ASSESSING PHONEMIC ACQUISITION: A NORMALIZATION-DEPENDENT PROCEDURE?

Adrien Méli

Univ Paris Diderot – Sorbonne Paris Cité
CLILLAC-ARP – EA 3967
adrienmeli@gmail.com

Keywords: automatic formant extraction, phonemic acquisition, normalization, interphonology

1. INTRODUCTION

Spontaneous speech learner corpora are likely to feature skewed word and phoneme frequencies that violate the requirements of commonly used methods to analyze phonetic data: vowel-extrinsic normalization methods such as Nearey [7], Lobanov [4] or Watt & Fabricius [12] require that acoustic measurements for all the vowels of a speaker's system be collected in roughly the same amount.

The question therefore arises of the influence that methods of normalization may have upon assessing the evolution of learners' phonemic acquisition over time.

2. DATA

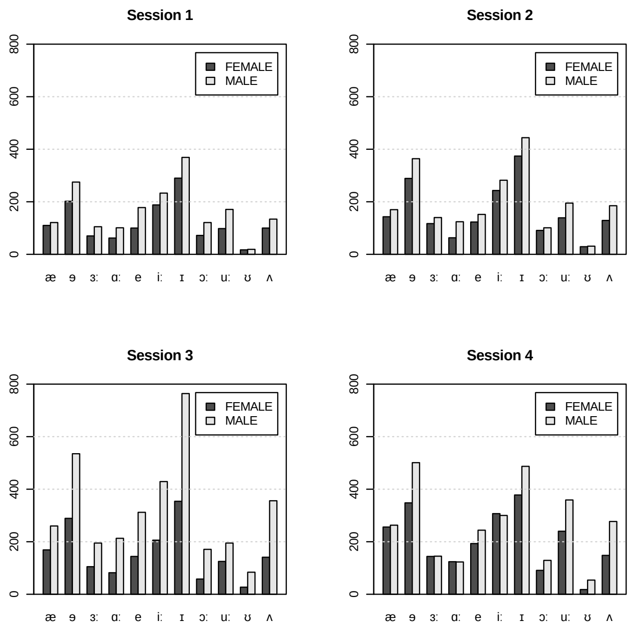
2.1. Corpus

The data used for this investigation come from a project by University Paris Diderot, in which 5 French native speakers, 3 male and 2 female students, were recorded in English four times over six-month intervals. 20 recordings were thus obtained. The interviews were conducted by a native English speaker, and recorded in an individual stereo 16-bit resolution sound file at a sampling rate of 44100 Hz captured in an uncompressed, pulse code modulation format using an Apex435 large diaphragm studio condenser microphone with cardioid polar pattern.

2.2. Procedure of acoustic extraction

The recordings were analyzed in the following fashion: the transcriptions of short, consistent sentences were aligned on a PRAAT TextGrid (Boersma [3]), which were then extracted and automatically aligned at the segmental level with SPPAS (Bigi [2]) using an American transcription of the CMU dictionary. For each vowel, a PRAAT script then collected, among other things, the mid-temporal F1, F2 & F3 values. 17,407 monophthongs were thus collected and their distribution can be found in Figure 1.

Figure 1: *Per-sex monophthong distribution across the four sessions.*



3. ANALYSIS

3.1. Normalization

The data was then normalized using the four following methods:

1. Traunmüller's [11] Bark method;
2. The Bark Difference Metric [10];
3. Nearey's [7] Extrinsic method (Nearey 2);
4. Lobanov's [4] z-score.

The computations were made using the statistical software R [9], and the PhonTools [1] package for the last two. As shown in Table 1, this set of methods features all possible combinations of Vowel/Formant ex-/in-trinsic computations. No speaker extrinsic method was adopted due to the differences in vowel

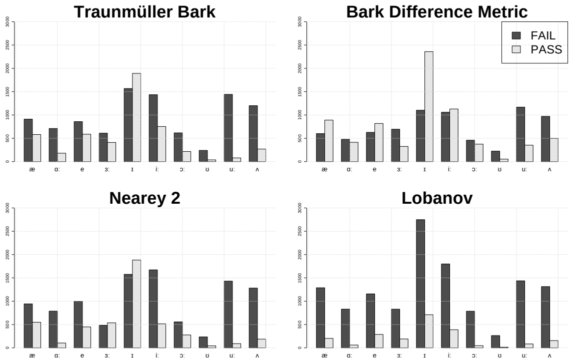
Table 1: *Reminder of the specificities of the normalization methods.*

Method	Vowel	Formant
Bark	Intrinsic	Intrinsic
BDM	Intrinsic	Extrinsic
Lobanov	Extrinsic	Intrinsic
Nearey 2	Extrinsic	Extrinsic

count from one speaker and one session to another. Because the evolution of pronunciation is a key research interest, the dataset was split into session-based subsets prior to normalization.

3.2. Phone-gating

Figure 2: *Cross-session, cross-speaker count of phonemes phone-gated as "FAIL" or "PASS" by method of normalization.*



A native dataset was then normalized. For this study, data from Peterson & Barney [8]), which comes with the PhonTools [1]) package, was used for convenience. Corresponding sex-dependent minima and maxima of the F1 and F2 formant values for each vowel and each method of normalization were then stored and extracted. This procedure returned a 10-row dataframe, one row for each vowel (/ə/ was excluded), with 8 columns: the male and female F1/F2 minima and maxima. The normalized F1 and F2 values from the learners' datasets were then checked against these extreme values, and flagged as correct ("PASS") if both their F1 and F2 values were comprised within the natives' range, "FAIL" otherwise.

4. DISCUSSION

Figure 2 shows the count of vowels flagged as "FAIL" or "PASS" for each session and each method of normalization. The total proportion of vowels flagged as accurate is displayed in Figure 3. Because the phone-gated values cannot be assumed to be really independent from one method of normalization to another, pair-wise McNemar's Chi-squared tests of independence with one degree of freedom were performed from the contingency tables in order to examine the relation between the method of normalization and the phone-gating results. The results of those pair-wise comparisons are displayed in Table 2. The extremely low *p*-values mean that the null hypothesis that no differences exist between the phone-gated results from one method to another can safely be rejected.

Figure 3: *Proportion of phonemes tagged as accurate by method of normalization.*

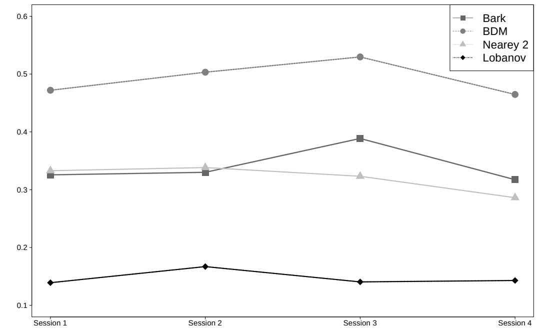


Table 2: *Pair-wise comparisons of the phone-gating results by normalization methods using McNemar's chi-squared tests.*

Method	χ^2	p value
Bark vs. BDM	454.96	$< 2^{16}$
Bark vs. Nearey 2	1639.21	$< 2^{16}$
Bark vs. Lobanov	3173.79	$< 2^{16}$
BDM vs. Nearey 2	439.77	$< 2^{16}$
BDM vs. Lobanov	1395.56	$< 2^{16}$
Nearey 2 vs. Lobanov	3577.21	$< 2^{16}$

Although these results might suggest a certain inadequacy between normalization constraints and the unbalanced characteristics of natural language (as opposed, for instance, to controlled lists of words), they may also reveal the limitations of automatic formant extractions. The extrinsic methods used in this paper rely on F3 measurements, whose proximity with F2 in vowels such as /ɪ/ may account for the wide gap in phone-gating results between the BDM and Lobanov methods for this vowel. One way to work around the relative unreliability of automatic formant extraction in high-dispersion corpora such as learners' corpora would be to study vowel-inherent spectral change (*cf.* Nearey (1986) [6] or Morrison (2007) [5]), rather than mid-temporal values.

5. REFERENCES

[1] Barreda, S. 2014. *phonTools: Functions for phonetics in R*. R package version 0.2-2.0.

[2] Bigi, B. 2012. Sppas: A tool for the phonetic segmentations of speech. LREC, , (ed), *Proc. of LREC 2012* 1748–1755.

[3] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10), 341–345.

[4] Lobanov, B. 1971. Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.* 49, 606–608.

[5] Morrison, G. S., Nearey, T. M. jul 2007. Testing theories of vowel inherent spectral changea). *J. Acoust. Soc. Am.* 122(1), EL15–EL22.

[6] Nearey, T., Assmann, P. 1986. Modeling the role of vowel inherent spectral change in vowel identification. *Journal of the Acoustical Society of America* 125, 2387-97.

[7] Nearey, T. M. 1978. *Phonetic Feature Systems for Vowels*. PhD thesis Indiana University Linguistics Club.

[8] Peterson, G. E., Barney, H. L. 1952. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America* 24(2), 175–184.

[9] R Core Team, 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[10] Syrdal, A. K., Gopal, H. S. 1986. A perceptual model of vowel recognition based on the auditory representation of american english vowels. *The Journal of the Acoustical Society of America* 79(4), 1086–1100.

[11] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America* 88(1), 97–100.

[12] Wand, D., Fabricius, A. 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1 ~ F2 plane. *Leeds Working Papers in Linguistics and Phonetics* 9, 159-173.