SYLLABLE-ONSET ACOUSTIC PROPERTIES ASSOCIATED WITH SYLLABLE-CODA VOICING

Noël Nguyen[1] *& Sarah Hawkins*[2]

1. Laboratory for Psycholinguistics, University of Geneva, Switzerland
2. Department of Linguistics, University of Cambridge, United Kingdom

ABSTRACT

This study investigates durational and spectral variation in syllable-onset /l/s dependent on voicing in the coda. 1560 pairs of (C)lVC monosyllables differing in the voicing of the final stop were read by 4 British English speakers. Onset /l/ was longer before voiced than voiceless codas, and darker (for 3 speakers) as measured by F2 frequency and spectral centre of gravity. Differences due to other variables (lexical status, isolation/carrier context, syllable onset, vowel quality and regional accent) are outlined. It is proposed that coda voicing is a feature associated with the whole syllable, phonetically implemented as a variety of properties spread throughout the syllabic domain. Implications for word recognition are outlined.

1. INTRODUCTION

It is well known that syllables have longer vowels when their codas are voiced rather than voiceless. Recent work [1] shows that syllable-onset /l/ and other sonorants can also be longer in syllables with voiced codas compared with voiceless codas. We asked how general this new finding is, and whether systematic spectral differences accompany durational differences in onset /l/ conditioned by the voicing of the coda. Impressionistically, onset /l/s seem darker before voiced than voiceless codas in many accents of English, but this could be an illusion since synthetic (intervocalic) /l/s sound darker when they are longer, even when there is no spectral difference [2]. A contrast in an onset segment caused by coda voicing has implications for linguistic and psycholinguistic models, as developed below.

Although regional accent is not our main concern, it provides additional interest. Accents of English maintain a clear/dark contrast with onset liquids (/l r/), but whereas most have clear /l/ and dark /r/, others have dark /l/ and clear /r/. Conceivably, then, voiced codas may condition relatively darker onset /l/s in most accents but clearer /l/s in others. So, whereas [1] used a single speaker of General American, we used 4 speakers chosen to represent 4 accents of Northern England thought to exhibit coda-conditioned spectral differences in onset /l/.

2. METHOD

2.1. Material

The experimental material consisted of 39 pairs of /(C)lVC/ monosyllables differing in the voicing of the final stop (voiced for one member of the pair and voiceless for the other). For 14 pairs, the voiced member was an English word (e.g. *load*) and the voiceless member a non-word (*loat*), while the reverse was true for the other pairs. (Lexical status was manipulated for a companion perceptual study, [3]). All words had frequencies of less than 50 per million (mean 4.41 in [4]), and frequency was matched in each pair. Syllable onsets were /l/, /bl/, /pl/, /gl/, /kl/, /fl/, /sl/; vowels were /i/, /ɪ/, /eɪ/, /eə/, /ɛ/, /æ/, /aɪ/, /ɜ/, /ɑ, /ʌ/, /ɒ/, /əʊ/, /u/.

2.2. Speakers

The 4 subjects were all native speakers of British English, with non-rhotic accents thought to exhibit systematic spectral and durational variations in onset /l/s. S1, the second author, a woman in her 40s, has lived in several different regions of England and the USA. For this experiment she maintained an accent reasonably described as Educated Northern British, with clear initial /l/. S2-S4 were males aged mid-20s to early 30s. S2 is mildly RP with Manchester attributes including dark initial /l/. S3 has a strong northeastern (Yorkshire) accent with fairly dark initial /l/. S4 has a strong regional accent from the north west (St. Helens, Lancashire). His is the only accent of the four to realise the contrast in darkness between onset /l/s and /r/s in the opposite way to that of most English accents: a relatively darker onset /l/ than /r/. All except S4 have some phonetic training, and all except S1 were naive as to the purpose of the study. All were teachers or students at Cambridge University.

2.3. Procedure

The recordings were made in a sound-treated room using high-quality equipment. Each speaker read the monosyllables in random order, five times in isolation, and five times in a carrier phrase. The carrier phrase (*Put up a ... above all*) was designed to minimize potential lingual coarticulatory interactions between the critical monosyllable and the flanking sounds. The material to be read appeared on a computer screen in front of the subject, one item or item-plus-carrier at a time. Speakers were told to speak naturally, but to keep the rate, f0 contour, and stress steady throughout the recording. (Nuclear accent was on *all*.) Each utterance was low-pass filtered, digitized at 16 kHz SR, and its waveform displayed in real time on a Silicon Graphics workstation. This, and listening, allowed us to ask for disfluent or otherwise unsatisfactory items to be repeated immediately.

2.4. Segmentation criteria

The data were segmented by hand from waveforms and wide-band spectrograms. For each item, we marked the onset of /l/, the start of its periodic part (where this was different from its onset, as when clustered with a voiceless obstruent e.g. /sl/), and its offset. Criteria for defining /l/ onset depended on the syllable onset, as follows. Single /l/ preceded by a vowel (i.e. in the carrier): the abrupt fall in overall intensity and formant frequencies. Single /l/ in isolated items: the start of periodicity. Stop+/l/ clusters: stop release burst. In /sl/ clusters: an abrupt decrease in intensity of aperiodic noise and change in spectral shape, sometimes immediately after a high-amplitude transient. In /fl/ clusters: typically, an abrupt increase in noise intensity and associated spectral changes. For all items, /l/ offset was at an abrupt rise in formant amplitudes and, usually, frequencies.

3. RESULTS

Preliminary analyses showed no differences in any measure due to lexical status, confirming our impression that the non-words were read fluently. The vowels were grouped into 4 classes as in Table 1, except that S1's /ʌ/ was classed as low/back-or-central.

Vowel category	Vowels
high/front	/i/, /ɪ/, /eɪ/, /aɪ/
low/front	/ɛ/, /eə/, /æ/
low/back-or-central	/ɜ/, /ɑ/
back/rounded	/ɒ/, /əʊ/, /u/, /ʌ/

Table 1: Vowel categories used in the statistical analyses

3.1. Duration

Durations of /l/ were analysed for each speaker (S) separately, using two repeated-measures ANOVAs, the vowel ANOVA and the syllable-onset ANOVA. Independent variables (IVs) in the vowel ANOVA were context (with or without carrier), vowel category (as in Table 1), and coda voicing (voiced or voiceless). IVs in the syllable-onset ANOVA were context, syllable onset, and coda voicing. Vowel Category and Syllable Onset could not be part of the same ANOVA because our corpus did not have all possible syllable-onset x vowel category combinations. In all ANOVAs, the monosyllabic pairs were a random factor.

speaker	voiced coda	voiceless coda	diff.
1	90.3	85.9	+4.4
2	78.8	74.6	+4.2
3	78.4	74.2	+4.2
4	86.8	83.0	+3.8

Table 2: Mean duration (ms) of onset /l/ in voiced-coda and voiceless-coda monosyllables, for each S. Rightmost column shows differences between contexts (voiced – voiceless coda).

Table 2 shows that, as expected, the mean duration of onset /l/ was longer before voiced codas than before voiceless ones. The mean difference was small, but highly significant in both ANOVAs, and independent of all other IVs (except for a significant interaction with syllable onset for S3).

As expected, /l/ was longer for monosyllables in isolation than the carrier phrase ($p < 0.001$ in both ANOVAs for S1, S3, & S4, and in the syllable-onset ANOVA for S2, $p<0.01$). The duration of /l/ depended on syllable onset ($p < 0.001$ for each S), interacting strongly with context for S1, S2, and S4 ($p < 0.001$).

3.2. F1 frequency

speaker	voiced coda	voiceless coda	diff.
1	370	370	0
2	275	276	−1
3	315	321	−6
4	306	313	−7

Table 3: Mean F1 frequency (Hz) of onset /l/ in voiced-coda and voiceless-coda monosyllables, for each S. "Diff" column shows differences between contexts v(voiced – voiceless coda).

F1 frequency was measured at the mid-point of the periodic part of /l/ using *xwaves*' automatic formant tracker with 49-ms and 100-ms windows. The value from the 49-ms window was used when the difference between the two measures was less than 50 Hz and both measures fell between 200 Hz and 1000 Hz. Measures which did not meet these criteria (22%) were discarded. Table 3 gives the mean frequency of F1 in voiced-coda and voiceless-coda monosyllables. Vowel and Syllable-onset ANOVAs confirmed that the small differences in F1 frequency dependent on coda voicing were not significant.

3.3. F2 frequency

speaker	voiced coda	voiceless coda	diff.
1	1534	1536	−2
2	1292	1306	−14
3	1002	1033	−31
4	1112	1108	+4

Table 4: Mean F2 frequency (Hz) of onset /l/ in voiced-coda and voiceless-coda monosyllables, for each S. "Diff" column shows differences between contexts (voiced – voiceless coda).

F2 frequency was interpreted as indicating degree of darkness. It was measured as described in §3.2, except that the automatically-tracked frequency was rejected if either value from the two window sizes was more than 1800 Hz. For the 33% of cases that did not meet the measurement criteria, F2 frequency was measured manually from DFT spectra supplemented by LPC spectra and spectrograms. Thus we had a complete set of measures for F2 frequency. Table 4 shows the mean F2 frequency at the midpoint of the voiced part of /l/.

For S1-3, /l/ was slightly darker in voiced-coda syllables than voiceless-coda syllables, the difference being significant for S2 and S3. (S2: vowel ANOVA, $F(1,35) = 6.788$, $p < 0.02$; syllable-onset ANOVA, $F(1,32) = 8.129$, $p < 0.01$. S3: vowel ANOVA, $F(1,35) = 36.658$, $p < 0.001$; syllable-onset ANOVA, $F(1,32) = 39.965$, $p < 0.001$). For S2 the syllable onset affected the magnitude of the difference slightly (Syllable Onset x Coda Voicing ($F(6,32) = 2.536$, $p < 0.05$). S1's data included some large differences between voiced-voiceless pairs, but there was more variability than in the other Ss' speech, reflected in bigger standard deviations and only a nonsignificant trend for F2 to be lower in voiced-coda contexts than in voiceless-coda contexts. It is interesting that S4's difference is in the opposite direction, albeit nonsignificantly. Since his accent is unusual in having relatively darker /l/ than /r/ onsets, his realisation of coda-conditioned differences in darkness amongst onset /l/s may also be the opposite to that of most accents of English.

3.4. Spectral centre of gravity

Slater and Coleman [5] found that temporal differences in the acoustic structure of a syllable as a function of coda voicing are not uniform throughout the syllable. They used dynamic time warping to compare pairs of monosyllabic words differing in the voicing of the final stop. For each of a number of acoustic parameters, they identified a region of maximal expansion in each voiced-coda word, relative to its voiceless-coda pair. Accordingly, we asked whether spectral differences in our data are confined to the onset /l/, or extend into the following vowel.

The spectral centre of gravity (COG) was computed at 3 places: the mid-point of the voiced part of /l/, /l/ offset, and 40 ms after /l/ offset (roughly at the end of the /l/-to-V transition). To calculate the COG, a DFT spectrum (49-ms Hanning window centred at the specified place) was converted into an auditory excitation pattern ([6], programmed by Zheng et al. [7]). The COG was computed from this excitation pattern in the range 50-3500 Hz. COG was used instead of F2 frequency because COG can be derived entirely automatically, whereas F2 needs manual checking (see §3.3), which is impractical for large datasets.

A lower COG within a voiced-voiceless pair was interpreted as reflecting darker /l/s and vowels. Table 5 shows the mean difference in the COG frequency between voiced-coda and voiceless-coda syllables at each of the three places. Differences that were significant ($p < 0.05$) in both vowel and syllable-onset ANOVAs are in bold.

Speaker	voiced /l/ mid-point	/l/ offset	onset of V stable part
1	**−7**	−4	−3
2	**−13**	+1	+2
3	**−14**	−2	+1
4	0	+4	**+4**

Table 5: Mean differences in COG (Hz) between voiced- and voiceless-coda syllables at three places in the syllable. Negative differences mean COG was lower when the coda was voiced.

Table 5 shows that at the mid-point of voiced /l/, the COG was significantly lower in frequency in voiced-coda syllables than in voiceless-coda syllables for S1-S3. In contrast, S4 had no difference in COG at the midpoint of the /l/, but a small positive difference appeared by the end of /l/, which achieved weak statistical significance 40 ms into the vowel. These opposing trends are consistent with our data on F2 frequency; they provide further evidence that, for most speakers, onset /l/s are darker in syllables ending in a voiced coda as opposed to a voiceless one. Table 5 also suggests that, for S1-S3, differences in COG dependent on coda voicing are essentially restricted to the medial part of onset /l/, and fade away in the transition between /l/ and the following vowel as shown in Fig. 1. Conversely, S4's distinction seems to develop gradually from the beginning of the vowel. These findings agree well with [5]'s hypothesis that spectral differences associated with stop voicing affect the preceding sonorant+vowel sequence unevenly, and may be confined to a specific temporal region of this sequence.

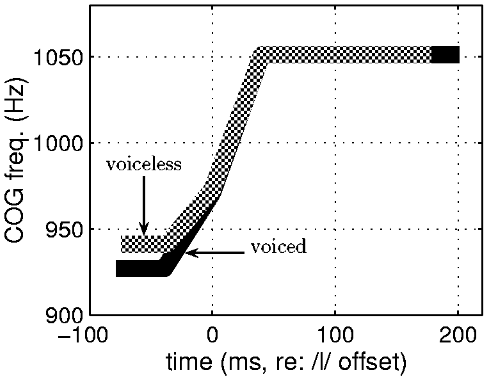

Figure 1: Mean spectral centre of gravity during onset /l/ and vowel, for monosyllables with voiced (dark line) and voiceless (light line) codas. Data for S3, averaged over all items, aligned at /l/ offset.

4. DISCUSSION

This study confirms other findings that onset /l/ differs in length and darkness depending on the voicing of the syllable coda [1], and that the changes are not evenly distributed, but localized to a particular region of the sonorant+vowel sequence [5]. It extends these observations from American English to four accents of British English (albeit one speaker each), and shows that the details of the distinction can differ with regional accent. The general pattern is for onset /l/s to be relatively long and dark when the coda of the same syllable is voiced, and relatively short and light when the coda is voiceless. For other accents, other patterns may hold. In this study, S4's /l/s were longer but slightly clearer before voiced codas, and the significant spectral differences appeared during the vowel rather than in onset /l/.

The spectral differences show that the impression of darkness does not arise only because the /l/ is longer (contra [2]). It has been argued, [8], that the main cause of the correlation between darkness and duration for *post-vocalic* /l/ is coarticulatory undershoot: when /l/ is short, the tongue has less time to take up a more extreme position. This explanation does not account for our data, for there seems no reason to assume coarticulatory undershoot in onset /l/s that are clearly spoken in citation form.

The properties we have identified are in non-adjacent segments. If the onset properties were articulatorily and acoustically independent of the coda properties, then this case would be quite different from better-known instances of nonadjacent influences, such as vowel-to-vowel coarticulation or spread of lip-rounding, in which essential attributes spread. Such an outcome would have far-reaching theoretical implications. Thus it is worth considering whether these onset properties are independent of the physical adjustments needed to produce voiced and voiceless obstruents i.e. that they do not result from anticipatory coarticulation of an essential attribute of the coda.

The articulatory independence of the durational adjustments and coda voicing seems indisputable. There is more room for doubt with the spectral adjustments. For example, whereas velarisation seems to be articulatorily independent of coda voicing, manoeuvres to enlarge the oral cavity need not be. On balance, enlargment of the oral cavity to anticipate coda voicing seems unlikely: changes in larynx height or jaw height in anticipation of coda voicing should affect F1 frequency, but no differences were found (§3.2). Vocal-tract tension might have been changed to produce fortis voiceless codas and lenis voiced codas. But fortis/lenis is arguably an enhancing rather than an essential dimension of obstruent voicing in English, and thus not best ascribed to anticipatory coarticulation; rather, it is an attribute of the syllable or its rhyme. Most tellingly, for S1-S3 the major acoustic differences lie in the /l/ and die out in the vowel, which does not point to anticipatory coarticulation of coda properties.

We tentatively conclude that coda voicing does not demand the observed changes in the /l/+vowel sequence. Further, we have shown elsewhere that listeners can use these coda-dependent differences in onset /l/ to recognize words faster [3]. If information about coda voicing is available to listeners in the syllabic onset, and is not due to anticipatory coarticulation of an essential property of the coda, then coda voicing could reasonably be modelled as a property of the entire syllable (assuming the sonorant is in the onset rather than the rhyme).

What are the implications of this conclusion for perceptual models? Both segmental (linear) and non-segmental (non-linear) approaches can offer descriptions, but in our view the non-linear approach is better. A segmental model assumes lexical items are represented as phoneme strings; it would describe our findings in terms of context-dependent variations allowing prediction of upcoming segments, and require coarticulatory effects to be "undone" in order to identify phoneme strings before a word can be recognized. A non-segmental model could describe these same findings in terms of features associated with units larger than phones (in this case, the syllable). The sensory signal is continuously evaluated, and the first contact phase of lexical access could involve mapping time-varying auditory patterns directly onto lexical items which themselves are relatively unanalysed time-varying auditory patterns built up along statistical principles from exposure to many instances.

Our preferred non-linear model is broadly compatible with psychoacoustic evidence that sound is structured by the perceptual system into auditory streams. It also shares interesting parallels with auditory enhancement theory (AET) [9], which proposes that acoustic properties combine to form complex Intermediate Perceptual Properties (IPPs), which in turn combine to contribute in varying proportions to the identification of distinctive features. Darkness before voiced codas might contribute to the low-frequency (LF) IPP. Or, since the LF property is seen as a psychoacoustic blending of high-amplitude energy in only the f0 and F1 range, darkness might more plausibly be a different IPP, or auditory stream, reflecting centre of gravity in the mid-frequency range (cf. arguments for brightness as a dimension of vowel quality [10]).

Our view, developed further in [3], describes word recognition in terms of activation and decay of excitation. It assumes words can be recognised from relatively weak auditory information spread across more than one acoustic-phonetic segment, as long as it is consistent [cf. 11, 12]. The non-segmental view is also compatible with the stronger claim that the initial contact phase of lexical access involves matching the speech signal, analysed into auditory streams but otherwise unsegmented, with words in the lexicon which, for initial contact, are themselves unanalysed patterns of auditory streams.

Acknowledgement: We thank Sebastian Heid for his help.

6. REFERENCES

1. van Santen, J.P.H., Coleman, J.S., and Randolph, M.A. (1992). "Effects of postvocalic voicing on the time course of vowels and diphthongs", *J. Acoust. Soc. Am.* **92**(4/2), 2444.
2. Newton, D.E. (1996). "The nature of resonance in English: an investigation into lateral articulations", *York Papers in Linguistics* 17, 167-190.
3. Hawkins, S., and Nguyen, N. (1998). "Effects on word recognition of syllable–onset cues to syllable–coda voicing", *LabPhon VI*, York, 2–4 July 1998.
4. http://wapsy.psy.uwa.edu.au/uwa_mrc.htm
5. Slater, A., and Coleman, J.S. (1996). "Non-segmental analysis and synthesis based on a speech database", *ICSLP 96*, Philadelphia, 3-6 October 1996, vol. 4, 2379-2382.
6. Moore, B., and Glasberg, B.R. (1987). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns", *Hearing Research* **28**, 209–225.
7. ftp://ling.ohio-state.edu/pub/phonetics/specmom/
8. Sproat, R., and Fujumura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291-311.
9. Kingston, J., and Diehl, R.L. (1995). "Intermediate properties in the perception of distinctive feature values." In B. Connell and A. Arvaniti (eds.) *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV.* Cambridge: Cambridge University Press, 7-27.
10. Fischer-Jørgensen, E. (1985). "Some basic vowel features, their articulatory correlates, and their explanatory power in phonology." In V. A. Fromkin (ed.) *Phonetic Linguistics.* Orlando, FL: Academic Press, 79-99.
11. Warren, P., and Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition", *Perception and Psychophysics* 41, 262-275.
12. Hawkins, S., and Warren, P. (1994). Implications for lexical access of phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics* 22, 493-511.