Evaluation of Three Automatic Alignment Tools for the Processing of Non-native French

Qian Zhou[1], *Mathilde Hutin*[2,3]

[1]LiDiLEM, Université Grenoble-Alpes, Grenoble, France
[2]ATILF, Université de Lorraine, CNRS, Nancy, France
[3]ILC, Université catholique de Louvain, FNRS, Louvain-la-Neuve, Belgium
qian.zhou@univ-grenoble-alpes.fr, mathilde.hutin@cnrs.fr

Abstract

The production of non-native speech is known to display "cross-language phonetic interference", which makes such speech uneasy to align and label automatically. Automatic phonetic alignment refers to an automated process whereby software synchronizes speech with its transcription, usually at the phone and word levels. This method has proven useful and reliable for native speech, yet this reliability usually does not extend to non-native speech. This paper proposes to test three major automatic aligners (WebMAUS, MFA and SPPAS) on non-native French uttered by two native speakers of Chinese by comparing them with two manual segmentations. This paper's goal is to offer non-computer linguists a preliminary investigation on which to rely when choosing a tool for their studies in non-native phonetics or language didactics. Results show that the best performing tool for labeling is SPPAS while the best performing tool for both word- and phone-segmentation overall is WebMAUS and MFA the worst.

Index Terms: speech recognition, automatic alignment, non-native French

1. Introduction

The production of non-native speech, i.e., speech by second-language learners, is known to display so-called "cross-language phonetic interference" [1], commonly referred to as "non-native (or "foreign") accent". Although the systematic investigation of such interferences dates back at least 60 years [2, 3, 4, 5], its *automatic* investigation remains difficult, as automatic phonetic alignment is relatively recent and mostly trained on native speech. This in turn makes investigations in non-native phonetics more challenging than in general phonetics.

Automatic phonetic alignment refers to an automated process whereby software synchronizes speech with its transcription, usually at the phone- and word-levels. To that extent, the aligner is provided with the audio file and its transcript, usually in the orthography of the language. It then uses a grapheme-to-phoneme converter and pronunciation dictionaries to predict the possible phonetic transcriptions, as well as acoustic models corresponding to the language's phonemic inventory to identify the probable phone in the audio. The output is a segmented file such as a Praat textgrid, where the beginning and end of words and phones are timestamped (for the alignment at the word and phone levels respectively), and intervals are labeled (with words or phones respectively).

This method has proven useful and reliable. For instance, phone labeling realized with the MAUS software tool in the 1990s has been shown to match manual phone labeling $\sim$88% of the time, and $\sim$60% of timestamps differ less than 10ms between manual and automatic boundaries [6, 7]. However, this reliability usually does not extend to dialectal variation or non-native speech, as speech in a second language usually causes an increase in word error rate [8, 9, 10, 11, 12, 13, 14, 15]. There have been many attempts to improve the performance of automatic aligners, but what about current, readily-available, open-access tools that non-computer linguists, specializing for instance in second language learning, may use?

In the present paper, we aim to test the performance of 3 open-access, automatic alignment tools processing non-native French: WebMAUS [16, 17, 18], the Montreal Forced Aligner (henceforth MFA) [19] and SPPAS [20][1]. Our goal is to provide a review of these three automatic aligners so that linguists who do not specialize in computer linguistics but wish to align non-native speech, can use the best available tool.

2. Data and methodology

2.1. Data

For this study, we use the recordings made on a ZOOM H1N professional recorder in a quiet environment by two female native speakers of the Xiāng dialect, a Chinese language mainly spoken in the Hunan and Sichuan provinces. Both participants are 20 years old and in their 3rd year of undergraduate studies. Their language of instruction is Standard Mandarin, their 2nd language. English is their 3rd language, and French is their 4th. Their levels of French are respectively A2 and B2 of CECRL.

The two speakers read a list of 125 French words and French-like non-words, an experimental design inspired by the IPFC protocol [22] and a specialized manual [23]. For the present study, we create for each speaker a unique sound file comprising a subset of 44 French words displayed in (1) (non-words are usually not supported by automatic aligners) which are members of minimal pairs illustrating the voiceless *vs* voiced series, an opposition that does not exist in Mandarin Chinese or exists partially in Xiāng and is thus expected to exhibit pronunciation difficulties. In total, the study investigates at least 232 phonemes (not counting schwas).

(1) goût, dans, bac, dune, thon, code, qui, rogue, tousse, tais, cap, temps, logo, tire, rade, court, codé, gant, guide, tout, roc, con, tas, locaux, don, douce, gai, rate, dire, doux, dais, camp, coût, kid, bague, Guy, tune, da, cote, gond, quai, coté, cab, gourre

2.2. Methodology

For the present study, we compare 5 alignments. Two are manual alignments by trained phoneticians – one by a native speaker of French, which will be used as reference, and one by a native

[1]We initially planned on also testing EasyAlign [21], however this tool failed providing any alignment that could be used in this study.

speaker of Xiāng Chinese, for control. The other three are the output of three open-access aligners.

2.2.1. Automatic aligners

WebMAUS [18] is the online version of the Munich AUtomatic Segmentation (MAUS) system [16, 17]. In general, MAUS creates a pronunciation hypothesis graph based on the orthographic transcript of the recording using a grapheme-to-phoneme converter. During this process, the orthographic transcription is converted into the Speech Assessment Methods Phonetic Alphabet (SAMPA [24]). The signal is then aligned with the hypothesis graph and the alignment with the highest probability is chosen. The present alignment is the output of the WebMAUS alignment for French (no multiple varieties available) with the default setting (Version 3.16), resulting in a three-tier textgrid (words in French orthography, words in SAMPA, and phones).

The Montreal Forced Aligner (MFA, [19]) is an update of the Prosodylab-Aligner [25]. For the present study, MFA for French (no multiple varieties available) is used with the default settings by downloading the pronunciation dictionary for French and the French MFA acoustic model (using Miniconda3-py311_23.11.0-2). The alignment yields a textgrid with phone transcriptions in International Phonetic Alphabet (IPA) and word transcriptions in the orthography of French.

SPPAS [20] is an open-source software tool allowing the alignment of audio data with the specificity of proposing a high degree of customizability. For the present study, the SPPAS (4.13) alignment is generated using the default settings for the IPU search, and Standard French is used as the target language for normalization and phonetization, and ultimately, alignment. The process yields a textgrid file with 3 tiers: phones in SAMPA, words in French orthography and words in SAMPA.

The two sound files and the five alignments are available at this OSF repository: `https://osf.io/fg6y3/`.

2.2.2. Comparison of the 5 alignments

After setting the word and phone boundaries and labeling the intervals manually using SAMPA, and generating an automatic alignment from each of the described tools, we first convert MFA's IPA labels into SAMPA and calculate an inter-annotator agreement score for the labels between the reference manual alignment (that by the native French speaker), the control manual alignment (that by the native Chinese speaker) and the 3 automatic ones. To that extent, following [7], we calculate the percentage of corresponding labels (CL) between two segmentations of the same 44 words using the formula *CL=2nc/(n1+n2)*, where *nc* is the number of corresponding labels between two segmentations, and *n1* and *n2* are the number of labels in each segmentation respectively.

We then compare the time differences between the boundaries set by each test-aligner (the human and the three automatic tools) and those set by the reference aligner (the native speaker of French). To that extent, all times for phone and word boundaries for each aligner are extracted by tabulating the phone and word tiers in Praat and the deviation of boundaries of each aligner from the reference is calculated. As there are transcription differences among the aligners, only the same (e.g., <k∼k>) or similar phones (e.g., <k∼c>) are included to measure boundary alignment at the phone level, i.e., 153 intervals for the A2 speaker and 155 for the B2 speaker, for a total amount of 308 boundaries. The significance of the differences is established using an ANOVA test comparing each test-segmentation with the reference segmentation.

3. Results

3.1. Labels

First, we compare the labels used by the reference aligner (the native French speaker) and the other four.

As can be seen in Table 1 for the A2 speaker, labels are similar in 84.98% (between WebMAUS and MFA) to 97.94% (between SPPAS and the reference) of the cases (δ=12.96%). None of the differences are statistically significant, indicating that all aligners perform similarly well regarding labeling. Compared to the reference (native speaker of French), the aligner with the most differences is WebMAUS, with 90.21% similar labels, the one with the less differences is SPPAS, with 97.94% similarity (δ=7.73%). Compared to the other human aligner (native speaker of Chinese), MFA provides the most different labels, with only 87.39% similarity, while SPPAS still provides the most similar results, with 94.21% similarity (δ=6.82%).

Table 1: *Rates of similar labels between each test aligner (Human, WebMAUS, MFA and SPPAS) and the Reference for the A2 speaker. The p-values are the outcomes of a paired t-test.*

	Human	WebMAUS	MFA	SPPAS
Reference	95.00	90.21	91.30	97.34
Human	100.00	88.89	87.39	94.21
WebMAUS	88.89	100.00	84.98	86.08
MFA	87.39	84.98	100.00	90.52
SPPAS	94.21	86.08	90.52	100.00
p-value	0.7056	0.5031	0.5275	0.6216

As can be seen in Table 2 for the B2 speaker, labels are similar in 82.20% (between WebMAUS and MFA, again) to 96.20% (between the two human annotators) of the cases (δ=14.00%). None of the differences are statistically significant, indicating that, again, all aligners perform similarly regarding labeling. Compared to the reference (native speaker of French), the aligner with the most differences is MFA, with 88.31% similar labels, the one with the less differences is the human, with 96.20% similarity (δ=7.89%). The best automatic tool however is still SPPAS, with 93.16% similar labels (δ=4.85% compared to MFA). Compared to the other human aligner, MFA still provides the fewest similar labels (88.89%), while SPPAS is the automatic aligner with the most similar labels (93.67%, δ=4.78%).

Table 2: *Rates of similar labels between each test aligner (Human, WebMAUS, MFA and SPPAS) and the Reference for the B2 speaker. The p-values are the outcomes of a paired t-test.*

	Human	WebMAUS	MFA	SPPAS
Reference	96.20	92.05	88.31	93.16
Human	100.00	89.26	88.89	93.67
WebMAUS	89.26	100.00	82.20	86.19
MFA	88.89	82.20	100.00	89.18
SPPAS	93.67	86.19	89.18	100.00
p-value	0.7224	0.4703	0.6587	0.9539

In conclusion, the automatic aligner providing labels most similar to a human aligner is SPPAS, and the worst performing one is WebMAUS for the A2 speaker and MFA for the B2 speaker. The differences lie mainly in the confusion of the vowels /ɛ/ and /e/, /ɔ/ and /o/ or in the potential addition of /ə/ at the end of words. For MFA, differences are also registered in the phonetization of /k/ and /g/ before the vowel /i/, where they are transcribed as <c> and <ɟ> respectively, which may be viewed as more fine-grained than plain /k/ and /g/ but then

requires more post-processing by the linguist analyzing these stops.

3.2. Alignment at the word level

In this subsection, we investigate the time alignment of word boundaries. The density plots in Fig. 1 and 2 show the time difference of the word boundaries assigned by the 4 aligners in comparison with our reference (native speaker of French) for the A2 and B2 speakers respectively.

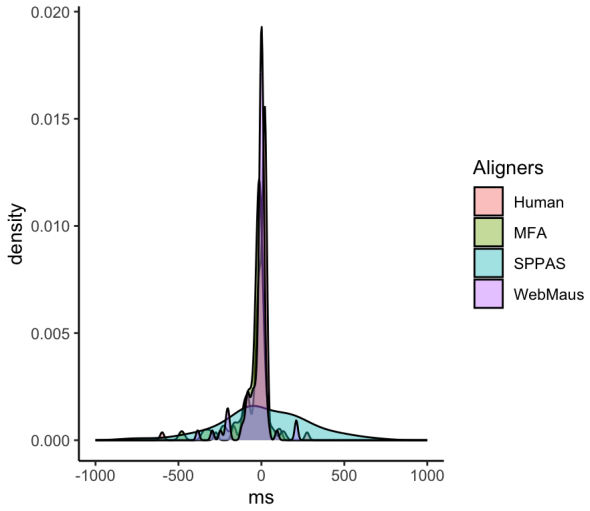

Figure 1: *Density plot for the distribution of word boundaries for each aligner (Human, MFA, SPPAS and WebMAUS) differing from the Reference as a function of time (from −1000 ms, i.e., the boundary by the aligner is set 1 second sooner than that set by the reference, to +1000 ms, i.e., the boundary by the aligner is set 1 second later than that set by the reference), based on the recording by the A2 speaker.*

In Fig. 1, it can be seen that aligners perform generally well, with a time difference between their boundaries and those of the reference ranging from −598 to +275 ms (δ=873, μ=−14.2, σ=89.02, p<0.0001) for the other human aligner, from −19,133 to +132 ms (δ=19,265, μ=−4778.96, σ=3917.64, p=0.7826) for MFA, from −797 to +706 ms (δ=1,503, μ=14.46, σ=273.10, p<0.0001) for SPPAS and from −384 to +212 ms (δ=596, μ=−29.49, σ=98.83, p<0.0001) for WebMAUS. The other human aligner (native speaker of Chinese) and WebMAUS perform extremely well, with respectively 88.64% and 87.5% of their word boundaries within a 0.1 second range from the reference boundaries. WebMAUS is thus the automatic aligner matching the reference the best, with a difference range of only 596 ms and 75% of boundaries set within a 0.05 second range of the reference boundaries. Because of some extreme outliers, MFA displays by far the largest difference range (19,265 ms) and matches the reference with only 55.68% of boundaries within the range from −200 to +200 ms, and displays the highest standard deviation (3,920 ms). SPPAS also performs less well than WebMAUS, with the second largest difference range after MFA (1,503 ms).

In Fig. 2, it can be seen that, again, aligners perform generally well, with a time difference between their boundaries and those of the reference ranging from −103 to +111 ms (δ=214, μ=7.3, σ=33.3, p<0.0001) for the other human aligner, from −9,282 to +134 ms (δ=9,416, μ=−533.04, σ=1732.6,

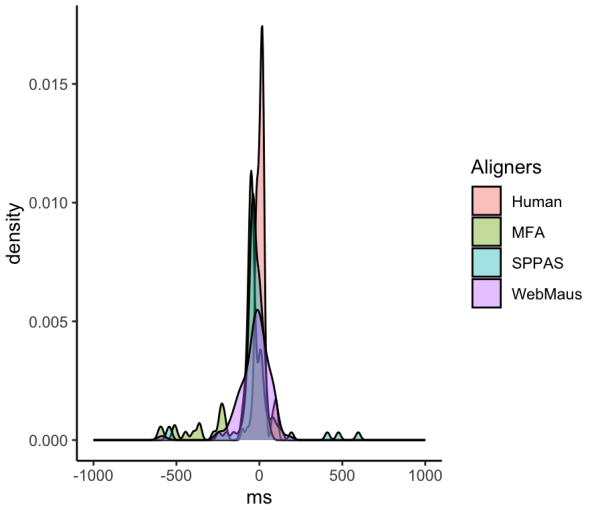

Figure 2: *Density plot for the distribution of word boundaries for each aligner (Human, MFA, SPPAS and WebMAUS) differing from the Reference as a function of time (from −1000 ms, i.e., the boundary by the aligner is set 1 second sooner than that set by the reference, to +1000 ms, i.e., the boundary by the aligner is set 1 second later than that set by the reference), based on the recording by the B2 speaker.*

p<0.0001) for MFA, from −552 to +2,624 ms (δ=3,176, μ=22.05, σ=334.92, p<0.0001) for SPPAS and from −250 to +104 ms (δ=354, μ=−29.5, σ=98.83, p<0.0001) for WebMAUS. In the order of their performance, the other human aligner, SPPAS, WebMaus and MFA all perform well, with respectively 96.59%, 86.36%, 78.41% and 72.73% of boundaries within a 0.1 second range from the reference boundaries. Of them, the human aligner is the aligner matching the reference the best, with 89.77% of word boundaries set within a 0.05 second range from the reference. Among the automatic tools, SPPAS performs the best (although less well than for the A2 speaker), with 68.18% of word boundaries within a 0.05 second range from the reference boundaries. WebMAUS also performs quite well, with 53.41% of word boundaries within a 0.05 second range of the reference boundaries, despite a few outliers around +0.5 second. Finally, MFA is still the tool with the worst performance, as it has the largest difference range (9,416 ms) and displays 13.64% of its boundaries between −9,282 and −250 ms.

To conclude, alignments at the word level by the non-native human or by the automatic tools are in general very close to the reference. WebMAUS and SPPAS are the best performing tools for the recordings by the A2 and the B2 speaker respectively, while MFA is the worst performing tool for both speakers.

3.3. Alignment at the phone level

In this subsection, we focus on the time alignment of phone boundaries. The density plots in Fig. 3 and 4 show the time difference of the phone boundaries assigned by the 4 aligners in comparison with our reference (native speaker of French) for the A2 and the B2 speakers respectively.

In Fig. 3, it can be seen that differences from the reference boundaries range from −139 to +182 ms (δ=321, μ=9.01, σ=44.96, p<0.001) for the second human aligner, from −18,612 to +93 ms (δ=18,705, μ=−4,025.36, σ=5,934.45,

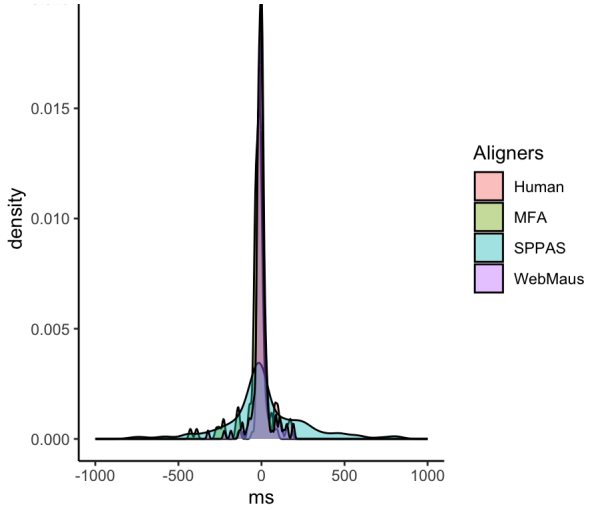

Figure 3: *Density plot for the distribution of phone boundaries for each aligner (Human, MFA, SPPAS and WebMAUS) differing from the Reference as a function of time (from −1000 ms, i.e., the boundary by the aligner is set 1 second sooner than that set by the reference, to +1000 ms, i.e., the boundary by the aligner is set 1 second later than that set by the reference), based on the recording by the A2 speaker.*

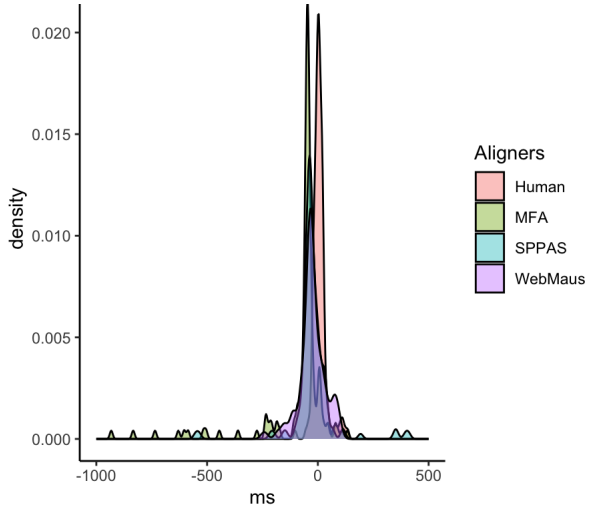

Figure 4: *Density plot for the distribution of phone boundaries for each aligner (Human, MFA, SPPAS and WebMAUS) differing from the Reference as a function of time (from −1000 ms, i.e., the boundary by the aligner is set 1 second sooner than that set by the reference, to +1000 ms, i.e., the boundary by the aligner is set 1 second later than that set by the reference), based on the recording by the B2 speaker.*

p<0.001) for MFA, from −768 to +815 ms (δ=1,583, μ=17.66, σ=246.93, p<0.001) for SPPAS and from −322 to +3,519 ms (δ=3,841, μ=18.59, σ=290.46, p<0.001) for WebMAUS. The human alignment is performing the best, with the smallest difference range, the smallest mean difference, and the smallest standard deviation, 87.58% of its boundaries within a 0.05 second range from the reference, and 100% of its boundaries within a 0.5 second range. Comparatively, WebMAUS has quite a high difference range of 3,841 ms but 83.66% of its boundaries are set within a 0.05 second range from the reference and 99.35% within a 0.5 second range. SPPAS and MFA perform less well, with only 37.25% and 51.63% of their boundaries within a 0.05 second range from the reference, but SPPAS has 93.46% of its boundaries set within a 0.5 second range from the reference while MFA has only 63.4%.

In Fig. 4, it can be seen that differences from the reference boundaries range from −103 to +136 ms (δ=239, μ=7.42, σ=30.02, p<0.001) for the second human aligner, from −5,565 to +134 ms (δ=5,699, μ=−355.57, σ=1138.73, p<0.001) for MFA, from −552 to +4,073 ms (δ=4,625, μ=109.86, σ=637.44, p<0.001) for SPPAS and from −250 to +104 ms (δ=354, μ=−26.49, σ=59.58, p<0.0001) for WebMAUS. The human alignment is the best performing one, with 92.90% of phone boundaries set within a 0.05 second range from the reference. Among the automatic aligners, WebMAUS and SPPAS seem to be the two matching the reference the best, with respectively 89.68% and 85.81% boundaries set within a 0.1 second range from the reference. In Fig. 4, MFA seems to be matching the reference quite well, yet only 78.71% of its boundaries are set within a 0.1 second range from the reference and it also displays many outliers in the negative range. If set within a 0.5 second range from the reference, the human alignment and WebMAUS perform best with 100% each, followed by SPPAS with 93.55% and MFA with 89.03%.

In conclusion, the phone boundaries appear to be more problematic, or at least more prone to variation, than word boundaries. In such a list-reading task, this is not surprising, since the words are separated by silences, while phones are pronounced as a sequence, which causes co-articulation, where it can be difficult to establish when a sound starts and stops. This indicates that connected speech may present less precise alignments than list readings for non-native speech. Among the automatic aligners, WebMAUS is the one performing the best at the phone-level and MFA the one performing the worst in general.

4. Discussion

We present here a preliminary evaluation of phone labeling and word- and phone-segmentation by 3 automatic aligners and 1 human aligner compared to a reference human aligner. Our results show that SPPAS is the tool providing labels closest to the human aligner and WebMAUS (although outperformed by SPPAS on word boundaries for the B2 speaker) is generally the one providing the closest word- and phone-boundaries, while MFA is the one providing the most different boundaries both at the word and phone levels. These results invite us to advise second-language or didactics specialists to align their non-native audio data with SPPAS if their interest lies with label precision or with WebMAUS if they need data best aligned in terms of timing.

This pilot study offers an initial benchmark for non-specialists to choose the right tool in exploring non-native speech, thus paving the way towards more inclusive phonetic studies. Future work will build on this methodology to explore connected speech from more native speakers of Chinese to confirm our results, but also from native speakers of other languages to test whether the present performances hold for French uttered by speakers of other native languages. We will also investigate the same tools with more fine-tuned parametarization.

5. Acknowledgements

This work was partially supported by an F.R.S.-FNRS research grant to Mathilde Hutin (project PPaDisM).

6. References

[1] J. E. Flege and R. Port, "Cross-Language Phonetic Interference: Arabic to English," *Language and Speech*, vol. 24, no. 2, pp. 125–146, 1981. [Online]. Available: https://doi.org/10.1177/002383098102400202

[2] F. Chreist, *Foreign Accent*, ser. Foundations of speech pathology series. Prentice-Hall, 1964. [Online]. Available: https://books.google.fr/books?id=m-i0AAAAIAAJ

[3] K. Wiik, *Finnish and English Vowels: A Comparison with Special Reference to the Learning Problems Met by Native Speakers of Finnish Learning English*, ser. Annales Universitatis Turkuensis. Turun Yliopisto, 1965, no. vol. 94. [Online]. Available: https://books.google.fr/books?id=fQ0HAQAAIAAJ

[4] E. J. Brière, "An investigation of phonological interference," *Language*, vol. 42, no. 4, pp. 768–796, 1966. [Online]. Available: http://www.jstor.org/stable/411832

[5] S. Singh and J. W. Black, "Study of Twenty-Six Intervocalic Consonants as Spoken and Recognized by Four Language Groups," *The Journal of the Acoustical Society of America*, vol. 39, no. 2, pp. 372–387, 02 1966. [Online]. Available: https://doi.org/10.1121/1.1909899

[6] M.-B. Wesenick and A. Kipp, "Estimating the quality of phonetic transcriptions and segmentations of speech signals," in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP 1996)*, 1996, pp. 129–132.

[7] A. Kipp, M.-B. Wesenick, and F. Schiel, "Automatic detection and segmentation of pronunciation variants in German speech corpora," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, vol. 1, 1996, pp. 106–109 vol.1.

[8] S. M. Witt, "Use of speech recognition in computer-assisted language learning," Ph.D. dissertation, Department of Engineering, University of Cambridge, 2000. [Online]. Available: https://www.repository.cam.ac.uk/handle/1810/251707

[9] Z. Wang, T. Schultz, and A. Waibel, "Comparison of acoustic model adaptation techniques on non-native speech," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, vol. 1, 2003, pp. I–I.

[10] G. Stemmer, S. Steidl, C. Hacker, and E. Nöth, "Adaptation in the pronunciation space for non-native speech recognition," in *Proc. Interspeech 2004*, 2004, pp. 2901–2904.

[11] Y. Liu and P. Fung, "Multi-accent Chinese speech recognition," in *Proceedings of Interspeech 2006*, 2006. [Online]. Available: https://doi.org/10.21437/Interspeech.2006-34

[12] Y. R. Oh, J. S. Yoon, and H. K. Kim, "Acoustic model adaptation based on pronunciation variability analysis for non-native speech recognition," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 1, 2006, pp. I–I.

[13] T.-P. Tan and L. Besacier, "A French Non-Native Corpus for Automatic Speech Recognition," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk, and D. Tapias, Eds. Genoa, Italy: European Language Resources Association (ELRA), May 2006. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2006/pdf/33_pdf.pdf

[14] H. Tsubaki and M. Kondo, "Analysis of L2 English speech corpus by automatic phoneme alignment," in *Proceedings of Speech and Language Technology in Education (SLaTE 2011)*, 2011, pp. 141–144.

[15] J. v. Doremalen, C. Cucchiarini, and H. Strik, "Automatic pronunciation error detection in non-native speech: The case of vowel errors in Dutch," *The Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. 1336–1347, 08 2013. [Online]. Available: https://doi.org/10.1121/1.4813304

[16] F. Schiel, "Automatic phonetic transcription of non-prompted speech," in *Proceedings of the XIVth International Congress of Phonetic Sciences : ICPhS 99 ; San Francisco, 1 - 7 August 1999*, J. J. Ohala, Ed., San Francisco, 1999, pp. 607 –610. [Online]. Available: http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-13682-6

[17] ——, "MAUS goes iterative," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, and R. Silva, Eds. Lisbon, Portugal: European Language Resources Association (ELRA), May 2004. [Online]. Available: https://aclanthology.org/L04-1058/

[18] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech Language*, vol. 45, pp. 326–347, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0885230816302418

[19] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Proc. Interspeech 2017*, 2017, pp. 498–502.

[20] B. Bigi, "SPPAS: A tool for the phonetic segmentations of speech," in *The eighth international conference on Language Resources and Evaluation*, Istanbul, Turkey, May 2012, pp. 1748–1755. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00983701

[21] J.-P. Goldman, "Easyalign: an automatic phonetic alignment tool under praat," in *Proc. Interspeech 2011*, 2011, pp. 3233–3236.

[22] I. Racine, S. Detey, F. Zay, and Y. Kawaguchi, "Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2: l'exemple du projet "Interphonologie du français contemporain" (IPFC)," *Recherches récentes en FLE*, pp. 1–19, 2012.

[23] D. Abry and J. Veldeman-Abry, *La phonétique: audition, prononciation, correction*. CLE International, 2007.

[24] J. Wells, *SAMPA computer readable phonetic alphabet*. Mouton de Gruyter, 1997, vol. Part IV, no. section B. [Online]. Available: www.phon.ucl.ac.uk/home/sampa

[25] K. Gorman, J. Howell, and M. Wagner, "Prosodylab-aligner: A tool for forced alignment of laboratory speech," *Department of Linguistics Faculty Scholarship and Creative Works*, vol. 2, 2011. [Online]. Available: https://digitalcommons.montclair.edu/linguistics-facpubs/2