Syllable-based Regional Swiss French Accent Identification using Prosodic Features

Alexandros Lazaridis[1], Jean-Philippe Goldman[2], Mathieu Avanzi[3] and Philip N. Garner[1]

[1]Idiap Research Institute, Martigny, Switzerland
[2]University of Geneva, Geneva, Switzerland
[3]LLF, UMR 7110, Paris Diderot, France
<alaza@idiap.ch>

Résumé

Il est question dans cette étude de reconnaissance automatique d'accent régional parmi 4 variétés de Suisse Romande. La variation dite régionale trouve son origine à la fois dans la prononciation des mots, soit le domaine segmental, mais également dans le style de parole, c'est-à-dire des propriétés prosodiques et plus précisément les propriétés rythmiques, mélodiques et d'intensité. Ce sont ces dernières que nous proposons d'exploiter pour une localisation basée sur l'unité syllabique et au moyen d'une méthode de classification connue que sont les support vector machines (SVM).

Keywords: *Swiss French, Regional Accent Identification, Support Vector Machines, syllable-based approach, prosodic features, Jitter-Shimmer features*

1. Introduction

In human communication, various linguistic and para-linguistic aspects of speech convey information about gender, age, emotions, emphasis, contrast and even the regional and social accents of the speaker (Laver, 1994). Humans through interaction with each other, over the years learn to some extent, to identify and interpret most of these aspects of speech. In research a lot of effort has been made to automatically, identify this kind of information from speech, such as emotion recognition (Schuller et al., 2003), gender and age recognition (Bocklet et al., 2008).

One of these aspects, embodied in speech, is the accent/dialect information. Dialect variations, in contrast to accent variations, are characterized by differences mainly in word selection and use of the grammar in a language. On the other hand, the main aspects that characterize accents are the diversities in pronunciation (phone sequence) and in the speaking style (rhythm, variation in pitch) (Racine et al., 2013; Lodge, 1993). Additionally, foreign and regional accents are the two subcategories of accent variations. The former characterizes the varia-

tions in speech uttered by non-native speakers speaking a foreign language. The pronunciation of a word might vary a lot depending on the level of the foreign language proficiency of the speaker and the native language of the speaker. The latter case, regional accents, refers to the changes in pronunciation but mainly in speaking style (Lodge, 1993; Racine et al., 2013; Woehrling & de Mareüil, 2006; Leemann, 2009) among native speakers of a language. This fact makes the tasks of differentiating them and identifying the origin/region of the speaker even more difficult.

A lot of research has been done over the last years in the field of automatic foreign accent and dialect recognition (Biadsy, 2011; Russell & Carey, 2013; Huang et al., 2007; Mporas et al., 2008). The main contribution of this work is for building robust automatic speech recognition (ASR) systems which are not influenced by the foreign accent of the speaker or are adapted to the dialect of the speaker (Humphries & Woodland, 1997; Biadsy et al., 2010). On the other hand, there is very limited research done on regional accent identification (RAI). Recently, in the work by Russell & Carey (2013), for identifying 14 British English regional accents, a framework of Gaussian mixture model - universal background model (GMM-UBM), a GMM-SVM model and GMM tokenization combined with n-gram language model (LM) were used. The evaluation results showed that the GMM-SVM approach achieved the highest identification accuracy score. Demarco & Cox (2012), in the same task, using the same database as the previous work, compared i-vectors to GMM-SVM concluding that no advantage was gained from the use of i-vectors. Regional accent identification (RAI) can help in personalizing synthetic speech of a text-to-speech (TTS) system according to a speaker of a specific regional accent. Consequently, RAI can also be beneficial for personalizing a speech-to-speech translation (S2ST) system for synthesizing the recognized and translated speech from one language to a specific regional accent in another language (Liang et al., 2010). To the extent of our knowledge, no previous work has been done on the regional accent identification task of French or Swiss French accents, apart from the recent work by Lazaridis et al. (2014), where they relied on a generative probabilistic framework for classification based on Gaussian mixture modelling (GMM) to automatically recognize the speaker's accent among regional Swiss French accents. Two different GMM-based algorithms were investigated: (1) the baseline technique of universal background modelling (UBM) followed by maximum-a-posteriori (MAP) adaptation, and (2) total variability (i-vector) modelling, with the i-vector-based system outperforming the baseline one.

Woehrling & de Mareüil (2006), conducted a study on human accent identification and how the background of the listeners affects their perception of the accents of 6 Francophone regions, i.e. Normandy,

Vendée, Romand Switzerland, Languedoc and Basque Country. The listeners from two different regions (Paris and Marseille) achieved an average of approximately 43% of accuracy on human regional accent identification, verifying the difficulty of the RAI task in French accents.

This paper is a preliminary work on attempting to automatically recognize the speaker's accent among regional Swiss French accents from four different regions of Switzerland, i.e. Geneva (GE), Martigny (MA), Neuchâtel (NE) and Nyon (NY). Among these regional accents, the variations in speech occur in both segmental and suprasegmental domains. These differences are subtle and thus can not be considered as phonological differences. For instance, some typical attested variations lie with the realisation of the primary accent. In the segmental side, some differences mainly concern the realisation of /o/, /R/ or some nasal vowels, but are very sporadic. In other words, the variations are mainly focused on the speaking style, i.e. different rhythm and pitch variations, rather than on the pronunciation of the words (Lodge, 1993; Racine et al., 2013; Woehrling & de Mareüil, 2006), making the task of regional accent identification even more difficult. To achieve this goal a syllable-based classification framework is implemented using prosodic features extracted from the speech signal. Since, among these regional accents, the variations in speech are mainly originated from the speaking style, i.e., different rhythm and pitch variations, rather than from the pronunciation of the words (Racine et al., 2013; Woehrling & de Mareüil, 2006), our hypothesis is that focusing mainly on features related to variations in pitch, intensity and rhythm, can be beneficial for the RAI task. For the classification task, a well known, widely used machine learning algorithm was used, i.e. support vector machines (SVM).

The rest of the paper is organized as follows. In section 2, the Swiss French speech database is described. The feature sets used in this work is presented in section 3. The experimental protocol and results are described in section 4. Finally the conclusions are given in section 5.

2. Swiss French Accent Database

Our material consists of speech samples extracted from the Phonologie du Français Contemporain (PFC) database (Durand et al., 2009). We analyzed a subset of a larger dataset processed in the frame of a project dealing prosodic variation in European French (Avanzi, 2014). Data from 32 speakers, 8 of each variety were selected. Care was taken to select four females and four males in each of the seven variants, and to control the speaker's age across the variants (ANOVA[1] tests reveals that the age of the speakers is similar among the 4 groups of speakers

[1] http://en.wikipedia.org/wiki/Analysis_of_variance

($F(3, 32) = 0.308$, n.s.), between male and female speakers ($F(1, 32) = 0.04$, n.s.) and between male and female speakers across the 4 groups ($F(5, 32) = 0.32$, n.s.).

Participants were instructed to read a journalistic text (the text used in the PFC project, which includes 398 words (22 sentences)) and to converse in pairs for 20-25 minutes. The entire reading text and a stretch of 3 minutes of spontaneous speech for each speaker were orthographically transcribed and automatically aligned within Praat (Boersma & Weenink, 2014) with the EasyAlign script (Goldman, 2011). All alignments were manually verified and corrected when necessary by inspecting both spectrogram and waveforms (i.e. boundary adjustments, segments deletion or addition in case of schwa, liaison, for example). All in all, the corpus is approximately 3 hours and 20 minutes long. Both types of speaking styles, i.e. read and conversational material, were used as unified speech for each speaker in our experimental setup. In Table 1 more information concerning the working dataset is presented.

Accent Groups	Age ranges	Mean Age (s.d.)	Dur. (s.d.)	Total Phones (s.d.)	Total Sylls (s.d.)	Total Tokens (s.d.)
GE	21-59	44.3 (17.9)	2946 (25.6)	23564 (129)	10386 (63)	7296 (82)
MA	22-79	48.8 (27.6)	2773 (31.2)	22421 (131)	9863 (57)	6845 (61)
NE	25-78	52.5 (24.1)	3289 (27.1)	22150 (135)	9679 (57)	6740 (58)
NY	30-70	46.2 (17.1)	3002 (27.1)	22243 (118)	9721 (44)	6799 (44)

Age: in years, Duration: in seconds

Table 1: *Database summary: Ranges and mean age, duration of speech, total number of phones, syllables and tokens for each of the 4 groups of speakers*

Furthermore, an experiment was conducted online to rate these 32 speakers with respect to their degree of regional accent. One sentence was chosen within the text and the corresponding audio was extracted for the 32 speakers. Theses extracts were randomly presented, in a website[2], to 37 subjects who were asked to rate the degree of accent of each speaker from *No accent* to *Marked accent* on a slider (with hidden values from 1 to 5). The mean value and standard deviation are shown by sites (accents) in Table 2. The degree of accent is different for the 4 groups ($F(3, 668) = 47.22$, $p < 0.001$). Post-hoc tests show significant difference between all groups except for the MA-NE pair.

3. Feature Set

In this work we are interested in focusing on syllable-based prosodic features in the task of RAI for identifying Swiss French regional accents. There are three main aspects of prosody, i.e. duration, pitch and inten-

[2] http://www.labguistic.com

Accent Groups	Mean Accent Degree (s.d)
GE	2.57 (1.09)
MA	3.32 (0.97)
NE	3.37 (0.89)
NY	3.75 (0.86)

Table 2: *Accents' Degree: Mean and standard deviation of degree of accent (on a scale from 1 to 5) rated by 37 subjects*

sity (Dutoit, 1997). The prosodic features are characterized as suprasegmental parameters since they are correlated with segments of speech larger than phones, i.e. syllables, words or even phrases. Additionally, syllable is a phonetic structure appropriate for properly modelling the prosodic events in speech (Atterer & Ladd, 2004; Xu & Wallace, 2004). Consequently, syllables were chosen as the basic units in our work.

Based on the three prosodic aspects mentioned above, we created two feature sets which were used for training the RAI models. The first feature set is constituted of the following nine features: two tilt parameters (Taylor, 2000) i.e. the amplitude tilt and the duration tilt. Since the amplitude tilt parameter cannot represent the absolute values of f0, the difference between the maximum and minimum values of f0 in a syllable was also used. Furthermore, in respect to duration, the following features were used: the number of voiced samples, the number of unvoiced samples, along with their respective ratios to all the samples of the syllable. Finally, the discrete (Legendre) orthogonal polynomial (DLOP) coefficients were used (Neuman & Schonbach, 1974). The DLOP coefficients have been shown to be capable of capturing speaker identity in speech synthesis (Hsia et al., 2010) and also have been proven to be very useful in the task of speaker verification (Dehak et al., 2007). Furthermore, in a recent work, the ability of achieving high accuracy pitch contour reconstruction on a recognition/synthesis very low bit rate speech coder with a combination of HMM-based phonetic vocoder, and a syllable-based pitch encoding technique based on DLOP was shown (Cernak et al., 2013). Based on some preliminary experiments, the first two polynomials of DLOP were used in order to approximate the pitch contour, i.e. two DLOP coefficients. The temporal information (of one previous and one next syllable) was also included in the feature set. Consequently, the first feature set contains 27 features.

The second feature set is constituted of the features included in the first feature set, along with five jitter features, six shimmer features and one intensity feature. Jitter and shimmer are measures of cycle-to-cycle variations of the fundamental frequency and the amplitude respectively. These features have been widely used for detecting voice

pathologies (Kreiman & Gerratt, 2005), but over the last years they have also been used for identifying different human speaking styles and emotions (Li et al., 2007), or age and gender identification (Sadeghi & Homayounpour, 2006) and also in speaker recognition and verification task (Farrus & Hernando, 2009). The five jitter features are the following:

– (i) the jitter(relative) which is the average absolute difference between consecutive periods, divided by the average period,

– (ii) jitter(absolute) which is the average absolute difference between consecutive periods in seconds,

– (iii) the jitter(rap) which is the relative average perturbation i.e. the average absolute difference between a period and the average of it and its neighbours, divided by the average period,

– (iv) the jitter(ppq5) which is the five-point period perturbation quotient i.e. the average absolute difference between a period and the average of it and its four closest neighbours, divided by the average period,

– (v) the jitter(ddp) which is the average absolute difference between consecutive differences between consecutive periods, divided by the average period (Boersma & Weenink, 2014).

The six shimmer features are the following:

– (i) the shimmer(relative) which is the average absolute difference between the amplitude of consecutive periods, divided by the average amplitude,

– (ii) the shimmer(dB) which is the average absolute base-10 logarithm of the difference between the amplitude of consecutive periods in seconds, multiplied by 20,

– (iii) the shimmer(apq3) which is the three-point amplitude perturbation quotient i.e. the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude,

– (iv) the shimmer(apq5) which is the five-point amplitude perturbation quotient i.e. the average absolute difference between the amplitude of a period and the average of the amplitude of it and its four closest neighbours, divided by the average amplitude,

– (v) the shimmer(apq11) which is the 11-point amplitude perturbation quotient i.e. the average absolute difference between the amplitude of a period and the average of the amplitude of it and its ten closest neighbours, divided by the average amplitude,

– (vi) the shimmer(ddp) which is the average absolute difference between consecutive differences between the amplitudes of consecu-

tive periods (Boersma & Weenink, 2014).

Finally the difference between the maximum and minimum values of the intensity was used. The temporal information (of one previous and one next syllable) was also included in the feature set. In this way, the second feature set contains 63 features.

4. Experiments

In this paper we are interested in validating two hypotheses. Firstly, that the syllable-based approach using prosodic features could be used in the regional accent identification task. Secondly, whether the jitter/shimmers features could be beneficial or not, in the RAI task.

4.1. Experimental Setup

The experimental evaluation is conducted on the PFC dataset using a cross-validation technique: out of the eight speakers of a specific regional accent, seven of them are selected for the training, and the remaining one is used for the testing (i.e. all the possible combinations $8^4 = 4096$ folds). Since this is a preliminary work in order to test our hypothesis, only 200 folds were randomly selected out of the 4096 ones. For the classification task, a well known and widely used machine learning algorithm was used, i.e. support vector machines (SVM).

4.1.1. Support Vector Machines

A Support Vector Machine (SVM) constructs a hyperplane in a high-dimensional space, which can be used for classification (SVM) and regression (SVR) tasks (Smola & Scholkopf, 1998). The basic idea governing the SVM is the production of a model that can be expressed through support vectors which define the hyperplane. A function is used to approximate the training instances by minimising the prediction error. A parameter ε defines the level of accuracy of the approximation function. In this tube the errors are ignored. The parameter ε controls how closely the function will fit the training data. The parameter C is the penalty for exceeding the allowed deviation defined by ε. The larger the C, the closer the approximation function can fit the data (Witten & Frank, 2005).

For our experiments the SVM model (Platt, 1999), which employs the sequential minimal optimization (SMO) algorithm for training a support vector classifier (Smola & Scholkopf, 1998), was used. Many kernel functions have been used in SVM such as the polynomial, the radial basis function (RBF) and the Gaussian functions (Scholkopfand & Smola, 2002), etc. In this paper, after some preliminary experiments, the polynomial kernel was selected (Scholkopfand & Smola, 2002).

The ε and C parameters, where $\varepsilon \geq 0$ is the maximum deviation allowed during training and $C > 0$ is the penalty parameter for exceed-

Feature Sets	GE	MA	NE	NY	Total Accuracy
Set1	28.09%	32.10%	24.66%	50.00%	32.90%
Set2	32.58%	34.57%	28.77%	56.25%	37.13%

Table 3: *Performance summary: This table reports the accuracy of the GMM and TV-SVM systems*

ing the allowed deviation, were set equal to 0.001 and 1.0 respectively. These values were selected after a grid search fine tuning (ε={0.0005, 0.001, 0.003, 0.005}, C={0.5, 1.0, 1.5, 10}) of the model after some preliminary experiments.

4.2. Experimental Results

In Table 3, the accuracy of the SVM using the two feature sets for each regional accent is shown, along with the total accuracy of each of the two cases. As can be seen, the model trained using the feature set2 outperforms the model trained on feature set1, achieving a relative improvement of 12.9% in the total accuracy. More precisely, the highest improvement can be seen in the cases of NE and GE regional accents, where a relative improvement of 16.7% and 16% was achieved. Finally, the smallest improvement was shown in the case of MA, reaching a 7.7% relative improvement in the identification accuracy.

Table 4, shows the accuracy of the second model, using the feature set2, in respect to the size of the test utterances. In this table, the accuracy in respect to the number of syllables can be seen for the cases of evaluating on utterances with equal or more syllables than: one (all test utterances of each fold, approximately 310 utterances), ten (approximately 210 utterances) and twenty (approximately 80 utterances). It is clearly shown that as the number of the syllables increases, *i.e.* not using utterances with less syllables than a threshold, the accent identification accuracy improves. This can be contributed to the fact that the small utterances do not convey enough accent information so as to be correctly identified by the system.

The experimental results confirmed our two hypotheses showing firstly that the syllable-based approach using prosodic features could be used in the regional accent identification task. Secondly, it was shown that the jitter/shimmer features could be beneficial, in the RAI task.

5. Conclusions and Discussion

The objective of this paper was to automatically recognize the speaker's accent among 4 regional Swiss French accents by using a syllable-based identification framework trained on prosodic features extracted from the speech signal. The experimental results confirmed our two hypotheses: (i) the syllable-based approach using prosodic features could

Num. of Syllables	≥1	≥10	≥20
Set2	37.13%	42.20%	45.72%

Table 4: *Total accuracy in terms of number of syllables: This table shows the accuracy rates of the system using feature set2, evaluating utterances with equal or bigger number of syllables than a threshold*

be used in the regional accent identification task and (ii) that the jitter/shimmer features could be beneficial in the RAI task.

As noted in the introduction, this is a preliminary work on Swiss French regional accent identification, that the authors are interested in further developing. One of the first issues that should be dealt is the lack of data. Additional speakers will be added to the existing database as soon as they are available in PFC project. Furthermore, an in depth analysis of the differences among the regional accents, concerning the phonetic and mainly prosodic characteristics of speech, will be conducted. Consequently, we will be able to identify and focus more on the specific differences among these regional accents and take advantage of them for identifying them more accurately. Finally, the acoustic-based framework which was used in a recent work of ours, could be combined with the syllable-based technique presented in this work, in order to take advantage of the benefits of both of these approaches.

Acknowledgements

The authors would like to thank Xingyu NA for providing the implementation of the discrete (Legendre) orthogonal polynomial (DLOP) algorithm.

This work has been conducted with the support of the Swiss NSF under grant CRSII2 141903: Spoken Interaction with Interpretation in Switzerland (SIWIS) and by the Swiss NSF under grant: Advanced Postdoc-Mobility n. P300P1-147781.

References

Atterer, M., & Ladd, D. (2004). On the phonetics and phonology of "segmental anchoring" of f0: evidence from german. *Journal of Phonetics*, *32*(2), 177–197.

Avanzi, M. (2014). A corpus-based approach to french regional prosodic variation. *Nouveaux cahiers de linguistique française*, *31*.

Biadsy, F. (2011). *Automatic dialect and accent recognition and its application to speech recognition* (Unpublished doctoral dissertation). Columbia University.

Biadsy, F., Hirschberg, J., & Collins, M. (2010). Dialect recognition using a phone-gmm-supervector-based svm kernel. In *Interspeech* (pp. 753–756).

Bocklet, T., Maier, A., Bauer, J., Burkhardt, F., & Noth, E. (2008). Age and gender recognition for telephone applications based on gmm supervectors and Support Vector Machines. In *IEEE ICASSP* (Vol. 1, pp. 1605–1608).

Boersma, P., & Weenink, D. (2014). Praat, v. 5.3. http://www.fon.hum.uva.nl/praat.

Cernak, M., Na, X., & Garner, P. N. (2013). Syllable-based pitch encoding for low bit rate speech coding with recognition/synthesis architecture. In *Proceedings of interspeech 2013.*

Dehak, N., Dumouchel, P., & Kenny, P. (2007). Modeling prosodic features with joint factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(7), 2095–2103.

Demarco, A., & Cox, S. (2012). Iterative classification of regional british accents in i-vector space. In *Machine learning in speech and language processing.*

Durand, J., Laks, B., & Lyche, C. (2009). *Phonologie, variation et accents du français*. Paris: Hermés.

Dutoit, T. (1997). *An introduction to text-to-speech synthesis*. Dordrecht: Kluwer Academic Publishers.

Farrus, M., & Hernando, J. (2009, July). Using jitter and shimmer in speaker verification. *Signal Processing, IET*, *3*(4), 247–257.

Goldman, J.-P. (2011). Easyalign: an automatic phonetic alignment tool under praat. In *Interspeech* (pp. 3233–3236).

Hsia, C.-C., Wu, C.-H., & Wu, J.-Y. (2010). Exploiting prosody hierarchy and dynamic features for pitch modeling and generation in hmm-based speech synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(8), 1994–2003.

Huang, R., Hansen, J., & Angkititrakul, P. (2007). Dialect/accent classification using unrestricted audio. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(2), 453–464.

Humphries, J., & Woodland, P. (1997). Identification of foreign-accented french using data-mining techniques. In *Interspeech* (pp. 2367–2370).

Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *The Journal of the Acoustical Society of America*, *117*(4), 2201–2211.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lazaridis, A., Khoury, E., Goldman, J.-P., Avanzi, M., Marcel, S., & Garner, P. N. (2014, June). Swiss French regional accent identification. In *Proceedings of odyssey 2014: The speaker and language recognition workshop.* Joensuu, Finland. (To appear)

Leemann, A. (2009). *Comparative analysis of voice fundamental frequency behavior of four swiss german dialects* (Unpublished doctoral dissertation). Bern Universität.

Li, X., Tao, J., Johnson, M., Soltis, J., Savage, A., Leong, K., & Newman, J. (2007). Stress and emotion classification using jitter and shimmer features. In *IEEE international conference on acoustics, speech and signal processing (ICASSP 2007)* (Vol. 4, pp. 1081–1084).

Liang, H., Dines, J., & Saheer, L. (2010). A comparison of supervised and unsupervised cross-lingual speaker adaptation approaches for HMM-based

speech synthesis. In *IEEE ICASSP* (pp. 4598–4601).

Lodge, A. (1993). *French. from dialect to standard*. London: Routledge.

Mporas, I., Ganchev, T., & Fakotakis, N. (2008). Phonotactic recognition of greek and cypriot dialects from telephone speech. In *SETN 2008, advances in artificial intelligence, lecture notes in computer science* (Vol. 5138, pp. 173–181). Berlin/Heidelberg: Springer.

Neuman, C. P., & Schonbach, D. I. (1974). Discrete (legendre) orthogonal polynomials. a survey. *International Journal for Numerical Methods in Engineering*, *8*(4), 743–770.

Platt, J. (1999). Fast training of Support Vector Machines using sequential minimal optimization. In B. Schölkopf, C. J. C. Burges, & A. J. Smola (Eds), *Advances in kernel methods* (pp. 185–208). Cambridge: MIT Press.

Racine, I., Schwab, S., & Detey, S. (2013). Accent(s) suisse(s) ou standard(s) suisse(s)? Approche perceptive dans quatre régions de Suisse romande. In A. Falkert (Ed.), *La perception des accents du français hors de france* (pp. 41–59). Mons: Éditions CIPA.

Russell, A. H. M., & Carey, M. (2013). Human and computer recognition of regional accents and ethnic groups from british english speech. *Computer Speech and Language*, *27*(1), 59-74.

Sadeghi, A., & Homayounpour, M. (2006). Speaker age interval and sex identification based on jitters, shimmers and mean MFCC using supervised and unsupervised discriminative classification methods. In *8th international conference on signal processing.*

Scholkopfand, B., & Smola, A. (2002). *Learning with kernels*. Cambridge: MIT Press.

Schuller, B., Rigoll, G., & Lang, M. (2003). Hidden Markov model-based speech emotion recognition. In *IEEE ICASSP* (Vol. 2, pp. 1–4).

Smola, A., & Scholkopf, B. (1998). *A tutorial on Support Vector Regression* (Tech. Rep. No. NeuroCOLT Tech. Rep. TR 1998-030). London: Royal Holloway College.

Taylor, P. (2000). Analysis and synthesis of intonation using the tilt model. *Journal of the Acoustical Society of America*, *107*, 1697–1714.

Witten, H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. Morgan Kauffman Publishing.

Woehrling, C., & de Mareüil, P. B. (2006). Identification of regional accents in French: perception and categorization. In *Interspeech* (pp. 1511–1514).

Xu, Y., & Wallace, A. (2004). Multiple effects of consonant manner of articulation and intonation type on F_0 in English. *Journal of the Acoustical Society of America*, *115*(5), 2397.