

Identifications of speaker-ethnicity: Attribution accuracy in changeable settings

Richard Todd

Department of Computer Science, University of Sheffield, UK

<https://doi.org/10.36505/ExLing-2011/04/0033/000202>

Abstract

Several studies have considered the auditory identification of foreign-accented/non-native speech. Here, a finer-grained alternative to the traditional definition of nativeness is used. The approach eliminates the shortcomings of selecting ‘native’ speech based on the usual criterion of birthplace alone, yet accommodates complexities which arise when dealing with speakers from countries which are historically diverse, ethnically. Using the foregoing construct, this study examines the potential for listeners to accurately group different speaker-types, with respect their ethnicity, in low/high quality transmission conditions. The findings of the Ethnic Group Attribution (EGA) task confirms overall human competence. The work’s large number of participants ($n = 120$) brings added generalisability to smaller-scale studies. It furthermore, allows a better understanding of contextualised performance, gender-wise.

Key words: ethnic group attribution (EGA), ethnicity, nativeness, identification

Introduction

Listeners’ ability to identify or characterise speakers in terms of their foreign-accentedness or ethnicity has attracted interest in a number of ways. This is evidenced by the investigations of Arslan and Hansen 1996; Flege 1988; Sebba 1993; Todd 1998; Walton & Orlikoff 1986, for example. It becomes clear from such works that delineating nativeness or ascribing ethnicity to voices heard using the same language is a non-trivial task. Furthermore, performing the latter challenge — Ethnic Group Attribution (EGA) — requires listeners to additionally attend to features associated with *speaker-types*, or *-groups*, rather than highly idiosyncratic qualities which, in forensic-phonetic situations, usually promote fixation on one speaker.

It is well-understood that factors such as locality, social networks/peer affiliations, and language familiarity may variously influence speech perception. Clopper et al. 2006; Kerswill et al. 2008 encapsulate the general affects of speaker|listener locale and social networks on identification. Clopper 2004; Sullivan & Schlichting 2000 further illustrate any benefits of on prior linguistic awareness. Notwithstanding the foregoing however, implicit ambiguities that may exist when speakers from countries/regions having a long-standing history of ethnic diversity (like, England or the USA) are considered for study. The traditional native|non-native criterion,

operating purely on lines of birthplace, appears to serve us well, initially. Inadequacies of the binary approach arise however, on realising speakers of variable descent (say, Italian, Ukrainian, etc.) may also be born and raised in the same locality as others who are ‘native’ via traditional definition.

The predicament means differing speaker-types may unwittingly populate a stimulus group. Devising a more refined framework allows speakers having shared overarching features to be more clearly disambiguated, especially for the sake of others wishing to further develop acoustic and auditory studies alike (cf. Kerswill et al. 2008). Figure 1, below, illustrates the how speaker-nativeness is nuanced. From both socio- and forensic-phonetic standpoints, it is believed this finer-grained approach may, respectively, permit greater openings for study and debate.

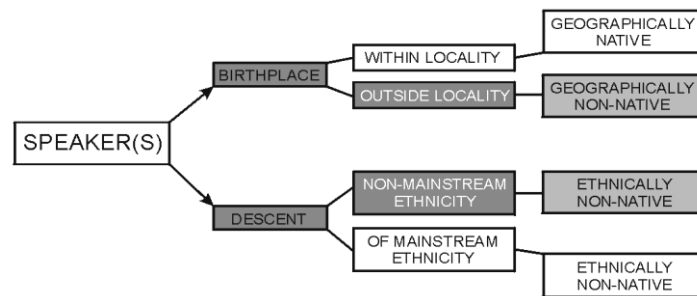


Figure 1. A new framework of Nativeness considers speakers’ ethnicity in addition to geographical origins, to allow finer-grained distinctions.

In this work, nativeness is vitally considered in the above terms, given that speaker-ethnicity and its identification is central themes.

The Study

Without doubt, work on the perceptions of non-native speech from an L2 perspective continues to steadily grow. Attention however, still seems to bypass the issue of whether listeners can reliably group speakers in terms of their respective ethnicities, especially if all (1) inhabit a common locality; (2) have routine/innate familiarity with those speech norms; and, (3) were raised with some other language/system(s) also in use (L_n).

Here, EGA potential in high- and low-quality listening conditions is examined while encompassing the three aforementioned points.

Participants and Method

Various negative constraints mean studies related to either the identification or perception speaker-accentedness or -ethnicity may feature a relatively modest number of participants. To illustrate, Yuang et al. (2010) considered the attributions of just 3 participants. However, Clopper (2004) and Kerswill et al. (2008) improve generalisations further (where, respectively, speaker types = 6 and 4; $n = 49$ and 68 subjects, respectively, from the same locale). Despite this, both of the latter works are problematic, if seeking to precisely determine the ethnicity of any speech considered.

An even greater number of participants were considered in this investigation ($n = 120$). All were adults (mean age = 34.4; male = 52; female = 68). While, overall, each speaker-group would be of the participants' locale, non-native ethnicities were South Asian, Caribbean, and East Asian (total $n = 45$). Research volunteers had to consider such speech (being lexically identical) while presented in a High-Quality setting (with face unseen; HQ), and Low-Quality setting (i.e., via telephone; LQ).

Results

Table 1. Summarised attribution potential (overall, and gender-wise).

Settings considered	Overall Ratings [†]			Female Ratings [†]			Male Ratings [†]		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Face Unseen	2.08 (sd 0.8)	1	4	2.22 (sd 0.83)	1	4	1.88 (sd 0.73)	1	3
Telephone	2.28 (sd 0.86)	1	5	2.43 (sd 0.85)	1	5	2.08 (sd 0.84)	1	4

[†] rating 1 = all the time; 2 = frequently; 3 = occasionally; 4 = hardly ever; 5 = virtually never

Besides gender-wise participant means, Table 1, above, shows highest/lowest ratings which, practically, were analogous to the accuracy and confidence of EG as being made. In terms of descending competence/accuracy, rating '1' = 75-100%; '2' = 50-75%; '3' = 25-50%; '4' = 10-25%; and '5' = 0-10%. Males yielded higher, yet less varied, performance ratings, suggesting the more assured identifications of speaker-ethnicity. Significant gender differences in performance outcomes existed. T-test p -values moved from $p = 0.022$ in the HQ (face unseen) setting, to $p = 0.027$ when making comparisons for the least reliable LQ (telephone) setting. Latterly, male and female EGA potentials dropped, to equal 73% and 64.25%, respectively (68%, overall). Previously this was 73% overall, where males showed 78% accuracy; females claimed 69.5%.

Discussion and Conclusions

The aim of the study was to bring more clarity and attention to an increasingly inadequate notion of speaker-nativeness, plus the under-researched area regarding the identifiability of speaker-ethnicity.

Despite limited space, this paper initially presents a framework which challenges, and may influence, current assumptions of nativeness. By featuring two stimulus qualities, the study allows better understanding of how extrinsic matters may influence listener perception, and thus, the potential outcome of attributions. Unlike other studies, participants contemplated speakers of a common locality, all producing like utterances, yet clearly differing in terms of direct ethnic descent. Results of this large-scale work in brief show listeners may, overall, claim '*frequently*' accurate performance in favourable conditions, despite gender differences.

References

- Arslan, L.M., Hansen, J. 1996. Foreign Accent Classification in American English. *Speech Communication*, 18, 353-367.
- Flege, J.E. 1988. Factors Affecting Degree of Perceived Foreign Accent in English Sentences. *Journal of the Acoustical Society of America*, 84, 70-79.
- Sebba, M. (1993) *London Jamaican: Language Systems in Interaction*. London: Longman.
- 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.
- Todd, R. 1998. 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*, 20, 6, 343-350.
- Todd, R. 2002. Speaker-Ethnicity: Attributions Based on the Use of Prosodic Cues. *Proceedings of the 1st International Conference on Speech Prosody*, *Speech Prosody 2002*, 663-666.
- Walton, Orlikoff. 1986. Speaker and Race Identification from Acoustic Cues in the Vocal Signal. *Journal of Speech and Hearing Research*, 37, 738-745.
- Clopper, C. G. (2004). *Linguistic Experience and the Perceptual Classification of Dialect Variation*. Doctoral dissertation, Indiana University.
- Clopper, C. G., Levi, S. V., & Pisoni, D. B. (2006). Perceptual similarity of regional varieties of American English. *Journal of the Acoustical Society of America*, 119, 566-574.
- Yuan, J., Jiang, Y., Song, Z., 2010. Perception of foreign accent in spontaneous L2 English speech, *Proceedings of Speech Prosody 2010*, 100884:1-4.
- Kerswill, P., Torgersen, E., Khan, A. and Fox, S. Perceiving ethnicity and place in Multicultural London English. Paper presented at NWAV37, Rice University, Houston, November 2008.