosives and positive VOTs. If major acoustic correlates are shared with German and English rather than with French, then plosive segments should be aligned with German or English models rather than with French ones. Detailed results are shown in the left panel of Figure 2. As expected, segments are distributed almost evenly among German and English models and only 10-20% of the segments used the French models. The Luxembourgish plosives /k/ and /b/ tend to prefer English models. When adding the pooled models (bottom, left), about 30% of the data switch to these models during alignment.

**(ii) approximate symbol mapping : diphthongs** The Luxembourgish phonemic repertoire includes a large number of diphthongs. Diphthongs are broadly represented in the various regional variants, but their use in "standard" Luxembourgish (as broadcasted through radio and TV) needs to be further studied. Figure 2 (mid) shows a less homogeneous picture than for the plosives. There is a tendancy to use the English models, Luxembourgish /ɛɪ̯ / segments were 90% aligned with English models, whereas the /ɔɪ̯ / segments show a clear preference for German (all occurrences stem from *euro* words and compounds). When adding the pooled models, /ɛɪ̯ / segments tend to move to the corresponding pooled model, whereas the rate of German models remains almost unchanged for the /ɔɪ̯ / segments. For these, the German model provides a good match, whereas the English model of /ɛɪ̯ / is superseded by the pooled model.

**(iii) approximate symbol mapping : nasal vowels** Nasal vowels may be used in Luxembourgish in French import words. Figure 2 (right) shows results for nasal and corresponding oral vowels. The rate of French model usage is very high for the ã and õ segments. When introducing pooled models, this rate tends to drop, indicating that there is a relatively weak match between Luxembourgish and French nasal vowels.

It is interesting to note the predominance of French models for the rounded front vowels (/y/ and /ø/) (not shown in Figure 2 due to lack of place). Similarly, German model rates are generally higher for tense vowels, whereas English models achieve higher rates for the lax vowels.

It can be seen that the pooled model (Figure 2 bottom) is used to align a large proportion of the observed phone segments (e.g. /ɛɪ̯ /) for some phonemes, whereas for other phonemes, the pooled model will only account for a relatively small percentage of data. These results thus give an indicative measure of match/mismatch between the acoustic realisations of a given pair of phonemes between source and target languages. In future work, we propose to use this measure to provide phonetic instructions for foreign language learning such as defining lists of the most difficult phonemes to learn, given an L1/L2 pair.

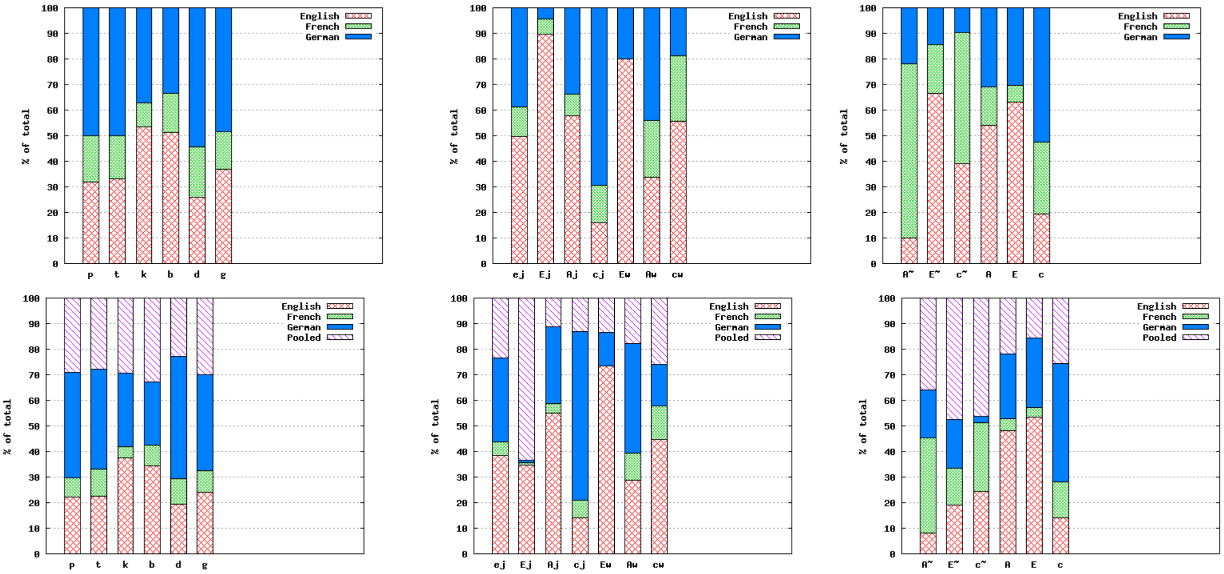



**Figure 2:** Alignment using the superset of monolingual English, French and German models (top) vs adding the pooled (bottom) models and optional model switch at phone boundaries. Model selection rates are shown per phoneme. IPA symbols in the caption are in the same order as the displayed symbols in the figures. Left: plosives (/p,t,k,b,d,g/ ; Mid: diphthongs /$e^j$,$ɛ^j$,$a^j$,$ɔ^j$,$ɛ^w$,$a^w$,$ɔ^w$/; Right: nasal /ɛ̃ ,ã,ɔ̃/ vs corresponding oral vowels /ɛ,a,ɔ/.

## 6. Summary and prospects

The present work focused on the issue of producing acoustic seed models for Luxembourgish, a language with strong Germanic and Romance influences and exploring their use in phonemic alignments. A phonemic inventory was defined and linked to inventories from major neighboring languages (German, French and English), using the IPA symbol set. For each of these languages, acoustic seed models were composed using either monolingual German, French or English acoustic model sets. In Luxembourgish speech alignments, a superset of multilingual acoustic seeds was used putting together the three language-dependent sets. The language-identity of the aligned acoustic models provides information about the overall acoustic adequacy of both the cross-language phonemic correspondances and the acoustic models. Furthermore, some information can be gleaned on inter-language distances. It was shown that the Germanic acoustic seed models provided the best match with 36% of the segments aligned using German seeds and 27% using the English ones. Only 10.5% used the French acoustic models. Since Luxembourgish is considered a Western Germanic language close to German, this result is in line with its linguistic typology. The remaining 26.5% of segments were used by the models estimated from the pooled multilingual audio data. The pooled model rates give an indicative measure of match/mismatch between the acoustic realisations of a given pair of phonemes between source and target languages. In particular, phoneme-based results revealed a particularly high match between Luxembourgish and English /ɛɪ̯ /, Luxembourgish and French /ø/. As a perspective, we propose to enhance this measure to provide help for foreign language learning, elaborating lists of potentially difficult phonemes given an L1/L2 pair.

Computational ASR investigations and corpus-based analyses will not only enhance the development of a more full-fledged ASR system for Luxembourgish, but can also be used to generate more specific predictions about lexical processing and phonetic learning in human listeners.

## 7. REFERENCES

[1] F. Schanen, *Parlons Luxembourgeois*, L'Harmattan, 2004.

[2] M. Adda-Decker, L. Lamel, N.D. Snoeren, "Comparing mono- & multilingual acoustic seed models for a low e-resourced language: a case-study of Luxembourgish.," in *Interspeech*, 2010.

[3] M. Adda-Decker, T. Pellegrini, E. Bilinski, and G. Adda, "Developments of Lëtzebuergesch resources for automatic speech processing and linguistic studies.," in *LREC*, 2008.

[4] C. Krummes, "Sinn si or si si? mobile-n deletion in luxembourgish," in *Papers in Linguistics from the University of Manchester: Proceedings of the 15th Postgraduate Conference in Linguistics*, Manchester, 2006.

[5] T. Schultz and A. Waibel, "Experiments on cross-language acoustic modeling," in *Proceedings of Eurospeech*, Aalborg, 2001.