DETERMINING LEARNER PRONUNCIATION MODELS WITH THE K-NN ALGORITHM

Nicolas Ballier

Univ. Paris Sorbonne Cité (Paris Diderot) / CLILLAC-ARP (EA3967)
nicolas.ballier@univ-paris-diderot.fr

Keywords: learner pronunciation models, vowels, French-English interphonology, K-NN algorithm,

1. INTRODUCTION

Most phonetic learner corpora are dictionary-based and transcribe a target / pronunciation model for learners [1], which is usually an American pronunciation model, because the electronic resources for British English are scarce. This paper proposes to investigate which reference pronunciation model learners aim to emulate, by comparing their realisations with two reference formant datasets, which represent pronunciation models for British and American speakers.

Reference varieties (and their putative rejection or adoption in the name of ELF) have been the topic of a heated debate, sometimes laden with political and economical considerations ([9], [12]). The whole debate ([5], [7], [10], [12]) and the economic issues surrounding prestige [9], as well as ideological and political implications have been commented upon, as well as the potential need for a "Lingua Franca Core" [12].

Various methods have been used to determine a variety of English on segmental criteria. Wieling et al. [13] have adopted another methodology to "evaluate the suitability of a computational pronunciation comparison method" and have used ready-made transcriptions from the ACCENT project archives and have shown the Levenshtein distance to be a good metrics for the measure of pronunciation distance, congruent with on-line native judgements.

2. METHOD

Our method is signal-based, analysing F1 and F2 learner performance for variables.

2.1. Reference datasets (training sets)

For our case study, we have focused on the two main competing pronunciation models for French students. French students are advised in the prescriptivist Agrégation reports to avoid a 'MidAtlantic' mix of British and American features (see also Cruttenden [6] discussing 'Amalgam English' and 'International English').

The reference value for American English used in this experiment were the ones included in the Phontool R package [3]. Data from [8, personal communication] were converted from barks to hertz to ensure comparability.

For the following assumed realisations of the nine vowels of the words *"had", "head", "heard", "heed", "hid", "hod", "hood","hud"* and *"who'd"*, datapoints were collected, yielding 273 comparable observations from [8] and 1,390 data points for [11].

Vowel formants were measured in the middle position of the vowel, the vowel interval boundaries were automatically determined by automatic pitch tracking of F0. This means that diphthongs were excluded from comparison.

2.2 The test set

The learner corpus is a longitudinal series of interviews of undergraduate students from the university of X. The formant values of the vowels of 13 speakers (3 males, 10 females) whose L1 is French were automatically extracted with a Praat [4] script (see [2] for a more detailed description of the protocol). We selected the male datapoints corresponding to the 9 monophthongs from our database, resulting in 6,354 tokens.

2.3. The k-nn algorithm

The k-nearest neighbour algorithm is a point-to-point classifier that assesses the Euclidian distance between each learner datapoint (for F1 and F2) and a variable number of neighbouring datapoints. For this paper, we have limited the experiment to a comparison between two varieties, but the principle can be extended to multiple classifications (and therefore as many reference pronunciation models as deemed sensible) as well as to other dimensions (F3, F4, vowel duration).

For each phonemic vowel type, F1 and F2 learner tokens were automatically compared to the k neighbouring datapoints established in the two reference studies (for Standard British English and for General American). For each datapoint (a vocalic realisation and its corresponding F1 and F2), the

distance between k nearest neighbours of each instance of the reference varieties (the training datasets, in our case, the result of British and American reference studies) was measured, and the system returned a score establishing whether the considered learner vowel was closer to a British or an American realisation. We have limited the investigation to male speakers (the only subjects studied in [8]) for better data comparability.

3. RESULTS

We used a tenth of the training set to check the consistency of the training data. Phoneme ellipses are notoriously messy and zones of overlaps exist (as evidenced in Figure 4 in [8]). In spite of these encroaching zones for the phoneme ellipsis within the two varieties, the F1/F2 values resulted in consistent results as to discrimination between varieties. With a 10-fold cross-validation, the confusion matrix for the training phase of the 1,663 data points reads as follows: only 7,39% of the tokens were misclassified. To minimise outliers, k is optimal with 12 neighbours.

```
=== Confusion Matrix ===
a    b   <-- classified as
174   99 |    a = SBE
24 1,366 |    b = GenAm
```

As to the test phase, male learner tokens were mostly classified as Gen Am (see percentages and row data in Table 1).

Table 1: Majority votes for the Gen Am reference variety per vowel per speaker

vowel	speaker1	speaker2	speaker3
had	77%(352)	64%(129)	78%(228)
head	78%(260)	50%(202)	78%(207)
hid	49.87%(794)	45%(655)	45%(655)
heed	49.17%(181)	31%(112)	26%(129)
heard	100 % (6)	61%(18)	0%(2)
hod	78%(360)	72%(201)	79%(295)
hood	37%(322)	25%(173)	17%(348)
who'd	33%(83)	24%(58)	15%(79)
hud	74%(241)	70%(156)	72%(196)

The two sets of problematic categorisations are *hid* vs. *heed* and *who'd* vs. *hood*. F3 values may refine the results for the latter. In two cases, the non-native datasets include CV vowel tokens, which are excluded from the hVd protocol; including CV tokens may have skewed the results, as the hVd reading lists in [8] and [11] preclude the comparison with CV realisational contexts.

We have submitted the recordings to six experts trained in phonetics to estimate the likely pronunciation model of each speaker to validate the assumptions. The six experts confirmed the attribution of the label (British/American) for the three speakers (k=1).

4. DISCUSSION AND FUTURE RESEARCH

Two main caveats need to be taken into account. Because students may not know some of the hVd words, we did not ask them to read those words, and instead relied on a more limited set of words for read speech in isolation. To increase the data, we relied on unscripted speech, where variation is more likely. The full version of the paper discusses:

- evolution of scores over the three year period
- applicability to female subjects using Vocal Tract Length Normalization (VTLN) techniques.
- the bias-variance trade-off and the optimisation of the hyper-parameter k (number of neighbours considered in the analysis)
- extension to other pronunciation models in multiclass comparisons (using supplementary data from [8]). For L2 speakers under British influence, one could fine-tune the accent detection between the thirteen accents under scrutiny in [8]
- the optimisation of the formant measurements in hertz, log(herz) and barks to better approximate speaker perception, the need to rescale duration variation (in ms) to make it compatible with bark/herz variation
- the strengths and weaknesses of the algorithm for this classification task / conformity metric.

The k-nn algorithm is dependent on the training sets, but more pronunciation models can be learnt by rote, so that finer-grained recognition of other reference accents can be taught. The algorithm can function as multiclass classification, eg judging between British, Singapore or American realisations, acquiring more expertise as formant reference values are fed onto the system as training datasets.

5. ACKNOWLEDGEMENT

Thanks are due to Emannuel Ferragne for granting access to their Ferragne & Pellegrino 2005 data and to Aurélie Fischer and Marie Candito for exchanges about the knn-algorithm.

Part of this research was carried out during a research leave granted by the French national accreditation board (CNU), for which grateful thanks are acknowledged.

6. REFERENCES

[1] Ballier, N. & Martin, Speech annotation of learner corpora. *Handbook of Learner Corpus Research*, CUP 107-134.

[2] Méli, A. & Ballier, N. Assessing L2 phonemic acquisition : a normalization-independent method. Submitted, Glasgow.

[3] Barreda, S. 2014. phonTools: Functions for phonetics in R. R package version 0.2-2.0. http://cran.r-project.org/web/packages/phonTools/phonTools.pdf

[4] Boersma, Paul & Weenink, David 2012. Praat: doing phonetics by computer [Computer program]. Version 5.3.19, retrieved 24 June 2012 from http://www.praat.org/

[5] Collins, B. S., & Mees, I. M. 2013. *Practical phonetics and phonology: a resource book for students*. Routledge.

[6] Cruttenden, A. 2014. *Gimson's Pronunciation of English*, Routledge.

[7] Dziubalska-Kołaczyk, K., Przedlacka, J. (eds.). 2008. *English pronunciation models: A changing scene* Bern: Peter Lang.

[8] Ferragne, E., Pellegrino, F. 2010. Formant frequencies of vowels in 13 accents of the British Isles. *JIPA* 40(1), 1–34.

[9] Fix, E., Hodges, J.L. 1951. Discriminatory analysis, nonparametric discrimination: Consistency properties. *Technical Report 4*, USAF School of Aviation Medicine, Randolph Field, Texas.

[10] Graddol, D. (2006). *English next*. London: British Council.

[11] Hillenbrand, J.M., Getty, L.A., Clark, M.J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *Journal of the Acoustical Society of America*, 97, 3099-3111.

[12] Jenkins, J. 1998. Which pronunciation norms and models for English as an International Language?. *ELT journal*, *52*(2), 119-126.

[13] Wieling, M., Bloem, J. Mignella, K. Timmermeister, M, Nerbonne, J. 2014. Measuring foreign accent strength in English: Validating Levenshtein distance as a measure. *Language Dynamics and Change*, 4(2), 253-269.