Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Phonetics 31 (2003) 321–339

Journal of Phonetics

www.elsevier.com/locate/phonetics

Variable domains and variable relevance: interpreting phonetic exponents

John Local*

Department of Language & Linguistic Science, University of York, Heslington, York Y010 5DD, UK

Received 26 September 2002; received in revised form 24 June 2003; accepted 2 July 2003

Abstract

Spoken language is a resource which is systematically deployed in the management of social interaction, its primary site of occurrence. The patterns and structures in language are emergent properties of, and shaped by, the contingencies and demands of social interaction. However, despite significant advances in modeling speech perception and understanding, and an increasing acknowledgment of the relevance of phonetic detail, there continues to be an overemphasis on issues of lexical distinctiveness and lexical access with the consequence that many kinds of systematically-controlled fine phonetic detail do not find their way into contemporary models. I argue that it is now timely to think more carefully about what it means to talk about linguistic–phonological contrast and distinctiveness and the relevance of phonetic detail. I argue that:

i. Lexical contrast is overvalued in speech perception and understanding;
ii. It is time to examine more closely the phonetic detail of talk-in-interaction;
iii. Particular phonetic details and phonetic variability are associated with particular interactional, grammatical and lexical systems and that this 'context-embeddedness' is both useful for and used in speech understanding.

1. Introduction

There is a long and continuing tradition in research in speech perception and understanding, linguistics, phonetics and psychology of focussing on lexical aspects of language and on so-called phonemic contrasts. See, for instance, the recent paper by Norris, McQueen, and Cutler (2000),

*Tel.: +44 1904 432658.
E-mail address: lang4@york.ac.uk (J. Local).

0095-4470/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0095-4470(03)00045-7

and the responses to it, in *Behavioral and Brain Sciences,* which takes as uncontroversial its starting point that 'sounds uttered by a speaker are converted to a *sequence of words* recognized by a listener' (my emphasis). In part, this tradition has arisen because practitioners in linguistics, phonetics, psychology and acoustics have shown a willingness to analyze language divorced from its natural site of occurrence, which is talk-in-interaction. One of the consequences of this focus has been an over-emphasis on lexical distinctions and on particular and rather localized ('punctual') aspects of the phonetic detail of speech related typically to 'phonemic contrasts' or localized aspects of 'allophonic' patternings (e.g. coronal assimilation effects). This in turn has drawn attention away from the presence of other kinds of meaningful linguistic information in the speech signal and stands in danger of distorting the relative importance of various processes in models of how speech is perceived and understood.

Nonetheless, recent work has produced a number of impressive and challenging approaches to speech perception, speech understanding and phonological representation (e.g., Goldinger, 1997; Goldinger & Azuma, 2003; Coleman, 1998; Boardman, Grossberg, Myers, & Cohen, 1999; Pierrehumbert, 2001; Grossberg, 2003) which now emphasize the importance of phonetic detail in modeling emergent, multi-functional, multi-category, distributed, cognitive representations. However, constructs such as 'phoneme' 'allophone' and 'segment' are still routinely mobilized in contemporary research despite what I detect is an uneasy sense that these categories are not really necessary or even helpful. For instance, as a preface to their elegant modeling of variable-rate speech categorization, Boardman et al. (1999) write 'The language units that are familiar to us from daily experience, such as phonemes, letters, and words, do not form appropriate levels in a language processing hierarchy' (1999, p. 2). It would seem that few, if any, models have truly freed themselves from the tyranny of the lexicon or of a belief in the primacy of short-domain, local phonetic phenomena.

In what follows I explore the nature and role of 'fine phonetic detail' in speech. I suggest that it is timely to rethink what orders of phonetic detail and senses of 'relevant'/'meaningful' contrast which are currently not being entertained should now be entertained. As I will show, there is a lot of structural information encoded in the fine phonetic detail of the signal. Meaning is much more than lexical meaning. If we want to construct a robust, integrated model of speech perception, speech understanding and phonological representation, we need to entertain richer ideas about the ways in which phonetic detail relates to the construction of meaning.

This paper is organized as follows. Section 2 presents some data from talk-in-interaction which provides evidence of systematic 'nonphonemic' phonetic detail and the way it functions informatively in everyday talk. Section 3 reviews some evidence which suggests that phonetic detail associated with phonological contrasts may extend over longer temporal windows than is usually assumed (nonlocal exponency). Section 4 examines some of the ways in which phonetic detail may relate differentially to grammatical and lexical structures.

2. Systematic phonetic detail and talk-in-interaction

Extensive analysis of data from talk-in-interaction challenges a number of conventional assumptions about the orders of phonetic detail which are produced and attended to in ascribing meaning (e.g. Local & Kelly, 1986; Schegloff, 1998; Wells & MacFarlane, 1998; Shriberg, 1999; Ogden, 2001). Four key results emerge from this enterprise which are important for an understanding of the kinds of process we may wish to model in speech perception and understanding:

i. Each part of the speech signal relates to several functions simultaneously.
ii. Some systematic differences in phonetic fine detail are relatively localized in the speech signal, others stretch over several syllables.
iii. No order of detail can be dismissed, a priori, as disorderly, accidental or irrelevant.
iv. Fine phonetic detail simultaneously provides interactional, grammatical and lexical information.

The observations made here about phonetic patterns in interaction are drawn from many hundreds of cases in over 20 h of recorded talk-in-interaction which includes face-to-face talk, telephone calls and radio-phone-ins. The recordings involve a range of speakers, in terms of age, sex and social class, a range of activities and settings, and a range of varieties of British and American English, including a number of nonstandard varieties.[1]

2.1. Phonetic detail and the joint production of turns-at-talk

Consider, as a first example, the following two stretches of talk produced in the course of everyday conversation. The first occurs during a telephone call, the second occurs in face-to-face interaction:

(1) (i) *but when we walk out of the class nobody knows what went on*
(ii) *we're not going to get back 'til like Monday morning*

There is nothing particularly remarkable about these two stretches of natural talk, which are produced without hesitations or other kinds of disruption, except that they are jointly accomplished by two speakers, rather than one. The relevant turns are indicated by arrows in data fragments (2) and (3)[2]

[1] The data fragment labeled 'Two girls' comes from a telephone call between two young American women (Schegloff, 1996, p. 57), recording provided by Emanuel Schegloff, Department of Sociology, UCLA. Those labeled 'Lab' were made in 2000 for course credit by Rachel Dawes and Gareth Walker, undergraduates at the University of York. Those labeled 'McN' come from recordings made during preparatory work for the Tyneside Linguistic Survey (Local et al., 1986; p. 415). Those labeled 'Call You and Yours' come from the BBC radio phone-in program of the same name, broadcast 28.3.2002. The fragment labeled 'Heritage' comes from a collection of telephone calls made in the 1980s and provided by John Heritage, Department of Sociology, UCLA.

[2] The transcriptions of interaction given here are based on the conventions used in Conversation Analysis (Atkinson & Heritage, 1984, pp. ix–xvi). Turns at talk are shown sequentially down the page. Speakers are identified, by initial, at the beginning of a line. Audible in-breaths are indicated by sequences of 'h'; increased duration is indicated by ':'. Intervals of no talk are timed in seconds and durations are shown within parentheses; '(.)' indicates a brief no-talk interval of around 0.1 s. Vertically-aligned left square brackets indicate the start of talk from one speaker which overlaps that of another. Vertically-aligned right square brackets indicate the point of offset of overlapping talk.

```
(2)     [Two Girls]

 B:      hhhh and we nod when he wants us to say yes
         a[n ]
 A:       [Ye]ah
 B:      we raise our hands when he wants to take a poll
 B:      'n[ : : ]
 A:         [Yeh ]
 B: →    hh you know but when we walk out of the class
 A: →    nobody knows what went on


(3)     [Lab]

 J:      and then like we don'thhh I mean I don't know when (0.4)
     →   we're not going to get back 'til lik:e [Monday morning] h
 R: →                                           [Monday morning]
           (.)
 R:      I  know
```

We can note, amongst other things, that in doing talk of this kind

i. participants attend to the moment-by-moment evolution of complexes of phonetic detail and what that detail encodes about other levels of linguistic organization so that they can locate the precise temporal moment to begin their talk. The talk is always syntactically and lexically well-fitted; indeed, we find instances where a second speaker will complete some *morphological* part of a word in another speaker's talk (Lerner, 1996);
ii. participants appear to be monitoring the detail (particularly the timing) of both their own talk and the talk of others. They can entrain the rate, rhythm, timing and also pitch range and loudness characteristics of their speech to that which has just been produced by another speaker (Local, 2000). One piece of evidence for this is that 'choral' talk produced by second speakers, such as that exemplified in fragment (3), is rather accurately timed with respect to that of the other speaker. Typically, such talk is timed not to extend beyond the end of that produced by the first speaker.

Sequences of the kinds illustrated above are not rare. Indeed, data from everyday interaction makes it clear that participants systematically produce and attend to many kinds of nonlexical ('subphonemic') phonetic detail in the on-line construction and understanding of what is being said, why it is being said and what sort of functions it has (e.g. Local & Kelly, 1986; Local, 1996; Ogden, 2001; Local & Walker, 2002). That this is so raises the question of whether such detail and variability play a key role in the way in which speech is parsed into chunks and how categories of various kinds are represented. Such an orientation to a variety of fine phonetic details suggests that current models of speech perception and speech understanding which concentrate on lexical distinctiveness neglect information-bearing properties of the speech signal. I suggest that this phonetic detail is just as 'linguistic' and meaning-bearing as those details of the speech signal that express lexical items, and that in modeling the real-time perceptual behavior of speakers and listeners we ignore it at our cost.

2.2. Phonetic detail and turn transition

One of the principal reasons for examining the ways in which phonetic details and phonetic variability are linked with different kinds of interactional activities (and their linguistic components) in talk is to gain a richer, more grounded understanding of 'context' and 'function'. In doing this we can also begin to reconfigure our understanding of the constitutive elements of phonetics and phonology which play a role in the architecture of models of perception and speech understanding.

Consider the following transcriptions of the English words *got, Vincent, that, Mick, amenities, took, back, toilet, kinds* and *tell* drawn from face-to-face spoken interaction between speakers of Tyneside English

gɒʔt˺	v̥ĩnən̺ˈʔ	n̺əʔ	mɪk̟˺	ɪmɪˈnɪ̂ʔt¯iz̥
tʰʊˈʰk̲ʰ	baʱk̲ʰ	tʰəilʲɪtʰ	kʰɛɪndz̥	tʰɛlʲ

The word-initial, medial and final voiceless plosives exhibit differences in their phonetic detail that typically involves differences in the co-ordination of oral and glottal activity and timing. There is particular variation in the word final plosives: *took, back* and *toilet* end in voiceless plosives which are all aspirated, while the word-final plosives in *got, Vincent, Mick* and *that* are not aspirated. Is this variation just happenstance? In treatments of the phonology of English it is often noted that while aspiration with voiceless plosives is lexically contrastive in initial position there is 'free variation' in final position (e.g. Kreidler, 1989; Davenport & Hannahs, 1998). That is, such plosives can be realized as aspirated, unexploded or in some circumstances (or accents) as simple glottal stops, all of which we find here. So is this just free variation? If we were only concerned with modeling lexical contrast we might answer 'yes'. But if the interactional context of these tokens is examined it is clear that word-final aspiration of voiceless plosives represents a particular kind of systematic lawful variation. All the words illustrated are taken from two fragments of a multi-party conversation ((4) and (5) below). There are seven speakers engaged in this conversation at various points: four men and three women. The word-final voiceless plosives which are aspirated are just those which also occur turn-finally in the following two sequences. (Turn-final aspirated plosives are indicated by underlining.)

(4) (McN)

A: Have you got your snaps Vincent that Mick took̲
N: No Connie's got them

(5) (McN)

P: and I says oh oh she's away round th? the back̲
M: aye (.) she's e[h
J: [gone to the toilet̲
N: all kinds of amenities I'll tell you

The audibly aspirated release of word-final plosives is being systematically employed to indicate that the speaker has reached the end of a turn at talk (Local, Kelly, & Wells, 1986). This variability is *interactionally* informative and conditioned by its place in the speaking turn. There is ample evidence that the presence of word-final aspiration is systematically produced (and oriented to by other participants) as part of a package of events which indicate turn-finality. Indeed, there is evidence from the interactional behavior of the participants that for Tyneside speakers, this aspiration, in conjunction with centralization of the vowel qualities in the turn-final foot, is criterial for signaling turn-transition (Local et al., 1986). The rule is, if there is a word at the end of a turn and that word ends in a voiceless plosive, it is aspirated. Voiceless plosives that are not word-initial and not turn-final are typically co-ordinately glottalized and show different types of variation; Docherty, Milroy, Milroy, and Walshaw (1997) offer a sophisticated socio-linguistic discussion of such variation. The temporal window for this distributed bundle of turn-final features may be quite large. For example, in the case of a disyllabic word like *toilet* the final aspiration features occur 260 ms after those turn-ending features located around the stressed vowel of the last foot of the turn.

In some 9 min of conversation from which these instances are drawn there are 206 words which end in voiceless plosives. Of these, 61 occur in turn-final position (/p/: 4, /t/: 45, /k/: 12) and 145 occur in turn-medial position (/p/: 8, /t/: 104, /k/: 33). Of the turn-final tokens only one exhibits no aspiration. Of the turn-medial tokens only four (3%) exhibit aspiration. Three of these aspirated turn-medial tokens are followed by words beginning with aspiration (*just <u>h</u>eard, about <u>h</u>alf, cannot <u>h</u>ear*) and the other is in the phrase *at all.* Docherty et al. (1997) report figures roughly consistent with this finding though with a somewhat higher percentage for glottalized/nonaspirated turn-final /t/ (9%). However, in their sample, 5% are produced by young working-class females. None of the women in the conversational data I have presented come from this social group.

2.3. Phonetic detail and 'discourse markers'

It has long been known that the phonetic detail of the same lexeme (e.g. *some, that, to*) can differ when it has different syntactic functions e.g. deictic *that* as opposed to complementizer *that* as in *<u>that</u> house is big* and *the house <u>that</u> Jack built* (Sweet, 1910; Jones, 1934). Similarly, forms such as *for* and *four* may show phonetic differences which can be related to their different linguistic functions (Lavoie, 2002). So too can phonetic detail differ between forms where one functions 'lexically' and one has a primarily 'interactional' function (Local & Kelly, 1986). Forms with 'interactional' function are sometimes referred to as 'discourse markers' (Schiffren, 1987). Such phonetic discriminability of forms prompts the questions of whether 'form' and 'function' may be rather more closely linked than is usually thought, and of the uses that may be made of this link by the perceptual system. For example, we regularly find forms of 'think', in the phrase *I think*, such as these taken from recordings of everyday talk:

(a) ɐθĩŋkʰ   ɐθĩŋʔ   ~   (b) əhĩŋk̞ʰ   ɐh̃ĩŋˑ   ɐ̃h̃ĩŋ̆ˑ

The (a) forms are what we might expect from a 'canonical' pronunciation of *think* and regularly occur in ordinary everyday talk. The (b) forms, however, are noticeably 'noncanonical' and show considerable variability in the temporal organization of their phonetic characteristics. These too regularly occur in everyday talk. One very obvious difference between the two sets of forms is that the (a) forms have voiceless dental friction at the beginning of *think* whereas the (b) forms do not. Rather, they begin with voiceless glottal friction, which may also have nasalization and co-occurring vocal-fold vibration. The (b) forms also have rather central vowels and an overall lax articulatory setting which may, for instance, result in a lack of complete closure for the final velar articulation. They are also usually shorter than the (a) forms and frequently have breathy phonation throughout. Notice here that there appear to be phonetic events which encompass and characterize the whole of the phrase *I think*. (The following observations are not meant to constitute an exhaustive analysis of the relationships between phonetic variability and meaning/functioning of *I think* in everyday talk.)

The (a) forms of *I think* regularly occur in sequences of talk such as the following (instances are underlined):

```
(6)      [Call you and yours]

BC:      I think it's very interesting that (0.2) uthat that uh people's .hhhhmh I mean
         whatever one thinks about the specific remarks

(7)      [Call you and yours]

WC:      uh? ?I? ?I think that people (0.3) have not yet woken up hhhhh er to (.) to
         what's going on
```

In cases such as these we do not appear to find tokens with glottal friction at onset of *think*. By contrast, the (b) forms of *think* are found in sequences such as:

```
(8) [Lab]

F1: the only I might have (.) going to buy some shoes I think

(9) [Heritage]

I:  so uh hh we'll see.h
        (0.8)
A:  mm hm
        (0.3)
I:  they should be here by the time you come out next weekend I think
```

The (a) forms represent 'lexical' uses of the verb *think*. The (b) forms do not. Unlike the (a) forms, the (b) forms of *think* do not co-occur with preceding second or third person pronouns. The (b) forms of *I think* do not carry lexical meaning. Rather, they represent a fused chunk of talk ('unitary epistemic phrases', Thompson & Mulac, 1991) which serves interactionally to 'hedge' the meaning of preceding talk and are regularly found turn-finally. Specifically the form without dental friction is used to indicate that the proposition expressed in the prior talk is not necessarily (going to be) the case. Notice, of course that this postpositioned hedge requires listeners to recast their understanding of any immediately prior talk which can be thought of as a very long temporal window within which meaning has to be computed. So, for instance in fragment (8) the relevant prior talk consists of six syllables with a total duration of 1 s; in fragment (9) the relevant prior talk consists of 13 syllables with a total duration of 2 s.

The cases of (*I*) *think* (and similar phonetic variability observed for *I mean*) are not isolated (Local, 2002). What we have here look like gestalts determined by their functional role in the sequential structure of interaction. In turn this suggests that part of perceptual categorization, recognition, and chunking may be facilitated by systematic patternings which relate particular instances of speech to particular places in interaction and particular kinds of function in interaction. These may well be grammaticalized chunks produced and processed as wholes and not via access to, or the composition of, separate lexical items.

3. Variable domains: temporal extents and phonological contrasts

Even if we do just focus on lexical-type contrast, there is a growing body of evidence that suggests there is more to the phonetic correlates of lexical phonological distinctions than conventional wisdom recognizes and than is usually included in perceptual models. A number of researchers have shown that phonetic detail which is associated with lexical meaning can be distributed as well as local (e.g., Kelly & Local, 1989: 218–262; Tunley, 1999; Heid & Hawkins, 2000; West, 2000; Coleman, 2003; Carter, 2003). These findings are important because such data bring into question a key assumption which seems to underlie much of the research into speech-perceptual processes: that phonetic details (exponents) associated with so-called 'phonemic' contrasts are local and punctual. Though it has long been known that particular phonetic parameters may spread over stretches longer than a 'segment' (e.g. nasalization (Krakow, 1993; Solé, 1995), lip-rounding (Benguerel & Cowan, 1974), vowel-to-vowel co-articulation (Alfonso & Baer, 1982; Recasens, 1989)) such spreading of phonetic properties that are essential to the conditioning segment has typically been seen as primarily due to natural, i.e. unconstrained physiological, processes. The more recent findings concerning the distributed nature of phonetic information associated with phonological distinctions draw attention to other kinds of phonetic detail which cannot be given such straightforward explanation. Nonetheless, if phonetic correlates of phonological distinctions *can* be rather widely distributed in the signal we need to look for these and model them just as systematically as any local phenomena.

3.1. Nonlocal exponency: assimilation in English

It is well documented that, in English, the place of articulation of word final consonants such as /n/ may assimilate to (or share) the place of articulation of a following word-initial consonant. So in *ran quickly* it is not uncommon to find a velar nasal, /ŋ/, at the end of *ran* rather than an alveolar nasal, /n/, as might be found in *ran down* or *ran around.* However, there is evidence to suggest that even where there is such assimilation to a velar nasal, the words are *not* identical with forms such as *rang* which have final citation-form velars (Kelly & Local, 1989, pp. 155–158). Whatever the nature of the /n/$\sim$/ŋ/ contrast is, for some speakers it would appear not to be straightforwardly neutralized in assimilation. The data also suggest that the phonetic correlates of the contrast are not simply restricted to the final consonants but may spread throughout the whole word.

A similar pattern of distributed phonetic exponents appears to be present in the speech elicited in Nolan's (1992) assimilation experiment, which was also used in a perception experiment by Kerswill and Wright (1989). Nolan's investigation focussed on whether assimilation is best treated as a gradual or discrete phenomenon. He employed seven word-pairs, differing only in whether their final consonant was /d/ or /g/, embedded in a sentence frame which, for the /d/-final words, could induce assimilation to a velar place of articulation. Fig. 1 illustrates some acoustic consequences of the kinds of patterns I have been describing. It plots the first three formants in Bark for two productions of *lead* (/lɛd/) and one of *leg* (/lɛg/), from Nolan's data. One production of *lead* ends in a 'full alveolar' consonant and one in a completely assimilated 'zero alveolar' consonant (Nolan, 1992; p. 267). The tokens were spoken in the frame *was the_covered, did you notice?* In the figure the 'full alveolar' token is labeled *FA.lead*, the 'zero alveolar' token is labeled *ZA.lead* and the 'velar token' (*leg*) is labeled *V.leg*.

Fig. 1 gives some indication of the acoustic differences and similarities between the cases. Comparison of the 'full alveolar' *lead* and the 'zero alveolar' *lead* indicates that the two cases are very similar; the main differences occur towards the end of the vocalic portion for F_1 and over the last third of the vocalic portion for F_3. In contrast, when 'full alveolar' *lead* and 'velar' *leg* are compared, we observe differences throughout the whole syllable for F_1 and F_2. 'Velar' *leg* has a lower F_1 throughout. This correlates with a percept of a closer articulation particularly in the vowel (as Nolan himself notes 1992, p. 272). Likewise, F_2 has a different trajectory throughout *leg* compared with 'full alveolar' *lead,* and also a difference in F_3-F_2 spacing at the end of the vowel (consonant closure). For 'zero alveolar' *lead* and 'velar' *leg* comparisons we observe a difference through the syllable for F_2, again with a different trajectory throughout, at the end of the vocalic portion for F_3 and in the mid-part of the vowel for F_1. These differences suggest that even in the case of a fully assimilated 'zero alveolar' token there may be differences preserved at other earlier points in the word, and possibly a less clearly velar articulation at the end. Likewise, the EPG

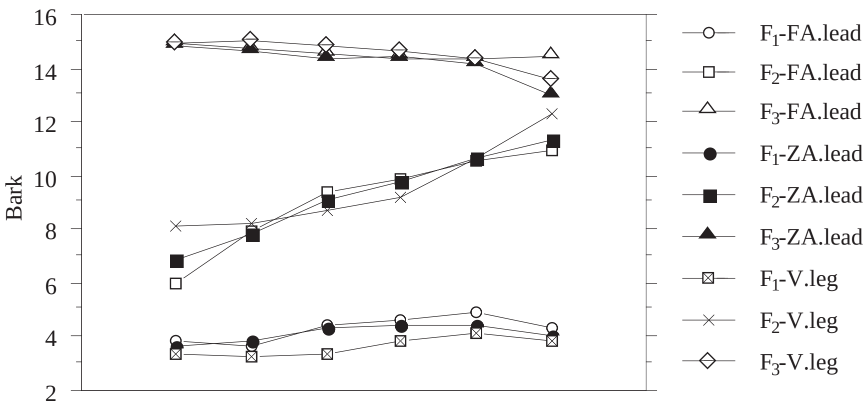

Fig. 1. Bark-scaled F_1, F_2, F_3 frequencies of single tokens of 'full alveolar' *lead*, 'zero alveolar' *lead* and velar *leg* produced by one speaker. Measurements taken at sonorant onset, midpoint, offset and vowel onset, midpoint and offset.

patterns for the final consonants in 'zero alveolar' *lead* and *bed* on one hand, and 'velar' *leg* and *beg* on the other (Nolan, 1992, p. 273) indicate different kinds of occlusion.

Nolan's results are suggestive rather than definitive. For each of the seven word-pairs in the data set there is only one token of the 'full alveolar', the 'zero alveolar' and the 'velar' form. However, if results such as these can be shown to be robust, the consequences for perceptual modeling are clear. It is inappropriate to model variation in final place of articulation as the only difference between full alveolars, assimilated alveolars and velars. The distributed acoustic cues are available to the perceptual system and models which employ 'impoverished' phonemic representation ignore them at their cost. The window for integration of information may be greater than a 'single segment'.

3.2. Nonlocal exponency: the voice–voiceless distinction in intervocalic /t/~/d/ in English

Evidence of even longer temporal windows for phonetic parameters corresponding to a 'segmental' contrast is presented by Scott (1984). She presents the well-known case of tapped articulations corresponding to intervocalic /t/ and /d/ in American English (e.g., in *metal/medal, rater/raider:* 'T-words' and 'D-words') in which the perceptual difference, where there is one, has been identified as a durational difference in the vowel preceding the tap with longer vowels before taps in D-words and a different vowel-tap duration ratio (Port, 1979; Port & Dalby, 1982). Scott suggests that the vowel duration difference is actually rather fragile as measurements indicate that vowels preceding taps in D-words are not necessarily longer than those in T-words. However, in a perceptual experiment she demonstrates that, even when there is no durational difference in the vowel preceding the taps in the two word types, listeners can reliably distinguish between the two classes of word. Scott's acoustic analysis of production data from two speakers shows that there are systematic differences in the spectral characteristics not only of the vowels preceding the taps in the two classes of word but, to a lesser extent, in the vowels which follow the taps. Some of these differences are illustrated in Fig. 2.

Scott's data indicate that the differences are primarily associated with F_1 and F_2 and are distributed to varying degrees throughout these words. Kelly and Local (1989, pp. 158–161) provide an impressionistic phonetic assessment of part of Scott's data. They examined 15 of the 40 T-/D-word-pairs in Scott's data and observed that the T-words had generally fronter, closer vowel qualities than did the D-words. They also noted that final consonants in these words displayed 'front' or 'back' qualities similar to those shown by the vocalic portions of the two sets of words. In some cases it was found that the consonantal portions of T-words had what sounded like advanced places of articulation as compared with their D-word congeners e.g., final laterals had 'clearer' resonance and the nasal portions at the ends of words like *putting* for both speakers were noticeably front of velar in contact as compared with these portions in *pudding*. In the absence of other (e.g. EMA) data it is not possible to confirm the articulatory basis of these impressionistic percepts. However, the acoustic data for final laterals (e.g., Fig. 2) do offer some support. Kelly and Local also suggest that for some of the word-pairs (e.g., *latter* and *ladder*) the initial consonants may have different auditory qualities. This observation predicts that an initial sonorant in a syllable with a voiced coda will show spectral (and possibly) temporal differences when compared to one with a voiceless coda. While Scott's data are not extensive, recent experimental work by Kwong and Stevens (1999) confirms Scott's and Kelly and Local's

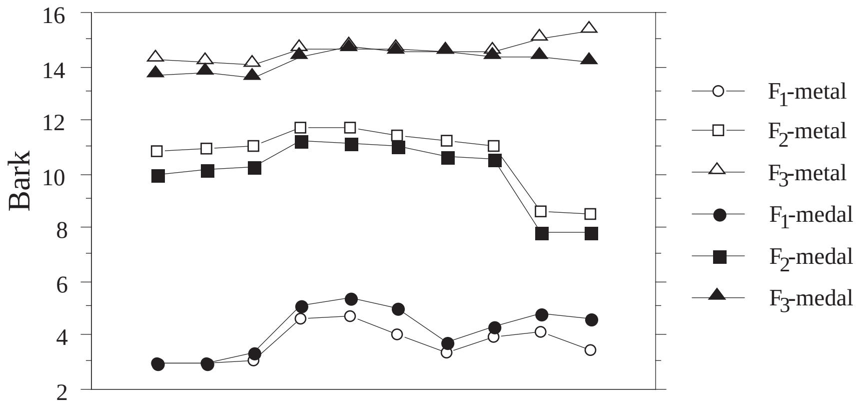

Fig. 2. Bark-scaled F_1, F_2, F_3 plot of *metal* and *medal* for the two speakers in Scott (1984). Data averaged over two productions of each word measured at sonorant onset, midpoint, offset; vowel-1 onset, midpoint and offset; tap midpoint; vowel-2 onset, midpoint and offset.

observations concerning the closer and fronter quality of the first vowel in 'writer', compared with 'rider' for American English speakers; the 'temporal spreading' of the coda voice–voiceless difference into syllable onsets has been confirmed experimentally by van Santen, Coleman, and Randolph (1992) and Hawkins and Nguyen (2004).

The assimilation and alveolar-tapping examples I have discussed raise the question of whether we have a 'punctual' phonological distinction with distributed cues (such as 'voicing' in coda position in English) or whether we have a different kind of phonological organization with a different structural domain. We could equally well say that the 'final voicing distinction' in a pair of syllables such as *dock* and *dog* has as its domain the rime of the word rather than the coda. Indeed, given the results reported in the preceding paragraph concerning the implications of coda voice–voiceless differences for syllable onsets, we may well need to reflect more radically on how best to express domains for phonological contrast. Nonetheless, what might look like a problem for a 'punctual' analysis (e.g., sameness at the end of assimilated *ran* and velar *rang*) is potentially resolved by an analysis in terms of distributed phonetic properties.

3.3. Nonlocal exponency: liquids in English

Kelly and Local (1986; 1989, pp. 203–217) describe another instance of nonlocal exponency which may extend beyond word boundaries. On the basis of impressionistic observation, they claim that liquid consonants in English, such as those at the beginning of *lip* and *rip*, may be differentially associated with 'clear' and 'dark' resonance in different varieties (see also Kelly, 1989). Thus, for example, some varieties of English have 'clear' initial and intervocalic /l/ and 'dark' /r/ while other varieties have dark initial and intervocalic /l/ and 'clear' /r/. One reasonably robust acoustic correlate of 'darkness' is relatively low F_2 (Fant 1960; Ladefoged & Maddieson, 1996; Recasens, Fontdevila, & Pallarès, 1996; West, 1999; Local & Carter, 2002). This is illustrated in Fig. 3 which gives spectrograms of a 'dark' /l/ and 'clear' /r/ in the words *belly* and *berry* spoken by a teenage female from Leeds (Local & Carter, 2002). The left panel in the figure exemplifies a relatively low F_2 throughout the first vowel and 'dark' liquid of *belly*, as compared

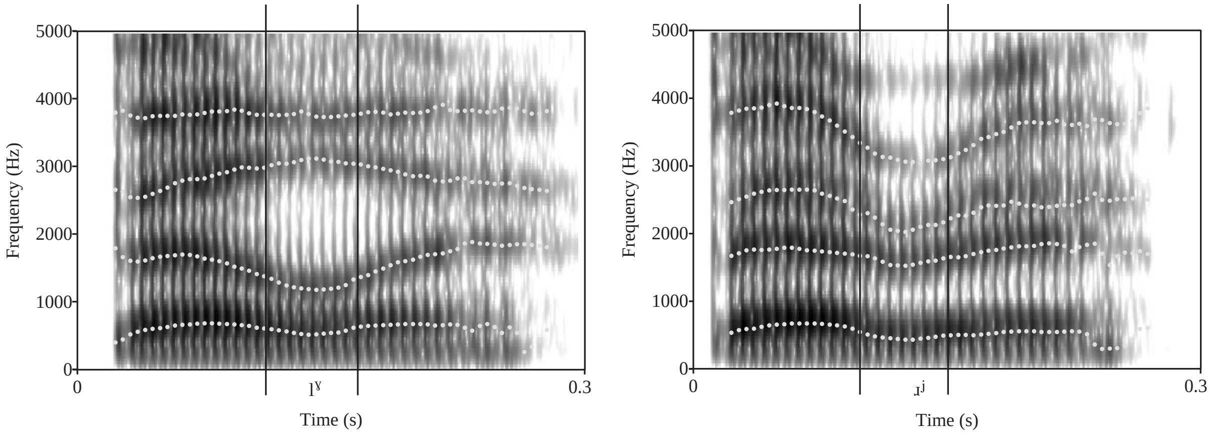

Fig. 3. Spectrograms of *belly* ('dark' /l/, left panel) and *berry* ('clear' /r/, right panel), produced by a teenage female speaker from Leeds, West Yorkshire, England. Tracks for F_1, F_2, F_3 and F_4 are overlaid.

with the first vowel and 'clear' liquid of *berry* (right panel). On average F_2 is 100 Hz (0.5 Bark) lower through the first vowel, and 300 Hz (1.5 Bark) lower at the mid-point of the liquid, for *belly* than for *berry*.

Kelly and Local (1986) also suggest that such consonant-associated resonances may extend to other consonants and vowels in words where these liquids occur, and may also be found in adjacent unstressed vowels in pairs of words such as in the vocalic and consonantal parts of the words and adjacent unstressed syllables. The effect which Kelly and Local identify is challenging partly because it is claimed to extend over rather long temporal windows and partly because it is subtle and poorly understood.

Since Kelly and Local's early and limited work, a number of researchers have replicated this effect and explicated its domain more precisely. Carter (2002, 2003) confirms the 'clear' and 'dark' polarities for liquids and, in addition, shows the sensitivity of the precise phonetic detail of 'clear' and 'dark' to the presence of other contrasting elements in the phonological structure. Local and Carter (2002) demonstrate that, for 16 speakers of two Northern varieties of English, initial and medial /l/ and /r/ associate differently with 'clear' and 'dark' polarities. They show that this /l/$\sim$ /r/ resonance-association is statistically significant regardless of whether the resonance of the speaker's /l/ is clear or dark. So, for example, F_2 is lower in Leeds 'dark' initial-/l/, than in 'clear' initial-/r/ for both males ($t(149.9) = 7.6$, $p < 0.0001$) and females ($t(149.6) = 7.0$, $p < 0.0001$) and in medial position for both males ($t(138.8) = 12.8$, $p < 0.0001$) and females ($t(146.6) = 13.3$, $p = < 0.0001$)[3]). Moreover, their results indicate that in words such as *belly* and *berry*, the resonance effect appears to be available early in the first vowel, before the F_2 transition for the liquid, and to persist to around half-way through the vowel in the second syllable (Local & Carter, 2002), typically about 60 ms before the onset of the intervocalic l/r-consonants in question and some 40 ms after the offset of those consonants. Local and Carter's analysis also reveals that the precise phonetic details of these resonance polarities vary with the position of the liquid in the syllable and the metrical structure of words; different ranges of resonance variability are observed

[3] A conservative estimate of statistical significance based on the Welch modification to the degrees of freedom.

for intervocalic liquids in trochaic contexts (e.g., *belly* and *berry*) as opposed to iambic contexts (e.g., *believe* and *bereave*). Tunley (1999) demonstrates that the temporal extent of 'dark' resonance associated with /r/ in southern British English varies with vowel height and the number of consonants in the syllable onset. Her results also indicate that 'dark' resonance may extend for at least two syllables on either side of the r-constriction itself as long as those syllables are unstressed and especially if they are in feet of three or more syllables. West (1999) provides EMA and EPG data which shows that the long-distance effects associated with these liquids may extend up to two syllables before the liquid. West (2000) additionally provides evidence of perceptual access to these effects. Heid and Hawkins (2000) in an acoustic study show that anticipatory effects of the 'clear' and 'dark' resonances associated with these liquid consonants can affect as many as *five* syllables (from 0.5 to 1 s) before the liquid and can also 'pass through' up to two stressed syllables. They identify a complex range of behaviors and interactions involving syllable stress, the presence of particular consonantal places of articulation and particular classes of vowel.

It is not at all clear what the role of these long-extent characteristics is but they certainly warrant more sophisticated and systematic investigation. For instance, it is not clear why liquids in English should be so strongly associated with them. It could be that there are many more of these effects around if we looked for them systematically, and that there is nothing special about the case of the liquids, although Coleman's (2003) work in this volume suggests that they are special. Of course, one of their functions may simply be as 'enhancement' properties of the signal making it more coherent in ways that are as yet poorly understood. Certainly, if resonance effects such as these, extending over more than one syllable, are modeled in synthesis, the intelligibility of synthetic speech in cafeteria noise can increase by around 15% (Hawkins & Slater, 1994; Tunley, 1999).

4. Grammar and phonetic detail: variable relevance

Section 3 focussed on some nonlocal exponents of lexical distinctions. Here we consider briefly some of the kinds of phonetic detail and variability which are associated with grammatical systems in language. The aim is to illustrate some of the other nonlexical ways in which the speech signal is rich with structural information which can guide the perceptual and understanding processes.

4.1. Labiality and nasality in English 'am'

Grammatical or function words have systematically different phonetic exponents from nonfunction words. To take a simple example from English, word-final /m/ does not usually assimilate to the place of articulation of a following consonant, while word-final /n/ does. However, in the grammatical chunk *I'm* … (= 'I am') assimilation regularly happens in everyday talk (Kelly & Local, 1989, pp. 190–202; Ogden, 1999). For *I'm* (when both stressed and unstressed) in utterances such as *I'm opening*, *I'm bringing*, *I'm fishing*, *I'm doing, I'm going*, and *I'm washing* we find forms such as:

aĩm ɐĩɱ aĩn aĩŋ aĩw̃

(There is a greater probability of shared place of articulation between the nasal of '*m* and a following consonant in forms with unstressed *I* where we often find a very reduced vowel or often no obvious vocalic portion; in such cases we find something like a syllabic nasal.) By contrast the labiality and nasality in forms such as *time*, *lime*, *mime*, *rhyme* in utterances such as, *the lime opens easily*, *the lime fills the glass*, *the lime never tastes right*, *the lime goes quickly*, *the lime was off*, is rather different and shows much more restricted variability. We get [m] preceding vowels and bilabial consonants and sporadically [ɱ] preceding labiodentals but examination of data from everyday talk suggests that in such cases we do not seem to get [n], [ŋ] or [w̃]. (However, in compound forms such as *timetable* or *sometimes* or the name of the BBC program *Crime Watch* such variation in the nasal may occur.)

One plausible reason for these differences in ranges of variability is that nasality and labiality of *I'm* is doing *grammatical* work (i.e., it is part of an auxiliary verb) whereas the nasality and labiality in *lime* is doing *lexical* work (distinguishing it from, say, *line*) and the component parameters of nasality and labiality are bound together in different ways in the two cases. The nasal part of *I'm*, which represents *am,* is in a system of contrast with the other grammatical patterns of pronoun + nonpast forms of the verb *be*: *are* in *you're*, *we're*, *they're*, on the one hand and *is* in *he's*, *she's* on the other; i.e., a three term system of contrast where the contrasts involve parameters of nasality (as in *I'<u>m</u>*), centrality (as in *you'<u>re</u>*) and friction (as in (*s/he/it'<u>s</u>*) rather than unitary composites such as /m/ or/z/. Word-final, citation-form /m/, on the other hand, enters into a different range of contrasts which include another nasal /n/ (*climb/cline, lime/line*). Thus, in *am*, place of articulation and nasality are not tied together in the same kind of way as they are in the /m/ in lexical cases. We might, of course, be observing a frequency effect here. Grammatical items are generally high frequency words and it is well-known that high-frequency words tend to exhibit distinct phonetic characteristics (Fosler-Lussier & Morgan, 1999). Whatever the cause or causes, these different patterns of variation are linguistically informative although they do not reflect simple 'phonemic contrast'.

4.2. Dentality, grammar and word-joins

In many languages, including English, the phonological structure of grammatical or function words is considerably more limited than that of content words. Moreover, function words exhibit rather different kinds of phonetic organization with respect to the temporal organization of phonetic parameters (Ogden, 1999). They may also be different with respect to the way they are phonetically joined up with surrounding words. For example many English function words begin with voice and dentality [ð] or [d̪] (e.g., *the, that, this, these, those, then, though,*) whereas lexical items/content words do not. Where dental friction occurs at the beginning of content words it is voiceless: [θ] (e.g., *think, throw, thistle, theme*). As in the *I'm/lime* case, the phonological ingredients of the beginnings of these grammatical and lexical items are bound together in different ways and this manifests itself in (informatively) different kinds of behavior in connected speech.

Grammatical words beginning with voice and dentality enter into particular kinds of relationships after word-final /l/ and /n/ that are not found for other /l/-fricative or /n/-fricative sequences. Thus chunks such as *ban these, in that, still the* and *all that* are commonly produced in many British and American varieties of English with long dental nasals or long dental laterals at the juncture of the two words with no sign of any (dental) friction (Manuel, 1995; Ogden, 1997):

banːiz ɪn̪ːat stɪl̪ːə ɔːl̪ːat

Other /n/-fricative and /l/-fricative sequences at word joins do not exhibit this assimilation but retain the initial fricative. So we do not find *ban thought*, *win thanks*, *still think*, or *all thanks* produced as

ban̪ːɔːt wɪn̪ːaŋks stɪl̪ːɪŋk ɔːl̪ːaŋks

The long articulations in [ban̪iːz], etc. are, unsurprisingly, acoustically different from those in stretches such as *ban gnats, in knees, ten no's, sell Lee's,* etc. The acoustic work by Manuel (1995) shows that after nasal /n/, at least, the phonetic realization of dentality and voice in English grammatical words is complex. For instance, the long dental nasal portion has a relatively low F_2 at the release of the consonant closure (compared with content-word alveolars) and the transitions out of the initial nasal in *the* (= [n̪ə]) are slower than those out of the nasal in, for example, *in a*. Moreover, Manuel shows that English listeners have perceptual access to the phonetic differences between [n] and [n̪] and are able to use these differences in a lexical-decision task.

The key result which emerges here is that different kinds of linguistic contrast make use of different combinations and relationships between phonetic exponents. The phonetic parameters of these grammatical items may be temporally associated in different ways (see also Ogden, 1999). Whatever notion of phonological contrast we may entertain, the relationship between initial dentality and friction in function words such as *the, this, that*, etc. is not the same as that which holds between initial dentality and friction in content words such as *thought, thin, thousand*, etc. The same is true for the *I'm/lime* cases. Examples such as these emphasize that particular combinations and alignments of the same parameters may be related to different strands of linguistic meaning at different places in the stream of speech. This in itself can carry important information about the linguistic structure of the talk as it unfolds in time. This clearly has implications for the kinds of things that we want our perceptual models to model. Presumably a robust perceptual model could make use of information about how strictly such phonetic parameters are associated with each other and aligned in the temporal dimension.

5. Conclusion

The data I have considered in this paper have implications for understanding how the perceptual system might handle different relationships between phonetic parameters and their different temporal alignments and extents. I have made three principal claims. The first is that the kinds of data which typically inform perceptual models (i.e., data drawn from sources other than naturally occurring talk-in-interaction) at best underdetermine, and at worst hinder, our understanding of phonetic detail and the meaningful uses to which it may be put. The second is that some phonetic details which might define or contribute to, say (lexical) phonological, category membership or 'meaning' in one context do not necessarily do so in another; that is, the same phonetic detail may have 'variable relevance'. The third is that phonological contrast operates over different pieces of linguistic structure; that is it has 'variable domains'.

While constructs such as segments and phonemes may seem to simplify a phonological or perceptual analysis, the data I have presented suggest that things are far more complex than segmentally-based approaches to phonology might lead us to believe. With a broader view of what might count as phonological contrast or linguistic meaning and the detail of its phonetic interpretation, the analysis is often actually simplified. One implication of this is that small temporal windows on their own will regularly fail to capture the richness of phonetic detail and its relation to systems of meaning. The perceptual system would seem to need access to windows of differing and often quite long duration (Pöppel, 1997).

The analytic position I have taken throughout this paper is shaped by Firthian prosodic analysis or FPA (Firth, 1948; Ogden & Local, 1994). In interpreting the functionality of the speech signal from this perspective I have endeavoured to demonstrate the value of the following principles:

i. *Parametric interpretation:* do not be restricted to the domains and categories superficially imposed by e.g., traditional phonetic description and classification. Pay close attention to component phonetic parameters, their relationships to each other and their synchronization in time.
ii. *Variable-domain interpretation:* give equal priority to the identification, description and analysis of features or sets of features over different domains (e.g., phrasal units, words, syllables, syllable constituents). Do not restrict attention to, for instance, 'segment-sized' units.
iii. *Variable-relevance interpretation:* be prepared to accept any phonetic parameter or group of parameters as of possible phonological/informational interest at a particular point in structure.
iv. *Polysystemic interpretation:* be sensitive to the possibility that different meaning systems operate at different places in linguistic structure.

It seems to me that these principles could provide a robust phonetic and phonological infrastructure for those approaches to speech perception that seek to make use of the rich phonetic detail available in the speech signal. (See Hawkins (2003); Hawkins and Smith (2001) provides an extended discussion of this issue.)

This paper began by looking at some speech data in its natural ecological setting: talk-in-interaction. Such material contrasts starkly with the 'semiotically sparse' experimental data which typically inform perceptual models. It draws attention to the ways in which phonetic detail can be deployed in producing and interpreting speech and suggests that it is appropriate to entertain a richer view of the role of phonetic detail in phonological organization and speech perception. Taking a richer view of how phonetic detail might relate to phonological contrast highlights, furthermore, that information relevant to the identity of the correlates of 'units of speech' tends to be widely distributed and nonlocal, as opposed to the much more local, punctual phenomena that have been the traditional focus of linguistics and speech perception. If we are to understand the workings of phonetic detail and its variability we need to relate, systematically and differentially, the phonetic detail of utterances to various categories and levels of analysis (e.g., interactional, grammatical, lexical) and not assume that the lexicon has a privileged status.

Acknowledgements

Many of the ideas (and some of the wording) presented here arise directly from conversations with Paul Carter, Sarah Hawkins, Eric Keller, John Kelly, Richard Ogden and Adrian Simpson. I thank them for their creativity. I thank two anonymous reviewers, Paul Carter and particularly Sarah Hawkins for sharpening up the text and my ideas. The usual disclaimer applies—none of them are responsible for any errors here present.

References

Alfonso, P. J., & Baer, T. (1982). Dynamics of vowel articulation. *Language and Speech*, *25*, 151–173.

Atkinson, J. M., & Heritage, J. (1984). *Structures of social action*. Cambridge: Cambridge University Press.

Benguerel, A-P., & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, *30*, 41–55.

Boardman, I., Grossberg, S., Myers, C., & Cohen, M. (1999). Neural dynamics of perceptual order and context effects for variable-rate speech syllables. *Perception and Psychophysics*, *61*, 1477–1500.

Carter, P. G. (2002). *Structured variation in British English liquids*: *The role of resonance*. Unpublished Ph.D. dissertation, University of York.

Carter, P. G. (2003). Extrinsic phonetic interpretation: Spectral variation in English liquids. In J. Local, R. A. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI: Phonetic interpretation* (pp. 234–249). Cambridge: Cambridge University Press.

Coleman, J. S. (1998). Cognitive reality and the phonological lexicon: A review. *Journal of Neurolinguistics*, *11*, 295–320.

Coleman, J. S. (2003). Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics*, *31*, doi:10.1016/j.wocn.2003.10.001

Davenport, M., & Hannahs, S. J. (1998). *Introducing phonetics & phonology*. London: Arnold.

Docherty, G. J., Milroy, J., Milroy, L., & Walshaw, D. (1997). Descriptive adequacy in phonology: A variationist perspective. *Journal of Linguistics*, *33*, 275–310.

Fant, C. G. M. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Firth, J. R. (1948). Sounds and Prosodies. *Transactions of the Philological Society*, 127–152.

Fosler-Lussier, F., & Morgan, N. (1999). Effects of speaking rate and word frequency on conversational pronunciations. *Speech Communication*, *29*(2–4), 137–158.

Goldinger, S. D. (1997). Words and voices: perception and production in an episodic lexicon. In K. Johnson, & J. W. Mullenix (Eds.), *Talker Variability in Speech Processing* (pp. 33–66). San Diego: Academic Press.

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics*, *31*, doi:10.1016/S0095-4470(03)00030-5

Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, *31*, doi:10.1016/S0095-4470(03)00051-2

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, doi:10.1016/j.wocn.2003.09.006

Hawkins, S., & Nguyen, N. (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics*, *32*, doi:10.1016/S0095-4470(03)00031-7

Hawkins, S., & Slater, A. (1994). Spread of CV and V-to-V coarticulation in British English: Implications for the intelligibility of synthetic speech. *Proceedings of the International Conference of Spoken Language Processing 94*, Vol. 1, (pp. 57–60).

Hawkins, S., & Smith, R. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics—Rivista di Linguistica*, *13*, 99–188.

Heid, S., & Hawkins, S. (2000). An acoustical study of long-domain /r/ and /l/ coarticulation. In *Proceedings of the fifth seminar on speech production: models and data* (*ISCA*). Kloster Seeon, Bavaria, Germany (pp. 77–80).

Jones, D. (1934). *An outline of English phonetics*. Cambridge: W. Heffer & Sons Ltd.

Kelly, J. (1989). Consonant-associated resonance in three varieties of English. In J. Windsor-Lewis (Ed.), *Studies in general and English phonetics* (pp. 335–349). London: Routledge.

Kelly, J., & Local, J. (1986). Long domain resonance patterns in English. *Proceedings of IEE Conference on speech input/output: Techniques and applications* (pp. 304–309). London: Institution of Electrical Engineers.

Kelly, J., & Local, J. (1989). *Doing phonology*. Manchester: Manchester University Press.

Kerswill, P., & Wright, S. (1989). On the limits of auditory transcription: A sociophonetic approach. *York Papers in Linguistics*, *14*, 35–59.

Krakow, R. A. (1993). Nonsegmental influences on velum movement patterns: Syllables, sentences, stress and speaking rate. In M. K. Huffman, & R. A. Krakow (Eds.), *Nasal, nasalization and the velum* (pp. 87–113). New York: Academic Press.

Kreidler, C. W. (1989). *The pronunciation of English: A course book in phonology*. Cambridge, MA: Basil Blackwell.

Kwong, K., & Stevens, K. N. (1999). On the voiced-voiceless distinctions for writer/rider, MIT-RLE. *Speech Communication Group working papers, 11*, 1–20.

Ladefoged, P., & Maddieson, I. (1996). *Sounds of the world's languages*. Oxford: Blackwell.

Lavoie, L. (2002). Some influences on the realisation of *for* and *four* in American English. *Journal of the International Phonetic Association*, *32*(2), 175–202.

Lerner, G. (1996). On the 'semi-permeable' character of grammatical units in conversation: Conditional entry into the turn space of another. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 238–276). Cambridge: Cambridge University Press.

Local, J. K. (1996). Conversational phonetics: Some aspects of news receipts in everyday talk. In E. Couper-Kuhlen, & M. Selting (Eds.), *Prosody in conversation: Ethnomethodological studies* (pp. 175–230). Cambridge: Cambridge University Press.

Local, J. K. (2000). Speaking other people's minds: The phonetics of Collaborative Completions. Patterns of speech sounds in unscripted communication, Sankelmark, Germany. Organised by IPDSP, Kiel, under the auspices of the International Phonetics Association. Handout and sound files downloadable from http://www-users.york.ac.uk/~lang4/John-Local-talks.html.

Local, J. K. (2002). Getting back to 'prior' talk: and-uh(m) as a skip connecting device. Special prosody panel, *International conference on conversation analysis*, Cophenhagen, Denmark. Handout and sound files downloadable from http://www-users.york.ac.uk/~lang4/John-Local-talks.html.

Local, J. K., & Carter, P. G. (2002). Structured variation in the phonetic interpretation of British English liquids. Final report for ESRC award R000223534.

Local, J. K., & Kelly, J. (1986). Projection and 'silences': Notes on phonetic and conversational structure. *Human Studies*, *9*(2–3), 185–204.

Local, J. K., Kelly, J., & Wells, W. H. G. (1986). Towards a phonology of conversation: Turn-taking in Tyneside English. *Journal of Linguistics*, *22*, 411–437.

Local, J. K., & Walker, G. (2002). On the sequential distribution of a phonetic event: Rushthroughs as a resource for the production of multi-unit, multi-action turns. *Euroconference on interactional linguistics*, Helsinki, Finland. http://www-users.york.ac.uk/~lang4/John-Local-talks.html.

Manuel, S. Y. (1995). Speakers nasalise /ð/ after /n/, but still hear /ð/. *Journal of Phonetics*, *23*, 453–476.

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In G. J. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II* (pp. 261–280). Cambridge: Cambridge University Press.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *3*, 299–325.

Ogden, R. A. (1997). Final report for ESRC award R000221880. *A declarative account of deletion phenomena in English phonetics and phonology*.

Ogden, R. A. (1999). A declarative account of strong and weak auxiliaries in English. *Phonology*, *16*, 55–92.

Ogden, R. A. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, *31*(1), 139–152.

Ogden, R. A., & Local, J. K. (1994). Disentangling autosegments from prosodies: A note on the misrepresentation of a research tradition in phonology. *Journal of Linguistics*, *30*, 477–498.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, *1*, 56–61.

Port, R. (1979). Influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics*, *7*, 45–56.

Port, R., & Dalby, J. (1982). Consonant/vowel ration as a cue for voicing in English. *Perception & Psychophysics*, *32*, 141–152.

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

Recasens, D., Fontdevila, J., & Pallarès, M. D. (1996). Linguopalatal coarticulation and alveolar-palatal correlations for velarized and non-velarized /l/. *Journal of Phonetics*, *24*, 165–185.

van Santen, J. P. H., Coleman, J. S., & Randolph, M. A. (1992). Effects of postvocalic voicing on the time course of vowels and diphthongs. *Journal of the Acoustical Society of America, 92*(4)(Part. 2), 2444.

Schegloff, E. A. (1996). Turn organisation: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 52–133). Cambridge: Cambridge University Press.

Schegloff, E. A. (1998). Reflections on study prosody in talk-in-interaction. *Language and Speech*, *41*(3–4), 235–263.

Schiffren, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Scott, D. (1984). More on the /t/:/d/ distinction in american alveolar flaps. *Journal of the Acoustical Society of America, 75*(suppl. 1), 566.

Shriberg, E. (1999). Phonetic consequences of speech dysfluency. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences*, Vol. 1 (pp. 619–622). Berkeley: University of California.

Solé, M-J. (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological organization. *Language and Speech*, *38*, 1–23.

Sweet, H. (1910). *The sounds of English*. Oxford: Clarendon Press.

Thompson, S. A., & Mulac, A. (1991). The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics*, *15*, 237–251.

Tunley, A. (1999). *Coarticulatory influences of liquids on vowels in English*. Unpublished Ph.D. dissertation, University of Cambridge.

Wells, B., & MacFarlane, S. (1998). Prosody as an interactional resource: Turn-projection and overlap. *Language and Speech*, *41*(3–4), 265–294.

West, P. (1999). The extent of coarticulation of English liquids: An acoustic and articulatory study. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences.*, Vol. 3 (pp. 1901–1904). Berkeley: University of California.

West, P. (2000). Perception of distributed coarticulatory properties of English /l/ and /ɹ/. *Journal of Phonetics*, *27*, 405–425.