Flexibility in the face of incompatible English VOT systems*

James M. Scobbie

The Voice Onset Time (VOT) cue to the /p/-/b/ voicing contrast in Shetland Isles English was found to demonstrate a high degree of interspeaker variation, which was non-arbitrary on two counts. First, there was an inverse relation between the amount of prevoicing for /b/ and the aspiration duration for /p/. Second, the most vernacular contrast, between typically prevoiced /b/ and relatively short lag /p/, was generally found in speakers whose parents were also native Shetlanders. These results suggest that indexical and phonological cues are simultaneously present, that phonetic targets for VOT are specified in fine detail by speakers, and that they are not restricted to using a fixed low number of VOT categories. The relevance of the results to exemplar models of the mental lexicon and the relationship of phonetics to phonology is discussed, and arguments are presented that both categorical and gradient approaches to phonology must be pursued.

1. Introduction

1.1. Language-specific and speaker-specific sound systems

Features remain central to theories of phonological categories. The idea that there is a universal set of features, however, becomes harder to defend the more it is discovered how much of learned linguistic competence is bound up in the specification of language-specific phonetic detail. In their review and preview of phonology as a laboratory science, Pierrehumbert, Beckman, and Ladd (2000: 285) state: “there are no two languages in which the implementation of analogous phonemes is exactly the same. When examined in sufficient detail, even the most common and stereotypical phonetic processes are found to differ.” For example, a recent study of the fortis vs. lenis vs. aspirated stop contrast in Korean concludes: “it would take a great deal of

procrustean effort to force Korean stops into the categories that have been developed for phonological descriptions of other languages" (Cho, Jun and Ladefoged, 2002: 222).

Subtle variation between cross-linguistically analogous phonemes is, however, a limited source of evidence against universal features. Each phoneme exists in the context of a wider system, and different languages have (by definition) their own syntax and lexicons and (in practice) their own phonotactic, allophonic and prosodic systems, even if they have absolutely identical phonemic inventories. Therefore, it would be advantageous to examine the phenomenon of fine-grained phonetic differences and their relevance to phonological categories not just in a cross-linguistic situation, but also in the context of linguistic systems that differ only in minimal ways. Such sister systems are, of course, dialects of the same language.

The new consensus – that significant amounts of phonetic detail are language-specific – can be dealt with theoretically in more than one way. A more conservative view holds true to the traditional modularity of phonology. It posits that the observed cross-linguistic differences in analogous phonemes and features reside outside phonological competence, say in phonetics, or in performance. Analogous phonemes may therefore be identical (i.e. have the same feature content), but are accompanied by a sophisticated language-specific phonetic grammar of phonetic realisation rules (e.g. among many, see Cho *et al.*, 2002). A modular approach to phonology and phonetics does not seem to provide the conceptual framework for understanding *how* the mass of language-specific fine phonetic knowledge interacts with universal feature systems, and it predicts that a discrete interface between the relevant modules should exist.

A more radical position is that language learners create their own phonological and phonetic (gradient and categorical) systems, prompted by statistical patterning in the input and universal linguistic and cognitive predispositions. On this view there is continuity between language-specific phonetics and phonology. Many aspects of such a view in child and adult studies are decades old (see the reviews in Vihman and Velleman, 2000; Docherty, 1992:55, 75). Recently, however, further research (e.g. Bybee, 2001; Pierrehumbert, 2002; Coleman, 2002; and see the papers in Bybee and Hopper, 2002) has drawn on psycholinguistic "exemplar" or "episodic" models of the mental lexicon (e.g. Goldinger, 1997; Mullenix, 1997; Pisoni, 1997; Johnson, 1997; Hawkins and Smith, 2001), in which multiple detailed exemplars of every word are stored, making the lexicon a structured mix of abstractions and detailed memories of previous speech events and contexts.

A major motivation for exemplar models is that a variety of psycholinguistic experiments have shown listeners can be simultaneously influenced not only by *what* is being said but also by *who* is talking. The implications of this, and of the fact that listeners perceive phonetic details subtle enough to convey both lexical contrast and linguistically non-contrastive indexical factors are that “the indexical and linguistic attributes of speech are not neatly partitioned” (Pisoni, 1997: 11) into different modules. The *same* phonetic parameter can convey *both* types of information.

Exemplar models are of particular interest to phonologists, sociolinguists and others because speakers, when they store traces of individual words, are forced (by the sheer numbers of such exemplars, perhaps) to generalise across them. Such generalisations must be relatively abstract because functionally-similar exemplars are not phonetically identical: variability “causes the need for abstraction” by the learner (Pierrehumbert, Beckman and Ladd, 2000: 292). Such abstractions are presumably made on the basis of similarity in sound and articulation on the one hand and/or similarity in function and context of use on the other (where the nature and limits of “similarity” crucially must be elucidated). If these reflect the systematic nature of language, basic phonological phenomena such as phonemic contrast will tend to induce abstractions similar to those posited by traditional phonological analyses. Such a model predicts the existence of “recurrent” features: fuzzy categories with a close family resemblance will tend to emerge without being universal. Phonemic and indexical functions can be conveyed along the same acoustic parameters and either may be distributed more or less categorically.

Many phonetic (e.g. Docherty, 1992; Browman and Goldstein, 1992; Lavoie, 2001) and sociophonetic studies (Thomas, 2002; Labov, 1994; Foulkes and Docherty, in press) have shown just how fine-grained the control of phonetic and phonological variation actually is – *within* what is often considered a single dialect. So, in addition to the continuum between indexical and contrastive functions, there is a continuum of indexicality, from idiolectal variation, through micro-dialectalisms, to variation characteristic of larger and more disparate groups of speakers. The exemplar type of model predicts that *sociolinguistic* expectations should also influence the perception of contrast, for the issue of “who” is talking relates to groups as well as to individuals.

Nevertheless, and despite a large body of research arguing for continuity from phonetics to phonology,¹ many linguists have tended to hold to the conservative modular position presented above: that while speaker-specific, dialectal, stylistic or phonetic differences vary along continuous scales, pho-

nological features are discrete and universal. Further evidence is presented here that distinctive features are recurrent, are created during acquisition as a response to complex and possibly competing environmental targets, and are intimately bound to indexical functions of language.

In exemplar models, different speakers' "sound systems"² end up being similar to the extent that their experiences are similar.³ Broad generalisations are just those that have previously been – and will therefore continue to be – internalised by large numbers of speakers. Individual speakers can vary from crosslinguistic norms, within statistical limits. In this paper the focus is on the systematicity and flexibility of such normal variation in speakers whose target language offers more than one possible feature analysis of a contrast which they all share.

1.2. Voice Onset Time from a cross-linguistic perspective

On the basis of a major cross-linguistic study, Lisker and Abramson (1964; 1967) proposed the universally-available parameter of Voice Onset Time (VOT), comprising three acoustic categories to cue VOICING contrasts for stops in word-initial position.⁴ The categories are voicing lead, short lag (voiceless unaspirated) and long lag (voiceless aspirated). More recently, Cho and Ladefoged explored a wider range of languages under a consistent methodology and concluded (exemplifying the point with specific values for velars): "it is not at all clear that there are just two phonetic categories [in addition to voicing lead] from which languages can choose... it would certainly be plausible to say that there are four phonetic categories, one around 30 ms representing unaspirated stops, another around 50 ms for slightly aspirated stops, a third for aspirated stops at around 90 ms, and a fourth for ... highly aspirated stops" (Cho and Ladefoged, 1999: 223).

Lack of any universal boundary between aspirated and unaspirated categories is independent of the questions of whether there *are* recurrent categories of Voice Onset Time, and if so, *how many* there are. If four modal values of positive VOT are observed crosslinguistically, Cho and Ladefoged take a rather agnostic stance on the implications for feature theory: "we consider what might appear to be phonetic categories as *at best* modal values within the continua formed by the physical scales – the parameters – that define each feature" (Cho and Ladefoged, 1999: 225 [emphasis mine]). In any case, further research is required to identify such general crosslinguistic tendencies for categorisation.

Cho and Ladefoged's survey compares only those phonological categories *distinguished by VOT "alone."* They adopt the standard phonological assumption that when languages have more than two homorganic stops with measurably distinct voicing lag, in all but two cases the VOT differences must be a "secondary" cue which "enhances" some other contrast, in voicing, place, constriction, or airstream mechanism. Cho and Ladefoged (1999) and Docherty (1992) are both quite clear there are not just three universal *phonetic* targets for VOT. Consequently, they seem to disagree with the position that there is a "fixed and universally specified set" (Keating, 1984: 289) of VOT targets relevant to phonetics and phonology alike. For example, Cho and Ladefoged show that VOT for EJECTIVE /k/ tends to be intermediate between UNASPIRATED /k/ and ASPIRATED /k/, but conclude that the intermediateness of EJECTIVE /k/ is not proof that there is a phonological category of "medium lag". Rather, VOT must be a secondary cue for one of the three types of /k/ (EJECTIVE in this case). Clearly, there is a danger of such argumentation being circular. A variation on this argument was proposed in a recent very detailed analysis of the three-way voiceless stop contrast in Korean (a system recognised as a challenge for the VOT system by Lisker and Abramson, 1964). Cho, Jun, and Ladefoged (2002) proposed that one of the three stops (the lenis one) is unspecified, and that its VOT target is specified *phonetically*.

In the exemplar model there is no a priori demarcation between primary and secondary: instead, functional reasons are expected for any patterns observed. For example, there is likely to be a diminishing reliance on VOT as a cue to the identity of all members of a set of homorganic stops if the VOT targets get to be too close together or overly numerous. This avoids the claim that languages specify a precise maximum of two positive VOT targets, and that a greater number of distinct VOT targets must be due to a categorical distinction between VOT as a primary vs. enhancing cue (or that at most two positive VOT targets can be specified by phonology and all others are specified by a different phonetic module).

It seems that Docherty (1992) was right to question the value of rigid phonetic classifications such as voiced, unaspirated and aspirated. It is theoretically more useful to present VOT values numerically in tandem with the relevant phonological contrasts. This may be particularly valuable for those languages that have no VOICING contrast for stops at all, including those in which phonetic voicing is contextually conditioned. One major outcome of Cho and Ladefoged's paper is that there are no clear demarcations in VOT to classify such stops in a universal or deterministic way.

1.3. VOT variation in standard and non-standard English

Previous studies of VOT within a single language have mainly investigated the influence on VOT of linguistic factors such as prosodic context, place of articulation, segmental context, organic factors such as sex, age, pathology, and intermediate factors such as speech rate and dialect in a very broad sense. But “arbitrary” interspeaker variation has, of course, also been observed. A recent example from English is a study of eight General American speakers which found VOT for post-pausal /p/ varied from about 60 ms to about 110 ms (Allen, Miller and DeSteno, 2003). Normalising for speech rate reduced the indexical interspeaker variation, but Allen et al.’s conclusion was nevertheless that VOT varies and may function indexically, in line with exemplar models. The source of the variation could not be identified, because the subjects were a homogenous group. One speaker’s normalised mean VOT for /ptk/ was, at 74 ms, distinct from the other more closely grouped seven speakers, but all the VOICELESS stops were what is normally called long lag.

Other studies have found variation in word-initial, post-pausal position in the occurrence of prevoiced and short lag variants of /bdg/ (Westbury, 1979; Docherty, 1992; Smith, 1978; Lisker and Abramson, 1964; Flege, 1982). Prevoicing is *normal* in this position in English as has pointed out by, for example, Westbury (1979) who found 62% of /b/ tokens were prevoiced; Smith (1978), 56%; and Flege (1982), 59%. As Lisker and Abramson (1967: 5,7) observe, in American English “it appears that while /ptk/ have distributions that are essentially unimodal, /bdg/ show values that fall into two discontinuous ranges, with modes at about –100 msec and near zero.” Their interpretation is that English speakers select from one or both of “two phonetic categories of /bdg/... In fact the relation between the two distribution modes for each member of the /bdg/ set is not *overtly* different from the relation between phonemically distinct categories in certain other languages” (Lisker and Abramson, 1967: footnote 13 [emphasis mine]). The cautious nature of this statement about a possible equivalence between English allophonic categories and contrastive categories in other languages has perhaps been overlooked in an enthusiastic search for a universal model of distinctive and redundant features.

The existence of positive and negative VOT values is taken as evidence that English uses two categories for VOICED stops. Further support comes from interspeaker variation in the use of prevoicing. Some speakers appear to use prevoicing, others to avoid it, and others to vary freely. Lisker and Abramson (1964; 1967) discuss such speaker-specific use of prevoicing in

some detail. Docherty (1992) also found such variation in his comprehensive and detailed study of obstruent voicing in five speakers of Southern Standard British English. Recently, age, race/ethnicity and sex have been shown to be conditioning factors for VOT, such that “there appear to be many more instances of prevoicing” than is normally acknowledged (Ryalls, Zipprer and Baldauff, 1997: 644).

Prevoicing is also a common characteristic of the speech of many English L2 speakers and natively bilingual English speakers, and their (grown-up) children.⁵ This is to be expected if sound systems are able to differ in very subtle ways. For example, in a detailed longitudinal study of English/Arabic bilingual pre-teenagers, Khattab found very fine differences in VOT between the analogous phonemes of each language. The bilinguals had very slightly different language-specific targets when using the “same” phonetic category. Her conclusion was that “there are important phonetic differences between English VOICED and Arabic VOICELESS stops involving divisions that are finer than the boundaries suggested by the three supposedly universal categories, and suggesting that neither of the two broad categories ‘short lag’ or ‘long lag’ can adequately describe Lebanese VOICELESS stops” (Khattab, 2003: 290).⁶

In order to combine these findings on very fine within-language variation with Cho and Ladefoged’s more dramatic cross-linguistic results, this paper examines the VOT parameter in a single speech community (the Shetland Isles of Scotland) in which the phonology and lexicon can be held fairly constant while VOT can be expected to vary across categories. The traditional Shetlandic system has a voicing lead target for /bdg/ and a short lag voice onset target for /ptk/, whereas the majority of standard English varieties have voicing lead and more often short lag voice onset for /bdg/ and a long lag voice onset target (i.e. aspiration) for /ptk/. On standard assumptions, these are *incompatible* VOT systems because short lag VOT is ambiguous as to whether it cues VOICED or VOICELESS stops. The implicit decisions taken by different individuals in such a situation should reveal something about the phonetic options that are available to all speakers.

2. Method: interspeaker variation as an experimental resource

The language reported in this paper is English, specifically those varieties of Scots and Scottish English spoken in the Shetland Isles.⁷ The Shetlands are an isolated archipelago of about 15 inhabited and more than 80 uninhabited

islands, about 300km (180 miles) north of the Scottish city of Aberdeen, a major transportation link, and 350km (220 miles) west of Bergen in Norway. The three biggest settlements are Lerwick (population circa 7,500), Scallovay (c. 1,000) and Sandwick (c. 900), all on the same island.

In recent decades, historic depopulation has been reversed in the Northern Isles (Orkney and Shetland). This has been particularly dramatic in Shetland, where there has been significant immigration due to the oil industry: the population swelled by about a third in the decade leading up to 1981. Though the population had fallen back from this peak by 1991, it has remained relatively steady (21,988 in the census of 2001). The net increase from 1971 is made up in part by incomers who settled in the islands and subsequently had children. Thus, at the time the data was collected (1999), a significant minority of young adult native Shetlanders had parents who themselves were incomers from England and elsewhere in Scotland.

Contemporary vernacular Shetlandic preserves many relic forms of older Scots, perhaps including the Scots instantiation of the VOICING contrast: “early authorities [from the 1930s and 1940s] are united as to the unaspirated nature of Scots voiceless stops in syllable onsets” (Johnston, 1997: 505). Scottish-accented Standard English on the other hand has long-lag VOT (aspiration). Traditional Scots VOT patterns can indeed still be heard in vernacular varieties throughout Scotland, though “aspirated ones equal to the equivalent /ptk/ allophones in other English dialects” are spreading into Scots (Johnston, 1980: 78). VOT in Southern British English /p/ VOT in post-pausal, word-initial position is around 47 ms (Hawkins, 1979), 40 ms (Suomi, 1980), or 46 ms (Docherty, 1992). There are no comparable published figures for Scottish Standard English, though Docherty’s review of VOT studies suggests that, in initial position, long lag /p/ is unlikely to be under 40 ms. However, consider the very small study of VOT in vernacular southern Scots (Masuya, 1997). Unexpectedly, only one Scottish English speaker (Sp19) out of 20 had unquestionably short lag /p/. However, since there are only two tokens per subject, this cannot be given much credence. Of greater reliability and relevance are the pooled results, which show that the Scottish subjects have a *slightly* lower VOT than the English controls: phrase-medially, lowland Scottish /p/ was 32 ms (n=29, s.d. 16 ms), while English /p/ was 48 ms (n=10, s.d. 17 ms). This Scottish mean is not easily classifiable as being either “short” or “long”.

Children growing up in the Shetlands in recent decades are likely to have been exposed to short lag and long lag variants of /p/, and short lag and prevoiced variants of /b/. Children growing up in families where the parents

are non-Shetlandic incomers would have had to face incompatible functions for the same VOT value. The meaning of short lag stops would vary depending on the speaker. As a result, wide interspeaker variation in VOT is to be expected in young Shetlanders. Yet all speakers would typically be regarded by linguists as having the same contrast at an abstract level: there is no *phonological* conflict to resolve in acquisition or in adult communication. The incompatibility of different systems is likely to lie in the location of the VOT targets and perhaps in whether VOT is a robust cue to the contrast at all. If human language provides only three universal categories of VOT, then we should expect to see categorical variation.

Exploration of such a situation demands a fundamental departure from normal practice: it requires subjects who might be *expected* to vary in their systemisation of the phonetic or phonological phenomena under consideration along partially predictable non-linguistic lines. Augmenting standard experimental phonetic methodology with *socially structured* groups of subjects offers a new basis for advancing phonological theory. Two linked strands of previous research have greatly influenced the adoption of such a methodology here. One has been the work of Docherty and colleagues (Docherty, Foulkes, Milroy, Milroy and Walshaw, 1997; Docherty and Foulkes, 2000; Foulkes and Docherty, 1999). The other has been collaborative work with Stuart-Smith (Scobbie, Timmins, Stuart-Smith, Tweedie, Hewlett and Turk, 2000). However, while many instrumental phonetic studies of VOT have addressed a wide range of sources of variation (see above), social stratification has tended to be seen as the preserve of sociolinguistics. This is despite the fact that social variation can be a more delicate tool than cross-linguistic variation for probing subtle differences in sound systems.

The analysis is based on wordlist recordings made by Marie Cluness, a native Shetlander (then aged 20). The wordlist comprised a random ordering of 270 single words, intended to enable study of a number of dialect variables (Cluness, 2000; Scobbie, 2005). On a number of grounds, the subjects might be expected to have *similar* accents: all had been born in the Shetlands and lived there all their lives; all lived in the same geographical area (West-side); six were final year students at the same high school; eleven were 16 to 22 years old; many were acquainted with each other (and Cluness). (Indeed, S9 and S11 are brothers.) However, one major structuring factor was introduced to provide a basis for possible interspeaker differences (Table 1). The subjects form three groups based on the geographical origin of their parents (and, therefore, by hypothesis, the parental accents). Crucially, one group (S1–S4) had parents who themselves were native Shetlanders and so were

likely to have broader, more vernacular accents than members of the other two groups.

Table 1. Subject identifiers and age details coded by sex and group.

Parents born and raised in	Shetland	Scotland	England
Male Shetlanders	S1: age 20	S5: age 16	S9: age 20
	S3: age 21	S7: age 17	S11: age 19
Female Shetlanders	S2: age 30	S6: age 17	S10: age 17
	S4: age 22	S8: age 17	S12: age 17

VOT is an acoustic measure which can be defined in a number of ways (cf. Fischer-Jorgensen and Hutters, 1981; Docherty, 1992: 24). Here it is the difference in time from the burst at the end of stop closure to the onset of the periodicity, which continues into the following vowel.

Figure 1. Annotation of VOT ranging from pre-voiced to long-lag aspirated. The x-axis in each panel is 200 ms, with 0 ms at the burst. The timing (in ms) and direction of VOT is indicated. The y-axis of each spectrogram is 0–5kHz. a. S4 brood [brʊd], b. S2 pod [pʰɒd], c. S10 both [peθ], d. S1 piece [pis].

If periodicity was detected following the release only, it was recorded as *positive* VOT (Figure 1b,c,d). Typically, VOT was easy to identify (Figure

1d). If the stop had a particularly energetic release, 5–10 ms of frication energy were present, perhaps obscuring the initiation of voicing. In common with other acoustic studies, this effect presumably boosts mean VOT in unaspirated stops by a few milliseconds. In 18 cases (1.5% of the data) a weak burst was simultaneous with the onset of periodicity, so a VOT of zero was assigned. These tokens were classified as examples of voicing lag. Of these tokens, 17 were /b/ and one was /p/, and two subjects (S7 and S6) provided 11 of the examples.

If periodicity was first detected at a point clearly before the stop release, it was recorded as *negative* VOT (Figure 1a). There was no requirement, however, for voicing to be maintained from this VOT point until the burst and beyond, though this occurred in most cases. In some tokens, periodicity died away again before the burst, and only began again at or just after the release.

3. Results

3.1. Pooled results

Consider mean VOT for /p/ and /b/ for all subjects pooled (Table 2). The VOICELESS stop /p/ has a long voicing lag. The VOICED stop /b/ shows voicing lead, i.e. is prevoiced. The mean difference in onset time cueing the VOICING contrast is 85 ms. Note the very large ranges and the large standard deviation for /b/ in particular, which will be explored in detail below.⁸ About half (52%) of the VOT measures for the VOICED stop /b/ were greater than or equal to zero. This proportion is broadly comparable to previous findings for /b/ discussed above (§1.3).⁹

Table 2. Voice Onset Time (ms) based on pooled data from all subjects.

	mean	s.d.	count	min	max
/p/	56	24	280	0	112
/b/	–29	51	334	–190	41

As discussed above in §1.3, the presence of positive and negative VOT in English has been taken as evidence that there are two different phonetic (acoustic) categories: voicing lead and short voicing lag. Stronger evidence for this conclusion is the apparently *bimodal* distribution of VOT in Figure

2, rather merely than a unimodal distribution spanning both negative and positive territory.

Figure 2. Histogram of Voice Onset Time (ms) for /b/, all subjects pooled.

The standard solution to this distribution is to separate VOT into positive and negative ranges. Such an approach provides two means for VOT: positive and negative. These are usually assumed to meaningfully reflect the VOT targets of the two modes of articulation (Docherty, 1992). The stop release (which defines this positive/negative split) initiates the emission of trapped intra-oral air, hence the *beginnings* of the increase in trans-glottal pressure differential which is necessary for phonation, so 0 ms VOT as the boundary point is articulatorily meaningful, but its use is based at least in part on convenience. An empirical alternative – selecting a boundary point based on the *actual* distribution of the data – is clearly less convenient, because it would be impossible with small sample sizes, might be nondeterministic, and would be specific to the speaker, task, style, place of articulation, language, and so on. On the basis of the 10 ms bins in Figure 2, it would be somewhere between –30 ms and –20 ms. Even if the adoption of 0 ms as a unique boundary point between voicing lead and voicing lag modes is less than ideal, using it has certain practical advantages: it lets us compare results with previous studies of English and can also be applied deterministically in individual cases or small samples.

Table 3 presents revised pooled results for /b/ on this basis. Standard deviations are, unsurprisingly, much lower than those in Table 2. Negative

VOT has a wider range and greater standard distribution than positive VOT, a characteristic visible in Figure 2. To the extent that the contrast between VOICED and VOICELESS stops in Shetlandic is cued by two degrees of positive VOT (short vs. long lag), a tight distribution of short lag VOT for VOICED stops would likely occur for perceptual reasons. Such factors are relevant even if half the VOICED stops have voicing lead, and even though the voicing lag /p/ and /b/ ranges overlap.

Table 3. Voice Onset Time (ms) for /b/ calculated from pooled negative VOT data for voicing lead /b/ and pooled positive VOT data for voicing lag /b/.

	mean	s.d.	count	%	min	max
Voicing lead /b/	-71	38	172	52	-190	-2
Voicing lag /b/	15	9	162	48	0	41

It is well-known that aspirated stops show large amounts of glottal abduction, which delays voice onset and makes the closure voiceless, but Flege (1982) is unusual in examining post-pausal word initial VOICED stops. He showed that such stops in American English may or may not be adducted. One subject in that study (his S8) produced tokens of /b/ which were nearly always acoustically voiceless yet with vocal cords adducted suitable for phonation, while some produced voiceless unaspirated /b/ without adduction. Three subjects varied between prevoicing and short lag, but there was no simple laryngeal/acoustic relationship: “much of the variation in the presence/absence of prevoicing observed in utterance-initial voiced stops is due to factors other than variation in laryngeal timing” (Flege, 1982: 189).

It is also well known that phonation is unlikely to begin towards the *end* of a stop closure: as seen in Figure 2 there are generally few tokens with small negative VOT. Flege is skeptical that this is simply due to aerodynamic conditions within the vocal tract tending to make phonation harder to initiate as the supraglottal pressure builds up. Rather, the discontinuous acoustic pattern may be due to a combination of factors such as the strength and timing of glottal gestures, allied to the extent to which a speaker suppresses “passive” devoicing. In the Shetland data, there are a relatively small number of cases in which prevoicing is initiated (at about -100 ms or so), only to become attenuated before the stop release. The lack of such cases argues against high intraoral pressure being solely responsible for the lack of voicing initiation just before the burst.

It is possible, however, that the bimodal distribution in Figure 2 may come in part from unimodal variation in the underlying articulatory strategies (cf. Pierrehumbert and Talkin, 1992). The bimodal nature of Figure 2 does not tell us whether the acoustic distribution reflects “intentional” targets or is the “accidental” output of other types of underlying variation that may even be unimodal. Either way, the acoustic distribution itself is an aspect of grammar – it is not *necessary* for speakers to produce prevoiced tokens, but some do. The observed VOT patterns are therefore linguistically relevant and must reflect the internalised grammar of speakers. This conclusion is in keeping with exemplar theory, which tolerates such ambiguity and non-determinism.

3.2. Individual results

The main benefits of considering individuals are: first, that we can determine how many categories each speaker uses, which lets us probe small systematic differences; and second, that categories *in use* can be compared within and across speakers in order to reveal patterns to replace the unstructured variation of the previous section. The analysis developed below will be based on preliminary individual results for three straightforwardly definable units, namely /p/, positive VOT /b/ and negative VOT /b/.

Table 4 shows that the great variability in the pooled results in Table 2 is due to individuals having their own VOT targets, including distributions above and below 0 ms for /b/. Note that *all* these individual categories have a lower standard deviation (not shown) than the pooled results. Two speakers stand out: S1 has a strict categorical opposition of prevoiced /b/ and short lag /p/, exemplifying the traditional Shetlandic VOT system. S12, on the other hand, has short lag /b/ vs. long lag /p/.

It might seem that in addition to these two, the ten other speakers use two categories for /b/, and that no more needs to be said. However, though all have positive and negative VOT, some speakers have only three or four tokens of one category but 20 or 25 in the other. Consequently, a threshold will be used to allow us to discover which categories are *most clearly* used. If five or fewer tokens are positive or negative (always <20%), they will be treated as noise and not evidence of a category. This threshold is justified by an obvious natural discontinuity in the data itself: the least populated category beyond the threshold is S10’s prevoiced /b/, with nine tokens (32%).¹⁰ Its location is arbitrary, but the use of a threshold reduces the arbitrariness

of the final analysis, because S1 and S12 are not treated as having a completely different sound system to the other subjects merely on the grounds of the presence or absence of a single example of positive or negative VOT. This analysis of the raw data means that S3 and S9 join S1 as speakers who prevoice /b/ and do not use short lag VOT much.¹¹ S5, S7 and S6 join S12 as speakers who do not have a clear prevoicing category. The remaining five present the strongest evidence that the use of both positive and negative VOT for /b/ is typical and systematic.

Table 4. Mean Voice Onset Time (ms), all subjects, for /p/, and for both positive and negative /b/. Values in italics are revised in the text. Subjects S1–S12 are presented in rank order of increasing mean VOT for /p/.

	S1	S4	S9	S3	S2	S11	S7	S6	S12	S8	S10	S5
/p/	22	33	36	37	47	55	61	69	73	78	81	83
/b/		9	<i>18</i>	<i>17</i>	14	16	<i>15</i>	<i>15</i>	17	23	5	17
/b/	–64	–60	–64	–60	–62	–87	–9	<i>–11</i>		–96	–121	–76

One final adjustment to the results for /b/, however, can be made. For two of the speakers with only a few negative VOT tokens (S7 and S6), all eight tokens were clustered between –4 ms and –15 ms. As noted above, these speakers were already unusual in that a number of tokens were excluded on the grounds that no release could be detected. It thus seems likely that they cue /b/ in part through lenition, or with a very low intra-oral pressure which enables voicing to begin just before the point identified as the release during annotation. Consequently, all their measured tokens are analysed as examples of short lag voicing, even though some are below 0 ms. The revised means (and standard deviations) for /b/ are therefore amended (slightly) to S7: 12 ms and S6: 11 ms. Their mean VOT is now a bit lower and a bit more variable, in keeping with their actual behaviour. Figure 3 presents the results of this analysis.

Each speaker has either one or two categories for /b/, as represented by positive and negative means. All speakers are assumed to have a single category for /p/, because speaker-specific histograms of VOT provide no strong evidence to the contrary; but such a technique really requires much more data.¹² Note, however, that the histograms seem to show that even the three subjects in the middle of the distribution of /p/ (S11, S7 and S6) are not varying between short and long lag targets. Despite the fact that these subjects have the relatively high standard deviation of 17 ms compared to the other

subjects' range of 10–14 ms, the minimum values for S7 and S6 are both above 35 ms, suggesting within-category variation. S11 has a single token at 14 ms, one at 27 ms, and the rest above 34 ms.

Figure 3. Mean Voice Onset Time (ms) relative to burst, with whiskers indicating one standard deviation for each speaker for the categories identified in the text. Circles indicate /p/, diamonds and triangles indicate positive and negative means for /b/.

What, then, of each speaker's /p/? The gradient increase in mean VOT for /p/ across the subjects means that there is no basis for allocating /p/ into short or long lag categories on the basis of data on /p/ alone. This conclusion is similar to that of Cho and Ladefoged (1999), but arises from within a language, so avoids the problems of interpretation attaching to crosslinguistic studies. It is also far more radical than the intercategory differences revealed by Allen et al. (2003). It is possible, however, to make some judgments about the categorical affiliation of /p/ on the basis of the distribution of VOT means for /b/, an analysis which neither was able to undertake. In particular, the hypothesis that there are only three universal categories of VOT predicts that the subjects with short lag /b/ *must have*, by definition, long lag /p/. This is not problematic for the contiguous eight-strong group of subjects between S2 and S5 in Figure 3 and Table 4. But S4 clearly has short lag /b/, so her /p/ on this basis should be long lag, though it is only 33 ms.¹³ This creates

problems for the analysis of S9 and S3. Are VOT means of 36 ms and 37 ms for /p/ (in opposition with heavily prevoiced /b/) to be categorised as short or long lag?

Consider now the claim that short lag stops are ambiguous. In Shetlandic a short lag stop can represent either a standard VOICED stop (Figure 1c) or a vernacular VOICELESS one (Figure 1d). However, VOT is slightly greater on average when it cues /p/ and less when it cues /b/. The lowest VOT for /p/ (S1, 22 ms) is right at the top end of the scale of the nine speakers with short lag /b/, whose means range downwards from 23 ms to 5 ms. This effect can neither be put down to cross-linguistic differences nor to experimental artifact. The small VOT differences between short lag /b/ and /p/ are observed in speakers of the *same language*, from the *same speech community*, using the *same phonemic contrast*, and reading the *same materials*. Koenig (personal communication) points out that, since phonetic models must be able to deal with the combination of phonetic cues, the VOICELESS specification of /p/ may make a short lag target for VOT a little longer (i.e. relatively more voiceless) than the “same” VOT target for /b/. The idea that the “same” value of short lag VOT could be either VOICED or VOICELESS may express no more than the idea that a VOICING contrast will never be cued *solely* by a tiny VOT difference.

VOT may well be predictable from the phonological specification of the phonemes concerned, but it is likely that it is a robust indicator of the contrast – it just happens to be in different parts of the continuum for different speakers. In Shetlandic English, the “/p/ vs. /b/” contrast seems to be shared at some abstract level, while individual speakers encode their own VOT targets. This offers a challenge to feature-based modular theories, in which VOT must be either distinctive or redundant, if phonological, or be indexical. The exemplar model, however, denies that such strict delineations are necessary, and assumes each part of the system is linked more or less directly to the other relevant cues that might pertain.

Furthermore, if short lag VOT is different in /b/ and /p/, then what is to be gained by positing at most three oppositions in VOT? For example, in the case of subject S4 there are two voicing lag targets very close together in the short lag range. *Both* /b/ and /p/ may be “short” lag, or /p/ may be a phonetically short “long” lag, if we are forced to assign categorical identities. But what surely matters is that short positive VOT may be less robust (not simply either distinctive or redundant) as a cue to the contrast for speaker S4. It is ultimately unhelpful to assume that a category *label* such as “short

lag" in the absence of (or even in addition to) the actual quantitative details captures all the important parts of the system, a point made strongly by Docherty (1992).

Shetlandic short-lag labial stops do not exist in isolation, but function as one pole of a binary opposition. We can hypothesise that the more prevoicing for /b/ that a speaker uses, the shorter their VOT for /p/ can be without causing perceptual confusion. Conversely, if a speaker uses very little prevoicing for /b/, their VOT target for /p/ is more phonetically distinctive if it is long lag. In Figure 4 we may be able to see some evidence to support these hypotheses. There is generally an inverse relationship between the rate of prevoicing and the mean VOT for /p/, although there are two subjects (S8 and S10) who contrast aspirated /p/ with a /b/ which is prevoiced in about 60% of cases.

This is not to rule out a simultaneous categorical analysis, so long as it can be approached flexibly. With reference both to the raw data plotted in Figure 4 and the analysis in Figure 3, it may be that S10 and S8 could be classified as having a mix of VOICED and VOICELESS /b/ plus ASPIRATED /p/. Subjects S5, S6, S7 and S12 could have VOICELESS /b/ and ASPIRATED /p/. Subjects S3, S9 and S1 could have UNASPIRATED /p/ with VOICED /b/. Subject S4 could have UNASPIRATED /p/ and vary in VOICING for /b/. This still leaves some doubt about S2 and S11.

Figure 4. Relationship of rate of prevoicing of /b/ (%) to mean Voice Onset Time (ms) of /p/ for all speakers (identified by code number), based on the raw data in Table 4.

Even if a categorical interpretation of Figure 4 were possible, so that all variation were either categorical feature switching or phonetically gradient, as the more conservative approaches to fine phonetic differences demand, three problems remain. First, the categorisation of some subjects such as S11 in particular remains highly ambiguous. Consider also S4. Since 40% of her /b/ are voiceless, perhaps her /p/ is VOICELESS ASPIRATED even though it is only 33 ms and even though S9 and S3 have a longer VOT for their /p/ which could be VOICELESS UNASPIRATED. Second, more than one binary feature would be required to encode /p/ vs. /b/ for the group of speakers in order that short lag /b/ is distinguished from short lag /p/. Third, there appears to be a gradient inverse relationship between prevoicing for /b/ and aspiration for /p/ that exists *within-category*, independently of any particular categorisation that might be imposed.

Within a modular framework, a categorical analysis ought to reveal the simpler structure underlying variable data, not replicate in an impoverished way the detailed linguistic distinctions and patterns of variation that the speakers make. An exemplar approach, on the other hand, predicts flexibility among speakers and permits flexibility in our linguistic analysis between gradient and categorical aspects of sound systems.

4. Summary and conclusions

As expected, the distinction between prevoiced /b/ and unaspirated /p/ is present as a phonemic contrast in Shetlandic English, for at least one speaker. Generally there is an inverse relationship between the rate of prevoicing for /b/ and the VOT target for /p/, though two speakers seem basically to oppose long lag /p/ with substantial amounts of prevoicing for /b/. It is unclear whether this inverse relationship reflects a gradient tendency within a single system of /p/ vs. /b/, or two inverse relationships within two categorically distinct systems (which we might call [b] vs. [p] and [p] vs. [p^h]). In a modular approach to phonology and language-specific phonetics a decision would have to be made on this issue, but we do not seem to have reliable methods for doing so, a difficulty reflected in previous discussions by Cho and colleagues. This may be a general methodological failing, but it may instead be seen as support for the view (e.g. Browman and Goldstein, 1991; Ohala, 1990) that there is no strict interface between phonetics and phonology – as support for “meta-gradiance.” This would mean that the distinction between gradient and categorical phenomena is itself gradient (Scobbie, in press).

VOT also seems to function indexically. Mean VOT for /p/ falls on a continuum, such that the four speakers with Shetlandic parents reflect the local vernacular (i.e., shorter VOT values for /p/, and pre-voiced /b/) more than most of the other subjects in this study. Interestingly, those whose parents are from elsewhere in Scotland (S5–S8) are at the other extreme on the continuum. English-parented subjects, including the two brothers, appear throughout the continuum. Speakers' systems therefore reflect to an extent the parental and community target systems in addition to arbitrary individual differences. VOT for /p/ and the rate of prevoicing for /b/ are therefore likely to be sociolinguistic variables.

In Shetlandic English, VOT seems to be a robust indicator to the VOICING contrast for each individual, despite the fact that the VOT values for /p/ and /b/ may be scattered through a large and apparently ambiguous region of the relevant phonetic space, when thinking of the community as a whole. The range of positive VOT values suggests that there is little to be gained from the theoretical supposition that phonetically meaningful features exist universally. Rather, speakers share two things: a comparable phonological VOICING contrast in the same lexical items, and the same abilities to respond to the functional demands of production, perception and acquisition of lexical contrast and other aspects of linguistic competence.

The VOICING contrasts observed range along a continuum reflecting traditional Scots and English, or draw on the extreme aspects of both. Such solutions are presumably typical of English-speaking communities the world over in both multidialectal and multilingual situations. These results and those of Khattab (2003) suggest that situations in which learners are faced with distinct – even incompatible – systems can lead to the acquisition of arbitrary targets for VOT, as exemplar theory allows. However, such stochastic models must not limit learners to acquiring a slavish recapitulation of the distribution of raw tokens to which they have been exposed. Higher-level abstraction and analysis by learners is also necessary.

Labov (1994: 25) expresses his desire “to reinforce the natural alliance of dialect geography, sociolinguistics, phonetics and historical linguistics – fields that share a common interest in objective [speech production] data.” We can add at least child language acquisition and speech pathology to this list. To date, most of the impetus for a rapprochement between sociolinguistics and laboratory-based phonetics and phonology has arisen due to the demands for quantitative phonetic data in sociolinguistic research. By definition, the field of laboratory phonology also focuses on the descriptive and theoretical importance of subtle phonetic and phonological distinctions.

What has not yet been fully appreciated by the experimental community is the potential value of the more vernacular varieties of English in their own right, or, moreover, the varied systematic relationships which are known to exist between different socially-distributed varieties of a language.

Unwanted variation is a problem for all experimental studies. Variation due to differences between subjects themselves is typically minimised rather than being controlled as an experimental factor. This is explicit in the design when subjects are only selected if they have *a priori* similar accents of a language, typically English. Often, experimental subjects are university colleagues or students: well-educated adults who almost always speak a standard variety. Thus an implicit homogenisation arises because study after study draws on a limited number of English accents. This paper rejects the assumption that the use of homogenous pools of subjects simplifies phonetic and phonological analysis. Instead, the approach to sound system analysis advocated here draws on a structured pool of subjects who can *a priori* be *expected* to vary linguistically. This structured heterogeneity reveals more about the underlying uniformities of the speakers' linguistic systems, as well as providing new challenges for the theorist.

Notes

- * Without Marie Cluness, Jane Stuart-Smith and the Economic and Social Research Council (Personal Research Fellowship R000271195), this research would never have happened. Many thanks to Cathi Best, Laura Koenig and an anonymous reviewer for very careful and helpful comments on an earlier draft, and additionally to Bob Ladd, Gerry Docherty, Paul Foulkes, Ben Matthews, Klaske van Leyden, Lisa Lavoie, and Susanne Fuchs for further comments, discussion, help and inspiration.
- 1. Consider the enormous influence of Ohala, Articulatory Phonology and research reported in this series. Of particular relevance are the instrumental sociophonetic studies of Docherty and colleagues (Docherty, Foulkes, Milroy, Milroy and Walshaw, 1997; Docherty and Foulkes, 2000), and acquisition research by Vihman and colleagues (Vihman and Velleman, 2000).
- 2. A sound system comprises all the learned aspects of phonetics and phonology.
- 3. Such experiences include internal psychological states and innate predispositions as well as real world interactions. Acquisition itself is an experience: the very first abstractions drawn by children influence those that come later, perhaps explaining why the developmental paths which children take are apparently even more varied than the resulting adult systems.

4. Capitalised terms (other than the acronym VOT) indicate possible phonological features but do not assume any specific set.
5. The implications of this are profound for research in acquisition, perception, bilingualism and second-language learning which often assumes, contrary to fact, that English speakers never produce or experience prevoicing. In the USA, the census of 2000 reports that more than 50 million people (18%) live in a household where a language other than English is spoken (in addition to or instead of English). More than half of these people live in a household in which Spanish (a language with prevoicing) is used. English is often acquired as a native language in a multilingual or multidialectal context, and the adult English used may be very different from what is assumed in the literature.
6. The rate of occurrence of prevoicing in the English speech of older bilingual children (English/Panjabi) has also been found to be intermediate relative to two control sets of monolinguals (Heselwood and McChrystal, 2000).
7. Scots, like English, is West Germanic, and is sometimes said, rather whimsically, to be the language most closely related to English.
8. In all 614 labial stops were measured. The minimum number of tokens analysed per subject for /p/ and /b/ were 17 and 21. All /p/ were singleton consonants, but 107 /b/ were in the clusters /bl/ and /br/. The means for singleton /b/ were -71 ms (s.d. 15 ms), for cluster /b/ -71 ms (s.d. 16 ms).
9. Lisker and Abramson (1964) and Docherty (1992), however, found a smaller proportion of negative VOT, due largely to a single speaker. Docherty argues that this suggests first that individual speakers can choose a particular VOT target and second that interspeaker variation is an important but neglected aspect of the description of a language.
10. Only 11 tokens are excluded (of 334), for S3, S9 and S5. Eight other tokens (from S7 and S6) fall below the threshold but will be reassigned, see the text.
11. Under a modular interpretation, brothers S9 and S11 would now be classed as having different systems, a rather arbitrary conclusion.
12. No speaker has a bimodal distribution for all their VOICELESS stops.
13. S4 does not have a bimodal distribution for /p/, nor do her distributions of /b/ and /p/ overlap, so her low mean for "long lag" /p/ is not due to intrasubject variation between short and long lag targets.

References

- Allen, J. S., Joanne L. Miller and David DeSteno
 2003 Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America* 113: 544–52.

- Browman, Catherine P. and Louis Goldstein
 1991 Gestural structures: distinctiveness, phonological processes, and historical change. In: I. G. Mattingly and Michael Studdert-Kennedy (eds.), *Modularity and the Motor Theory of Speech Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, 331–338.
 1992 Articulatory Phonology: an overview. *Phonetica* 9: 155–80.
- Bybee, Joan
 2001 *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Bybee, Joan and Paul Hopper (eds.)
 2002 *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.
- Cho, Taehong, Sun A. Jun and Peter Ladefoged
 2002 Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30: 193–228.
- Cho, Taehong and Peter Ladefoged
 1999 Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27: 207–29.
- Cluness, Marie
 2000 An investigation into the parental influence on dialect acquisition. Queen Margaret University College, Edinburgh.
- Coleman, John
 2002 Phonetic representations in the mental lexicon. In: Jacques Durand and Bernard Lax (eds.), *Phonetics, Phonology, and Cognition*. Oxford: Oxford University Press, 96–130.
- Docherty, Gerard J.
 1992 *The timing of voicing in British English obstruents*. Berlin: Foris Publications. Netherlands Phonetic Archives IX.
- Docherty, Gerard J. and Paul Foulkes
 2000 Speaker, speech and knowledge of sounds. In: Noel Burton-Roberts, Philip Carr and Gerard J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press, 105–129.
- Docherty, Gerard J., Paul Foulkes, James Milroy, Lesley Milroy and David Walshaw
 1997 Descriptive adequacy in phonology: a variationist perspective. *Journal of Linguistics* 33: 275–310.
- Fischer-Jorgensen, E. and B. Hutter
 1981 Aspirated stop consonants before low vowels: a problem of delimitation – its causes and consequences. *Annual Report, Institute of Phonetics, University of Copenhagen* : 77–102.

- Flege, James E.
 1982 Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics* 10: 177–92.
- Foulkes, Paul and Gerard J. Docherty (eds.)
 1999 *Urban Voices: Accent Studies in the British Isles*. London: Arnold.
- Foulkes, Paul and Gerard J. Docherty
 in press The social life of phonetics and phonology. *Journal of Phonetics*.
- Goldinger, Stephen D.
 1997 Words and voices: Perception and production in an episodic lexicon. In: Keith Johnson and John W. Mullenix (eds.), *Talker Variability in Speech Processing*. London: Academic Press, 33–66.
- Hawkins, Sarah
 1979 Temporal coordination of consonants in the speech of children: further data. *Journal of Phonetics* 7: 235–267.
- Hawkins, Sarah and Rachel Smith
 2001 Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian journal of Linguistics / Rivista di Linguistica* 13: 99–188.
- Heselwood, Barry and Louise McChrystal
 2000 Gender, accent features and voicing in Panjabi-English bilingual children. In: Nelson, Diane and Foulkes, Paul (Eds.) *Leeds Working Papers in Linguistics and Phonetics* 8: 45–70.
- Johnson, Keith
 1997 Speech perception without talker normalization. In: Keith Johnson and John W. Mullenix (eds.), *Talker Variability in Speech Processing*. London: Academic Press, 146–165.
- Johnston, Paul A.
 1980 A synchronic and historical view of border area bimoraic vowel systems. University of Edinburgh.
 1997 Regional variation. In: Charles Jones (ed.), *The Edinburgh History of the Scots Language*. Edinburgh: Edinburgh University Press, 433–513.
- Keating, Patricia A.
 1984 Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286–319.
- Khattab, Ghada
 2003 Sociolinguistic competence and the bilingual's adoption of phonetic variants: Auditory and instrumental data from English-Arabic bilinguals. University of Leeds.
- Labov, William
 1994 *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

- Lavoie, Lisa M.
 2001 *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. New York and London: Garland Publishing Inc. Outstanding Dissertations in Linguistics.
- Lisker, Leigh and Arthur A. Abramson
 1964 A cross-language study of voicing in initial stops: acoustic measurements. *Word* 20: 384–422.
 1967 Some effects of context on voice onset time in English stops. *Language and Speech* 10: 1–28.
- Masuya, Yoshihiro
 1997 Voice Onset Time of the syllable-initial /p/, /t/ and /k/ followed by an accented vowel in Lowland Scottish English. In: Yoshihiro Masuya (ed.), *Phonetics and Phonology*. Tokyo: Koban-Shobo, 139–172.
- Mullenix, John W.
 1997 On the nature of perceptual adjustments to voice. In: Keith Johnson and John W. Mullenix (eds.), *Talker Variability in Speech Processing*. London: Academic Press, 67–84.
- Ohala, John J.
 1990 There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics* 18: 153–171.
- Pierrehumbert, Janet B.
 2002 Exemplar dynamics: Word frequency, lenition and contrast. In: Joan Bybee and Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 137–157.
- Pierrehumbert, Janet B., Mary E. Beckman and D. R. Ladd
 2000 Conceptual foundations of phonology as a laboratory science. In: Noel Burton-Roberts, Philip Carr and Gerard J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press, 273–303.
- Pierrehumbert, Janet B. and Deborah Talkin
 1992 Lenition of /h/ and glottal stop. In: Gerard J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II*. Cambridge: Cambridge University Press, 90–117.
- Pisoni, David B.
 1997 Some thoughts on “normalisation” in speech perception. In: Keith Johnson and John W. Mullenix (eds.), *Talker Variability in Speech Processing*. London: Academic Press, 9–32.
- Ryalls, Jack, Allison Zipprer and Penelope Baldauff
 1997 A preliminary investigation of the effects of gender and race on Voice Onset Time. *Journal of Speech Language Hearing Research* 40: 642–645.

Scobbie, James M.

- 2005 Interspeaker variation among Shetland Islanders as the long term outcome of dialectally varied input: speech production evidence for fine-grained linguistic plasticity. *QMUC Speech Science Research Centre Working Papers WP-2*.

Scobbie, James M.

- in press Interface and overlap in phonetics and phonology. In: Charles Reiss and Gillian Ramchand (eds.): *Handbook on Interface Research in Linguistics*. Oxford: Oxford University Press.

Scobbie, James M., Claire Timmins, Jane Stuart-Smith, Fiona Tweedie, Nigel Hewlett and Alice E. Turk

- 2000 Fieldwork in the urban jungle: an empirical phonological study of Glasgow English. *The VIIIth Conference on Laboratory Phonology*. Nijmegen.

Smith, Bruce L.

- 1978 Temporal aspects of English speech production: a developmental perspective. *Journal of Phonetics* 6: 37–67.

Suomi, K.

- 1980 Voicing in English and Finnish stops. *Publications of the Department of Finnish and General Linguistics of the University of Turku* 10.

Thomas, Erik R.

- 2002 Instrumental phonetics. In: Jack K. Chambers, Peter Trudgill and Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell Publishers Ltd, 168–200.

Vihman, Marilyn M. and Shelley L. Velleman

- 2000 Phonetics and the origin of phonology. In: Noel Burton-Roberts, Philip Carr and Gerard J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press, 305–339.

Westbury, John R.

- 1979 Aspects of the temporal control of voicing in consonant clusters in English. *Texas Linguistic Forum* 14: 1–304.