ETHNIC GROUP ATTRIBUTION: IS OUR RELIABILITY CONSTRAINED BY TIME SPENT WITH OTHERS?

Richard Todd

Speech and Hearing Research Group, Dept. of Computer Science, University of Sheffield, UK
R.Todd@dcs.shef.ac.uk

ABSTRACT

The advancement of socio-phonetics has helped the steady growth of research relating to characterising non-native or minority group speech over the years. While some studies focus on differences in the *production* of sounds by native and non-native speakers, others have paid more attention to the *perception* or *identification* of such varieties by listeners of various kinds.

This work examines the idea of whether routine exposure to other speaker-types, from a standpoint of differing ethnicities, leads to an increased capability for such correct identification perceptually, i.e., as in Ethnic Group Attribution (EGA) tasks. It achieves this by capitalising on relevant social data from a large-scale investigation that, perhaps for the sake of avoiding controversy, others have apparently left somewhat unattended.

Overall, findings of the examination corroborate those previously suggesting listener performance generally reaches a plateau and does not continually heighten by repeated (task) exposure. It furthermore uncovers that EGA potential is not so predictable, in terms of outcome, when listeners' interactions are considered using either of two distinct socialisation measures.

Keywords: Ethnic Group Attribution (EGA), speaker-ethnicity, perception and identification, accuracy, social exposure

1. INTRODUCTION

1.1. General identifiability

It is well-understood that factors such as locality, social status, personal affiliations, age, and self-image may influence speech *production* or listener *perception* to various extents [3, 4, 12, 13, 15]. The long-standing interest in listeners' capacity to identify or otherwise characterise speakers in terms of their ethnicity or nativeness is evidenced by the various early investigations (e.g., Arslan and Hansen [2]; Flege [6]; Sebba [11]; Todd [14]; Walton and Orlikoff [18]). It becomes clear from such work that ascribing an identity to a voice in this way — i.e., via Ethnic Group Attribution (EGA) — is a non-trivial task. Its performance requires a listener to be additionally mindful of discriminating a *speaker-type*, or *-group*, rather than only individualised features, as the need accordingly arises, when utterance ambiguities or limitations are apparent.

1.2. Identifiability and limited context

Foulkes and Barron [8] add to the above by acknowledging how matters related to common group-belonging can influence identifiability. Meanwhile, the empirical findings of Purnell, Isardi and Baugh [10], then Todd [16] variously illustrate that limited utterance context and/or duration does not always negatively impact attribution accuracy if listeners, for whatever motive, consciously focus on lower-level content to facilitate identification (like group-specific pronunciations). This finding was most particularly apparent for cases involving the same phrases or focus-words/segments produced by multiple speakers, as Andics, McQueen and Tourenout [1] go on to confirm in another identification task.

1.3. Identifiability and prosodic cues

A later study also reveals that some listeners tend to perform EGA tasks better than others even when explicitly known to rely on the same cue, such as intonation [17], which reinforces other empirical evidence of speaker-group differences being discernible at the prosodic level (cf. Foreman [7]; Lass, *et al.* [9]).

Overall, the foregoing research goes some way to collectively demonstrate that the auditory identification of non-native speech and speaker-types may be reliably performed, even if not invariably so.

2. THE STUDY

Little attention has been paid to clarifying whether listeners' potential, in terms of EGA accuracy, is predictably diminished or increased in line any with repeated exposure to ethnically-differentiated speakers they encounter.

This work therefore aims to bring some clarity and attention to an under-researched area, which must inevitably enter and influence both socio- and forensic-phonetic assumptions or knowledgebases with respect speaker- or listener-nativeness.

3. METHOD

3.1. Research participants

Unlike the works of Figueiredo [5], with 28 participants; or Sebba [11], where $n = 34$, data from a vastly increased number of participants were gathered for this investigation ($n = 120$).

There was a reasonable balance of respondents, gender-wise (Females = 68; Males = 52). All of them were adults living in the UK, being aged over 18 years old (overall mean age = 34.4 years old).

3.2. Increased scope

In addition to a larger number of participants, EGA capacity was investigated in a more diversified manner than with the earlier work of [14], for example. While speakers would be of the participants' locale, the non-native ethnicities were South Asian, Caribbean, and East Asian (total $n = 45$) being the UK's largest visually distinct minority groups. Unlike in [14] no other limited minority group demarcation was applied or used. Participants had to consider such speech presented in randomised order; being lexically identical; with typical utterances durations being < 5 seconds.

Participants' respective potential was graded along a 5-point scale representing the overall range of identification/attribution accuracy. A rating of '1' was assigned to those able to claim the most reliable performance EGA-wise (75-100%); '2' = 50-75% accuracy; '3' = 25-50%; while ratings '4' and '5' were respectively for those flagged with the least potential (10-25%, then lower still, at the 0-10% accuracy range).

The other data, which crucially bring added meaning to the performance-related aspect of this investigation, were obtained on the following rationale. Namely, an adult gains her/his maximal exposure to ethnic counterparts via from either (a) work-like obligations; and/or (b) within self-determined, free time/choice. Further to considering minority group speech attribution-wise, respondents were required to reveal to what degree they spent time (per typical week) with such people of a different ethnic group from their own. The methods of exposure were measured in two ways, accordingly: $Tdiff_1$, being for (a) work-based; and $Tdiff_2$, being for (b) during free time/choice. In each case, 5-point scales were employed, as with the EGA accuracy ratings previously. Doing this, facilitated grouping participants by exposure levels. It moreover, finally helped determine whether there was (partial) evidence to support or reject a hypothesis of 'increased exposure $\rightarrow$ increased EGA performance', or vice-versa, for either social context.

4. RESULTS

4.1. Work-based exposure to out-groups

With respect the variability of typical inter-ethnic exposure within the population, examinations revealed that 27.5% of individuals shared the same maximal $Tdiff_1$ level of '1' ($n = 33$, at 75-100% weekly).

If taken in isolation, the largest group of participants finding ethnic inter-mixing to be a rather permanent feature of their work settings, would have provided a misleading illustration of what was to come, exposure-wise. Rather than the respondent data showing a gradual decrease in $Tdiff_1$ levels, a U-shaped distribution became evident, as Figure 1, below, shows.

Figure 1: Inter-ethnic exposure was surprisingly polarised in the work-based setting.
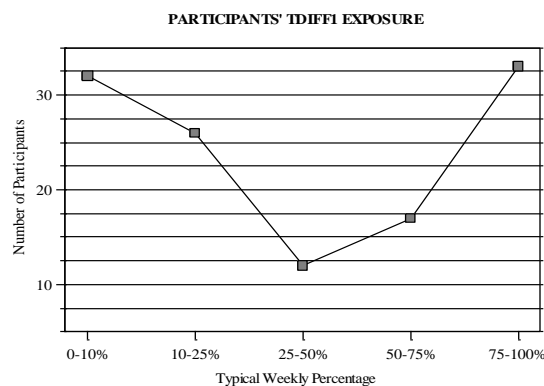
The pattern observed somewhat differed from that expected for such a multi-ethnic European population as the UK, generally, and England, in particular. Indeed, the data showed there were much the same percentage of participants who received virtually no inter-ethnic exposure from

the workplace (i.e., 26.7%, where $Tdiff_1$ = '5'). This negating impact was further reinforced by those having just 10-25% weekly exposure levels outnumbering participants which enjoyed 50-75% of their typical time with other ethnicities by some 50% (thus, $Tdiff_1$ ratings of '4' and '2' were respectively 21.7%, compared to 14.2%).

4.2. Freely-chosen exposure to out-groups

Once the constraints of work-related inter-mixing were removed, the time participants were apart from ethnic peers actually increased. On examining sub-group data, a drastic reduction of participants constantly inter-mixing among other ethnicities was evident. Those previously at $Tdiff_2$ level '1' exposure plummeted by some 27%. There was much less change for those which mixed with others 50-75% of their free time, however (a decrease of 5.8% from the $Tdiff_1$ level '2'). The greatest downward shift of all was seen for participants which were, in any case, only ever modestly exposed to other ethnicities 10-25% of the time in working scenarios (where $n = 17$ for participants at $Tdiff_2$ level '4'; decrease = 34.6%).

A paired t-test showed the overall mean inter-setting fall in ethnic socialisation, from 48.5% to 39.8%, was significant ($p = 0.007$). The above Figure, 2, additionally illustrates how the availability of free time/choice changed the outcome of the population's exposure to those who were ethnically differentiated.

Figure 2: Overall, subjects only mixed inter-ethnically for about 40% of a typical week given free-choice.

**PARTICIPANTS' TDIFF2 EXPOSURE**
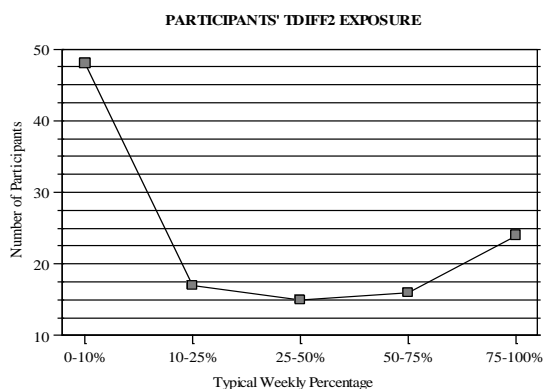
4.3. Exposure types and EGA potential

Data associated with the performance-related aspect of this investigation revealed participants, overall, were able to claim a mean score of 71% accuracy (thus, 5-point scale performance

equivalent rating = 2.16; s.d. 0.69). Figure 3, below, highlights how subjects performed, in terms of mean identifiability, across the entire span of both $Tdiff_1$ and $Tdiff_2$ inter-ethnic socialisation ratings (i.e., zero to full-time exposure, weekly).

Figure 3: The data dispel the idea of EGA outcomes reliably increasing with adult participants' exposure.

ETHNIC GROUP ATTRIBUTION ACCURACY
(Under Work-based and Free Time/Choice Constraints)
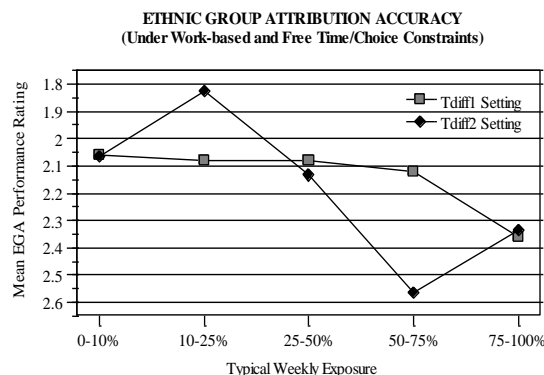
The analysis revealed that those from ethnically-restricted working environments boasted the greatest potential, in terms of EGA accuracy. This was contrary to any hypotheses favouring more exposed participants as the better performers, given their increased contact with any group in question, etc. Indeed, the mean performance ratings of the least and most exposed participants equated to a difference in accuracy of some 7.5%. However, the latter sub-group's large standard deviation meant this failed to reach significance at the 0.05 level ($p = 0.39$; mean sub-group scores = 73.44% versus 65.91% accuracy, respectively). The equivalent EGA potential of the remaining sub-groups, considered via $Tdiff_1$, was more closely arranged, being between 72 and 73.4%.

Considering participants' inter-ethnic exposure during free time ($Tdiff_2$), those at the 10-25% weekly level ('4') were found to perform most accurately, at 79.4%. The difference between this and both the very- and second-lowest intra-setting scores were significant ($p \leq 0.001$, then $p = 0.012$, at equivalent accuracies of 59.3%, then 66.6%). These low scores were again derived from the most inter-ethnically exposed participants, thus again, weakening the 'increased exposure $\rightarrow$ increased EGA performance' theory.

5. DISCUSSION AND CONCLUSION

We previously have learned:
- auditory- and acoustic-phonetically, EGA under same-language conditions is a workable task [2, 10, 15, 18];

- it still can remain so even if speaker-nativeness is finely nuanced and stimuli have minimal phonetic content [17]; and
- awareness of a given language system (L_n) can, to a point, improve general speaker identifiability in non-native settings [12, 13].

We additionally become aware that recurrent speaker-specific hearings have limited influence on listeners' sensitivity to other voices of that kind [19], further suggesting benefits prior experience/priming ultimately levels off.

Despite the above, there was no immediately apparent work exclusively dealing with how, if at all, ongoing exposure to other (ethnic) speaker-types increases their overall identifiability. This work has thus made an opening in that direction, where, perhaps, either oversight or fear of rousing controversy on racial/language ideological lines previously left a gap in our understanding.

This work argues that developing robust phonetic knowledgebases is dependent on ascertaining factors underpinning speaker|listener variability. The latter are thought to be influenced by in- or out-group interactions and contacts at times. Thus, the need for a study having such explicit focus was inevitable, as further underscored by the fact discussions on cross-cultural interactions are now being sought in a wider (linguistic) sense.

Finally, this work's scale shows: (1) populations can be much less inter-mixed than envisaged; yet (2) constrained EGA outcomes, evidently, can not be predicted by these factors of social interaction alone, for all such listeners.

6. REFERENCES

[1] Andics, A., McQueen, J., van Tourenout, M. 2007. Phonetic content influences voice discriminability. *Proc. 16th ICPhS* Saarbrücken*, 1829-1832.

[2] Arslan, L.M., Hansen, J. 1996. Foreign accent classification in American English. *Speech Communication* 18, 353-367.

[3] Docherty, G. 2003. Speaker, community, identity: Empirical and theoretical perspectives on sociophonetic variation. *Proc. 15th ICPhS* Barcelona, 11-16.

[4] Docherty, G. 2007. Speech in its natural habitat: Accounting for social factors in phonetic variability. In Ignacio Hualde, J., Cole, J. (eds.), *Laboratory Phonology* Berlin: Mouton de Gruyter, 1-35.

[5] Figueiredo, R.M. 1999. Perceptual recognition of celebrity voices using random spliced speech. *Proc. 14th ICPhS* San Francisco, 161-162.

[6] Flege, J.E. 1988. Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America* 84, 70-79.

[7] Foreman, C. 1999. Dialect identification from prosodic cues. *Proc. 14th ICPhS* San Francisco, 1237-1240.

[8] Foulkes, P., Barron, A. 2000. Telephone speaker recognition amongst members of a close social network. *International Journal of Speech Language and the Law* 7(2), 180-198.

[9] Lass, R., Almerino, C.A., Jordan, L.F., Walsh, J.M. 1980. The effects of filtered speech on speaker race and sex identifications. *Journal of Phonetics* 7, 105-118.

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

[11] Sebba, M. 1993. *London Jamaican: Language Systems in Interaction*. London: Longman.

[12] Sullivan, K., Kügler, F. 2001. Was the knowledge of the second language or the age difference the determining factor? *International Journal of Speech Language and the Law* 8(2), 1-8.

[13] Sullivan, K., Schlichting, F. 2000. Speaker discrimination in a foreign language: First language environment, second language learners. *International Journal of Speech Language and the Law* 7(1), 95-112.

[14] Todd, R. 1998a. Auditory perception and ethnic group attribution of unknown voices: Assessing the robustness of experienced listeners' ratings when confronted with non-native but proficient English speech. *Proceedings of the Institute of Acoustics* 6, 343-350.

[15] Todd, R. 1998b. *Perceptions of Speaker-Ethnicity in Non-native English Speech*. Unpublished Ph.D. Thesis, University of Sheffield.

[16] Todd, R. 2000. *Salience of Speaker-Ethnicity and Group-Belonging: Observations of an Attribution Task*. Paper disseminated at the *Congress of the International Association of Forensic Phonetics* Rome, Italy.

[17] Todd, R. 2002. Speaker-ethnicity: Attributions based on the use of prosodic cues. *Proceedings of the 1st International Conference on Speech Prosody*, 663-666.

[18] Walton, J.H., Orlikoff, R.F. 1986. Speaker and race identification from acoustic cues in the vocal signal. *Journal of Speech and Hearing Research* 37, 738-745.

[19] Wretling, P., Sullivan, K., Schlichting, F. 1999. Does repeated exposure to a target voice reduce the impact of a similar voice? *Proc. 14th ICPhS* San Francisco, 1385-1388.