
Article:
Docherty, Gerard, Langstrof, Christian and Foulkes, Paul orcid.org/0000-0001-9481-1004 (2013) Listener evaluation of sociophonetic variability:probing constraints and capabilities. Linguistics. pp. 355-380. ISSN 0024-3949

https://doi.org/10.1515/ling-2013-0014

DE GRUYTER MOUTON
DOI 10.1515/ling-2013-0014 — Linguistics 2013; 51(2): 355–380

Gerard J. Docherty, Christian Langstrof and Paul Foulkes

Listener evaluation of sociophonetic variability: Probing constraints and capabilities*

Abstract: This paper reports the results of an experimental study designed to investigate how listeners learn to create new associations between phonetic properties of the speech signal and external social referents. Very little is known of how this learning takes place in children, and it is a particularly challenging area to study given the difficulty in controlling some of the variables which are likely to be important factors in children's learning of the productive and interpretative dimensions of social-indexical phonetic variation. Thus, in this study, we focus on adult listeners in order to develop a sense of how adults might approach this learning task, and also to test out a method for probing this form of learning in a controlled fashion. 49 participants were trained on new patterns of social-indexical variability and, in a subsequent test phase, we assessed the extent to which this training led the listeners to acquire new associations between specific realizational variants and the social categories with which they have been associated in the training material. Results are reported from four experimental conditions which provided listeners with a range of different learning tasks. Our findings suggest that learning of novel sociophonetic associations can be achieved as the result of a relatively short amount of exposure to training material incorporating the new association, but that the success with which learning takes place is dependent on a number of factors such as the nature of the criterial variable and individual learner variation.

Keywords: sociophonetic variation, speech pattern learning, individual variation, British English

Gerard J. Docherty: School of ECLS, King George VI Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom. E-mail: g.j.docherty@ncl.ac.uk
Christian Langstrof: Albert-Ludwigs-Universität Freiburg
Paul Foulkes: University of York

* This research was supported by the program *Apprentissages, connaissances et société*, funded by the ANR (French national agency for research). It has benefited from comments received at the 2008 Colloquium of the British Association of Academic Phoneticians (University of Sheffield), the 2009 Production, Perception, Attitude meeting (Katholieke Universiteit, Leuven), the 11th Laboratory Phonology meeting (Victoria University, Wellington), and at a seminar delivered by the first author to the Linguistics Department at Queen Mary University London (May 2010).

1 Introduction

In recent years there has been a progressive convergence of thinking across what have historically been rather high walls separating the communities of researchers in sociolinguistics and phonetics (Docherty 2007; Foulkes 2010; Foulkes et al. 2010). For example, many sociolinguistic studies now incorporate more detailed and sophisticated phonetic analysis of speaker performance (e.g., Stuart-Smith 1999, 2007; Thomas and Carter 2006), while a number of recent phonetic studies have examined heterogeneous speaker samples in theoretical frameworks developed in sociology and sociolinguistics (e.g., Local 2003; Scobbie 2006; Drager 2009, 2010). Moreover, there is now a rapidly-developing strand of research on how models of speech processing can handle the sort of social-indexical phonetic realization that has been the mainstay of sociolinguistic research over many decades (Kraljic et al. 2008; Samuel and Kraljic 2009; Hay et al. 2006a, 2006b). One focus of the interaction between these two fields that has had relatively little attention, however, relates to how children, as they acquire knowledge of the phonological patterning of their native language, learn to produce and perceptually evaluate the social-indexical properties of speech which are relevant to their speech community. The evidence that is available suggests that children learn to manipulate and interpret these properties in a way integral to phonological learning more generally (Foulkes et al. 2005; Khattab 2007; Smith et al. 2007; Barbu et al. this issue). However, there has been very little progress made in tracking the emergence of such properties within children's speech or their developing sensitivities to the sociophonetic variability to which they are exposed.

One of the obstacles to progress in this respect is that we know very little indeed about the learning mechanism that underpins children's ability to map meaning (of all sorts) onto properties of the substance of speech that they experience and reproduce. The dominant generative approach to phonological analysis has enhanced our understanding of many aspects of the acquisition of lexical phonological contrast (e.g., Smith 1973; papers in Archibald 1994; and papers in Kager et al. 2004) but this approach simply does not take social-indexical variability into account at all (Docherty and Foulkes 2000; Foulkes and Docherty 2006). It therefore focuses only on the referential strand of meaning contained within the speech signal, and has little or nothing to say about indexical or pragmatic strands of meaning, for example. Likewise, the prominent corpus of research focusing on how very young children tune in to the phonetic categories of their ambient language (Vihman 1996; Jusczyk 2000; Werker and Yeung 2005; Maye et al. 2002) has brought to light the importance for phonological learning of exposure to the distributional properties of ambient phonetic substance. But,

here too, this work has had little to say about the social-indexical dimensions inherent to the phonetic characteristics which children appear to adapt to (although work by Jusczyk on infants' sensitivity to familiar vs. unfamiliar voices is not unconnected to the issues which are the focus of the present paper; Jusczyk et al. [1993]). A different approach which has emerged over the last decade (building on initial contributions by Goldinger 1997, 1998; Pisoni 1997; Johnson 1997; and further developed by inter alia Pierrehumbert 2001, 2002, 2003, 2006; Hawkins 2003; Lachs et al. 2003; Wedel 2006; Hay et al. 2006a; Foulkes and Docherty 2006; Drager 2009; Munson 2010) invokes an exemplar-based phonological representation to offer an in-principle account of how social-indexical properties can be integrated within phonological acquisition more generally. In the most recent thinking around this approach, the representation of the phonological shape of words in memory takes a hybrid form, consisting in part of phonemic representations akin to those conventionally postulated in many phonological models, and in part of memory traces of tokens of those words experienced by listeners in such a way that gradient phonetic detail can be encoded alongside a range of contextual factors intrinsic to particular exemplars (e.g., details of speaker, place, context, situation, etc.). However, while this model does provide a principled basis for integrating learning of the different types of meaning contained within the speech signal, it too fails to say very much at all about how such learning occurs in children or in adults, and there remain many aspects of this model which are in need of refinement and testing (Pierrehumbert 2003, 2006; Docherty and Foulkes forthcoming; Foulkes 2010).

Understanding how children acquire knowledge of the social-indexical phonetic properties relevant to their native language poses significant methodological challenges. These arise in part from the fact that there is so much inter- and intra-speaker variability in the speech performance of children, and in part also from the difficulty in controlling for factors which are likely to be important in determining how such learning unfolds (e.g., extent and style of parental spoken input, the role of any siblings or other close family members, and differential rates of cognitive and/or motor development). Given difficulties such as these, one way to shed light on this issue is therefore to look at how *adults* acquire novel associations between phonetic variability and real world referents – the sort of situation that would occur when someone moves to an area with a different accent, or when people are exposed to innovative forms in their own speech community, or a change in the way in which such forms are indexed to social meaning (cf. Dyer 2002). This is the approach adopted in the present study. We set out to test one specific hypothesis that is predicted through an exemplar-based stochastic model of learning: exposure to novel phonetic variability that is sociolinguistically structured (i.e. such that particular phonetic forms which listeners have not

previously encountered are associated entirely or predominantly with particular social groups) should lead listeners, over time, to form associations in memory between those forms and the relevant social category.

In previous work (Foulkes et al. 2010) we tested this hypothesis using natural data. Our starting point was previous research on adult speakers in Newcastle upon Tyne (Docherty and Foulkes 1999) which demonstrated that pre-aspirated variants of /t/ were much more frequent in the speech of women than men. Thus we can infer that members of this dialect community would likely hear more pre-aspirated /t/ from female speakers than from male speakers, and thus that they may come to learn an association between frequency of pre-aspirated /t/ and speaker gender/sex. In subsequent work examining children's speech production in the same city (Foulkes et al. 2005; Docherty et al. 2006), we found abundant tokens of plain, glottalized, and pre-aspirated /t/ in the performance of both boys and girls, with the first signs of gender differentiation emerging in the performance of the older children in our cohort (those aged 3;6 from a cross-sectional cohort of 39 children aged between 2;0 and 4;0). Our question then was whether adult listeners would be led by their experience of hearing pre-aspirated /t/ mainly from female talkers to make gender judgments about children's voices in line with the particular realizations of /t/ produced in individual word tokens. Specifically, our hypothesis was that samples of children's speech containing pre-aspirated /t/ would be more likely to be judged as having been produced by girls. We tested this hypothesis with listeners from Newcastle, and as a control also with listeners from elsewhere in the UK and from the USA who we assumed would have no knowledge of the association between pre-aspirated /t/ and speaker gender. While the results were largely compatible with the hypothesis, there were a number of difficulties with the method adopted and with extending that particular design. While our results showed that Newcastle listeners did have additional sensitivity to pre-aspirated tokens as indexical of female speech, the responses from listeners were dominated by the gender-differential effects of relative loudness, f0, and rate, such that the subtle effects of different phonetic realizations were difficult to discern.

In view of the difficulties inherent in this previous investigation, in the present study we adopted a different approach in order to investigate the process through which participants learn sociophonetic variability. Our study involved a training phase in which participants were exposed to isolated word stimuli providing evidence for novel patterns of association between realizational variants and social category labels. In a subsequent test phase, we assessed the extent to which this training led participants to generate new associations between those variants and social categories. In some cases the association was implemented by a 100% correspondence between social category *x* and phonetic variant *y*,

whereas in other cases the association arose from the phonetic variant being only *predominantly* associated with relevant social category. This basic design was deployed with different implementations across four experimental conditions, described below. Overall, our method enabled us to address issues such as whether certain social-indexical properties of speech are easier to become attuned to than others, how much exposure is needed for an individual to link a particular pattern of variation to a novel social category, how categorical a phonetic variant/social category association has to be in order for it to be learned, and how consistent is cross-individual performance in this sort of learning. Our study also allowed us to evaluate the fitness for purpose of a laboratory-based experimental approach for shedding light on a learning process which is fundamentally embedded within the context of natural spoken interaction. As is pointed out in the discussion, there is no doubt that this particular approach is in need of further refinement, but while caveats need to be applied, the findings reported below do suggest that further exploration and development of this approach is indeed warranted.

2 Method

2.1 Participants

Forty-nine participants were recruited to take part as listeners in this experiment. They were all native speakers of British English in the age range 18–30, and students at either Newcastle University or the University of York in the UK. No further controls were applied in relation to the listeners' accent background or place of residence. As explained below, all of the key variables in the training and test stimuli are ones which it could reasonably be expected that all participants were familiar with (even though some of the realizational variants are not ones which they themselves would produce in their own speech performance). Participants reported normal hearing and were paid a nominal sum for their involvement in the study.

2.2 Training material

For each participant, the training material comprised 320 single-word stimuli, 160 of which encapsulated a systematic alignment of realizational variant and social group, and 160 of which were control stimuli of similar phonological shape but with no such alignment. See the Appendix for a list of the words which were

employed. The words from which the stimuli series were subsequently generated were produced by four speakers. The speakers were all phonetically-trained and thus were able to produce the variants described below with high consistency and accuracy. The stimuli materials were compiled using four repetitions of each recorded word, and the order of the stimuli was fully randomized across criterial and control categories. In addition, materials produced by the four speakers were cross-balanced across all of the criterial and control training words (i.e. the speakers were split evenly across each set of stimuli such that the stimuli contained no association between particular phonetic realizations and individual talkers' voices). The materials were recorded in a quiet recording studio and digitized using Praat at a sampling rate of 22.05 kHz. Each single word was stored in a separate file to enable subsequent preparation of the test stimuli.

Using widely available software for presenting trains of audio-visual stimuli at fixed intervals (DMDX and MS PowerPoint), the audio recording of each word stimulus was presented together with a picture file containing a graphic representation of the word in order to facilitate semantic processing. Within each DMDX/Powerpoint slide a visual indication was also given of which of two novel social groups the stimulus was associated with. Clearly, it was crucial that the social group was not one that participants had any prior experience with, as this could have pre-disposed their learning and perceptual response to the training. Thus, in order to achieve maximum neutrality, the slide for each training stimulus also contained one of the two labels "tribe1" or "tribe2" (see Figure 1 for an example of how one training stimulus was presented). Audio-visual prompts were presented at 4 second intervals and the listening material was delivered in three equal-sized blocks with a short pause between each block.

Within these general constraints, listeners were trained (and subsequently tested) in four experimental conditions as follows. In Condition 1, (in which 6 participants took part with stimuli presented via DMDX), the criterial stimuli were disyllabic words with intervocalic /t/, e.g., *butter*. The stimuli designated as associated with tribe1 always had a plain alveolar [t] realization for the medial plosive, while those associated with tribe2 always had the medial plosive realized as [ʔ]. The control stimuli were all disyllabic words with intervocalic stops other than /t/ and were presented with tribe1/tribe2 labels randomly assigned. The variants in Condition 1 were chosen as a benchmark for the subsequent experimental conditions; i.e. we reasoned that since [t]/[ʔ] variation is highly prominent in British English it ought to provide a good basis for discovering if the method worked at all as a means of capturing the learning of a new sociophonetic association. We also reasoned that this task would provide a basis for comparison with more challenging tasks to be set in the other experimental conditions. This also explained the smaller sample of listeners.

Fig. 1: One example of the presentation of one item in the training phase of the experiment (the visual presentation was simultaneously accompanied by the appropriate auditory stimulus [an utterance of the word *key*])

In Condition 2, for which there were 15 participants, with stimuli presented via DMDX, the material were identical to Condition 1 except that 80% of the tribe1 stimuli were produced with a medial [t] and the remaining 20% were produced with [ʔ]. The converse applied to the tribe2 stimuli. This was designed to provide a somewhat more challenging task than Condition 1 and to enable a test of the extent to which listeners could learn a new sociophonetic association from material where the association was not categorical.

Condition 3 investigated a different type of phonetic variable. In this case, tested on 9 participants with stimuli presented via MS PowerPoint, the criterial stimuli were all monosyllabic words corresponding to the FLEECE lexical set (Wells 1982). This particular lexical set was chosen as it enabled the testing of whether learning could be observed when the criterial sociophonetic variants were vocalic and of a particularly fine-grained nature. Thus, all of the stimuli associated with tribe1 were produced with a monophthongal [iː] vowel, while tribe2 stimuli were all produced with a slightly diphthongized variant [ɪi] (thus capturing an aspect of realizational variation which is prevalent in many current UK varieties of English; e.g., Tollfree [1999]; Williams and Kerswill [1999]). In choosing the stimuli for this condition, care was taken by the investigators to exclude excessively diphthongized tokens; as a result, the [ɪi] variants were less strongly diphthongal than the variants heard commonly in Australian English. The control words were all monosyllabic items containing monophthongs other than FLEECE.

For Condition 4 (on which 19 participants were tested with stimuli presented via MS PowerPoint) all of the criterial stimuli were monosyllabic words corresponding to the FACE lexical set (Wells 1982). In 80% of the stimuli associated with tribe1, the stimuli were produced with a monophthongal [eː] realization, with the remaining 20% produced with a diphthong [eɪ]. The converse distribution applied to the tribe2 stimuli. The control words were all monosyllabic items from lexical sets other than FACE produced with a range of monophthongs. The choice of this particular lexical set for Condition 4 was driven by the need to test whether the findings of Condition 3 were the result of the criterial variation being vocalic (as opposed to consonantal, as in Conditions 1 and 2), and whether there was a difference between vocalic variation which was sociolinguistically relatively prominent, as in Condition 4, versus that which is much less so, as in Condition 3 – see further below).

2.3 Test material

The material for the test phase of the experiment was identical to that used in the training phase but with the label indicating an association with tribe1 or tribe2 removed. Listeners were asked to respond to each stimulus indicating which social category they believed the speaker producing the stimulus belonged to (either by a left or right mouse click within DMDX or by ticking the box on a scoresheet for the tests run within MS PowerPoint). The test phase was delivered consecutively with the training, allowing for a period of time for the test materials to be set up on the lap-top, and for the investigator to explain the nature of task. For each participant, the test phase was preceded by a short set of examples to ensure that they had understood the task that they were being asked to undertake. Stimuli were presented over headphones, and the training and test phases of the experimental lasted approximately 40 minutes each. A small number of cases where participants failed to respond to a particular audio-visual stimulus were discarded.

2.4 Evaluation of the method adopted

There are a number of aspects of the design of this study which it is useful to highlight before moving on to the findings. It is reasonable to assume that in all four experimental conditions the phonetic variants that were manipulated were familiar to subjects. All are very commonly occurring features of many contemporary varieties of British English and are regularly encountered in conversational

interactions and via the media on television, films and radio. We chose not to control for the match between listeners' own varieties and the variants they were exposed to, nor did we attempt to gauge the extent to which listeners lived in areas where particular variants are more or less prevalent, reasoning that while participants may have more or less experience with the variants concerned, it is not self-evident that this would affect their ability to learn the novel sociophonetic associations embedded within the training material. Having said this, it is clear that the responses might well have been influenced by some top down features; for example, the variants in Conditions 1, 2 ([t] vs. [ʔ]), and 4 (FACE) are relatively prominent differences auditorily, whereas the variants in Condition 3 (FLEECE) are phonetically more subtle and bear less overt sociolinguistic marking in British English than word-medial /t/-glottaling or monophthong realizations of FACE. For example, /t/-glottaling is often described as a stigmatized realization of /t/ in the UK, although it has to be said that for the generation of listeners involved in this study the overt stigma appears to have diminished markedly in recent years, and may in fact not be stigmatized at all for some younger speakers (Fabricius 2002). And in similar vein, a monophthongal variant of FACE is strongly indexical of northern British English. The variability of FLEECE deployed in Condition 3 is, as noted, much less marked. Time constraints meant that there was no opportunity to check the extent to which listeners could discriminate between the monopthongal and slightly dipthongized variants of the FLEECE tokens, but to the experienced ears of the investigators all of the diphthongal variants were clearly auditorily distinct from the monophthongs. Needless to say, there was no basis on which to expect any pre-association of any of the variants with either of the social category labels.

3 Results

3.1 Condition 1

Figure 2 shows the pooled listeners' responses to the test material in Condition 1 (100% [t] aligned to tribe1 and 100% [ʔ] aligned to tribe2); the bars indicate the percentage of tokens with a medial alveolar or glottal stop identified as tribe1 or tribe2, and a comparison is shown with the responses to the control material (i.e. stimuli that did not contain medial /t/). Overall, establishing the connection between the particular realizational variant and the associated label does not appear to have been a particularly challenging task in this condition. While not every token is correctly assigned there is a strong response in the correct direction

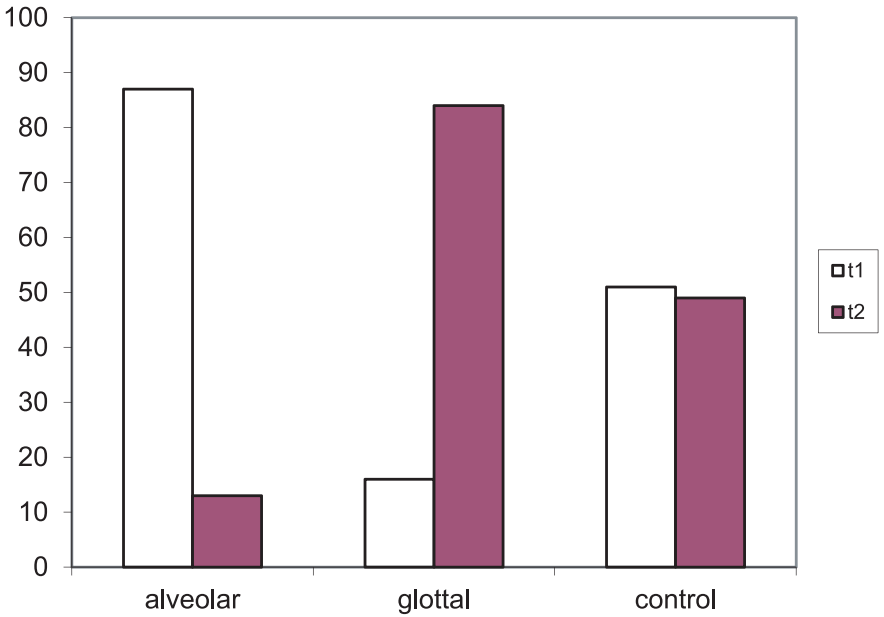

Fig. 2: Condition 1 – alveolar and glottal stimuli, 100/0% distribution (tribe1/t1) vs. 0/100% distribution (tribe2/t2); 6 test subjects

for both the alveolar and glottal stimuli, and these both differ significantly from the control material ($\chi^2(1, N = 688) = 104.61$, $p < 0.001$, and $\chi^2(1, N = 681) = 90.42$, $p < 0.001$ respectively) where listeners' responses do not differ from chance ($\chi^2(1, N = 430) = 0.01$, $p = 0.923$).

The results for each of the six participants are shown in Figure 3. It can be seen that five of the six subjects showed a high degree of learning of the association, with one subject (#1) appearing to fail to make the connection at all. It is also evident that, for each subject, performance in respect of the alveolar/tribe1 connection is mirrored by performance on the glottal/tribe2 connection. This finding suggests that, however learning is taking place, for this particular type of variation, both variants are equally effective for inducing listeners to acquire novel sociophonetic associations.

3.2 Condition 2

Figure 4 shows the pooled listeners' responses to the test material in Condition 2 (80% [t] aligned to tribe1 and 80% [ʔ] aligned to tribe2); the bars indicate the

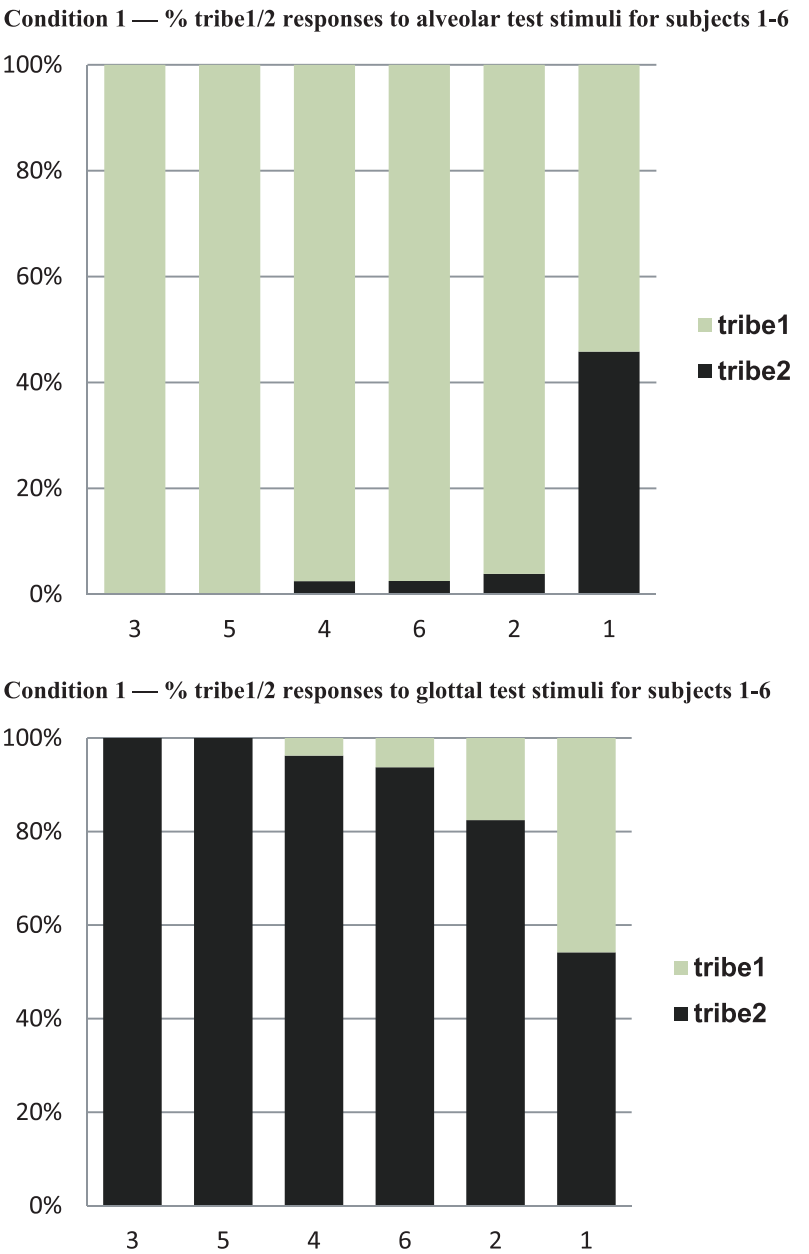

Fig. 3: The results for each of the six participants in Condition 1

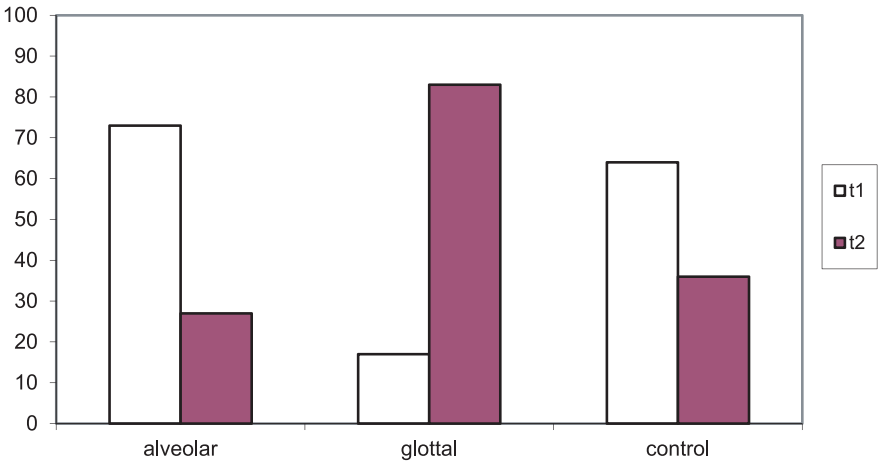

Fig. 4: Condition 2 – alveolar and glottal stimuli, 80/20% distribution (tribe1/t1) vs. 20/80% distribution (tribe2/t2); 15 test subjects

percentage of tokens with a medial alveolar or glottal stop identified as tribe1 or tribe2, and a comparison is shown to the responses to the control material. The results for this condition are somewhat less straightforward than those for Condition 1. While the overall responses seem to replicate the distribution of the criterial variants rather closely, this masks a good deal of inter-listener variability (discussed further below). There is also an unanticipated difference found in the control tokens with a skewing towards tribe1 responses resulting in an overall profile of responses which differs significantly from chance ($\chi^2(1, N = 3730) = 72.2$, $p < 0.001$). This latter finding is difficult to explain, especially as there was a complete cross-balancing of the material across the speakers who produced the material and an entirely random assignment of tribe1/2 to the control material. We note that there is also a significant difference between the responses to the alveolar tokens and control material ($\chi^2(1, N = 2821) = 24.91$, $p < 0.001$) suggesting that whatever underpins the response pattern for the control material is not the same as what is driving the response to the criterial alveolar tokens (but leaving open the question of how to account for the distribution of control responses). It is of further note that in the other conditions, with similar material and an identical task, the responses to the control material were distributed more evenly, as expected, albeit with a slight tendency for tribe1 responses to be more numerous than tribe2).

The results for each of the 15 participants are shown in Figure 5. It can be seen that there is a range of different response patterns across the group, with some

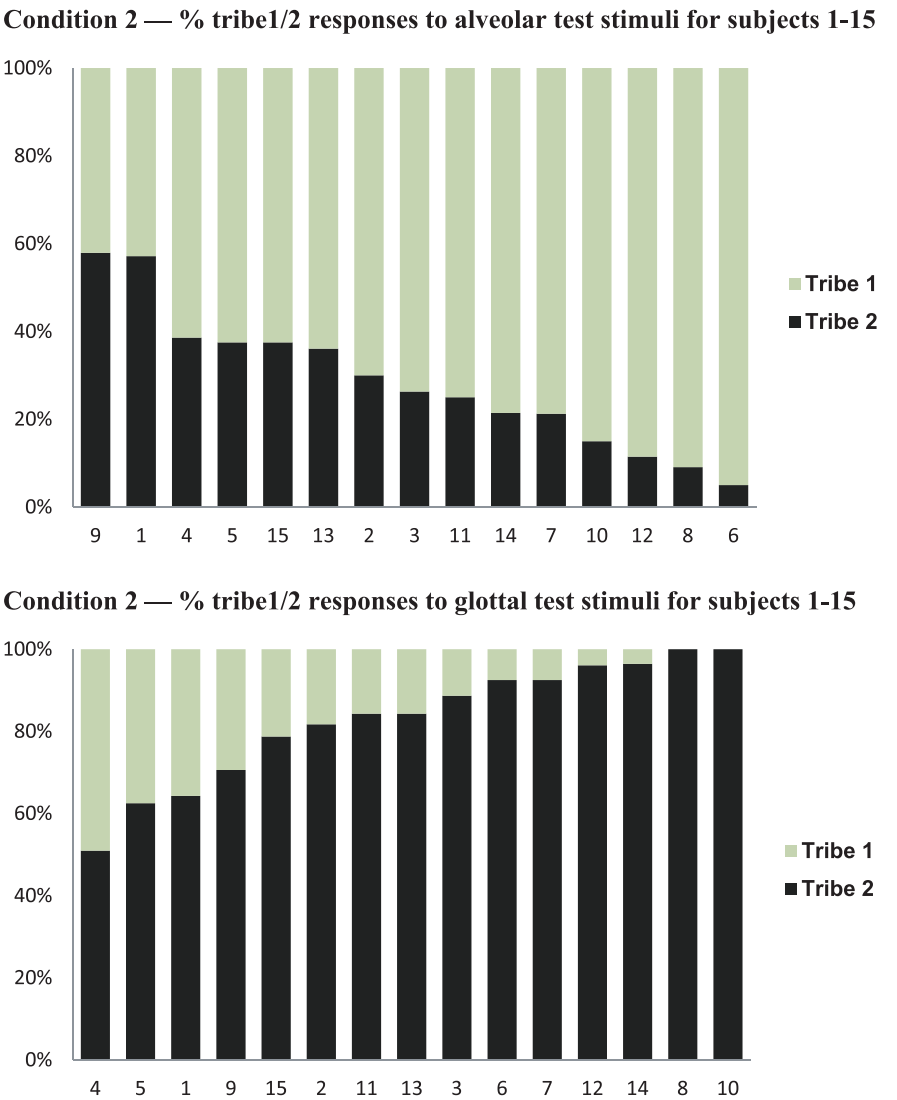

Fig. 5: The results for each of the 15 participants in Condition 2

listeners responding correctly in an almost categorical fashion, others having a pattern of responses which approximates the distribution of the criterial variants in the test material at circa 80:20, and others (a minority) appearing not to learn the tribe1/2 connection at all. (It should be borne in mind that, for individual speakers and with the sample size concerned, a response profile of >63% would

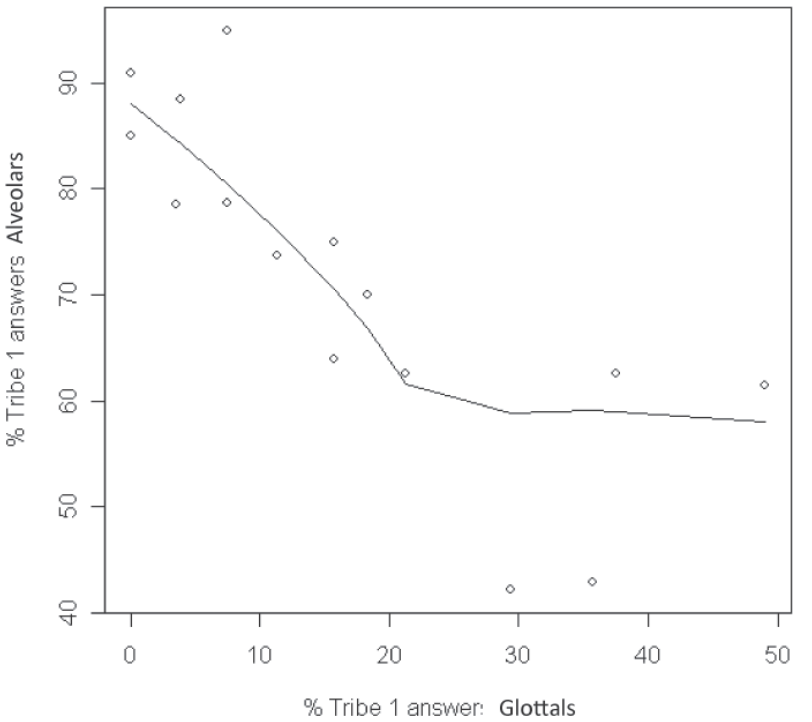

Fig. 6: Condition 2 – scatter plot of the % alveolar/tribe1 responses for each listener against the % glottal/tribe1 responses

be necessary for any of the criterial responses to differ from chance.) Figure 6 shows a scatter plot of the proportion of alveolar/tribe1 responses for each listener against the proportion of glottal/tribe1 responses. As in Condition 1, there is a strong relationship between the two for those speakers who show signs of learning the connection between the two variants and the tribe1/2 labels; specifically, speakers who are more likely to assign alveolar stimuli to tribe1 are also the ones less likely to assign the glottals to tribe1. This suggests that there could be a common underlying factor (i.e. exposure to the training material and learning of the novel patterns which is contains) governing the pooled distribution of tribe1 responses with respect to the two variants. Crucially, no such correlation exists with the control tokens: subjects more likely to assign an alveolar token to tribe1 are *not* those more likely to assign control tokens to tribe1, thus underlining the point made above that the tribe1 skew in the control material is unlikely to be due to the same factor as the tribe1 skew in the alveolar material.

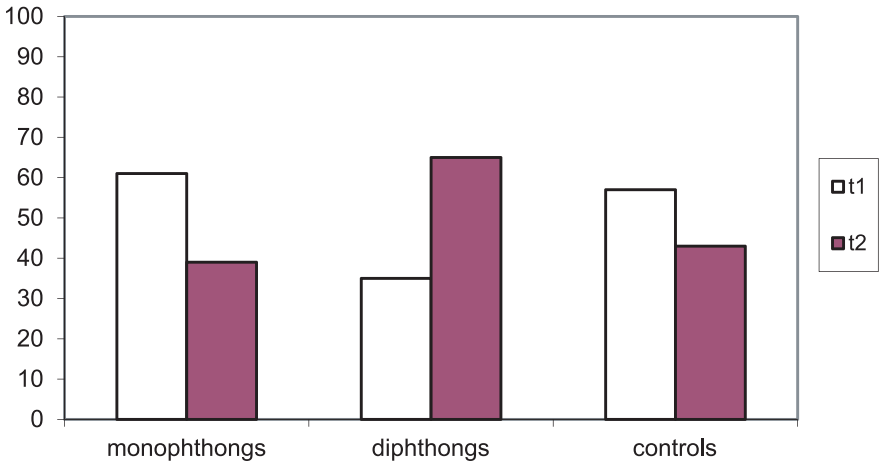

Fig. 7: Condition 3 – monophthongal and diphthongal variants of FLEECE, 100/0% distribution (tribe1/t1) vs. 0/100% distribution (tribe2/t2); 9 test subjects

3.3 Condition 3

Figure 7 shows the pooled listeners' responses to the test material in Condition 3 (100% monopthongal [iː] aligned to tribe1 and 100% diphthongal [ɪi] aligned to tribe2); the bars indicate the percentage of tokens with a monophthong or diphthong identified as tribe1 or tribe2, and a comparison is again shown to the responses to the control material. The monophthong responses did not differ significantly from the responses to the control material ($\chi^2(1, N = 1134) = 1.39$, $p = 0.238$), but the diphthong responses did yield a significant difference ($\chi^2(1, N = 1151) = 45.77$, $p < 0.001$). As in Condition 2, the control material responses differed from chance ($\chi^2(1, N = 1594) = 7.08$, $p < 0.008$) with a slight preference for tribe1 responses. Overall, these results point to a skewing in the responses to the test material in the anticipated direction, but they yield relatively weak evidence of learning. This is reinforced by the data from the nine individual participants (see Figure 8). Few subjects show any learning at all, and none of those whose results are skewed in the expected direction respond in categorical fashion (i.e. the responses do not reflect the categorical distribution of variants in the training material). Nevertheless, there are participants whose responses are in line with the associations in the training materials at a level which is significantly better than chance (participant #4, for example, achieves this for both the monophthong and diphthong tokens, participant #7 for the former, and participants #2 and #3 for the latter). In general, then, learning of the sociophonetic associations

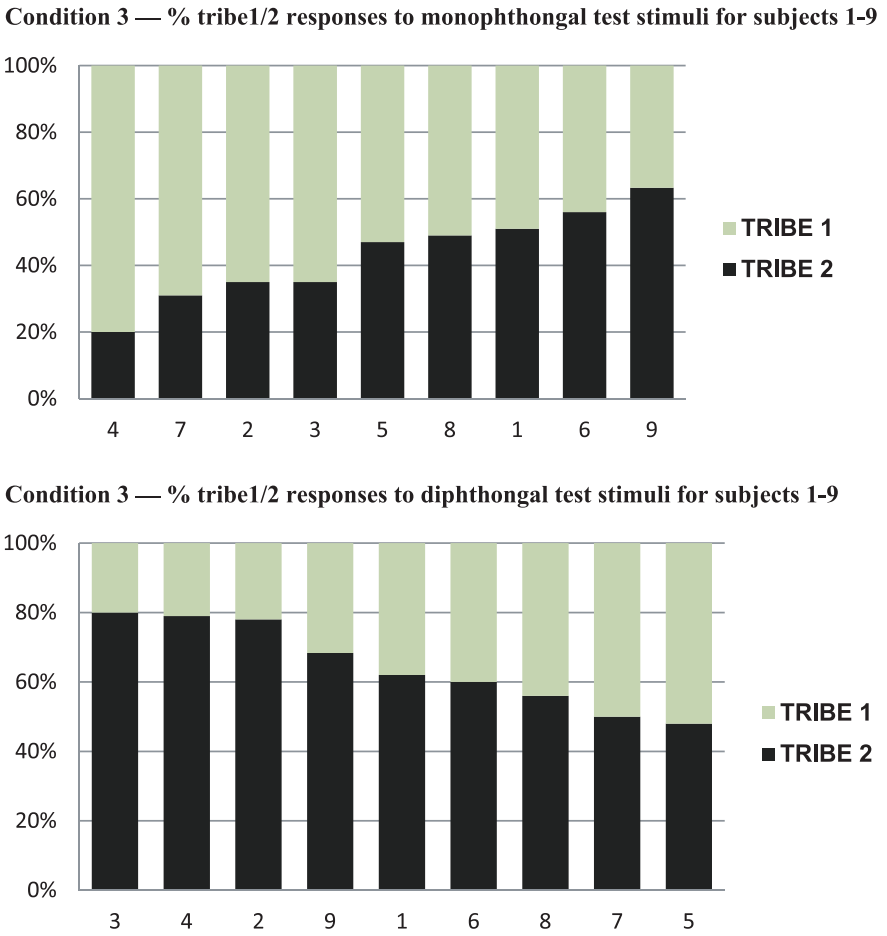

Fig. 8: The results for each of the 9 participants in Condition 3

in this case was less evident across speakers than in Conditions 1 and 2, but it is important to highlight that it is not absent altogether, with some subjects showing signs of tuning in to the patterns embedded in the training material relating to one or other of the variants concerned, or in the case of one subject to both.

3.4 Condition 4

Figure 9 shows the pooled listeners' responses to the test material in Condition 4 (80% diphthongal [eɪ] aligned to tribe1 and 80% monophthongal [e:] aligned to

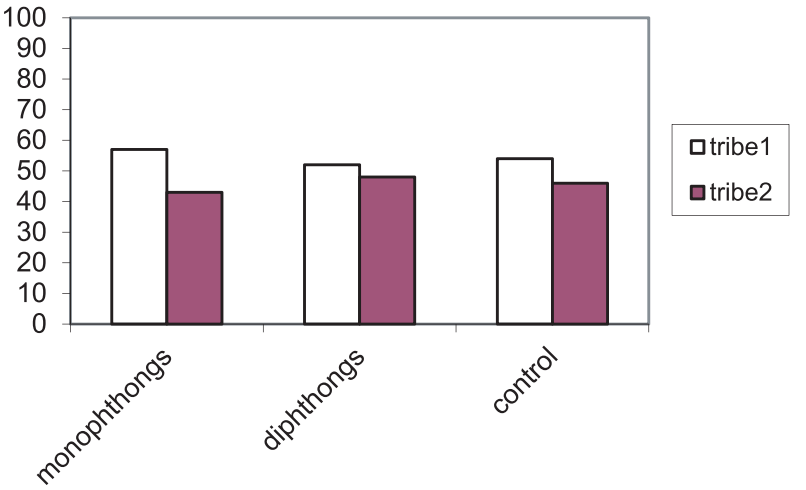

Fig. 9: Condition 4 – monophthongal and diphthongal variants of FACE, 80/20% distribution (tribe1/t1) vs. 20/80% distribution (tribe2/t2); 19 test subjects

tribe2); the bars indicate the percentage of tokens with a monophthong or diphthong identified as tribe1 or tribe2, and responses to the control material are once again shown for comparison. The overall responses for the monophthong variants differ significantly from chance ($\chi^2(1, N = 2862) = 12.67, p < 0.001$) in the direction expected if learning is taking place, but the responses to the diphthong stimuli do not ($\chi^2(1, N = 2916) = 1.1, p = 0.293$). Neither the monophthong nor the diphthong responses differ significantly from the responses given to the control material ($\chi^2(1, N = 4281) = 2.44, p = 0.119$ and $\chi^2(1, N = 4308) = 1.65, p = 0.199$ respectively). Overall, these results suggest that, in general, listeners have not been able to learn the association present in the training material. While this is found to be the case for all of the listeners when considered individually, there are some signs that some listeners may be tuning in to the pattern in the test material; thus, Figure 10 shows the responses for three subjects who do appear to associate monophthongal variants with tribe1 (although they do not do so consistently), but their responses to the diphthongal and control variants are at or around chance level.

4 Discussion

The results from Conditions 1 and 2 suggest that novel associations of phonetic variants with non-linguistic categories can be acquired fairly easily and

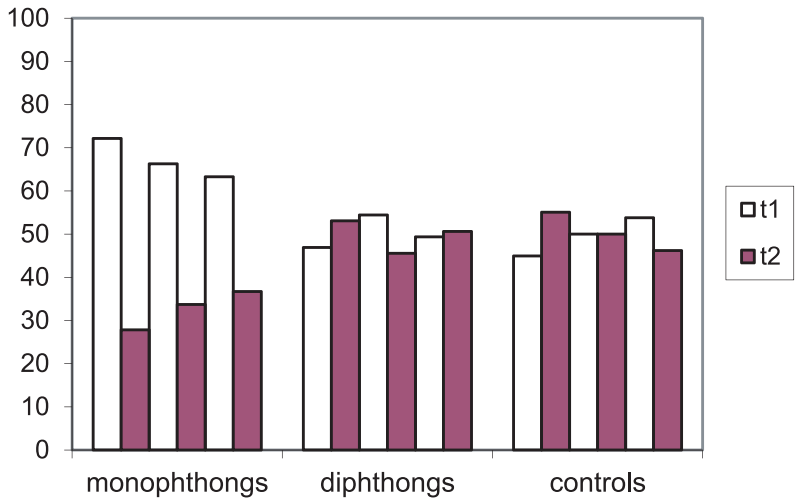

Fig. 10: Condition 4 – monophthongal and diphthongal variants of FACE, 80/20% distribution (tribe1/t1) vs. 20/80% distribution (tribe2/t2); individual results for three "partial learners"

consistently on the basis of exposure to material which embeds that association. This conclusion holds whether the association in the training is categorical (as in Condition 1) or if it takes the form of a (strong) statistical tendency (as in Condition 2). The latter condition more closely reflects the typical patterning of real sociophonetic variation, whereby particular variants are statistically *more likely* to be associated with specific individuals or social groups, rather than being categorically associated with discrete social groups. An interesting question to pursue in light of these findings is what weighting of the test variants of /t/ would be needed for an association such as that tested in Condition 2 to be learned robustly (70%/30%? 40%/60%?). This is a matter for future investigation (see also Labov et al. 2006 for discussion of evidence relating to listeners' sensitivity to the relative frequency of particular variants within ambient speech).

In Conditions 3 and 4, listeners were trained on variability that was equally systematic in terms of its encapsulation of novel sociophonetic associations, but in these cases learning was patchy, with evidence of some participants responding in line with the patterns embedded in the training material, while others apparently did not tune in to the patterns at all. The differences between the /t/ and vowel conditions (1/2 vs. 3/4) might stem from the fact that the vowel variables are less sociolinguistically marked than the [t]/[ʔ] variation (as mentioned above, the latter undoubtedly comes with a good deal of prominence in terms of the stigma or covert prestige associated with the use of [ʔ]). Although it is worth recalling

that Condition 4 tests a variable which *is* strongly indexical within the UK (of "northern-ness" as discussed above), this variable is arguably less ideologically marked than the glottal realizations of /t/.

While participants appear to have "learned" the patterns in Conditions 3/4 less effectively than those in Condition 1/2 with the same amount of training, it is notable that there was some evidence of participants tuning in to the vowel associations. This finding suggests that the fact that the variable concerned was a vowel does not rule out learning within this type of task, but learning is less effective across listeners/participants. And of course, the fact that some speakers appear to have tuned in to even the quite subtle variation manipulated within Condition 3 suggests that relative auditory salience need not be an obstacle to listeners forming new associations between phonetic variation and indexical categories. Further investigation is now required on a different set of variables in order to ascertain how the sort of learning which we have observed is influenced by the differing sociolinguistic salience of the variation concerned (for example, as defined in Labov's (1972) terms of whether a variable is a stereotype, marker or indicator), by different levels of phonetic prominence associated with particular variants, and by how these two factors interact.

While our results showed that subjects were less successful in making the connection between variants and tribe1/2 in Conditions 3 and 4, the fact that these variants are widespread in the performance of speakers of British English suggests that variability such as that tested in Conditions 3/4 must in principle also be learnable, at least in childhood. The question then arises as to what would be necessary within the context of the current experimental approach for learning to emerge in a consistently robust fashion? One possibility is whether a task with greater ecological validity would make a difference. While our decision to deploy tribe1/2 as the social referents in this task was able to provide complete neutrality in respect of those referents, it is clear that "tribe1/tribe2" are not meaningful social categories for the participants, but labels acting as a proxy for social categories. It is possible therefore that a more natural learning situation might facilitate learning of variants which are less auditorily salient and/or ideologically marked (see Wedel and Volkinburg (n.d.) for discussion of how computer simulation can be used as an alternative approach to creating and evaluating a natural learning situation). A similar point applies to the training material. The training task involving listening to a string of 320 single words might have been too laborious, and made it harder for participants to learn the sociophonetic associations for anything other than the most strongly marked variants (and in Conditions 1/2 even this was not uniformly achieved across all listeners); although informal discussions with participants following the test phase of the experimental tasks did not suggest that fatigue was in fact an issue. There is also

a question about whether the nature of the indexical category might affect the ease with which learning can take place. We have already commented above on how the need to preclude participants deploying their predispositions to interpret the training material led us to choose "tribe1/2" as a proxy social category simply as a means of testing for proof-of-concept of the basic notion that the learning of novel social-indexical phonetic associations could occur in a task such as this. However, it would be valuable in future research to probe the extent to which pervasiveness of a social category impacts on ease of learning (see Foulkes 2010 for further discussion).

Another factor which may underpin the findings is the amount of exposure to the novel variants provided in the training. Our results suggest that with a relatively short amount of training based on single word material from multiple speakers, participants were able to learn (or show signs of beginning to learn) the new sociophonetic patterns which they were exposed to. The question arises whether with further training, listeners would have started to tune in more consistently and robustly to the structured variation embedded in the Condition 3/4 materials (as shown in Figure 9 above, there is a suggestion that three listeners were perhaps starting to tune in to the association which they were trained on in Condition 4). But equally, we can ask how *little* exposure is needed before listeners can tune in to associations of this sort. For example, the training material allowed many listeners in Conditions 1/2 to tune in to the novel association of [t]/[ʔ] with tribe1/2, but the question arises what amount of training material (or in a natural setting, exposure to particular new pattern of sociophonetic variation) is needed before patterns of this sort can be identified.

This latter issue is germane to our understanding of how an exemplar-based model might work, some accounts of which give the impression that the auditory system behaves to all intents and purposes like a recording device which is continuously switched on, such that on-going experience continuously augments the exemplar store. Furthermore, the impression tends to be given that this is essentially a passive/implicit process (in fact this passivity is also built into some of the accounts of perceptual learning emerging from the speech processing literature whereby phonetic categories are skewed as a result of exposure to new tokens incorporating subtle phonetic differences; e.g., Evans and Iverson 2007). And indeed this conceptual approach finds support in studies which appear to show the influence of passive exposure on individual's performance in production/perception (e.g., Delvaux and Soquet 2007), and others showing cross-speaker entrainment in conversations (Pardo 2006) and reconfiguration of gestural timing in line with exposure to different patterns of articulatory coordination in the ambient language (Sancier and Fowler 1997). On the other hand, other investigators suggest that exemplar-based learning is not simply driven by what is con-

tained within the speech signal, but rather is mediated by a range of different pre-existing knowledge which the listener brings "top-down" to the process of interpreting the input from the auditory system (see Docherty and Foulkes forthcoming for further discussion). Thus Goldinger (2007) points out that "each stored exemplar is actually a product of perceptual input combined with prior knowledge", and perhaps more radically, Pierrehumbert (2006: 525) suggests that "[e]xemplar models are not sensitive to frequencies of ambient events per se, but rather to frequencies of memories. In between physical experience and memory lies a process of attention, recognition, and coding which is not crudely reflective of frequency".

The findings also point to other lines of inquiry that could be pursued by deploying variants of this basic train/test paradigm. For example, it is clear that our results show quite substantial cross-participant differences which certainly should be investigated in further experimental work. Of course, variation across individual participants is the norm in experimental work, although it rarely attracts comment because researchers tend to focus on group patterns. While this tendency is understandable, a focus on the individual may shed considerable light on cognitive processes underlying indexical learning (Docherty 2007). Taking account of individuals' own backgrounds and expectations may be the key to understanding what linguistic features do and do not reach the attention of listeners, and what social categories they identify. Relevant research can be found in the one field where understanding individual behavior is essential – forensic speaker (or voice) identification. Experiments with non-linguists, designed to mirror events in which witnesses may overhear the voice of a criminal, show considerable variation in performance across individuals, and have revealed a wide range of factors influencing performance (Bull and Clifford 1984, 1999).

A further line of future investigation relates to the type of phonological variables used in a learning task such as that reported here. The current experimental tasks used variables and variants from English that participants are familiar with, and it would be interesting to test them on variants with which they are less familiar (and, as discussed above, with different degrees of sociolinguistic salience and phonetic prominence). This would be particularly useful in shedding light on the role of the listener's experience in the learning task probed in this study and how this is weighted vis-à-vis the role of "bottom-up" or signal-dependent (Lindblom 1990) processing referred to above.

There are also issues to be explored about the persistence of learning. By varying the interval between training and testing we could investigate if there is any degradation over time in any learning that takes place. And by re-testing listeners after a period of time, it would be possible to ascertain the extent to which there has been any attrition in the learning of novel associations as a result of the

original training. This relates to another aspect of the exemplar model which remains to be fully-fleshed out, namely the role of memory attrition and the relationship between the level of attention which enables a new sociophonetic association to be registered in the first place (as discussed above) and the extent to which that association is subsequently reinforced.

5 Conclusion

In this study we set out to investigate how individuals can learn about the social-indexical meaning of particular patterns of phonetic realization. Our primary interest is in developing an account of such learning which can apply to both children and adults, but the difficulties of undertaking such research with children led us in the first instance to look at what can be learned from adult learners. A second key objective was to test the parameters of a particular methodology for simulating the learning process within an experimental/laboratory context.

Our findings suggest that learning of novel sociophonetic associations can be achieved as the result of a relatively short amount of exposure to training material incorporating the new association, but that the success with which learning takes place is dependent on a number of factors such as the nature of the criterial variable and individual learner variation (the precise nature of which remains to be elucidated). These results are but the first step in delving in to what is clearly quite a complex learning task, and one which we know relatively little about. With regard to our methodological objective, the results suggest that the experimental approach which we have adopted does warrant further development and has the potential to shed light on some of the key follow-on questions which arise from this study.

References

Archibald, John. (ed.) 1994. Phonological acquisition and phonological theory. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bull, Raymond & Brian Clifford. 1984. Earwitness voice recognition accuracy. In Gary Wells & Elizabeth Loftus (eds.), *Eyewitness testimony: Psychological perspectives*. New York: Cambridge University Press.

Bull, Raymond & Brian Clifford. 1999. Earwitness testimony. In Anthony Heaton-Armstrong, Eric Shepherd, Gisli Gudjonsson & David Wolchover (eds.), *Witness testimony: Psychological, investigative and evidential perspectives*. London: Blackstone.

Delvaux, Véronique & Alain Soquet. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64. 145–173.

Docherty, Gerard J. 2007. Speech in its natural habitat: accounting for social factors in phonetic variability. In Jennifer Cole & Jose-Ignacio Hualde (eds.), *Laboratory phonology 9,* 1–35. Berlin & New York: Mouton de Gruyter.

Docherty Gerard J. & Paul Foulkes. 1999. Newcastle upon Tyne and Derby: instrumental phonetics and variationist studies. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban voices: Accent studies in the British Isles,* 47–71. London: Arnold.

Docherty Gerard J. & Paul Foulkes. 2000. Speaker, speech, and knowledge of sounds. In Noel Burton-Roberts, Philip Carr & Gerard J. Docherty (eds.), *Phonological knowledge: conceptual and empirical issues,* 105–129. Oxford: Oxford University Press.

Docherty, Gerard J. & Paul Foulkes. forthcoming. An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*.

Docherty, Gerard J., Paul Foulkes, Jenny Tillotson & Dominic Watt. 2006. On the scope of phonological learning: issues arising from socially structured variation. In Louis Goldstein, Doug Whalen & Catherine Best (eds.), *Laboratory phonology 8,* 393–421. Berlin & New York: Walter de Gruyter.

Drager, Katie. 2009. *A sociophonetic ethnography of Selwyn Girls' High.* Christchurch: University of Canterbury dissertation.

Drager, Katie. 2010. Sensitivity to grammatical and sociophonetic variability in perception. *Laboratory Phonology* 1. 93–120.

Dyer, Judy. 2002. 'We all speak the same round here': Dialect levelling in a Scottish-English community. *Journal of Sociolinguistics* 6. 99–116.

Evans, Bronwen. G. & Paul Iverson. 2007. Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America* 121(6). 3814–3826.

Fabricius, Anne. 2002. Ongoing change in modern RP: Evidence for the disappearing stigma of t-glottalling. *English World-Wide* 23. 115–136.

Foulkes, Paul. 2010. Exploring social-indexical knowledge: a long past but a short history. *Laboratory Phonology* 1(1). 5–39.

Foulkes, Paul & Gerard J. Docherty. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34(4). 409–438.

Foulkes, Paul, Gerard J. Docherty & Dominic Watt. 2005. Phonological variation in child directed speech. *Language* 81. 177–206.

Foulkes, Paul, James Scobbie & Dominic Watt. 2010. Sociophonetics. In William Hardcastle & John Laver (eds.), *Handbook of phonetic sciences*, 2nd edn., 557–572. Oxford: Blackwell.

Foulkes, Paul, Gerard J. Docherty, Ghada Khattab & Malcah Yaeger-Dror. 2010. Sound judgements: Perception of indexical features in children's speech. In Denis Preston & Nancy Niedzielski (eds.), *A reader in sociophonetics,* 327–356. Berlin & New York: De Gruyter Mouton.

Goldinger, Stephen. 1997. Words and voices: perception and production in an episodic lexicon. In Keith Johnson & John Mullennix (eds.), *Talker variability in speech processing,* 33–66. San Diego, CA: Academic Press.

Goldinger, Stephen. 1998. Echoes of Echoes: an episodic theory of lexical access. *Psychological Review* 105. 251–279.

Goldinger, Stephen. 2007. A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 17th International Congress of Phonetic Sciences*, 49–54.

Hawkins, Sarah. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31. 373–405.

Hay, Jennifer, Aaron Nolan & Katie Drager. 2006a. From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review* 23. 351–379.
Hay, Jennifer, Paul Warren & Katie Drager, K. 2006b. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34. 458–84.
Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John Mullennix (eds.), *Talker variability in speech processing,* 145–165. San Diego: Academic Press.
Jusczyk, Peter. 2000. *The discovery of spoken language.* Cambridge, MA: MIT Press.
Jusczyk, Peter, Elizabeth Hohne, Ann Marie Jusczyk & Nancy Redanz. 1993. Do infants remember voices? *Journal of the Acoustical Society of America* 93. 2373.
Kager, René, Joe Pater & Wim Zonneveld. (eds.). 2004. *Fixing priorities: constraints in phonological acquisition*. Cambridge: Cambridge University Press.
Khattab, Ghada. 2007. Variation in vowel production by English-Arabic bilinguals. In Jennifer Cole & Jose Ignacio Hualde (eds.), *Laboratory Phonology 9*, 383–410. Berlin: Mouton de Gruyter.
Kraljic, Tanya, Susan Brennan & Arthur Samuel. 2008. Accommodating variation: dialects, idiolects and speech processing. *Cognition* 107. 54–81.
Labov, William. 1972. *Language in the inner city: Studies in the black English vernacular.* Philadelphia, PA: University of Pennsylvania Press.
Labov, William, Sharon Ash, Maya Ravindranath, Tracey Weldon, Maciej Baranowski & Naomi Nagy. 2006. Listeners' sensitivity to the frequency of sociolinguistic variables. *University of Pennsylvania Working Papers in Linguistics* 12. 105–129.
Lachs, Lorin, Kipp McMichael & David Pisoni. 2003. Speech perception and implicit memory: Evidence for detailed episodic encoding. In Jeffrey Bowers & Chad Marsolek (eds.), *Rethinking implicit memory,* 215–235. Oxford: Oxford University Press.
Lindblom, Björn. 1990. Explaining phonetic variation: a sketch of the H and H theory. In William Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 403–439. Amsterdam: Kluwer.
Local, John. 2003. Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics* 31. 321–339.
Maye, Jessica, Janet Werker & LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82(3). B101–B111.
Munson, Benjamin. 2010. Levels of phonological abstraction and knowledge of socially motivated speech-sound variation: a review, a proposal, and a commentary on the Papers by Clopper, Pierrehumbert, and Tamati; Drager; Foulkes; Mack; and Smith, Hall, and Munson. *Laboratory Phonology* 1(1). 157–177.
Pardo, Jennifer. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119. 2382–2393.
Pierrehumbert, Janet. 2001. Stochastic phonology. *GLOT* 5(6). 1–13.
Pierrehumbert, Janet. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7,* 101–139. Berlin: Mouton de Gruyter.
Pierrehumbert, Janet. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46. 115–154.
Pierrehumbert, Janet. 2006. The next toolkit. *Journal of Phonetics* 34. 516–530.
Pisoni, David. 1997. Some thoughts on 'normalization' in speech perception. In Keith Johnson & John Mullennix (eds.), *Talker variability in speech processing,* 9–32. San Diego: Academic Press.

Samuel, Arthur & Tanya Kraljic. 2009. Perceptual learning for speech. *Attention, Perception & Psychophysics* 71. 1207–1218.
Sancier, Michele & Carol Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25. 421–436.
Scobbie, James. 2006. Flexibility in the face of incompatible English VOT systems. In Louis Goldstein, Doug Whalen & Catherine Best (eds.), *Laboratory Phonology 8*, 367–392. Berlin & New York: Mouton de Gruyter.
Smith, Neil. 1973. *The acquisition of phonology: A case study.* Cambridge: Cambridge University Press.
Smith, Jennifer, Mercedes Durham & Liane Fortune. 2007. 'Mam, my trousers is fa'in doon!' Community, caregiver and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19. 63–99.
Stuart-Smith, Jane. 1999. Glasgow: accent and voice quality. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban voices: Accent studies in the British Isles,* 203–222. London: Arnold.
Stuart-Smith, Jane. 2007. Empirical evidence for gendered speech production: /s/ in Glaswegian. In Jennifer Cole & Jose-Ignacio Hualde (eds.), *Laboratory Phonology 9*, 65–86. Berlin: Mouton de Gruyter.
Thomas, Erik & Phillip Carter. 2006. Rhythm and African American English. *English World-Wide* 27. 331–55.
Tollfree, Laura. 1999. South-east London English: discrete versus continuous modelling of consonantal reduction. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban voices: Accent studies in the British Isles,* 163–184. London: Arnold.
Vihman, Marilyn. 1996. *Phonological Development.* Oxford: Blackwell.
Wedel, Andrew. 2006. Exemplar models, evolution and language change. *The Linguistic* Review 23. 247–274.
Wedel Andrew & Heather van Volkinburg. (n.d.) Modeling simultaneous convergence and divergence of linguistic features between differently-identifying groups in contact. University of Arizona. http://dingo.sbs.arizona.edu/~wedel/publications/PDF/Wedel_VanVolkinburgSneetches.pdf (accessed 8 October 2012)
Wells, John C. 1982. *Accents of English.* Cambridge: Cambridge University Press.
Werker, Janet & H. Henny Yeung. 2005. Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences* 9. 519–527.
Williams, Ann & Paul Kerswill. 1999. Dialect levelling: Continuity vs. change in Milton Keynes, Reading and Hull. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban voices: Accent studies in the British Isles,* 141–162. London: Arnold.

Appendix. Word lists

A Conditions 1/2 word list

Criterial
butter, city, fatty, heater, kettle, letter, mortar, party, scooter, putter.

Control
cheddar, floppy, harpist, hippy, ladder, leader, pudding, puppy, robber, rubber.

B Condition 3 word list

Criterial
cheek, cheese, feet, geek, sea, seat, seed, sheet tea, teach.

Control
chalk, cross, harp, rich, sad, short, suit, sword.

C Condition 4 word list

Criterial
bathe, braid, cake, cave, Craig, face, gate, lake, maze, tape.

Control
cross, goat, mouth, rich, ride, road, sad, suit, sword.