CHAPTER 4

SPEAKER-RELATED VARIATION– SOCIOPHONETIC FACTORS

GERARD DOCHERTY AND
NORMA MENDOZA-DENTON

In this chapter the authors provide a concise and rich review of the sociolinguistic literature on variation, including developments in the interpretation of such variation and the methods used to study it. It is argued that the concerns of sociolinguists and laboratory phonologists are increasingly converging on a deeper appreciation of the ways that "the social-indexical channel is embedded within speech processing and representation" (p. 56).

4.1 Introduction

Since building up momentum in the late 1980s, laboratory phonology has stood not for a unitary theoretical position, but rather for an approach which draws together a diverse group of scholars united in the belief that understanding of spoken communication can best be developed by integrating methods and concepts from research traditions (e.g. phonetics, phonology, psycholinguistics) which in the past have not always been effectively joined up (Cohn 2010; Pierrehumbert and Clopper 2010). A perusal of the LabPhon volumes over two decades provides ample evidence of the many facets of the laboratory phonology enterprise and highlights the defining theoretical questions which have driven the integration of methods and concepts characterizing its development. These include, for example, the nature of lexical and phonological representation, the extent to which representation of sounds and words in memory is governed by/reflects what we know about the processes of speech production, perception, and acquisition, and how much of what is observed empirically can be accounted for by general principles of cognition, motor control, physiology (and what aspects require some special modality-specific explanatory framework).

A common thread through work presented under the laboratory phonology banner is that advances in theory have been rigorously grounded on quantitative analysis of the performance of individuals, either as speakers or as listeners. However, it is only in the latter years of the development of laboratory phonology that members of this community have paid significant attention to the fact that the vast majority of utterances produced naturally by speakers and processed by listeners are situated in an interactional context in which the substance of speech is shaped by the social factors playing out within that interaction as well as by the propositional content transmitted between the interlocutors (see overview contributions by Local 2003; Docherty 2007a; Mendoza-Denton 2007; Huffman 2007; Foulkes 2010; Munson 2010; Foulkes et al. 2010).

The integration of a sociophonetic dimension into the laboratory phonology "project" raises a profound (but, as we suggest below, not irresolvable) tension. In general, the more we discover about socially-situated speech, the more we are confronted with the central role that the social-indexical channel plays in the natural performance of speakers/listeners, but equally, the more conspicuous becomes the absence of an account of how this channel of information is integrated into speech alongside the lexical-propositional channel which, to date, has been predominant in the development of theoretical stances around speech production, perception, and learning (though see McMurray and Farris-Trimble, this volume). This chapter presents a critical overview of these issues, highlighting some of the key ways in which social factors impact on the performance of speakers and listeners, and then

reviewing how this social-indexical dimension is starting to shape thinking within the laboratory phonology community.

4.2 SPEAKING THE SAME LANGUAGE

One of the symptoms of the long-standing disconnect between investigators working on socially correlated aspects of phonological variation and those from within the subject areas drawn together under the laboratory phonology banner is that a good deal of what we know about speech as a social phenomenon has been developed with a different frame of reference, terminology, and analytic methodology from that which applies more generally within the laboratory phonology community. One example of this is the predominant role played by Varbrul analysis in sociolinguistic studies of phonological variation (Sankoff and Labov 1979; Sankoff 1988; and critically evaluated by Mendoza-Denton et al. 2003; Pierrehumbert 2006a; Johnson 2009; and Coetzee, this volume). But perhaps the clearest example of this is the notion of the *phonological variable* which is deployed as an analytic tool within a great deal of sociolinguistic research (see Chambers 1995 and Milroy and Gordon 2003 for a thorough evaluation of the application of this method). In the sociolinguistic analysis of phonological variation and change, phonological variables are segmental loci of socially structured variability, broadly equating to a phonemic level of abstraction (and in the case of vowels defined by Labov 2001: xvii as "abstract phonological elements that define historical word classes"). Unlike a conventional analysis of allophonic realization which targets what is hypothesized to be the "same" phonemic element across different contexts (e.g. word-initial vs. word-final), in the sociolinguistic analysis of a phonological variable the aim is to systematically track within- and across-speaker variability in a single context with a view to identifying the extent to which such variability is governed by diverse social factors. Examples of variables which have been the focus of relatively recent studies of English are (t)[1] (e.g. Docherty et al. 1997), (ing) (Labov 2001), (th) (Stuart-Smith and Timmins 2006), and a range of vowel variables including (ay), (aw), and (aeh) (Labov 2001).

[1] Note the use of parentheses to denote a phonological variable within sociolinguistic research; although this usage is not consistently applied. For vowel variables there is a difference between investigators such as Labov (1994, 2001) who do use this notation, and others who refer to vocalic variables by using the "lexical sets" proposed by Wells (1982); thus Labov (1994) refers to (ay), whereas Kerswill et al. (2008) refer to the PRICE lexical set.

In order to track the variants of a phonological variable, in many cases the analysis proceeds by scoring the occurrence of a set of auditorily identified variants; for example, in investigations of -t/-d deletion, investigators typically track the presence/absence of the plosive; for (ing)—whether the nasal is alveolar or velar; for (th)—whether the fricative is dental or bilabial. In many instances, and particularly for consonantal variables, investigators approach this analysis task with a preconceived set of variants (based on previous studies or from a pilot investigation), but it is incumbent on them to identify every variant which is encountered in a particular context including those which are unexpected or which occur with only very low frequency. (For example Docherty et al. 1997 unexpectedly found voiced [t̬] tokens in their study of (t) in Tyneside English, and for some speakers this turned out to be a salient characteristic of their realization of (t).) While the use of acoustic methods to analyze consonantal variables has increased over recent years (Docherty and Foulkes 1999; Stuart-Smith 2007b; Foulkes et al. 2010), they have for many years been established as the conventional method for analyzing vocalic variables (at least as far back as Labov et al. 1972). Typically, vowel variants are plotted in F1/F2 space usually following some form of normalization in order to minimize the risk of any cross-speaker differences being attributed sociolinguistic significance when they might simply arise from differences in vocal tract length. The relative merits of different types of normalization for tracking sociophonetic variability in vowel production are amply discussed by Labov (2001, 2006), Adank et al. (2004), and Watt and Fabricius (2002). On the whole, consonantal variables have tended to be looked at independently of one another, whereas for vowels there has been a greater attempt to consider a number of variables alongside one another in order to identify any mutual interaction (driven largely by the hypothesis that variation and change in one part of the vowel space can give rise to a chain reaction of shifts in vowel quality for particular lexical sets, a view which is strongly encapsulated in Labov's extensive analyses of vowel chain-shifts; 1994, 2001).

This overall methodology is well established and has undoubtedly enabled many fundamental insights into socially correlated phonological variability in speech performance. However, when viewed in light of what we know from experimental phonetic research about the properties of speech production and perception and how these shape phonological systems, there are dimensions of this work which are potentially problematic. One salient issue relates to whether, when investigators track discrete consonantal variants of a particular variable, the actual speech behavior being tracked is really as discrete as the analyst-imposed categories being deployed in the analysis. For example, in British English there are a number of accounts of the variable realization of (r) as an alveolar or labial variant, but acoustic analysis of these various realizations (e.g. Foulkes and Docherty 2000) suggests that the variability is best captured by considering any particular token to be positioned in a continuous (acoustic, and therefore articulatory) space between [ɹ] and [ʋ]; meaning that it would be a simplification to base any theoretical development on

this variable if it was being handled simply as having binary realizations.[2] This issue of the status of discrete segmental categories is of course a very familiar one to members of the laboratory phonology community, having been a key focus of work carried out under the rubric of Articulatory Phonology (Browman and Goldstein 1986; Pierrehumbert and Talkin 1992), and it represents an important source of tension between the perspectives on variation in phonetic realization provided by the fields of study being considered here.

A second issue is that the large volume of studies of socially correlated phonological variation is heavily skewed towards the analysis of one language (English), and recurrently focuses on a subset of variables, partly as a consequence of the adoption of the phonological variable methodology (leading effectively to a focus on segmental variation), and partly arising from the fact that certain variables (e.g. -t/-d deletion, (ing), and certain vowel variables) have been recurrently investigated in order to test particular hypotheses regarding variation and change, such as the notion of variable rules or vowel chain-shifting. It is true that particularly in recent years there have been studies which point to a wider range of phonetic parameters taking on a social-marking role (e.g. Esling 1978; Stuart-Smith 1999 on voice quality; Daly and Warren 2001 on the realization of f0 contours), but, in contrast to the understanding built up over many decades of the phonetic parameters deployed across languages in support of lexical-phonological contrast (e.g. Ladefoged and Maddieson 1996), we are still a long way short of a similar understanding of the cross-language range of phonetic parameters associated with the social-indexical channel.

A third issue relates to the analysis of vowel variables. While it is positive that acoustic methods are the standard in sociolinguistic studies of vowels (including for some investigators normalization into auditory space in order to gauge the perceptual relevance of differentiation in vowel realizations—e.g. Warren et al. 2007), there is some ambiguity about the extent to which many investigators equate the two-dimensional space within which tokens of vowels are typically plotted and compared with the multi-dimensional articulatory space within which vowels are articulated. Thus, a set of vowels which is distributed such that F2 is higher than another set will often be referred to as more "fronted," and likewise if the difference is an overall lower F1, the vowels will said to be "raised." These terms may well simply be serving as a means of capturing relative positioning within acoustic/auditory vowel space, but what they plainly cannot do is to reflect the complex relationship between articulatory and acoustic properties of vowels, and the danger is that they are interpreted as relating to the latter when in fact they can only reliably relate to the former. For example, realizational variants in vowel quality are only relatively

[2] Note too that discrete variables are the only type which investigators have been able to accommodate within the predominant analytic tool Varbrul, although this constraint has recently been loosened through Johnson's (2009) recent work on the development of R-Brul.

rarely attributed to different degrees of lip-rounding/protrusion, even though this articulatory parameter can have a major influence on formant frequencies, and does indeed participate in socially correlated variation (see Kerswill et al.'s 2008 analysis of the fronting and loss of rounding on GOAT vowels in British English). As pointed out by Foulkes et al. (2010), the focus on F1/F2 space has also drawn attention away from interesting sociophonetic differences in duration (e.g. Scobbie et al.'s 1999 work on the Scottish Vowel Lengthening Rule), formant dynamics, and contributions to perceived vowel quality made by F3 and the higher formants.

As indicated above, sociolinguistic analysis of phonological variables attempts to minimize positionally generated variation in order to capture significant inter-/intra-speaker variation in the same context. However, a further area in need of elaboration is the extent to which this form of analysis of realization variants is sufficiently sensitive to the range of phrasal and other prosodic features which work in the laboratory phonology community has shown to be closely associated with the magnitude and timing of articulatory gestures (e.g. Keating et al. 2003; Cho and McQueen 2005; Keating 2006). In general, the analysis of socially correlated variation has not controlled for factors such as prosodic constituency, the structure of conversational interaction, or speech rate. That factors such as these may well be important in tracking socially correlated variation is evident in studies such as Docherty et al. (1997) who found significant differences in word-final (t) realization depending on whether the token was in pre-pausal position or not; Local (2003), who described how phonetic detail can be used to denote key landmarks within a conversational interaction (such as turn transitions or conversational repair); and Docherty's (2007b) finding that speech rate influences cross-speaker variability in the realization of (t), even in a fairly formal reading style.

The areas just identified as potentially problematic with respect to conventional approaches to the analysis of socially correlated realization variation suggest that there is much to be gained from the greater methodological and theoretical refinement which would be engendered by a more productive dialogue across the laboratory phonology and sociolinguistics communities, and indeed, this process is well under way. (See, for example, the 2006 *Journal of Phonetics* special issue on Modelling Sociophonetic Variation, and the thematic orientation of the eleventh LabPhon Conference towards "Social information in the lexicon," dealing with questions such as "Is phonetic information in the lexicon accompanied by social information?", "How do social expectations about a speaker affect speech perception?", "Is speaker-specific detail stored in the lexicon?").

Thus, notwithstanding the need for this dialogue to develop much further in order to refine our understanding of sociophonetic variation, there is now a clear recognition of the significance of the social-indexical channel for speakers/listeners and for models of how speakers and listeners plan and execute their participation in spoken communication. The present discussion now moves on to explore some of

the key features of sociophonetic variability and its key points of contact to debates within laboratory phonology.

4.3 SOCIAL-INDEXICAL VARIATION

The history of quantitative sociolinguistics is largely an attempt to understand how patterns characterizing a speech community emerge from and relate to individuals' linguistic production and perception, specifically in relation to phonological and other linguistic variables. A classic definition of the speech community reads as follows:

> The speech community has been defined as an aggregate of speakers who share a set of norms for the interpretation of language, as reflected in their treatment of linguistic variables: patterns of social stratification, style-shifting, and subjective evaluations. This orderly heterogeneity normally rests on a uniform structural base: the underlying phrase structure, the grammatical categories, the inventory of phonemes, and the distribution of that inventory in the lexicon. (Labov 1989a: 2)

This definition encompasses many of the defining characteristics and central assumptions of mainstream sociolinguistic research. Historically, the speech community was defined primarily by shared patterns of subjective evaluation (Labov 1972a), then redefined as "sharing a set of norms," which has largely been interpreted as *using* variables in similar ways rather than just *assessing* them in the same way—thereby eliding some of the complexity of production-perception relationships (Keating 1987; Johnson, Flemming, and Wright 1993; Liberman and Whalen 2000). Key in the sharing of norms are consistent patterns of social stratification, meaning that all segments of the population evaluate and use a particular form as more prestigious than another; methodologically, this entails that a community (already aggregated) must ordinarily be stratified by class, age, ethnicity, gender, etc., in order to be studied. In the above definition, style-shifting also emerges as central, albeit with a definition of style that posits a continuum between formal (word list) and informal (conversational) styles based on attention paid to speech: the more attention is paid, the more formal speech becomes (Chambers 1995). Making sense of "orderly heterogeneity" then becomes *the* puzzle in itself, since it is assumed that any stratified population replicates the history of language change and carries within it the seeds of further development. There is also an assumption that underneath the heterogeneity, the population is quite uniform and shares a "grammar," in the classic, generative sense of the term. Accordingly, if they didn't share a grammar (or a phonological inventory, or a set of evaluations) they would be a different speech community altogether. This in turn leads to claims that an individual's patterns of variation mirror those of her community grammar

in terms of the statistical ordering of factors that determine the variation (Guy 1980; Poplack and Tagliamonte 1991; Poplack 2001), and that if different factor orderings obtain among different subgroups for a specific linguistic variable in a population, then we must be dealing with different speech communities. Auger and Villeneuve (2008), for instance, make exactly this claim in their argument that Picard and French, two neighboring varieties, are distinct languages because their constraints on the morphophonological factors affecting *ne* deletion are differently ordered.

Most of the studies in this vein are quantitative (using the Varbrul method referred to above), but are not laboratory-based. The use of a stable, replicable methodology has nevertheless enabled a measure of control in the study of naturalistic speech. A canonical sociolinguistic interview divides its time between (1) demographic questions and background information; (2) a series of question-prompts that aims for relaxation and involvement as measured by unself-conscious storytelling on the part of the interviewee, with some fairly set topics (see Feagin 2002) that work more or less cross-culturally (though Wolfson 1976 and Mendoza-Denton 2008 offer some critiques); (3) a word list; and (4) a minimal-pair reading task. These last two tasks attempt to elicit the most self-conscious and formal genres in the speaker's repertoire, while the second storytelling task aims for the opposite: to capture the interviewee's speech at its most "natural" and "relaxed." It is these different levels of questioning that provide a control for interviewing protocols, and which define the styles (formal vs. informal) that are compared across subjects, interviewers, and even dialects/languages. Thus the methodology itself provides control across many different interviewing situations occurring in different cultural contexts (though many have subsequently noted crucial interviewer and contextual priming effects: Rickford and McNair-Knox 1994; Hay et al. 2009).

Traditionally, the emphasis in sociodemographically based sociolinguistics (Mendoza-Denton 2002) has been on understanding how language change arises from linguistic variation (the classical problem in Weinreich et al. 1968: actuation, transmission and diffusion of change in the speech community). Early studies were already tilted toward stratification in terms of sociological attributes and the styles (formal vs. informal) elicited as the independent variables. This iterative division in the samples yielded sociodemographically based correlations with linguistic variable use (e.g. as a function of social class, age, ethnicity, and gender),

Some of classical sociolinguistics' most notable findings (overwhelmingly driven by the study of phonology) include what Labov (1972a) called the Lower Middle Class crossover effect: the finding that in a population stratified by class, use of a phonological variable by the lower middle class will overshoot the norm of the upper class in the most formal styles. In Labov's case, this was demonstrated with (r) in New York City, and replicated early on by Trudgill (1974) in Norwich, England with the variable (ing). Curvilinear patterns showing that a group in

the center of the socioeconomic hierarchy is leading in a linguistic *change in progress* have been found in Philadelphia (Labov 2001), New York City (Labov 1966), Norwich (Trudgill 1974), Panama City (Cedergren 1973), and Cairo (Haeri 1996), *inter alia*. The explanation of the crossover effect is that the variable in question has gained an association with some desirable social distinction, so that status-sensitive groups overshoot what might otherwise pattern as parallel variation by class and style (known as *stable variation)*. As Labov (2002) remarks, "Sociolinguistic variation is parasitic upon linguistic variation. It is an opportunistic process that reinforces social distinctions by associating them with particular linguistic variant."

One of the most durable constructs in sociolinguistics is that of *apparent time* (Bailey et al. 1991), the assumption that if one slices the population into age brackets, the resulting distribution of variation will show changes in progress spreading through the population, with the speech of the youngest speakers reflecting the most innovative version of the community grammar and the speech of the oldest reflecting a more conservative version. The hypothesis of apparent time has two strong assumptions: one is that of the critical period, where it is assumed that speakers' phonology has been acquired and has stabilized by the teenage years (see Flege 2006); the other assumption is that speakers' linguistic systems are relatively stable and do not change as they age (but see Harrington, Palethorpe, and Watson 2000). One of the most complete studies so far to test the apparent time construct against real-time panel data, following the same speakers, is Sankoff and Blondeau (2007), who analyzed the community shift in the pronunciation of /r/ in Montreal French by comparing data collected in 1971 and 1984. They concluded that: "To the extent that older speakers change in the direction of change in progress during their adult lives, apparent time underestimates the rate of change" (Sankoff and Blondeau 2007: 582).

Findings in the area of ethnicity have tended to focus on the convergence/divergence question of black and white vernaculars in the USA (Labov and Harris 1986; Ash and Myhill 1986; Wolfram and Thomas 2002) and in transplant African-American communities such as those of Nova Scotia and Samaná (Poplack and Tagliamonte 1991; Poplack 2001), though the latter tend to focus on syntax rather than phonology. The bulk of sociophonetics work on ethnicity in the USA has historically concentrated on African Americans, immigrant non-whites, and their relationship to the changes in progress taking place in the majority community. In Europe, only recently have studies of bilingual immigrant communities taken a greater role as demographic changes show increasing participation of previously unexamined groups in the creation of new forms of the vernacular (Hewitt 1986; Rampton 1995; Kotsinas 1998; Heselwood and McChrystal 2000; Khan 2006; Khattab 2007; Cheshire et al. 2008; Alam 2009; Jannedy and Martins 2008). Other studies involving a single language with an "ethnicity" dimension (e.g. studies of language use in Northern Ireland by Milroy 1987a, McCafferty 1998; of varieties

of Bahraini Arabic by Holes 1986; and of varieties in Russia by Kochetov 2006c; and in China by Zhang 2005) are typically couched in terms of religion, culture, region, or other kinds of affiliation, and are similar to more canonical in-migration studies such as Kerswill (1994), Lane (2000), and Dyer (2002). Where ethnicity with concomitant multilingualism is the norm, studies tend to be classified as "language contact" (Flege 2006 on bilingual accommodation; Mesthrie 1992 on South African English; Devonish 2007 on Jamaican English; Holmes 1997 on New Zealand Maori English; papers in Meyerhoff and Nagy 2008).

The interpretation of gender differences in speech communities has been an enduring source of debate within sociolinguistic research. Eckert (1989) challenged the then-prevailing notion (linked primarily to Labov and Trudgill) that sound changes from above the level of consciousness were led by linguistically conservative, status-conscious women (see for instance Holmquist 1985), whereas sound changes from below the level of consciousness were started by working-class, covertly prestigeful men, and then taken over by women who became the leaders of change (for the ensuing debate, see Labov 1990; Coates 1993; Gordon 1997). A number of reasons had been suggested for what was perceived as social fact: because of status differentials, women were more linguistically insecure than men. (Note that this is also the kind of account that was used to explain why the lower middle class had the crossover effect.) Based on ethnographic work in an ethnically homogeneous high school close to Detroit, Eckert showed that social class and gender interacted within specific social structures in the field setting: at Belten High, the social landscape was dominated on the one hand by *jocks*, who were both establishment- and supra-locally oriented, had middle-class backgrounds, and a school-based social life; and on the other hand by *burnouts*, who were of working-class background, were locally oriented, rebelled against the school's *in loco parentis* role, and did not take part in school activities. Participation in the Northern Cities Chain Shift was led by burnouts, but within that category, it was the burnout girls who surpassed the burnout boys in iconic changes such as raising and backing of the nucleus of /ay/, while among the jocks the girls trailed the boys in this change. Clearly, more subtle explanations were needed than simply lumping all men and all women together in their participation in linguistic change.

This work paved the way for a major shift in the understanding of how change proceeds in communities and the role of individuals, and led to new ways of thinking about language and social meaning, especially the social indexicality of variables. By opening up the inner workings of communities in the late 1980s and early 1990s, both the social networks (Milroy 1987a; Milroy and Milroy 1985, 1993) and communities of practice frameworks (Lave and Wegner 1991; Eckert and McConnell-Ginet 1992) contributed to a sea change in how we understand the spread of variation and what it means to the speakers who are adopting it. Instead of looking at large communities from above, as disembodied analysts cutting up the social landscape into census tracts, researchers began trying to understand

communities from the participants' insider perspectives, and trying to uncover the social categories that may be meaningful within the community. An early harbinger of the problems in the traditional concepts of class, for example, was the work of Rickford (1986) who called for new, conflict- and power-based understandings of class because the social structure in Canewalk, Guyana, was not easily divided into composite class indices traditional in consensus-based sociolinguistics models. In Canewalk, language variation was governed by the categories of estate- and non-estate class, social divisions that were the result of Guyana's plantation history. Studies of class and gender as conglomerates of practices have led researchers deeper into the social histories of communities and life histories of individuals (Milroy and Milroy 1985; Johnstone and Bean 1997; Labov 2001; Mendoza-Denton 2008) to predict which individuals are the leaders in language change, and how that change might be structured in terms of language use, individuals' phonological systems, and the deployment of phonetic detail.

During the 1980s another large change took place in the way that sociophoneticians thought about style/register. As mentioned above, early studies linked individual style to community-wide stratification through the construct of attention paid to speech (Labov 1972a), but later studies such as Bell (1984) and Coupland (1980) looked at individuals' deployment of sociophonetic variables and posited that individuals, in crafting their temporary implementation of their linguistic styles, were responding to specific audiences, and often matching their phonological production to present, implied, or imagined audiences (this perspective is broadly known as audience design; Bell 1984).

The breaking apart of (a) strict demographic categories and (b) the perceivedly linear stratification of style has had significant consequences for sociophonetic research. In the area of gender, for instance, researchers are looking at gendered expectations in speech perception (Johnson et al. 1999); at the production of gendered speech toward children (Foulkes and Docherty 2006); and at non-binary gender situations (Pierrehumbert et al. 2004; Crocker and Munson 2006). In matters of style, laboratory phonology researchers in the communities of practice approach (exponents include Eckert 2000; Zhang 2005; Rose 2006; Stuart-Smith 2007b; Mendoza-Denton 2008; Alam 2009; Drager 2009; Lawson 2009) continue to develop ideas of styles as practices and constellations of behaviors (Eckert 2005; Podesva 2007, 2008), and of iconic personae that bring these styles together into salience and relevance in communities (see for instance Zhang's 2008 study of Beijing "smooth operator" speech, which involves strong rhotacization, the description of which is deeply rooted in Chinese literature—going all the way back to the Qing Dynasty).

Work in this vein suggests that for individual speakers the motivation for adopting particular socially marked patterns of phonetic realization seems to be chiefly about the construction and performance of identity or identities relating both to themselves as individuals and to their affiliation to (or dissociation from) the

diverse social groups with whom they interact. Where identities and ideologies regarding the use of language coincide within a community of speakers, this can provide the conditions conducive to the sorts of collective patterns of phonetic realization identified in conventional sociolinguistic studies. But where consistent differential patterning *is* observed across socially defined groups of speakers, it is perhaps not surprising that the distributions of variants across groups are rarely if ever categorical, given the readiness with which speakers adapt their performance to meet what they perceive as the social-indexical demands of particular communicative situations.

The overall picture emerging from sociolinguistic studies of phonological variation, then, is of individuals drawing on a wide range of phonetic parameters to index social affiliation and differentiation, and being able to do so flexibly on a moment-to-moment basis in line with the perceived demands of a particular communicative situation. While there is some evidence (e.g. Sangster 2002[3]) that stylistic adaptations can be under the conscious control of speakers, research into inter-speaker accommodation and convergence (Giles 1984; Bell 1984; Giles et al. 1991a; Coupland 2007) and anecdotal observation suggests that, more typically, shifting of this sort takes place without an explicit intention being formulated on the part of the speaker. And, of course, this all appears to be underpinned by a very significant process of learning and understanding of the community-specific social-indexical value of phonetic variation, and an ability to make instantaneous interpretations of the same.

4.4 THE INTERSECTION OF SOCIOPHONETICS AND LABORATORY PHONOLOGY

As indicated at the start of this chapter, these aspects of speech communication have until relatively recently not been seen as a central concern of laboratory

[3] Sangster studied phonological variation in the performance of undergraduate students who had relocated to Oxford University from Liverpool in the North-West of England (an area with particularly marked accent features—Watson 2007), uncovering the ways in which such variation was tied in to how individuals (in some cases quite overtly) managed their identity as Liverpudlians in an environment in which there were very few people from Liverpool. For example, one student is quoted as follows: "When I first came here I was more broad than I was normally because when you get there and everyone's like [posh voice] "oh yes I come from wherever" and then when you hear people speak like that—I think it's an unconscious thing that you just make yourself sound more Scouse because they like it, and almost everyone speaks the same, and it's good to be different, it's not a different bad-different, it's a good, happy sort of everyone-likes-it different" (Sangster 2002).

phonology; in part reflecting the walls around subdisciplines which affect linguistics research no less than other areas, but also reflecting an orientation to speech production/perception/learning which has focused predominantly on the lexical-contrastive information carried by the speech signal. However, what renders this disconnect particularly problematic is the fact that the speech signal is the channel through which, at one and the same time, speakers phonetically realize the phonological system acting as the foundation for lexical contrast *and* project the social-indexical features appropriate for particular communicative contexts (Docherty et al. 2006). Thus, from the point of view of the individual speaker-listener, the transmission of the lexical and social-indexical channels of meaning appears to be an integrated process, which suggests that any attempt to account for or model one of these channels without accounting for how it integrates with the other will necessarily be incomplete. Thus, as pointed out above, for the sociolinguistics community, there are gains to be made by giving greater consideration to a number of hitherto largely neglected factors which may well have an influence on the distribution of variants found across a sample of speakers or which might paint a more realistic view of the nature of the variants themselves. Likewise, for the laboratory phonology community, one of the key impacts of sociolinguistic studies of phonological variation is the realization that in drawing on the details of speech performance as a means of refining theories of (for example) lexical representation, it is simply not possible to filter out the social-indexical meaning which will also be conveyed within a particular utterance and which constitutes a key factor responsible for the phonetic shape of an utterance.

This message has perhaps come home most strongly for the laboratory phonology community as a consequence of an increasing number of speech perception studies which have shown that social factors shape the processing and interpretation of speech signals in ways which are not foreseen within conventional models which (if not by design, certainly by default) have not made any allowance for the social-indexical channel in production/perception (see Nguyen, this volume, for further discussion and details). For example, Lachs et al. (2003) and Nygaard (2005) review evidence showing how speaker-specific characteristics influence listeners' responses in various types of listening task. Two other particularly insightful studies are those by Strand (1999) showing that gender stereotypes shape listeners' responses to an [s ~ ʃ] continuum (concluding that "higher level relatively complex social expectations might have an influence on such low-level basic processes as phonological categorization of the speech signal"; p. 93), and by Niedzielski (1999) showing that Detroit listeners' judgments of vowel quality in the same stimuli are dependent on whether they believe the speakers are from Detroit or Canada. More recently, a set of similar studies has been carried out by Hay and colleagues (e.g. Hay, Nolan, and Drager 2006) showing differential perception of the same stimulus material by listeners depending on the (implicit) beliefs that they have about social factors relating to the material that they are being asked to respond to. Crucially,

as well as showing that speech perception is not independent of social-indexical information, these studies also highlight that listeners have acquired knowledge of the typical associations between specific features of speech performance and the characteristics of individual speakers, and, perhaps most relevant for this chapter, of groups of speakers of various sorts (e.g. males vs. females, young vs. old, USA vs. Canada, etc.).

With evidence pointing to the importance of integrating social-indexical phonetic properties into accounts of production and perception, there is no doubt that we also need to consider their role in phonological acquisition, not only from the point of view of how a child begins to learn the value of and make use of the socially governed variants within her/his speech community, but also from the point of view of understanding the social-indexical properties of child-directed speech and how this differs from adult-directed speech within the same community (bearing in mind that from a very young child's point of view, the speech community may well be made up simply of the immediate family). For further discussion of this area, see Foulkes et al. (2005), Foulkes and Docherty (2006), Foulkes (2010). A key question is to what extent, when setting out on the path of acquiring knowledge of the sound pattern of the ambient language, a child can separate out from within the input that she/he is exposed to from birth those features of the speech signal which are lexically contrastive and those which are social-indexical. While this may well happen at a later stage of development (as described by Foulkes 2010), it seems likely (Foulkes et al. 2005; Docherty et al. 2006) that the process of phonological acquisition is at one and the same time a means for learning the building blocks of the native language lexicon and for learning how to sound like a member of the immediate speech community. It is an empirical question how this integration is achieved and for how long it is maintained.

In sum, as mentioned above, the more we learn about the social-indexical channel in speech performance and the extent to which it is integral to the performance of speakers and the processing of speech by listeners, the stronger becomes the need to account for how the social-indexical channel is embedded within speech processing and representation. Historically, this was simply not possible given that most models of the latter had a clear focus on seeking to sustain the hypothesis of underlying representational invariance in the face of abundant surface variability. An example of this approach can be seen in the work on relational invariance underpinning the quantal theory of speech production (e.g. Stevens and Blumstein 1978; Stevens 2002; Stevens and Keyser 2010; see Hawkins 2004 for an overview) in which variability is cast as "noise" which needs to be minimized in order for the underlying invariants to be discerned (see also Lahiri, this volume, for an overview of the Featurally Underspecified Lexicon model which applies some of the same principles). This approach is also reflected in the quantitative methodologies adopted by many researchers in which conclusions are drawn from reports of central tendencies characterizing a sample of speakers as a whole without reporting

either at all, or in any detail, the extent to which the overall findings reflect the performance of individual members of the sample.

At the simplest level, progress towards bridging this gulf can begin to be made in quantitative production/perception/learning studies by simply factoring in to experimental designs some of the key factors which sociolinguistic research has shown to be relevant in accounting for speaker performance. For example, there is a growing awareness of the fact that in describing the accent background of experimental subjects, it is not sufficient to simply refer to the name of the language spoken by those subjects (e.g. "10 speakers of American English," or "10 speakers of French," etc.); at the very least there is a need to know something about the geographical provenance of the speakers, about the nature of the particular variety or varieties which are represented in the sample of data, and about the extent to which any cross-speaker variation is likely to impact on the focus of the study. But it is also possible to begin to develop "joined-up" accounts of speaker performance by factoring into the analysis non-linguistic factors which are relevant in accounting for the criterial dependent variables. A good example is Scobbie's (2006) study of VOT in the Shetland variety of English. This study involved a word-list task with twelve subjects aged 16–30, six males/six females, all born in the Shetlands, all of whom had lived there all their lives, all from the same geographical area on the Islands, half attending the same school, and mostly known to each other (i.e. a highly controlled group of speakers which might not unreasonably be assumed to provide a homogeneous sample). Overall VOT distribution was extremely variable across speakers (/p/ ranged from 0 ms to 112 ms, /b/ from −190 ms to 41 ms). However, analysis of individuals' performance revealed individual realization strategies which imposed some structure on the group findings, but also brought to light that an important factor in accounting for the variability found in the VOT results was the place of origin of the speakers' parents; parents of Shetland origin were associated with shorter VOTs for /p/ and more pre-voiced /b/s; other Scottish parents with higher VOTs for /p/ and fewer cases of pre-voicing; English-parented subjects were more variable across the VOT continuum. Thus, despite having sampled speakers in a way that by almost any measure would appear to be a very good basis for generalizing across speakers of Shetland English, it was only by considering differences between individual members of that community that a more informed and theoretically more challenging account emerged.

A key development over the last decade has been the elaboration of a more far-reaching theoretical platform for addressing the issues identified above. The exemplar approach to phonological representation, in which knowledge of phonological patterning is based on a multifaceted, phonetically rich representation in memory derived from and continually shaped by an individual's experiences as a speaker-listener, has opened the doors to a model in which the integration of the lexical and social-indexical would be entirely predictable and natural (Goldinger 1997; Johnson 1997b; Pierrehumbert 2001a, 2006a; Hawkins 2003; Foulkes and Docherty 2006).

Notwithstanding that there are a number of dimensions of this particular approach which remain to be developed (Docherty and Foulkes forthcoming), a key characteristic is its assertion that, in building up knowledge of the systematic aspects of sound patterning from their experience with spoken communication, speaker-listeners automatically and simultaneously map out associations between signal properties and both linguistic and non-linguistic aspects of experienced stimuli (Johnson 1997b; Foulkes and Docherty 2006; Pierrehumbert 2006a; Hay, Warren, and Drager 2006; Mendoza-Denton 2007; Foulkes 2010). Since social-indexical information is systematically intertwined with other channels of meaning within the speech signal, by hypothesizing an integrated, probabilistic, and experience-driven representation the exemplar approach provides a conceptualization of how these various channels can be fundamentally integrated in speech processing and representation as suggested above. And in doing so, its advocates argue that it is not incompatible with the sorts of abstract phonological representations which have predominated to date in work on phonological representation (Pierrehumbert 2006a). Indeed, there is an emerging consensus behind the concept of a hybrid model of representation incorporating both abstract and exemplar representations with the balance between the two now constituting something of a new focus of experimental work (Goldinger 2007). McLennan (2007: 69) summarizes this debate saying that "the field has entered into a new phase in which, rather than debating over abstract versus episodic representations, efforts are now focused on determining the ideal framework that can account for their coexistence."

Tellingly, Pierrehumbert (2006a) refers to the conceptual framework offered by a phonetically rich probabilistic representation as a "toolkit," correctly reflecting the fact that while there is some way to go before the details of this framework are fully tested and evaluated, nevertheless it does allow for the framing of questions which hitherto would have struggled to find a theoretical "hook." For example, there has been a growth in interest in the dynamic nature of an individual's phonological knowledge; an exemplar model of representation predicts that phonological knowledge continues to evolve through the life span, shaped by individual experience, contrasting with the conventional view that the acquisition of phonology is focused in the early years of development and is from that point stable across speakers of the "same" variety. This is exemplified by Harrington and colleagues' study of the phonetic characteristics of the British Queen over fifty years' recordings of the annual Christmas Day Queen's Speech (Harrington et al. 2000, 2005; Harrington, this volume) which provides a particularly detailed real-time account of life-span changes in the speech performance of an individual positioned very much at the conservative pole of language use. As pointed out by Labov (2006), however, while evidence of change in adulthood such as this does indeed reinforce the view that phonological knowledge can continue to evolve through life, there is a need to devise an explanation for this which also accounts for the fact that this dynamism appears to be much less marked in adults than it is in younger speakers. This is

presumably in large part due to the connection between phonological patterning and identity formation referred to above, but this is one area which is in need of much further investigation.

A closely connected question is that of what takes place when speakers with different sociolects come into contact. An integrated theoretical framework should make it possible to devise an elegant interface between sociolinguistic models of dialect contact (e.g. Trudgill 1986; Britain and Trudgill 1999; Kerswill 2002) and what we know of how individual listeners' phonological representations are influenced by exposure to phonetic realizations which they previously had little experience of, as revealed, for example, in studies of "perceptual learning" following exposure to novel phonetic realizations (Norris et al. 2003; Kraljic et al. 2008; Cutler et al. 2010) and by work on the plasticity of phonological categories and how this can be associated to speakers' different levels of exposure to particular varieties of English (e.g. Evans and Iverson 2004). And increasingly there is potential to imbue models of dialect contact/change/formation with an understanding of the factors pertaining to conversational interaction which influence the behavior of individual speakers of different varieties when they interact (e.g. see Delvaux and Soquet's 2007 account of passive speech imitation in speakers of Flemish, studies by Pardo 2006 and Babel 2009 of phonetic convergence between interlocutors, and work by Wedel and Volkinburg 2009 and Pierrehumbert, this volume to model the consequences for a community of speakers of this sort of inter-interlocutor phonetic entrainment).

What these studies exemplify is that, while it remains an area of intense debate, the exemplar-model "toolkit" has brought to the fore the question of how speaker-listeners manage the multiple channels of information interwoven into the phonetic properties of the speech signal. This is now a central area of theoretical development and debate, and one which sits straightforwardly alongside the other questions which have drawn together the laboratory phonology community. Thus, the historical gap between models of production/perception/acquisition and what we know of how social-indexical meaning is conveyed and interpreted within speech has started to be bridged. And with this line of investigation showing every sign of developing further, the importance of addressing some of the methodological issues referred to earlier in this chapter cannot be overestimated.

4.5 Concluding comments

As indicated in the introduction to this chapter, in recent years there has been a significant and quite rapid change in the extent to which the social-indexical properties of speech have figured within debates on the key questions around which

the laboratory phonology community is unified. Arguably the most important factor in this development has been the postulation that phonological knowledge is phonetically rich and is shaped and defined by an individual's experience, thereby emphasizing an intimate connection between the properties of language and the speech performance and processing by users of language. This view is amply explored by Pierrehumbert (2006a: 516) who begins by asserting that "language is a collective behavior" and that it is formed "in populations, as people match their language systems to each other, and group themselves into social networks of people who share the same language." In its formative years, not least under the influence of Ohala's groundbreaking work (e.g. 1983, 1990c), the laboratory phonology community readily embraced the need to understand what aspects of phonological patterning could be accounted for by factors such as vocal tract physiology or general principles of motor control. But there is now a firmly established strand of activity within the laboratory phonology "project" which extends this to consider how factors arising from the social orientation of users of phonological systems account for the nature of those systems. This strand of work is at a relatively early stage of development, and, as pointed out above, is in need of further methodological and theoretical refinement. But, notwithstanding these points, a key attraction of this vein of research is that of allowing, in due course, the emergence of models of phonological knowledge with a more rounded understanding of the role of speakers and listeners in the acquisition and maintenance of that knowledge.

It is also important to emphasize that none of the above is necessarily out of line with the stance taken by many theoretical phonologists. Coetzee (this volume, p. 62) points out that, as the result of a growing interest in variation on the part of theoretical phonologists (largely driven by exploring the extent to which particular theoretical frameworks can deal with realizational gradience and variability), "phonology is now more ready than ever to integrate the apparently disparate approaches of theoretical and laboratory phonology." While, to date, sociophonetic variation has taken a somewhat secondary role in the exploration of this integration, the work reviewed here suggests that it is a nettle which needs to be grasped more firmly and which has the potential to deepen our understanding of how individuals' orientation to the social context of "real-life" speech communication impacts on the nature and characteristics of the speech signal from which that phonological knowledge is derived, and ultimately how it shapes the nature of phonological representation itself.