Perception of Coda Voicing from Properties of the Onset and Nucleus of *led* and *let*

Sarah Hawkins[†] *and Noël Nguyen*[‡]

[†]Department of Linguistics, University of Cambridge, UK
[‡] Laboratoire Parole et Langage, Université de Provence, Aix-en-Provence, France
sh110@cam.ac.uk; Noel.Nguyen@lpl.univ-aix.fr

Abstract

Syllable-onset /l/ in British English is longer and often has different (usually lower) F2 frequency before a voiced coda. Five experiments explore the perceptual power of these properties and of f0. In each experiment, listeners identified as *led* or *let* synthetic syllables whose latter half was replaced by noise. The most reliable cue was /l/ duration; F2 frequency in the /l/ was influential mainly when the vowel quality was held constant. However, listeners learn which cues are most effective, and some choose /l/ duration rather than spectral properties relatively late in the procedure. The results support word recognition models with non-segmental lexical representation that is sensitive to systematic variation in phonetic fine detail.

1. Introduction

Theories and computational models of spoken word understanding typically give weight to acoustic properties that provide strong cues to feature or phoneme identity, especially when they arise in the vicinity of the acoustic segment most closely associated with the phoneme in question. This is a sensible approach when acoustic properties unambiguously cue a particular feature or phoneme, or are complex and poorly understood. It is also congruent with the common assumption that the acoustic signal is converted into an abstract linguistic form before word recognition.

However, there are alternative views, and advances in research on categorisation and memory ([2][3][11][13]) and computational modeling of word recognition (e.g. [1][12]) converge to make it worthwhile to pursue them now.

The work described here is part of a program of research on the perceptual salience of fine phonetic detail that varies systematically with linguistic structure. We hypothesize that even very subtle acoustic-phonetic properties can be salient perceptually as long as they indicate linguistic structure. Both classical and recent experiments suggest that most systematically varying properties of speech will enhance perception in some situations ([10]:179-181). Such systematic subtle phonetic variation will not necessarily provide strong perceptual information, but, by adding natural variation, it will increase the perceptual coherence of the speech, making it easier to understand in adverse conditions ([4][10]).

One focus of our work is on correlates of phonemes that spread over several segments or syllables, because evidence that these acoustic properties are perceptually salient has implications for modeling speech understanding, encouraging us to relate phonetic variation to linguistic structure rather than to linear phoneme strings. Ideally, we want to show that acoustic properties in a segment that systematically varies with the linguistic-phonological attributes of a *non-adjacent phoneme* can enhance perception of that phoneme (see [9]). The chosen phonological contrast is voicing in syllable-coda obstruents; the systematic phonetic variation is in the (non-adjacent) onset of the same syllable.

Acoustic-phonetic differences associated with the voiced-voiceless distinction for English stops have been extensively researched, partly because they provide a rich test-bed for investigating speech perception, as many different acoustic properties can cue the distinction. For stops in coda position, the focus in looking for these properties is usually in the closure or burst of the stop itself, and in the preceding vowel duration (see any textbook). Earlier parts of the syllable can also cue coda voicing, though this is less well known, and the cues are not as powerful as the strongest of those at the end of the syllable ([15][16]).

We, [5], likewise showed systematic acoustic variation in syllable-onset /l/ contingent on coda voicing: in British English, onset /l/ is often longer and darker (lower spectral center of gravity and F2 frequency) before voiced codas than voiceless ones. Thus /l/ is longer and darker in *led* than in *let*.

A lexical decision task with cross-spliced naturally-spoken (C)lV(C)C pairs showed that the greater the acoustic change that resulted from cross-splicing, the slower listeners were to recognize that a stimulus was a real word, especially for voiceless codas [7]. Nonsense words showed no such effect. Thus, acoustic properties of onset /l/ can affect lexical decisions that depend on a phonological contrast in the syllable coda, i.e. in a non-adjacent phoneme.

Although we should expect only a weak perceptual relationship between the acoustic properties of onset /l/ and phonological voicing of the coda, the effect we found was too small to be convincing by itself. The experiments described in this paper directly explore the influence of systematic and unsystematic variation in the /l/ and vowel onset of synthetic *led* and *let* stimuli, using a forced-choice response.

2. Experiment 1

2.1. Introduction

Experiment 1 varied 3 properties of onset /l/: its duration, F2 frequency, and f0 (introduced because it was noticed that stimuli sounded more *led*-like when f0 was lower in /l/).

2.2. Method

Sensyn was used to synthesize 24 versions of *led*, identical except for the properties of the initial /l/ segment, which was one of 6 durations (range 70-170 ms), 2 F2 frequencies (1850 or 1740 Hz), and 2 f0 starting frequencies (180 or 168 Hz).

Other properties, illustrated in Fig. 1, are described in [6]. The impressionistic quality was good.

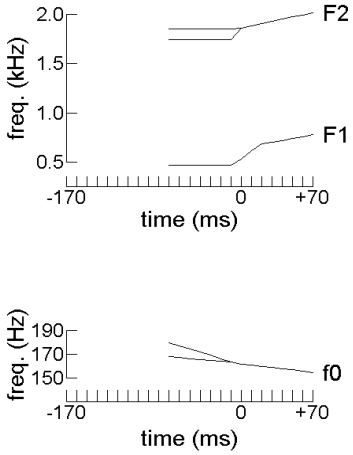

Figure 1: Expt. 1: f0, F1 & F2 trajectories up to 70 ms after vowel onset for the stimuli with 70-ms onset /l/s.

Each stimulus was truncated 80 ms after vowel onset. 300 ms of white noise (rise time 20 ms, decay time 50 ms, s/n ratio: 5 dB) was added, starting 70 ms from vowel onset. Thus the end of each stimulus seemed to be obliterated by noise: only its initial /l/ and 70-80 ms of the vowel could be heard.

Stimuli were randomized in 10 blocks of 24, preceded by 16 practice items, and played to 36 native speakers of English, who pressed one of two buttons depending on whether they thought the word was *let* or *led*. ISI = 3 s.

2.3. Results and discussion

Fig. 2 shows that shorter /l/s and higher f0 produced more *let* responses. Duration had the greatest influence, with a mean of 72% *let* responses when /l/ was shortest, falling to 44% when /l/ was longest ($F(5,165) = 26.38$, $p < 0.001$). F0 had a smaller effect: 60% vs. 56% *let* to stimuli with high vs. low f0 onset respectively ($F(1,33) = 9.21$, $p = 0.005$). F2 frequency did not significantly affect responses.

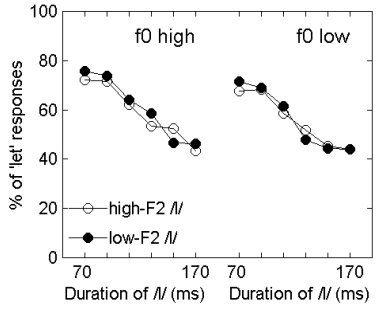

Figure 2: Expt 1: Mean percentage of let responses as a function of duration, f0, & F2 frequency of onset /l/.

Listeners gave more *let* than *led* responses: stimuli with the longest /l/s still had 45% *let* responses. This bias might be because the vowel was short, with no F1 offset transition, and followed by noise. Backward masking of the vowel by the noise might have made the vowel sound still shorter.

Contrary to predictions, a 90-Hz difference in F2 in /l/ did not affect how listeners reconstructed the missing syllable-final stop; but the 12-Hz difference in onset frequency of f0 did. F0 may have acted as a stronger cue to coda voicing than F2 frequency, overriding any effect that F2 might have had on coda identification. Yet f0 did not differ with coda voicing in natural speech [5], so its perceptual role was less central to our interests than that of F2. The low S/N ratio was a further potential confound: some subjects found the noise distracting. Expt. 2 assessed whether listeners use F2 to predict coda voicing when f0 is not varied and the added noise is quieter.

3. Experiment 2

3.1. Introduction

Of the five ways in which stimuli in Expts. 1 and 2 differed, the most important were slower /l/-vowel transitions in syllables with voiced codas, and coda-dependent spectral differences in the nucleus to reflect tendencies in natural speech [5]. We wanted to achieve spectral integrity of the whole syllable, to see if spectral differences early in the syllable are ever used in this task. If lexical representations include fine phonetic detail, to neutralize fine differences at a segment boundary could fatally disrupt perceptual decisions. Also, acoustic correlates of onset /l/ include abrupt changes in spectral shape and amplitude in the 10-20 ms after the acoustic segment boundary, and tongue-body changes due to the lateral may extend for 100 ms in a word like *let* [14].

3.2. Method

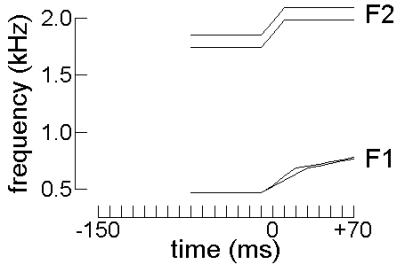

Figure 3: Expt. 2: F1 and F2 trajectories up to 70 ms after vowel onset for the stimuli with 70-ms onset /l/s.

Stimuli differed from Expt. 1 in 5 ways (Fig. 3). (1) F0 was 162 Hz through /l/. (2) There were only 5 /l/ durations (70-150 ms). (3) S/N ratio was 10 dB to reduce masking. (4) F2 frequency differed by 240 Hz throughout the CV: 1850 Hz in /l/ rising to 2090 Hz; and 1740 Hz rising to 1980 Hz. (5) F1 transitions at /l/-vowel boundary were 30 and 40 ms in high-F2 and low-F2 stimuli respectively. The procedure was as for Expt. 1, except that, to avoid carryover effects, each stimulus followed each other stimulus once (so there were 9 blocks of the 10 stimuli). There were 20 listeners and 20 practice items.

3.3. Results and discussion

Fig. 4 shows mean percent *let* responses as a function of onset duration and F2 frequency. As in Expt. 1, short-/l/ stimuli were labeled *let* more than long-/l/ ones ($F(4,76) = 7.12$, $p < 0.001$). Unlike in Expt. 1, there were 5% more *let* responses to high-F2 than to low-F2 stimuli ($F(1,19) = 4.59$, $p < 0.05$).

The two factors were independent of one another. Listeners were thus sensitive to both factors.

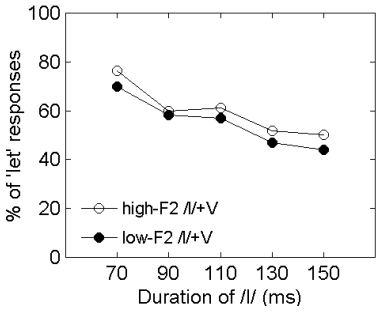

Figure 4: Expt. 2: Mean percentage of *let* responses as a function of onset duration & F2 freq. in /l/ + vowel.

However, this experiment does not show that spectral variation in the /l/ segment alone can help cue coda voicing. It may be that either a consistent difference throughout the syllable is needed, or simply a difference in the vowel alone, perhaps because it is heard last. The next three experiments investigated these questions.

4. Experiments 3a, 3b, and 4

4.1. Introduction

In order to see whether the spectral structure of onset /l/ and that of the nucleus are used as independent sources of information in the identification of coda voicing, Expts. 3a, 3b, and 4 used essentially identical stimuli. They were like the stimuli of Expt 2 except that F2 frequency in the /l/ segment and in the vowel varied independently rather than together, making 20 rather than 10 stimuli. These experiments differ in whether the stimuli were all presented in a single session, or blocked by F2 frequency in the vowel in order to more precisely assess the effect of varying F2 frequency in /l/ alone.

4.2. Method

Stimulus parameter values were as for Fig. 2 except that at each /l/ duration there were 4 rather than 2 stimuli varying in F2 frequency: ll = low in /l/ and vowel; lh = low in /l/, high in vowel; hl = high in /l/, low in vowel; hh = high in /l/ and vowel. F1 transition duration was always 40 ms.

All stimuli were presented in a single session in Expts 3a & 3b, and in two sessions, blocked by F2 frequency in the vowel, in Expt 4: hh with lh; and ll with hl, with block order counterbalanced across subjects. Stimuli were randomised in 3b and 4 so that each stimulus followed each other stimulus, and itself, once only (thus 20 repetitions); 3a followed the same principles, but, due to an error, had only 17 repetitions.

Practice varied. 3a: 20 items, no feedback. 3b: 28, with feedback on the first 6 items, which were the most extreme stimuli (hh with shortest /l/; and ll with longest /l/). 4: 18, with feedback on the first 6, which were the most extreme (high-F2 /l/ with shortest /l/; low-F2 /l/ and longest /l/).

Expts 2 & 3a had 11 out of 20 subjects in common. Expts 3a and 3b used similar presentation conditions but all different subjects. Expt. 3b thus replicates 3a, but it was run *after* Expt. 4, in the same session and with the same 20 subjects as Expt. 4.

4.3. Results and discussion

Figure 5 shows results for all three experiments. As before, shorter /l/s produced more *let* responses in all 3 experiments ($p < 0.001$). The range of *let* responses across duration was much greater than in Expts 1 and 2, especially in Expts. 3b and 4. The relevance of this point will become clear below.

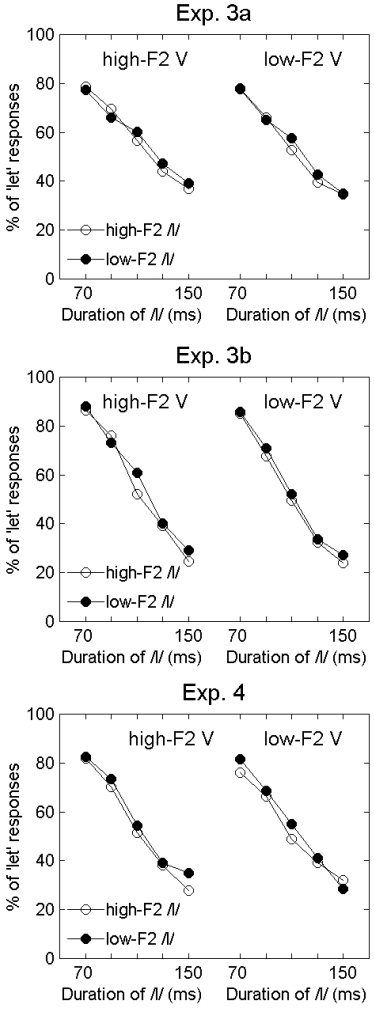

Figure 5: Expts. 3a, 3b, 4: Mean percentage of *let* responses as a function of onset duration, F2 frequency in onset /l/, and F2 frequency in the vowel.

Responses due to spectral variation were more complex. In Expts 3a and b, higher F2 frequency in the vowel led to more *let* responses, as predicted, but in 3-factor repeated measures ANOVAs this was only significant for 3b (5% difference**,** $F(1,19) = 7.71$, $p = 0.012$; in 3a, 2.5% difference, $F(1,19) = 3.01$, $p = 0.099$). But the difference was significant ($p < 0.04$) for 3a in a logistic regression. In contrast, F2 frequency in /l/ was not significant (3a: $F(1,19) = 0.97$, $p = 0.34$; 3b: $F(1,19) = 2.87$, $p = 0.12$). There were no significant interactions.

Thus in Expts 3a & b listeners were only slightly sensitive to F2 frequency in the vowel (especially in 3a), and not at all sensitive to it in /l/. This was surprising because half the stimuli (ll, hh) were essentially identical to those of Expt. 2, which did show a significant F2 effect. Presumably, when F2 frequency varies independently in onset /l/ and nucleus,

spectral variation is effectively random within the stimulus set, so spectral properties are not experienced as a reliable cue, even when they are consistent within ll and hh syllables and are used as cues to coda voicing in other circumstances.

Expt 4, which confined spectral variation in each stimulus set to /l/ and gave feedback during practice, partly supports this interpretation, and highlights Ss' sensitivity to properties of the stimulus set when cues to phoneme distinctions are ambiguous. Thus, whereas in Expts 3a & b high-F2 vowels produced more *let* responses and /l/ F2 had no effect, in Expt 4, high-F2 /l/ produced more *led*s ($F(1,18) = 5.70$, $p = 0.028$) and vowel F2 had no effect ($F(1,18) = 2.85$, $p = 0.11$).

However, there is evidence from the training data and the individual response patterns in each (vowel-F2) block of the main experiment that subjects changed strategies during Expt. 4 [8]. Specifically, as a group they took some time to learn about the durational cue: means for the first training trials show no sensitivity to duration, whereas those of the second training trials are strongly durational and very like Fig. 5. This pattern holds no matter whether the first training was with the low-vowel-F2 stimuli or the high-vowel-F2 stimuli. In the main experiment, although most listeners used duration (with or without spectral cues) in both vowel-F2 blocks, some used spectral cues without durational cues in one or both vowel-F2 blocks. Crucially, everyone who used duration in the first vowel block also used it in the second; some who used spectral cues in the first vowel block used duration in the second, with or without spectral cues; and relatively few subjects responded spectrally in the second vowel block. Thus people took different lengths of time to learn about the durational cue, but 'once durational, always durational'.

In an attempt to factor out the effect of subjects learning to use duration, we compared only earlier responses, namely those in the first vowel-F2 block each subject heard. A duration x /l/-F2 x 1st vowel-F2 ANOVA showed that the effect of vowel-F2 was marginally significant in the predicted direction ($F(1,18) = 4.20$, $p = 0.055$), and confirmed the significance of the difference in the unpredicted direction for /l/-F2 ($F(1,18) = 7.07$, $p = 0.016$), with no /l/ x vowel interaction ($F(1,18) = 0.28$, $p = 0.60$). Thus, in the first vowel block, when the stimuli were still relatively unfamiliar and several subjects had not yet learned to use duration as the main cue to coda voicing, there were more *led* responses when /l/-F2 frequency was high, yet more *let* responses when vowel-F2 frequency was high.

5. Concluding remarks

This work shows that perceptual cues to coda voicing are distributed across the entire syllable, not just the rhyme. Listeners are sensitive to whether spectral properties of the onset and nucleus vary systematically and naturally: they attended less to F2 frequency when it was manipulated independently in the onset and nucleus. The complexity and sensitivity of the decision-making process is hinted at: training data [8] suggest that listeners first experienced long-/l/ stimuli as spoken more slowly than short-/l/ stimuli and attended to their spectral properties, but gradually turned their attention away from spectral properties to focus on onset duration, which they perceived as more reliable. More work is needed to assess the effect of spectral variation when /l/ duration varies little. These conclusions suggest that the listener is highly sensitive to the distribution of fine-grained acoustic details in speech, and that the perceptual weight of each detail is constantly readjusted depending on how well it allows a target word to be differentiated from its competitors (see [7][8][9]).

6. Acknowledgements

Funded in part by grants from the Swiss National Science Foundation (83BC-056471) and the EPSRC (GR/N19595). We thank E. Fixmer, S. Heid, J. Local and especially R. Smith. Experiments 1 & 2 were presented at SWAP (ISCA Conference on Spoken Word Access Processes, 1999).

7. References

[1] Boardman, I., S. Grossberg, C. Myers and M. Cohen "Neural dynamics of perceptual order and context effects for variable-rate speech syllables", *Perception and Psychophysics*, 6:1477-1500, 1999.

[2] Coleman, J., "Cognitive reality and the phonological lexicon: a review", *J. Neurolinguistics*, 11:295-320, 1998.

[3] Goldinger, S.D., "Words and voices: Perception and production in an episodic lexicon". In Johnson, K. and Mullennix, J.W. (Eds.), *Talker Variability in Speech Processing* (pp.33-66). San Diego, Academic, 1997.

[4] Hawkins, S. "Arguments for a nonsegmental view of speech perception." In K. Elenius and P. Branderud (eds.) *Proc. ICPhS*, 3:18-25, 1995.

[5] Hawkins, S. and Nguyen, N., (under revision), "Acoustic properties of syllable-onset /l/ dependent on syllable-coda voicing". Accepted subject to revision, *J. Phonetics*.

[6] Hawkins, S. and Nguyen, N., "Predicting syllable-coda voicing from the acoustic properties of syllable onsets", *Proc. of the Workshop on Spoken Word Access Processes*, Nijmegen, 167-170, 2000.

[7] http://www.lpl.univ-aix.fr/~nguyen/labphon6.pdf

[8] http://www.lpl.univ-aix.fr/~nguyen/let-led.html

[9] Nguyen, N. and Hawkins, S., "Implications for word recognition of phonetic dependencies between syllable onsets & codas", *Proceedings of the 14th ICPhS*, 1999.

[10] Ogden, R., Hawkins, S., House, J., Huckvale, M., Local, J., Carter, P., Dankovicová, J., and Heid, S. "ProSynth: An integrated prosodic approach to device-independent, natural-sounding speech synthesis", *Computer Speech and Language*, 14:177-210, 2000.

[11] Pisoni, D.B. "Some thoughts on 'normalization' in speech perception". In Johnson, K. and Mullennix, J.W. (Eds.), *Talker Variability in Speech Processing* (pp.9-32). San Diego, Academic Press, 1997.

[12] Plaut, D.C., and Kello, C.T. "The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach". In MacWhinney, B. (Ed.), *The Emergence of Language* (pp 381-415). Lawrence Erlbaum, Mahwah, NJ, 1999.

[13] Pulvermüller, F., "Words in the brain's language", *Behavioral and Brain Sciences*, 22:253-336, 1999.

[14] Stevens, K.N. *Acoustic Phonetics* (pp. 543-554), MIT Press, Cambridge, MA, 1998.

[15] Summers, W.V., "F1 structure provides information for final-consonant voicing", *J. Acoust. Soc. Am.*, 84:485-492, 1988.

[16] Wolf, C.G., "Voicing cues in English final stops", *J. Phonetics*, 6:299-309, 1978.