Cite as Turton, Danielle. 2015. Determining categoricity in English /l/-darkening: A Principal Component Analysis of ultrasound spline data. Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015), 10-14 August 2015, Glasgow: University of Glasgow.

DETERMINING CATEGORICITY IN ENGLISH /L/-DARKENING: A PRINCIPAL COMPONENT ANALYSIS OF ULTRASOUND SPLINE DATA

Danielle Turton

Newcastle University
danielle.turton@newcastle.ac.uk

ABSTRACT

Although syllable-based accounts of /l/-darkening in English state that light [l] occurs in onsets (e.g. 'leap') and a dark [ɫ] in codas (e.g. 'peel'), analyses of the process from several phonetic studies have led to some arguing against an allophonic distinction altogether, stating that the difference between light and dark variants is merely two extremes of one continuum. The current paper attempts to address this debate using ultrasound tongue imaging, in particular, a Principal Component Analysis (PCA) of tongue spline data. Although PCA of spline contours may be seen as a relatively rough method when compared with analysis of raw pixel images, it is argued that the simplicity of attaching one individual figure to a contour is a highly efficient and convenient method for observing general patterns in the data.

Keywords: Ultrasound, Principal Component Analysis, /l/-darkening, English

1. INTRODUCTION

English /l/ is often shown to display varying realisations depending on its syllabic position. The variant is said to occur in coda position, where /l/ is realised with a delayed and/or reduced tongue-tip gesture and backed tongue body relative to its light counterpart.

The categorical nature of an allophonic distinction has been called into question, largely by studies which focus on articulatory realisation [12, 18]. Such studies claim that speakers do not have a categorical difference in the two realisations, but instead a continuum of darkness based on factors such as duration [18] or morphological affiliation [12]. Note that both of these studies are based on American English /l/.

This paper will compare tongue splines of /l/ in ten phonological contexts from several varieties of English spoken in England. The varieties in question are selected to demonstrate that, in fact, the categoricity of /l/-darkening processes seems to vary from dialect to dialect. Some speakers display bimodally distributed tongue body patterns between onset and coda position, whilst others show very little difference at all. Furthermore, we have some speakers who show a difference between the extreme syllabic positions, but with a gradual degree of darkness in intermediate contexts. These patterns may or may not be clear from a qualitative approach using spline contour data, but PCA can help to diagnose between categoricity and gradience in a quantitative fashion.

1.1. /l/ in varieties of English

Reports of allophonic variation in English /l/ vary to a remarkable extent in the existing literature. For this paper, three varieties with a range of reported patterns were selected for analysis: Received Pronunciation (henceforth RP), Manchester English and Newcastle English.

RP, the standard accent of English in England, is described as displaying the typical allophonic distinction, with light [l] prevocalically, but dark [ɫ] pre-pausally or pre-consonantally [7]. The variety spoken in Manchester (North-West England) is of interest, as the existing literature reports that Manchester /l/s are dark in all positions [5, 7, 11]. This is in direct contrast with Newcastle (North-East England), which is usually described as having light [l] in all positions [5, 6, 9, 21].

1.2. Diagnosing between categoricity and gradience

Techniques for objectively arguing whether a difference between two phonetic realisations is categorically phonological are not uncontroversial. Some observations work in one direction only. For example, we could argue that a consistently large phonetic difference, in terms of articulatory magnitude or acoustic range, is likely evidence of two allophones, but not vice versa. That is, an absence of a large phonetic difference does not necessarily entail an absence of categoricity.

Some have argued that a sensible solution for determining whether a sound pattern is categorical is by considering the phonetic facts, whilst conducting a thorough analysis of the statistical distribution from a quantitative perspective [4, 15]. Whereas a phonetically gradient pattern may be indicated by a continuous distribution, a phonological one may show some discontinuity, cut-off or bimodality between two categories [4, 14].

Following [20], the following diagnostics for categoricity are suggested:

1. Articulatory discontinuity between two (or more) sets of data points.
2. Articulatory consistency within these sets.
3. Bimodality in the quantitative analysis of the articulatory data.

2. METHODOLOGY

2.1. Experimental procedure

Speakers of RP, and Manchester and Newcastle English were recorded producing /l/ in the phonological contexts shown in Table 1 (henceforth referred to by example token).

Table 1: The ten phonological environments studied with example tokens

Context	Example
1. word-initial	***l**eap*
2. stem-medial pretonic	*be**l**ieve*
3. suffix-initial	*free-**l**y*
4. intervocalic	*he**l**ix*
5. stem-final presuffixal	*pee**l**-ing*
6. compound boundary	*pee**l**-instrument*
7. word-final phrase-medial	*hea**l**#* V
8. phrase-final	*hea**l**#*, V
9. utterance final	*pee**l***
10. word-final pre-consonantal	*pee**l** bananas*

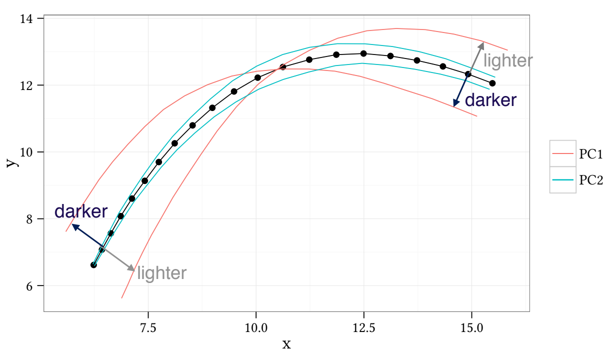
Subjects wore an Articulate Instruments probe stabilisation headset [16] whilst reading lists of words and phrases containing the target stimuli five times on a Mindray DP2200 ultrasound machine (frame rate 60 fps deinterlaced), with acoustics recorded through an Audio-Technica ATR-3350 microphone. /l/s were flanked by front high vowels in all contexts. The data were collected using Articulate Assistant Advanced (henceforth AAA; [2]). The audio and video were synchronised with the aid of a SyncBrightUp unit [1]. Sound files were exported out for acoustic segmentation in Praat, before being imported back to AAA where splines were hand-drawn for all frames within /l/ boundaries and flanking vowels. Spline coordinates (over 42 points) were extracted for contextual comparison and mean midpoint values plotted in R's `ggplot2` package [22] (plots are not to scale). Palate traces were taken for each speaker and are shown on the plots for reference. Midpoints only are analysed in this paper. The right side of the image shows the tongue-tip, and the left side the tongue-root.

2.2. PCA

Although visual inspection of spline contours is one way of observing the three categoricity diagnostics, a quantitative analysis may provide more solid evidence. The PCA provides a way of boiling down raw tongue contour data in order to extract areas of variance, and has been often used with articulatory data [10, 17, 19]. The output is a value for each principal component (PC). For this kind of data, we only get one valid PC per spline, for reasons explained below.

The PCA was conducted using `princomp` function in R's default `stats` package. Figure 1 shows the loadings plots of the PCA computed for /l/ midpoints only, for the RP speaker. The loadings matrix gives a score indicating how much of each principal component can be found in the /l/s tongue shape. A high PC1 represents a lighter /l/, and a low PC1 a darker /l/, and its range is represented in the pink splines in Figure 1. The black spline is the overall average. As can be seen, PC2 accounts for a negligible amount of the variation. Loadings plots for the other two speakers show the same pattern.

Figure 1: PCA on midpoint /l/s only for RP speaker



A drawback to this method is that when NAs ('not available' i.e. tokens which are missing in the dataset) occur in the dataset, all values for that particular column of data have to be removed from the analysis. NAs may occur not only when the ultrasound image is poor, but also when the tongue does not cross that particular fan point during a particular articulation. This results in the extreme edges of the points not being included in the PCA, instead giving a "tunnel-vision" outlook on the spline variation.

However, this drawback also has its benefits, which lie in the simplicity of the output. Because of the limited viewpoint, the result we get from the PCA of such tongue contour data is just one PC emerging as significant (following Baayen [3], PCs which account for less than 5% variance or have a clear discontinuity between the previous PC are dismissed as insignificant. This results in just one PC in all speakers). Therefore, we end up with one quantitative value per spline which neatly summarises both height and backness. Although it may be less detailed than a PCA of pixels, the drawback of these more complex analyses is that they can often output upwards of 9 PCs which do not obviously represent anything concrete in the articulation.

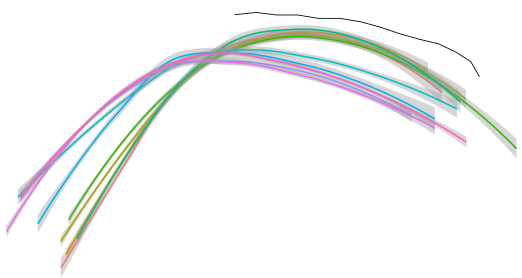
2.3. Bimodality

Another benefit of single figure representation of spline contours is the ability to enter this number into various statistical tests. Point 3 on the categoricity diagnostics is quantitative evidence for bimodality. Bimodality can be tested using Hartigan's dip statistic [8], using the PC1 spline value for each spline. The test works by measuring bimodality in the data, assessing potential 'dips' in the distribution, and outputs the dip statistic with an accompanying p-value indicating whether or not the dip is due to chance. The dip test is conducted on the overall PC1 values per speaker. R's `diptest` package [13] was used to calculate this.

3. RESULTS

The reported description of RP is confirmed in the spline data in Figure 2: the backed tongue body, reduced tongue-tip gesture, and retracted tongue root typical of a dark [ɫ] is only found in word-final position. Contexts 1-5 (in which /l/ is syllable initial)

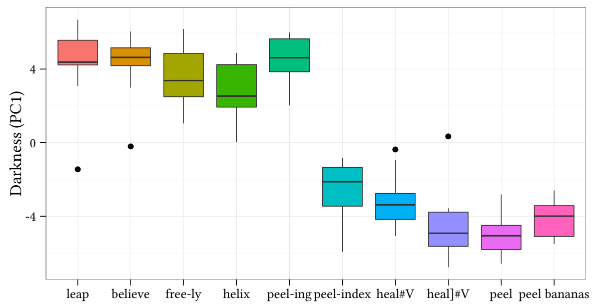
Figure 2: Midpoint /l/s for RP speaker across 10 phonological contexts



show a typical light articulation in comparison to Contexts 6-10 (where /l/ is word-final). Note the sharp, seemingly bimodal, distinction between these two extremes.

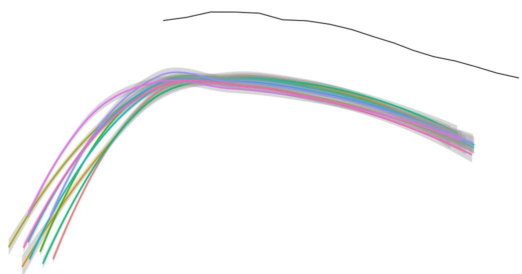
Plotting the results of the PCA accentuates this difference as shown in Figure 3, with lighter tokens showing a high PC1. The discontinuity between the two groups is clear, as is the general consistency between groups where the ranges are overlapping. Hartigan's dip test provides further support for the bimodality of this distribution on input of the PC1 values, giving a significant dip value of 0.078 ($p < 0.001$).

Figure 3: PC1 for RP speaker



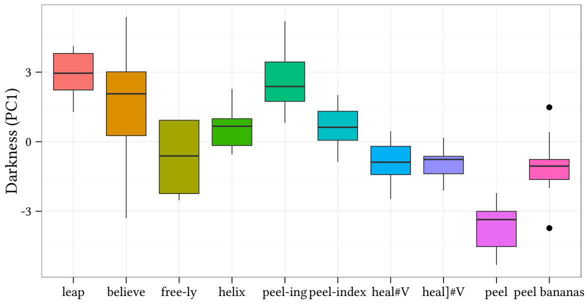
The Manchester splines in Figure 4 seem to corroborate the claims of only one category in this dialect. All tokens have a lowered tongue tip and tongue body.

Figure 4: Midpoint /l/s for Manchester speaker across 10 phonological contexts



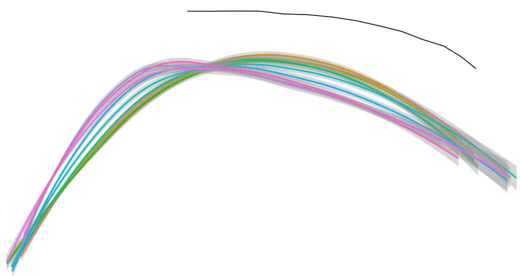
The PCA results in Figure 5 support this pattern showing, for the most part, that all contexts display overlapping confidence intervals. This unmodality is also confirmed by a non-significant dip statistic (D= 0.025; p= 0.97). However, the PC1 plot in Figure 5 does make it clear that the phrase-final *peel*-type tokens seem to be significantly darker than the rest, something which is not particularly obvious from the spline plot. This extra darkening is likely a side-effect of longer duration in utterance final position (see [20]), an effect discussed in detail in Sproat and Fujimura's /l/-darkening study [18].

Figure 5: PC1 for Manchester speaker



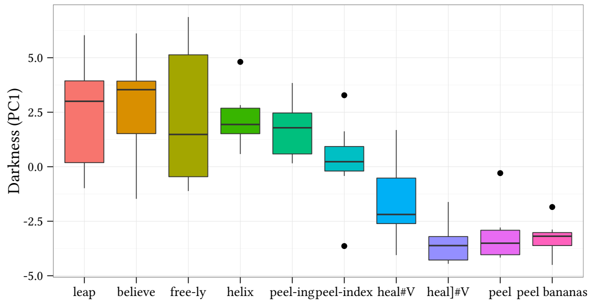
In Figure 6 we can see that the Newcastle splines are very tightly clustered, like Manchester, but also show a small but distinct split between some of the environments, as observed for RP.

Figure 6: Midpoint /l/s for Newcastle speaker across 10 phonological contexts



In the PC1 plot (Figure 7) we observe a similar trend to RP, but with much more overlap. There are signs of an articulatory discontinuity, but this is not as convincing as for RP. In Figure 7, we can observe intermediate tokens like *peel-index* types showing an in-between average in comparison to the lightest and darkest tokens. However, the individual tokens do seem to either be very light or very dark (as can be seen by the outliers in Figure 7), showing that this context is intermediate in terms of variance, not gradience. Despite this, under the three diagnostics for categoricity, we cannot claim that Newcastle has two allophones. Within the possible 'light' and 'dark' categories, the splines are not consistently the same. Hartigan's dip test shows that the potential bimodal difference just misses the significant cut-off point (D = 0.052; p =0.067). Although we cannot conclude there is evidence for two allophones, it would not be wise to conclude there is only gradience here. Further experimentation such as temporal analyses, or more quantitative analysis may help shed more light on these differences.

Figure 7: PC1 for Newcastle speaker



4. CONCLUSION

This paper has presented just three of the possible typologies of /l/-darkening systems in English, demonstrating that the potential for allophonic variation varies from dialect to dialect. We have seen that some varieties, such as RP, have a convincing categorical distinction between initial and final /l/s, as shown through both qualitative observation of discontinuity in tongue splines, as well as quantitative measures of bimodality from the results of the PCA. However, we have also seen that such categoricity may not be found in all varieties of English, with other accents showing little variation across all contexts (such as Manchester), or patterns which are more gradient in nature (such as Newcastle).

The contribution of this paper is twofold: firstly it has been shown that /l/-darkening patterns display a remarkable amount of variation across varieties and that the dismissal of categoricity based on one dialect may be missing the full picture. Secondly, this paper has shown that a principal components analysis of spline contour data is an effective technique for interpreting individual splines in a quantitative manner. Although this method may be more rudimentary than a complex pixel analysis, we have demonstrated that it does have the added benefit of simplicity, and although some more peripheral data points may be lost, the general pattern is nicely preserved.

5. REFERENCES

[1] Articulate Instruments Ltd, 2010. Syncbrightup users manual: Revision 1.10. Edinburgh, UK.

[2] Articulate Instruments Ltd, 2012. Articulate Assistant Advanced ultrasound module user guide, version 2.14.

[3] Baayen, H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: CUP.

[4] Bermúdez-Otero, R., Trousdale, G. 2012. Cycles and continua: on unidirectionality and gradualness in language change. In: Nevalainen, T., Traugott, E. C., (eds), *The Oxford handbook of the history of English*. Oxford: Oxford University Press 691–720.

[5] Carter, P. 2002. *Structured variation in British English liquids*. PhD thesis University of York.

[6] Carter, P. 2003. Extrinsic phonetic interpretation: spectral variation in English liquids. In: John Local, R. O., Temple, R., (eds), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press 237–252.

[7] Cruttenden, A., (ed) 2008. *Gimson's pronunciation of English*. London: Hodder 7th edition.

[8] Hartigan, J. A., Hartigan, P. 1985. The dip test of unimodality. *The Annals of Statistics* 70–84.

[9] Hughes, A., Trudgill, P., Watt, D. 2012. *English accents and dialects: an introduction to social and regional varieties of English in the British Isles*. Routledge.

[10] Johnson, K. 2011. *Quantitative methods in linguistics*. John Wiley & Sons.

[11] Kelly, J., Local, J. 1986. Long-domain resonance patterns in English. *Institute of Electronic Engineers: Proceedings of the International Conference on Speech Input/Output* 304–308.

[12] Lee-Kim, S.-I., Davidson, L., Hwang, S. 2013. Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology* 4(2), 475–511.

[13] Maechler, M. 2013. *diptest: Hartigan's dip test statistic for unimodality - corrected code*. R package version 0.75-5.

[14] Maye, J., Werker, J., Gerken, L. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, 1001–1111.

[15] Scobbie, J. M. 2005. The phonetics phonology overlap. *QMU Speech Science Research Centre Working Papers, WP-1*.

[16] Scobbie, J. M., Wrench, A. A., van der Linden, M. 2008. Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. *Proceedings of the 8th International Seminar on Speech Production* 373–376.

[17] Slud, E., Stone, M., Smith, P. J., Goldstein Jr, M. 2005. Principal components representation of the two-dimensional coronal tongue surface. *Phonetica* 59(2-3), 108–133.

[18] Sproat, R., Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291–311.

[19] Stone, M. 2005. A guide to analysing tongue motion from ultrasound images. *Clinical linguistics & phonetics* 19(6-7), 455–501.

[20] Turton, D. 2014. *Variation in English /l/: Synchronic reflections of the life cycle of phonological processes*. PhD thesis University of Manchester.

[21] Watt, D., Milroy, L. 1999. Variation in three Tyneside vowels: is this dialect levelling?'. In: Foulkes, P., Docherty, G., (eds), *Urban Voices: Accent Studies in the British Isles*. London: Arnold 25–46.

[22] Wickham, H. 2009. *ggplot2: elegant graphics for data analysis*. Springer Publishing Company, Incorporated.