THE EFFECT OF PHONETIC DETAIL ON PERCEIVED SPEAKER AGE AND SOCIAL CLASS

Abby Walker

University of Canterbury
ajw129@student.canterbury.ac.nz

ABSTRACT

It is well documented that the phonetic realization of a sociolinguistic variable can systematically differ according to the social attributes of a speaker, such as their age, class or ethnicity. What is less understood is the degree to which listeners routinely exploit this systematicity in order to make social judgments about speakers. This study uses speech synthesis to examine whether subtle changes to the phonetic realization of sociolinguistic variables in a sentence can alter the perceived age and social class of a speaker[1].

Keywords: sociophonetics, speech perception

1. INTRODUCTION AND BACKGROUND

Listeners are fairly accurate in making judgments about a speaker's identity, whether it be identifying a particular person or simply identifying that the speaker belongs to a certain group in society. This sort of identification is generally considered to be based on voice quality, but recent work has suggested that listeners are also using phonetic cues.

In a study by Remez et al. [7], participants were asked to identify people they knew from sine-wave signals where voice quality was lost but phonetic information retained. Their positive results led the researchers to conclude that listeners could recognise speakers "using only information about the linguistically governed articulation" (p.663).

Work by Purnell et al. [6] suggests that speakers can also use phonetic information to identify social attributes of an unknown speaker's identity. Participants in their experiment were able to identify the ethnicity of different speakers based on the word *hello* alone. A post-hoc analysis of the stimuli showed the most significant difference between how the different ethnicities had produced

the greeting was found in the value of the F2 in DRESS[2]. Similarly, participants in an experiment by Gordon [3] were almost in full agreement in labelling a broader accent as belonging to a lower class speaker. The vehicle for the accent was a letter filled with phonemes that showed sociophonetic variation in New Zealand English (NZE), and which were assumedly being used by listeners in making their judgments.

In a more controlled study, Drager [2] played participants in New Zealand words containing DRESS or TRAP, which vary in their pronunciation in relation to age. As they listened to a word they were shown a photo of a person of a certain age, whom they were told was the speaker. After hearing the word they were asked to match the nucleus vowel with one from a continuum of synthesised vowels. The age of the people in the photographs significantly influenced which variant participants believed they had heard.

This experiment will effectively test the converse. I hypothesize that different realizations of a variable can influence the perceived social attributes of a speaker. By synthetically altering phonemes that exhibit social variation, I predict that the perceived age and social class of a speaker can change, and the direction in which they change will mirror the social distribution of the particular realization in the dialect, namely NZE. If the manipulation works, it will show that listeners can use phonetic detail in identifying speaker attributes in relatively natural speech.

2. METHODOLOGY

This experiment was designed as a precursor to a second experiment, which will test the effect of phonetic detail on the grammaticality ratings of non-standard morphosyntactic constructions in NZE. Therefore all the sentences in this experiment contain either *come* or *done* in preterite constructions or *have-got* in constructions denoting

possession. All of these constructions have a production bias towards younger speakers, and the preterites are heavily associated with working-class speech. This means that the experiment has a built in bias towards the speakers being perceived as younger, non-professionals.

Two variables that exhibit sociophonetic variation in NZE were selected for manipulation: intrusive-/r/ after MOUTH and phrase-final /t/. Hay and Maclagan [5] have shown that intrusive-/r/ following MOUTH occurs more frequently in the speech of younger, working class speakers. Moreover, when individuals from lower social classes produce intrusive /r/, they tend to produce it with a lower F3 than when speakers from higher social classes produce it. Docherty et al. [1] show that phrase-final /t/ is most likely to be produced without release by all speakers, but if it is released it will most likely be by middle-class females. They also showed that many of the released tokens also contained glottalisation.

There were seventy unique sentences in total, and examples of some are given in 1-4. Thirty contained the *come* construction, thirty contained *done* and ten contained *have-got*. Ten of each of the preterite sentences contained an environment for phrase final /t/ (1), ten contained environments where intrusive /r/ could occur after MOUTH (2) and ten were designed to contain both variables (3). All ten of the *have-got* sentences had *got* phrase finally (4), so that the /t/ in this word could be manipulated.

1. I called her name, and she come down from the *hut*.
2. Sam done it yesterday, but *now it's* broken again.
3. He come over yesterday and *now it's* all sorted *out*.
4. It's the only home Jane's *got*.

Five females with theatrical experience were recorded reading all of the sentences using a head-mounted mic connected to a laptop via a USBPre into Sony SoundForge. They were all native speakers of NZE and ranged from 19 to 27 years of age. They were told to make the sentences sound as natural and conversational as possible. As none of the females natively produced intrusive-/r/ after MOUTH while reading the sentences, they had to be

trained how to do this and were also coached into consistently producing /t/ with release phrase finally.

For each different sentence type (e.g., *come* with just phrase final /t/), two unique sentences were taken from the recordings of each female. Sentences were chosen where the speaker had produced a natural sounding intrusive /r/ and/or a phrase-final /t/ which exhibited glottalisation followed by a release.

Using a Praat script designed by Paul Warren, the F3 of all occurrences of intrusive /r/ were both raised and lowered, so that two new files were created. The lowered F3 resulted in a more /r/-like, innovative realisation whilst the raised F3 created a less-/r/ like, more conservative token, and the two values were at least 500HZ apart. For the /t/s, the release would simply be cut from the end of the recording, so that the original version was used as the conservative form and the cut one, with only glottalisation to indicate the /t/, was the innovative realisation.

With both an innovative and conservative version of each sentence, there were now 140 utterances for participants to listen to. These were ordered so that both the speakers and sentence types rotated in a regularised fashion and variables alternated from conservative to innovative. To minimise the effect that order might have on the experiment, two trials were conducted that were identical except that all conservative and innovative tokens were inverted.

Seventeen people in their 20s participated in the experiment, which was conducted over headphones on a laptop running Media Lab experimental software. For each sentence they heard, participants were asked to rate the age and social class of the speaker. Age was given in five year blocks going from 15 to 40+, making a six point scale (1=youngest, 6=oldest). Social class was given on a four point scale with 1 indicating that the person was of a lower social class and 4 indicating that they were of a higher one. Participants were told that the speakers were actors reading lines from a script. It was stressed to participants that they should therefore base their judgments on how the person sounded and not by what they said, and also that they needed to be as quick as possible.

3. RESULTS

The raw data[3] was analysed using linear regression models in R.2.0.0, using Harrell's Design Library. Two linear regression models were fit by hand – one modelling perceived age, and one modelling perceived social class. Both were significantly affected by whether the participant had just heard a more conservative or innovative form of the variable. The conservative variant significantly increased the perceived age of the speaker (Fig 1 – left panel) in a model which also included who the speaker was, the order of the sentences, the construction (*come*, *done* or *got*), and the trial. Tables 1 and 2 show the coefficients and anova statistics for the model.

Table 1: Model Coefficients for the age ratings of speakers over all variables.

	Value	Pr(>|t|)
Intercept	2.5697816	0
realisation=innovative	-0.1401688	<0.01
speaker=E	-0.3334207	<0.001
speaker=K	1.8826945	0
speaker=L	0.3826948	<0.001
speaker=S	0.2615589	<0.001
order	0.0008626	<0.01
trial	-0.2410546	0.237
construction=done	0.0568065	<0.001
construction=got	0.2850571	0

Table 2: Anova model for the age ratings of speakers over all variables

Factor	d.f.	F	P
realisation (i/c)	1	10.03	<0.002
speaker	4	286.23	<.0001
order	1	9.26	<0.003
trial	1	29.54	<.0001
construction	2	8.85	<.0001
REGRESSION	9	144.27	<.0001
ERROR	2132		

Perceived social class also increased with the conservative variant (Tables 3&4). This model included speaker, order and the construction type. There was also an interaction between the form and the type of variable, so that the class rating increased with the conservative variable when the type of variable was /t/, or was both /t/ and /r/, but not when it was /r/ alone, as shown in Fig 1 (right panel). Note that /t/ is rated better overall, a fact that proved statistically relevant in a Wilcoxon test when the variable was paired with /r/ (p=<0.001).

Table 3: Model Coefficients for the class ratings of speakers over all variables.

	Value	Pr(>|t|)
Intercept	2.519161	0
realisation=innovative	-0.176418	0.01
variable=r	-0.150000	<0.02
variable=t	0.079464	0.157
speaker=E	-0.102300	0.04
speaker=K	0.477214	0
speaker=L	0.228423	<0.001
speaker=S	-0.038711	0.436
Order	0.000448	0.023
trial	-0.118682	<0.001
construction=done	-0.185500	<0.001
construction=got	0.324897	<0.001
realisation=i*variable=r	0.241176	<0.01
realisation=i*variable=t	0.105777	0.163

Table 4: Anova model for the class ratings of speakers over all variables.

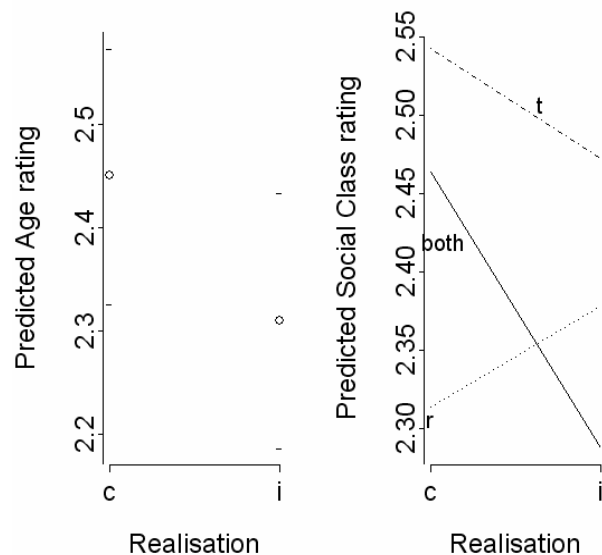| Factor	d.f.	F	P
realisation	3	4.13	<0.01
All Interactions	2	4.24	0.02
variable	4	6.42	<0.001
All Interactions	2	4.24	<0.02
speaker	4	46.26	<.0001
order	1	5.2	0.02
trial	1	14.23	<0.001
construction	2	48.64	<.0001
realisation*variable	2	4.24	<0.02
REGRESSION	12	27.87	<.0001
ERROR	2336		

Figure 1: The predicted age and social class ratings of speakers for conservative (c) and innovative (i) variants. The left panel shows age ratings. In the right panel we see the interaction between the realisation and the type of variable for class ratings.

If we compare the mean responses for the different realizations of each variable, as shown in Table 5, we see that the conservative form consistently receives a higher age and social class rating, excepting alone the class rating of intrusive-/r/, which reflects the interaction seen in Fig 1. In Wilcoxon tests, the difference between the two forms was significant for /t/ when age was being rated ($p=<0.02$) and for the sentences with both variables for both class and age ($p=<0.05$).

Table 5: Mean age and class responses to conservative (C) and innovative (I) realizations.

	r-age	r-class	t-age	t-class	both-age	both-class
C	2.70	2.27	2.96	2.64	2.71	2.42
I	2.65	2.34	2.78	2.57	2.54	2.24

4. DISCUSSION

Both the perceived age and social class of a speaker were successfully manipulated by altering the phonetic information of a single variant within a sentence. This happened even when the sentences contained constructions fairly stigmatized in NZE. When participants heard a more conservative realization, they would rate the speaker as older and from a higher social class, and conversely for when the realization was innovative. The fact that the direction of the results mirrors the documented distribution of these variables adds to mounting evidence that encountered realizations of phonemes are stored away complete with speaker information [see 4 for a review].

The manipulation of intrusive-/r/ was less successful than that of phrase-final /t/ or of both variables. There are three possible reasons why this may be. Firstly, there were more instances of /t/ than /r/ because of the *got* sentences, therefore there may simply not be enough data for /r/ to produce significant results. Secondly, phrase-final /t/ may be audibly more salient to listeners because of the release, and also because of its phrase final position. Thirdly, all of the original sentences with intrusive-/r/ environments had a salient /r/. It is quite likely that despite the synthesis, traces of the original /r/ remained in the conservative tokens, and that any evidence of an /r/ is subject to a floor effect, such that it will automatically receive a lower class rating. This last hypothesis receives support from the Wilcoxan test showing that sentences in the intrusive-/r/ manipulation received lower overall social class ratings than sentences in the phrase-final /t/ manipulation.

5. CONCLUSION

The results of this study show that listeners can and do use specific phonetic detail in judging aspects of a speaker's social identity. Furthermore, this detail can be embedded in a full sentence filled with various other features and still have an effect. This is particularly impressive when the sentences already contain constructions that carry strong social information, and indicates that subtly altering phonetic detail may also be able to alter the perception of such constructions, which will be the focus of my future work.

REFERENCES

[1] Docherty, G., Hay, J. and Walker, A. 2006. Sociophonetic patterning of phrase-final /t/ in New Zealand English. In P. Warren & C. Watson (eds*.), Proc 11th Australasian International Conference on Speech Science & Technology*, Auckland.

[2] Drager, K. 2005. The Influence of Social Characteristics on Speech Perception. MA Thesis. University of Canterbury.

[3] Gordon, E. 1997. Sex, speech and stereotypes: why women use prestige forms more than men. *Language in Society* 26, 47-63.

[4] Hay, J, and Drager, K. 2007 (forthcoming) Sociophonetics. Manuscript submitted to the *Annual Review of Anthropology*.

[5] Hay, J. and Maclagan, M. (forthcoming) Social and phonetic conditioners on the frequency and degree of intrusive /r/ in New Zealand English. To appear in Dennis Preston and Nancy Niedzielski (eds) *Methods in Sociophonetics*.

[6] Purnell, T., Idsardi, W. and Baugh J. 1999. Perceptual and phonetic experiments in American English dialect identification. *Journal of Language and Social Psychology* 18, 10–30.

[7] Remez, R.E., Fellowes, J. M. and Rubin, P.E. 1997. Talker Identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance 23*, 651-666.

[8] Wells, J.C. 1982. *Accents of English (vol. 1)*. Cambridge: Cambridge University Press.

[2] I use Well's [8] lexical set instead of phonemic transcription to refer to the vowel inside of the word in small caps.

[3] An error led to the conservative and innovative versions of 7 sentences being completely removed from the analysis.