LONG-DISTANCE COARTICULATORY EFFECTS OF BRITISH ENGLISH /l/ AND /r/: AN EMA, EPG AND ACOUSTIC STUDY

Paula West
Oxford University Phonetics Laboratory
paula.west@phon.ox.ac.uk

ABSTRACT

This paper demonstrates the existence of long-domain coarticulatory patterns associated with English /l/ and /r/, describing the extent and nature of differences in articulation and acoustics. Three speakers of Southern British English were recorded using simultaneous EMA and EPG. They produced six l/r minimal word pairs in a frame sentence. Strong local coarticulatory effects were found in the vowels adjacent to the liquid. Non-local differences in F_3, lip and tongue position were found in vowels not adjacent to the liquid, with lower F_3, more lip rounding and backer or higher tongue position surrounding an /r/. EPG data shows significant differences in contact patterns for consonants up to two syllables before and after the liquid.

1 INTRODUCTION

The coarticulatory effects of vowels on adjacent segments and across consonants are well documented (e.g. [9], [10], [15]) and most theories of coarticulation are based on these findings (e.g. [6] and [4]). Coarticulatory effects of consonants have not been similarly studied, despite the claim that consonants such as /l/ and /r/ exert long-distance coarticulatory effects, or "resonances" [7]. In fact, consonantal coarticulatory effects (C to V) are generally acknowledged to be smaller, with a more restricted temporal range than V to C effects [3]. However, the importance of the coarticulatory effects of liquids for speech perception has been demonstrated. Including coarticulatory effects in synthetic speech improves perception by up to 15% ([5], [11]) and coarticulatory information aids recognition of an /l/ or /r/ which has been replaced by noise in a progressive replacement task [12]. The perceptual relevance of these coarticulatory effects and the existence of conflicting claims about their extent motivates this study: an investigation of the long-distance coarticulatory effects of the English liquids /l/ and /r/, using simultaneous Electromagnetic Articulography (EMA) and Electropalatography (EPG).

2 EXPERIMENTAL METHOD

2.1 Stimuli and subjects

Three right handed male speakers of standard Southern British English (S1, S2 and S3), aged 21, 20 and 31, were recruited for the experiment. They were pre-recorded to ensure that they made long-domain acoustic distinctions between /l/ and /r/ sentences. These recordings were then used to check that their speech had not been unduly disrupted in the articulatory experiment. The stimuli sentences were six l/r pairs (leap/reap, lip/rip, lap/wrap, lope/rope, lobe/robe, lob/rob) placed in the frame sentence 'Have you uttered a — at home?'. The stimuli were randomised and recorded amidst distracter sentences, with each subject recording 144 stimuli sentences (12 repeats of each sentence). Twelve repeats of each utterance were obtained and analysed for S2 and S3, 7 for S1 (the speaker reported in [13]).

2.2 Data collection and processing

Simultaneous acoustic, EMA and EPG recordings were made at Queen Margaret College, Edinburgh, using the Carstens AG100 Articulograph (sampling rate 500 Hz) and Reading EPG system (sampling rate 200 Hz). The audio signal was recorded at 16 kHz, using a SoundBlaster analogue-to-digital converter in a PC. The audio, EPG and EMA signals are synchronised to within the 5 ms accuracy imposed by the EPG sampling rate.

Three EMA coils were placed on each subject's tongue, and one each on the upper lip, lower lip and the gum beneath the lower incisors. Reference coils were placed on the bridge of the nose and the gum above the upper incisors. The data were corrected for head movement and the post-processed data rotated so that the x axis was parallel to the subject's occlusal plane, as estimated by a plastic T-bar [14]. The origin of the system was set at the junction of the central-maxillary incisor diastema and the incisors' exposed tips.

The EMA and EPG data were processed in *Matlab*, using modifications of a set of *Matlab* macros written by Noel Nguyen [8], and routines written by the author. EMA and EPG data were extracted from the relevant data matrices at points of interest identified from the acoustic signal. The first three formant frequencies and (x,y) co-ordinates of all the coils were measured at the midpoint of the vowels adjacent to the liquid (local effects) and the schwas of 'uttered' and 'at' (non-local). EPG contact data was examined for the alveolar consonants in 'uttered' and 'at'. Acoustic measurements, unless otherwise stated, were made in *Waves* using 18 pole Burg spectra with a 50 ms Hanning window, checked manually against wide-band spectrograms and DFT spectra.

2.3 Statistical analysis

Most of the EMA and acoustic data is modelled using multivariate general linear models (GLMs) in *SPSS* as the variables are inter-correlated. Measurements are assumed to be independent, i.e. the error term is assumed to be a vector of independent errors. Equivalence of the variance-covariance matrices was checked where possible, using Box's M test, supplemented by Levene's test. Multivariate normality of distribution is difficult to check and the test is robust provided the data exhibits symmetry, so each variable was tested separately for the normal distribution, with the one-sample Kolmogorov-Smirnov test. The data modelled by multivariate GLMs in this paper met the required assumptions as far as they could be tested, and the results of the tests will not be reported here. The F statistics used for main effects tests are estimated by Pillai's test. The EPG data were analysed in SPSS, using Kolmogorov-Smirnov 2-independent samples non-parametric tests; Z scores from this test are reported.

3 RESULTS

3.1 Local coarticulation

Strong coarticulatory effects were found in vowels adjacent to the liquid, for both the acoustic and EMA data. EMA measurements (lip, tongue and jaw coil positions) and acoustic measurements (F_1, F_2 and F_3) were taken at the midpoint of the vowels surrounding the liquid. Multivariate GLMs constructed for these data with liquid (i.e. /l/ vs. /r/) and word pair as factors show liquid a significant factor for all speakers ($p < 0.001$). Consistent effects (significant for $\alpha = 0.01$) are lowered F_3, rounded lips (deduced from lip protrusion and aperture) and retracted tongue in /r/ contexts, reflecting the speakers' productions of /r/ with labialization and a strongly retracted tongue position. The effects appear equally strong in the perseverative and anticipatory directions. All speakers also have a significantly lower F_1 in the vowel preceding an /r/ than that preceding an /l/. This effect is not found for the vowel following the liquid. Two of the speakers (S1 and S3) also have a consistently raised tongue in the vowels surrounding an /r/. The differences in mean articulator position range from 0.4 mm to 8 mm, with smaller values for lips and larger values for tongue placement differences. As local coarticulatory effects were expected and are of lesser interest than long-distance effects, they will not be discussed in detail here. EPG data was not examined at these points, due to minimal tongue-palate contact.

3.2 Long-distance coarticulation in vowels

To examine long-distance coarticulation in vowels, EMA measurements were taken at the midpoint of the schwas in "uttered" and "at" of the phrase "Have you uttered a — at home?". For each subject, a multivariate general linear model was constructed for the articulatory and acoustic data from each of the schwas separately, with word and liquid as factors. Results for the schwa of "uttered" will be discussed under anticipatory effects, those for the schwa of "at" under perseverative effects. The GLMs show liquid as a highly significant factor in all cases; the most consistent effects are lip rounding, tongue raising and lowered F_3 in the /r/ context.

3.2.1 Anticipatory effects. For S1, both factors liquid and word pair were highly significant ($F(15,53) = 5.638$, $p < 0.001$ and $F(75,285) = 1.742$, $p < 0.001$), and there was no interaction between them. The significance of the factor liquid is the result of differences in F_3 ($p < 0.004$), tongue mid y ($p < 0.011$), upper lip x ($p < 0.001$) and upper lip y ($p < 0.046$). For S2, the factor liquid is significant ($F(15, 118) = 3.131$, $p < 0.001$), word pair and the interaction between liquid and word pair are not. Significance of the factor liquid is due to lower lip y ($p < 0.004$), upper lip x ($p < 0.044$), F_1 ($p < 0.045$) and F_3 ($p < 0.027$). For S3, the multivariate GLM shows liquid alone as a significant factor ($F(15, 108) = 3.500$, $p < 0.001$). Significance of the factor liquid is due to tongue back y ($p < 0.022$), tongue middle x and y ($p < 0.014$ and 0.018), tongue tip x ($p < 0.030$) and F_3 ($p < 0.001$).

There are some clear similarities between the three subjects: each subject has a lower F_3 in the /r/ context (by 18, 56 and 91 Hz). S2 also has a lower F_1 value in the /r/ context, although this is a very small difference in means (7 Hz) and may well be a spurious result. S1 and S2 have significant differences in lip placement: both have a fronter or more protruded upper lip in the /r/ context, S1 has a lower upper lip and S2 a higher lower lip in the /r/ context. This is suggestive of anticipatory lip rounding in the /r/ context. Differences in tongue placement are found for S1 and S3. For S1 and S3, tongue mid is higher in /r/ than in /l/ contexts. S3 also shows a higher tongue back and retracted tongue middle and tip in the /r/ context. These differences anticipate the articulatory configuration that these two subjects show for /r/: a retracted and raised tongue body. The articulatory differences are all less than 1mm in magnitude, although the acoustic differences range from 18 Hz for S2 to 100 Hz for S3.

3.2.2 Perseverative effects. For S1, the factors liquid and word pair are highly significant ($F(15,53) = 5.330$, $p < 0.001$ and $F(75,285) = 5.131$, $p < 0.001$) and there is an interaction between word pair and liquid ($F(75,285) = 1.686$, $p < 0.001$). The interaction is significant for the variables lower incisor y ($p < 0.004$), tongue mid x ($p < 0.028$) and F_2 ($p < 0.005$). The factor liquid is significant for three articulatory variables: tongue tip and mid x ($p < 0.001$ for both) and tongue back y ($p < 0.008$), and the factor word pair for all variables except lower incisor x. To explore the interaction between liquid and word pair, independent samples t-tests were conducted for the variable which showed an interaction: tongue mid x, for each word pair separately. Tongue mid x was significant for three word pairs, and the difference in means (around 0.6 mm) was in the same direction for all significant pairs: fronter tongue mid in the /r/ context.

For S2, the GLM has liquid and word pair as significant factors ($F(15,116) = 6.165$, $p < 0.001$ and $F(75,600) \approx 6.709$, $p < 0.001$), with a significant interaction between them ($F(75,600) \approx 1.941$, $p < 0.001$). The interaction term is due to tongue mid x and y ($p < 0.015$ and 0.001) and F_3 ($p < 0.001$). Word pair is significant for all variables except lower incisor, lower lip and upper lip x. Liquid is significant for lower incisor x ($p < 0.034$) and tongue back y ($p < 0.015$), and highly significant for tongue mid x ($p < 0.006$), tongue mid y ($p < 0.001$) and F_3 ($p < 0.001$). To explore the interaction between liquid and word pair, an independent samples t-test was conducted for tongue mid x and y and F_3 for each word pair separately with liquid as factor. Tongue mid x was significant for 'leap/reap' and 'lob/rob', tongue mid y for 'lip/rip', 'lobe/robe' and 'lob/rob' and F_3 for 'leap/reap', 'lip/rip' and 'lob/rob'. As was found for S1, the differences in means for the significant word pairs are all in the same direction (lower F_3, higher and backer tongue mid in the /r/ context).

S3's GLM has both liquid and word pair significant ($F(15,83) = 3.995$, $p < 0.001$ and $F(75,435) \approx 3.710$, $p < 0.001$), but the interaction between them is not. Word pair is significant for all variables except tongue mid and back x and lower incisor y. Differences due to liquid are found in tongue back y ($p < 0.040$), tongue mid y ($p < 0.028$), tongue tip x ($p < 0.012$) and F_3 ($p < 0.001$).

Means for all significant variables show small differences between /l/ and /r/ contexts. For S1 the tongue back is higher (by 0.66 mm) for /r/ contexts than /l/ contexts and the tip and mid are significantly fronter (by 0.84 and 0.56 mm respectively). S2 has lower F_3 in /r/ contexts (by about 70 Hz), higher tongue back and mid (about 80 mm), backer tongue mid and fronter lower incisor (by about 30 mm). S3 shows higher tongue back and mid (27 and 50 mm), backer tongue tip (76 mm) and lower F_3 (about 60 Hz) in the /r/ context. Main results are a lower F_3 in the /r/ context (S2 and S3) and raised tongue position for all three subjects. S2 and S3 have retracted tongue position (tip and mid respectively) in the /r/ context, whereas S1 has a fronter tongue (tip) position. These results are not entirely in accordance with our expectations: a small amount of perseverative tongue tip retraction after an /r/ might have been expected for S1 and S3, who produce retracted /r/s, but not necessarily for S2, where we might have expected the opposite pattern.

3.3 Long-distance coarticulation in consonants

The EPG data from all subjects were examined for long-domain coarticulatory effects in consonants. Tongue-palate contact could be examined effectively in the consonants of syllables remote from the liquid, but not in vowels, as they show very little tongue-palate contact. Various summary measures were used to determine where differences in contact patterns occurred: the total number of contacts, the total for each row, the number of front contacts (rows 1 to 3), back contacts (rows 4 to 8), lateral contacts (2 positions at the end of each row) and central contacts (total minus lateral contact). Each consonant was analysed separately. For each subject, the corresponding EMA data for the consonants were also examined for long-domain differences.

3.3.1 Anticipatory effects. Anticipatory coarticulation was examined at two points: the points of maximal tongue-palate contact in the /t/ and /d/ of 'uttered'. These were extracted by searching for maxima in the sum of EPG contacts over the relevant stretches of speech signal. The waveform was primarily used to identify the closure stretch, aided by the tongue tip y trace.

For S1, several differences in contact patterns for the /t/ of 'uttered' in different liquid contexts were found. Row two, row three, total and number of front contacts are significantly different across liquid contexts (Z = 1.418, 2,182, 1,964 and 1,865, $p < 0.036$, 0.001, 0.001 and 0.002). In all cases there was more contact in the /l/ than in the /r/ context. The difference in means is small, no greater than 2 contacts and once less than 1. Nevertheless the results are significant. Analysis of the /d/ of 'uttered' shows significant differences between the /l/ and /r/ context, but in a more limited area. Number of central contacts and contacts in row two proved significant (Z = 1.637, $p < 0.009$ for both), this indicates that the difference lies in the central contacts in row two. The means of these measures show that there is on average one more contact in the /r/ context. S2 shows fewer significant results: for the /t/ of 'uttered' there are significant difference in lateral and possibly back contact (Z = 1.4 and 1.708 $p < 0.006$ and 0.04), but the /d/ shows no significant differences at all. S3 shows significant differences in the /t/ of 'uttered': row three and front are significantly different (Z =1.417, $p < 0.036$ for both), with one more contact on average in the /l/ context. The /d/ of 'uttered' shows the same pattern for this subject: one or two more contacts in the /l/ context in row 2 and front (Z =1.667 and 1.417, $p < 0.008$ and 0.036).

The EPG results are interpreted as follows: S1 and S3 show the same pattern for the /t/ in the /l/ context, more contact between the front of the tongue and the front part of the palate (in the central contacts for S1). This may be a result of fronter tongue position preceding an /l/. S2 shows fewer contacts in the /l/ context in back and lateral contacts, which may amount to the same difference: fronter and more central tongue palate contact in the /l/ context. The data for the /d/ is less straightforward. S2 has no significant differences and S1 and S3 have differences in different directions. S3 has more contact in the front part of the palate and row two in the /l/ context, which fits with his other data, but S1 has less central contact and contact in row two in the /l/ context. In general S1 has considerably more contact in the front part of the palate than S3 does, as can be seen from the means for front contact (13 to 15 for S1 vs. 1 to 4 for S3 in the /t/ of 'uttered'). Thus greater contact in row two for S1 may represent a process of tongue retraction and a backer closure position (coarticulation with a following /r/), and thus a fronter closure position in the /l/ context, which is almost certainly what S3's data points to.

The corresponding EMA data were examined. For S1, the model for the /t/ of 'uttered' had liquid as the only significant variable (F(12,59) = 2.866, $p < 0.004$). This difference is a result of differences in upper lip, lower lip and lower incisor protrusion ($p < 0.007$, $p < 0.019$ and $p < 0.004$) and tongue back height ($p < 0.031$) with more protruded lip and jaw in the /r/ context and a higher tongue back. Means range from 0.5 to 0.8 mm. For /d/, liquid is again the only significant factor (F(12,59) = 3.168, $p < 0.002$) with differences attributable to upper lip x ($p < 0.008$) and y ($p < 0.048$), with the lip more protruded and lower in the /r/ context (by roughly 0.5 and 0.4 mm). Note that these differences are as expected, and indicative of lip rounding in the /r/ context. There are no differences in tongue tip placement, which shows that EMA data is complementary to EPG data; they measure different things. For S2, the model for /t/ has neither liquid nor word pair significant. The model for /d/ has liquid significant (F(12,120) = 3.363, $p < 0.001$) as a result of a difference in lower lip y ($p < 0.001$) which is higher in the /r/ context by approximately 0.64 mm. The factor word pair is not significant. For S3, the factor liquid is significant for /t/. The differences lie in tongue back and mid y ($p < 0.011$ and 0.016) and mid and tip x ($p < 0.028$ and 0.006). The tongue is retracted and raised in the /r/ context by more than 0.5mm in each case. For /d/ liquid is again the only significant factor (F(12,121) = 4.306, $p < 0.001$), with differences in tongue back and mid y, tongue mid and tip x and upper lip y ($p < 0.001$ for all). The differences in means for the tongue data are all greater than 0.7 mm and show retraction and raising in the /r/ context. The upper lip is lower by about 0.22 mm in the /r/ context. The EMA data confirm the interpretation of the EPG data: closure is fronter in the /l/ context for this subject.

3.3.2 Perseverative effects. EPG data were extracted at the point of maximum contact for the /t/ of 'at'. For S1 there were no significant differences here. S2 had one significant difference: row eight (Z = 1.620, $p < 0.011$) with slightly more contact in the /r/ context (only 1/2 a contact difference in means), suggesting backer tongue position. For S3, row two, row three, front and central contact are all different (Z = 1.750, 1.667, 1.833 and 1.667, $p < 0.004$, 0.008, 0.002 and 0.008 respectively). There are one or two less contacts in the /r/ context in all cases, clear evidence of tongue retraction.

The EMA data for the three subjects were analysed by constructing GLMs with word pair and liquid as factors. For S1, the GLM had word pair a significant factor (F(60,315) =1.636, $p < 0.004$), but not liquid. For S2, both liquid and word pair are significant in the model (F(12,120) = 3.435, $p < 0.001$ and F(60,620) = 3.392, $p < 0.001$). Differences in liquid are due to tongue tip x ($p < 0.004$) which is retracted by over 0.5 mm in the /r/ context. For S3, the model has both liquid and word pair and the interaction between them significant (F(12,121) = 5.680, $p < 0.001$, F(60,625) = 3.258, $p < 0.001$ and F(60,625) = 1.519, $p < 0.009$). The interaction term is due to differences in tongue back and mid y ($p < 0.036$ and 0.022) and tongue mid x ($p < 0.023$). Differences in liquid are due to lower incisor y ($p < 0.001$), lower lip y ($p < 0.023$), tongue tip x ($p < 0.001$) and upper lip y ($p < 0.028$). Tongue mid x is also significantly retracted in the /r/ context by about 0.5 mm ($p < 0.015$), but this needs to be treated with caution as this variable is significant for the interaction term. Closer inspection shows that the difference in means is in the same direction for all but the 'lobe/robe' pair. The other differences in means show a raised lower lip, upper lip and incisor in the /r/ context (by 0.13 to 0.34 mm) and a retracted tongue tip (0.88 mm).

4 SUMMARY

The acoustic and articulatory data showed retracted and/or raised tongue position, lip rounding and F_3 lowering in the /r/ relative to the /l/ context, up to two syllables remote from the liquid, for all speakers (Figure 1).

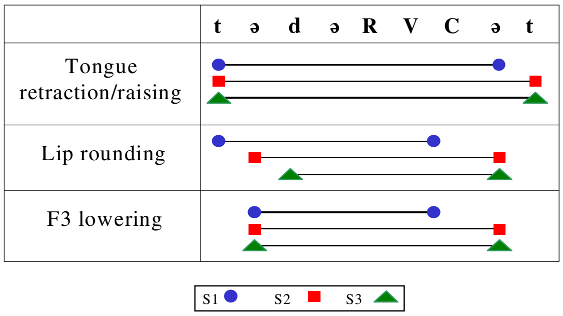

Figure 1: Extent of coarticulatory effects associated with /r/ (tongue retraction and/or raising, lip rounding and F_3 lowering) for S1, S2 and S3.

Some individual variation in temporal extent was apparent, with S1 showing less perseverative coarticulation, and others varying in the extent of lip rounding in the /r/ context. However the overall picture is one of surprising consistency, with l/r coarticulatory effects distributed across several syllables, and all speakers showing anticipatory coarticulation two syllables before the liquid. Long-distance coarticulatory effects were found in both tongue and lip position for all speakers studied.

5 MODELLING

These data pose problems for most recent theories of coarticulation, which do not attempt to model such long-distance effects. In particular, the school of Articulatory Phonology and its associated dynamic gestural modelling considers coarticulation to be the result of overlap or blending of gestures [4]. To explore this, a model of gestures as the result of damped second order differential equations has been implemented (based on [1] and [2]). Gestures are modelled as two halves (an upward/closing part and a downward/opening part) with parameters controlling the amplitude, frequency and equilibrium points of the half gestures, as well as damping and overlap with following gestures.

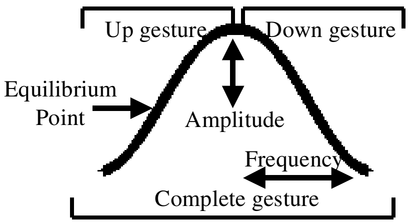

Figure 2: Some model parameters for a complete gesture

To date, EMA traces for S1 have been accurately modelled by manual adjustment to within the reported measurement error of the EMA system. Tongue mid *y* and upper lip *x* are modelled as a series of 9 and 7 whole gestures respectively. Multi-dimensional scaling was used to reduce the dimensionality of the data and find gestures which differ between /l/ and /r/ contexts. 2 sample Kolmogorov-Smirnov tests were conducted on the parameters of gestures of interest, for each word pair separately, with liquid as grouping variable. For tongue mid *y*, in all word pairs except 'leap/reap', parameters describing gestures non-adjacent to the liquid differ significantly ($\alpha = 0.05$). Effects are found in both the perseverative and anticipatory directions. Amplitude, frequency, equilibrium point and overlap parameters differ with l/r context, and word pair. The upper lip *x* models show a more restricted range of effects. All word pairs differ significantly for the l/r gesture (greater amplitude, longer gestures with higher equilibrium points for the /r/), all except 'leap/reap' differ in at least one adjacent gesture. Differences are again not restricted to the overlap parameter: overlap is insufficient to model these coarticulatory effects. Thus, gestural overlap does not account for these coarticulatory effects: the long-domain differences remain to be explained.

REFERENCES

[1] Browman, C. P. and L. Goldstein. 1985. Dynamic modeling of phonetic structure. In Fromkin, V. ed. *Phonetic Linguistics*. New York: Academic Press. 35–53.

[2] Browman, C. P. and L. Goldstein. 1990. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18, 299–320.

[3] Farnetani, E. 1990. V-C-V lingual coarticulation and its spatiotemporal domain. In W.J. Hardcastle and A. Marchal. eds. *Speech Production and Speech Modelling*. Dordrecht: Kluwer.

[4] Fowler, C. A. and E. Saltzman. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36, 171–195.

[5] Hawkins, S. and A. Slater. 1994. Spread of CV and V-to-V coarticulation in British English: Implications for the intelligibility of synthetic speech. *ICSLP 94*.

[6] Keating, P. 1990. The window model of coarticulation: articulatory evidence. In J. Kingston and M. E. Beckman. eds. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: CUP. 451–470.

[7] Kelly, J. and J. Local. 1986. 'Long-domain resonance patterns in English' International Conference on Speech Input/Output; Techniques and Applications. Conference publication no. 258. London: IEE. 56–59.

[8] Nguyen, N. 1996. A Matlab package for the analysis of EMA/EPG data. *QMC Working Papers in Speech and Language Sciences* 1: 5–12.

[9] Öhman, S.E.G. 1966. Coarticulation in VCV utterances: spectrographic measurements. *JASA* 24: 979–998.

[10] Recasens, D. 1989. Long-range coarticulation effects for tongue-dorsum contact in VCVCV sequences. *Speech Communication* 8: 293–307.

[11] Tunley, A. 1999. *Coarticulatory influences of liquids on vowels in English*. PhD thesis. University of Cambridge.

[12] West, P. (in press) Perception of distributed coarticulatory properties of English /l/ and /r/. *Journal of Phonetics*.

[13] West, P. 1999. The extent of coarticulation of English liquids: an acoustic and articulatory study. *ICPHS 99*, 3: 1901–1904.

[14] Westbury, J. R. 1994. On coordinate systems and the representation of articulatory movements. *JASA* 95: 2271–3.

[15] Whalen, D. 1990. Coarticulation is largely planned. *Journal of Phonetics* 18: 3–35.