RESPONDING TO ACCENTS AFTER EXPERIENCING INTERACTIVE OR MEDIATED SPEECH

XXX

XXX
XXX

ABSTRACT

Very little known is about how speakers learn about and/or respond to speech experienced without the possibility for interaction. This paper reports an experiment which considers the effects of two kinds of exposure to speech (*interactive* or non-interactive *mediated*) on Scottish English speakers' responses to another accent (Southern British English), for two processing tasks, phonological awareness and speech production. Only marginal group effects are found according to exposure type. The main findings show a difference between subjects according to exposure type before exposure, and individual shifts in responses to speech according to exposure type.

Keywords: sociophonetics, interaction, media influence, accents of English, individuals

1. INTRODUCTION

Experiencing speech in interaction is assumed to lead to short-term and long-term effects for speakers and communities. Small short-term shifts in speech production towards the accent of an interlocutor, 'phonetic convergence' are noted anecdotally and evidenced empirically [11, 2]. The sociolinguistic theory of 'accommodation' explains such short-term shifting in terms of the social-psychological dynamics of interpersonal interaction [4]. Longer-term shifts which become language changes, are explained as the result of small accommodatory shifts towards another's accent over time becoming accepted at the community level [9]. A corollary of this assumption is that language change is presumed to result solely from speakers experiencing live social interaction (and this is often related to the observation that infants seem only to be able to acquire linguistic contrasts when exposed to speech during live interaction [8]).

However there has been little research on the nature of short-term shifting, and what might provoke or constrain it; the relationship of short-term shifting to long-term community change is complex and poorly understood [1]; and sometimes experiencing speech without the possibility for interaction, e.g. engaging with a favourite TV show, is related to language change, e.g. [15], and/or hypothesized and expected, e.g. [6, 5]. It is also clear that many kinds of rapid perceptual learning are possible from non-interactive recorded speech. Exposure to non-interactive recorded speech can trigger short-term shifts in speech perception and production [7]. Speakers can also show short-term shifting even towards a virtual interlocutor [14].

At the same time we might expect live interaction to be special, and to engage attention more fully through: experiencing richer visual/physical sensory as well as auditory information; sharing common communicative and collaborative goals; and through psychological and physiological alignment and entrainment processes, e.g. [12]. As a result we might expect experiencing speech in live interaction to give rise to stronger memory traces, and hence to show more observable effects on short-term responses, than speech experienced non-interactively. But in the absence of empirical evidence, this remains hypothetical. This paper reports the first experiment to attempt to investigate these issues directly. Our main research question is: are there observable differences in responding to speech according to whether it is experienced with or without the possibility for interaction?

2. OUTLINE OF THE EXPERIMENT

The experiment was designed to investigate whether responses of speakers of one accent (Scottish English, SE) to a specific aspect of another accent (Southern British English, SBE) would be affected by whether their experience of the other accent was interactive, by playing a game with them, or non-interactive mediated, by watching a video of a game being played. The linguistic feature selected was the SBE /a ɑ/ vowel

distinction in e.g. *Pam* vs *palm*. In contrast with Scottish English, which has a single vowel /a/ distributed across one very large lexical set [16], SBE has two vowels distributed across three lexical sets, TRAP, which always shows /a/, and BATH/PALM which always show /ɑ/ (for the purposes of this paper, the results for /ɑ/ are discussed without separating PALM from BATH). Pilot testing confirmed our informal observations that Scottish English speakers, who typically show only allophonic differences between [a] and [ɑ] (e.g. before /r/), find it very difficult to predict which words in SBE take /a/ and which take /ɑ/.

40 male students, aged 18-30 years, were recruited from subjects other than Linguistics, as participants for the experiment. All were born and raised in/around Glasgow, and were monolingual speakers of Glaswegian Vernacular. All were assumed to have some experience of Southern British English through differing personal opportunities for dialect contact, and/or exposure/engagement with Southern English media. The experimenter was a 24-year old female speaker of Southern British English from Sussex, with a near-RP accent.

2.1. Exposure Type

We used a word game, based on *Articulate!*, containing 24 experimental words and 26 fillers, during which each player takes turns to define a word, or guess what is defined by the other player. Either the Experimenter (E) elicited the word from the Subject (S), e.g. E: *erm, what Catholics do together when they go…*S: *Oh,* ***mass****?* Or, the Experimenter would say the word during the definition, e.g. E: *um, it's the meat of a* ***calf*** P: *er, veal?* The 20 *interactive* subjects played the game with the experimenter. The 20 *mediated* subjects each watched a film of one of the interactive participants playing the word game with the experimenter (yoked controls design).

2.2. Experimental Tasks

We used two tasks to investigate how speakers might respond to a different accent before and after exposure. The first task was a phonological awareness task, cf e.g. [6], which looked at potential shifts in metalinguistic ability to categorize sounds lexically. Subjects completed a Rhyme Judgement Test (RJT), *before* and *after* exposure, in which they had to judge whether an imagined speaker from London would pronounce a set of 48 experimental words containing /a/ and /ɑ/, as rhyming with 'cat' /a/, or 'bath' /ɑ/.

The second task was a speech production task: subjects read a passage *before* and *after* exposure. Using Praat, we extracted the 41 experimental words, measured F1 and F2 at the temporal midpoint of the vowel, and normalised the Hz data in NORM using the Bark Difference Metric. (The scales of normalised F1/F2 are inverted with respect to Hz.)

2.3. Predictions

We expected all subjects to improve at the phonological awareness task after exposure, but we expected to find more improvement after interactive than mediated exposure. In the production task we expected that all participants would show short-term shifts in vowel quality, and again more shifting after interactive than mediated exposure.

Responding to phonological/lexical categories was embedded into the phonological awareness task, since subjects had to assign /a ɑ/ to 'cat' or 'bath'. Shifts in production could manifest in several different possible ways, from overall shifts in quality, to divergent/convergent shifts according to whether words in SBE take /a/ ('cat') or /ɑ/ ('bath').

2.4. Statistical Analysis

To look at between group behaviours in the phonological awareness task, we used Signal Detection analyses, cf [10], which calculates the proportion of 'hits' (correct response as 'cat') and 'false alarms' ('cat' or 'other' given in error for 'bath'), and from which two dependent variables were derived: d-prime, Sensitivity (giving 'cat' correctly and not giving 'cat' inappropriately for 'bath') and beta, Bias (giving 'cat' regardless of whether it was correct or inappropriate). Response Consistency allowed us to look at within subject behaviour, and specifically at the proportion of responses that were consistent before and after exposure.

The dependent variables for the speech production task were: normalised F1, F2, and the difference of normalised F2-F1. Generalized linear mixed-effects modelling in R was used to test our hypotheses. The independent variables were Exposure Type (interactive/mediated) and

Test/Passage (before or after exposure). We also ran Pearson correlations between the task results.

3. RESULTS

3.1. Phonological awareness

The results for Sensitivity and Bias are given in Table 1. No significant difference for d-prime was found according to Exposure Type or Test, suggesting no overall improvement in sensitivity to the SBE lexical categorization of /a ɑ/.

Beta scores were significantly affected by Exposure Type ($F(1,78)=6.50$, $p=0.013$) showing that mediated subjects were more biased to respond 'cat', even before exposure. There was no significant interaction of Exposure Type with Test, but the difference between the groups tends to be larger after exposure, particularly the mediated subjects, who increased their bias to respond 'cat' after watching the video.

Table 1: Mean and standard deviations of d-prime and beta by Exposure Type and Rhyme Judgement Test

Measure	**Test**	**Exposure Type**	
		Interactive	*Mediated*
d-prime (Sensitivity)	***before***	1.282 (*1.118*)	1.174 (*1.169*)
	after	1.238 (*1.323*)	1.340 (*1.487*)
beta (Bias)	***before***	0.058 (*0.350*)	0.174 (*0.237*)
	after	0.070 (*0.342*)	0.348 (*0.492*)

There was a marginally significant effect of Exposure Type (χ^2 (1)=3.527, p = 0.0603) for Response Consistency, such that interactive subjects are more consistent (69.8%) than mediated ones (66%). Being less consistent can mean improving or worsening after exposure. Inspection of the data suggests that interactive participants tended to improve at assigning /ɑ/ to 'bath', whilst mediated subjects were more likely to improve at assigning /a/ to 'cat', but are genuinely inconsistent for /ɑ/. The mediated subjects' tendency to improve at categorizing /a/ is in line with their increased bias to respond 'cat'.

3.2. Speech production

Results for normalized F1 and F2 are shown in Table 2. nF1 showed a strongly significant effect of Exposure Type (χ^2 (1)=64.501, p = 0.0001), with all subjects showing a higher Bark value, reflecting closer vowels after exposure.

Table 2: Mean and standard deviations for normalized F1 and F2 by Exposure Type and Passage.

	Passage	**Exposure Type**			
		Interactive		*Mediated*	
nF1	***before***	8.31	(0.89)	8.29	(1.02)
	after	8.49	(0.97)	8.54	(1.00)
nF2	***before***	4.13	(0.83)	4.21	(1.03)
	after	4.07	(0.85)	4.21	(1.05)

We also calculated the difference of nF2-nF1, which indicates shifting along both dimensions. This measure showed a significant effect of Passage (χ^2 (1) = 48.035, p < 0.0001**),** and a marginally significant interaction of Exposure Type with Passage (χ^2 (2) = 5.5741, p = 0.0616); see Table 3. There is a shift to more negative values, reflecting closer and fronter vowels after exposure, but the shift is more pronounced in mediated subjects.

Table 3: Mean and standard deviations of nF2-nF1 by Exposure Type and Passage

Passage	**Exposure Type**			
	interactive		*mediated*	
before	-4.24	(0.99)	-4.08	(1.03)
after	-4.36	(1.06)	-4.33	(0.99)

3.3. Correlations between task results

Inspection of the speech production results for individual subjects confirmed our expectation that some speakers might respond differently according to lexical category. We therefore calculated all vowel measures separately for each individual for 'cat'/'bath', and also the difference between them.

Increased sensitivity in the RJT after exposure was significantly correlated with a less negative value of nF2-nF1 (a more open/retracted vowel) for 'cat' words before exposure. Also, the more that subjects showed a change in the separation of 'cat' and 'bath' in nF2-nF1 before and after exposure, (a) the less bias they showed to respond 'cat' in the second RJT, and (b) the bias to respond 'cat' decreased after exposure. The latter correlation shows an effect of Exposure Type, since this only holds for interactive subjects, but not for mediated ones.

4. DISCUSSION

In phonological awareness and short-term shifts in speech production we found: a persistent difference between interactive and mediated

participants *before* exposure; two marginal, group effects relating to experiencing the experimenter's speech differently; and a difference in response according to Exposure Type at the level of the individual.

The difference between interactive and mediated subjects before the exposure task may reflect an initial effect of differences in data collection. Hay et al [7] found significant differences in subject responses based only on a short recorded instruction in two different accents. Here the interactive subjects could see the camera, and knew that they were going to take part in a game with the experimenter. Did they also develop a motivation for a rapport to enable this task to succeed? The experimenter thinks that she did need to engage with the interactive subjects because she needed to play the game with them. Staum Casasanto [14] found more accommodation in speech rate by those speakers who judged themselves to be more similar to the Virtual Interlocutor. Less initial rapport between fieldworker and the mediated subjects might also relate to their showing an increased tendency to generalize their own vowel/lexical category even in the first RJT. It is also possible that the interactive group, who responded first to the recruitment for subjects, reflected differences in personality and/or individual differences in readiness to take part in the experiment.

We had expected more differences according to how our subjects experienced the experimenter's speech, especially for the interactive subjects. It is possible that the 12-20 minute exposure period was too short, though other studies have observed learning or adaptation after similar lengths to ours e.g. [3]. We have already noted that all subjects had some interaction with the experimenter from the outset, and that this may have had a fundamental immediate effect on responses. Nevertheless the subtle results according to exposure are as we might expect: for example, the mediated subjects are less consistent by improving at generalizing their own /a/ category, and being inconsistent with /ɑ/, the new category, cf [10], whereas interactive subjects are more consistent because they are better at assigning 'new' /ɑ/.

But the one result of Exposure Type which was not marginal, namely the correlation between shift in bias and shift in separation of vowel quality, may also highlight the additional essential factor for understanding responses linked to differential exposure to speech – individual variation in behaviour. It was only when we considered the results from the point of view not of groups but of individuals, that a clearer difference emerged. Different subjects showed a range of different responses in phonological awareness and production; the variability in these responses seems to relate better to exposure than assuming group shifts in one direction or another. While individual differences are well recognized in speech production and perception, our finding may also relate to the observations and predictions of diffusion research, which anticipates different individual responses to innovations according to channel (interpersonal or mass media) [13].

5. REFERENCES

[1] Auer, P. and Hinskens, F. 2005. The role of interpersonal accommodation in a theory of language change. In: Auer, P., Hinskens, F., and Kerswill, P. (eds.), *Dialect Change*. Cambridge: CUP, 335–357.

[2] Babel, M. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society.* 39, 437-56.

[3] Clarke, C. and Garrett. 2004. Rapid adaptation to foreign-accented English. *JASA*, 116, 3647-3658.

[4] Coupland, N. 1980. Style-shifting in a Cardiff work-setting. *Language in Society.* 9, 145-204.

[5] Evans, B. and Iverson, P. 2004. Vowel normalization for accent. *JASA*, 115, 352-61.

[6] Hay, J., Warren, P., and Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics.* 34, 458-484

[7] Hay, J., Drager, K. and Warren, P. 2006. Cross-dialectal exemplar priming. Poster, *LabPhon 10*, Paris

[8] Kuhl, P., Tsao, F.-M., and Liu, H.-M. 2003. Foreign-language experience in infancy. *Proc Nat Academy of Sciences*, 100, 9096-9101.

[9] Labov, W. 2001. *Principles of Linguistic Change: II Social Factors*, Oxford: Blackwell.

[10] Otake, T. and Cutler, A. 1999. Perception of suprasegmental structure in a non-native dialect. *Journal of Phonetics*, 27, 229-53.

[11] Pardo, J. 2006. On phonetic convergence during conversational interaction. *JASA*, 119, 2382-2393.

[12] Pickering, M. and Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Science.s* 27, 169-225.

[13] Rogers, E. (2003), *Diffusion of innovations*, Fifth edition, New York: Free Press.

[14] Staum Casasanto, L, Jasmin, K., Casasanto, D. 2010. Virtual accommodating: Speech rate accommodation to a virtual interlocutor. In: Ohlsson, S., Catrambone, R. (eds), *Proc. 32nd Ann Conf Cog Sci Soc.* Austin, 127-32.

[15] Stuart-Smith, J. and Timmins, C. 2009. The role of the individual in language change. In Llamas, C. and Watt, D. (eds), *Language and Identity*. Edinburgh: EUP, 39-54.

[16] Wells, J. 1982. *Accents of English*. Cambridge: Cambridge University Press.