Proceedings of the Institute of Acoustics

ACOUSTIC-PHONETIC QUALITIES OF ASIAN- AND CARIBBEAN-ENGLISH CONSONANT CLUSTERS

Richard Todd

Speech and Hearing Research Group
Department of Computer Science, University of Sheffield, England, S1 4DP, UK.
E-mail: R.Todd@dcs.shef.ac.uk

1. INTRODUCTION

It is understood that many computational research activities focused on the recognition of the spoken word rely on some prior knowledge of the target language. Typically — during the early stages of such work — canonical forms of linguistic data are used for system development, which are then in turn modified to accommodate some non-ideal input (e.g., occluded signal end-points, background noise, etc.).

Recently, there has been an increased level of interest in the characterisation of data that is speaker-generated due to either some (external) stress factor or a difference in the speaker's accent. As was the case in [1], the work presented here is concerned with the latter issue, but the primary focus is on exemplifying any phonetic manifestation.

2. DELINEATING THE NON-NATIVE SPEAKER

The ill-defined concept of speaker-nativeness is highlighted in a companion paper [2]. Furthermore, an argument has been made that being native by birthplace alone, is in no way a guarantee of a speaker's linguistic concordance at the phonological level. To quote:

> '... in Britain and the USA, many of the so-called native citizens are descendants of people born in the Caribbean, Indian Sub-Continent, or Latin America which, in turn have language systems other than Standard English in use for everyday verbal communication. Since the passing-on of features belonging to any such co-occuring language system may give rise to differences in analytical findings, ethnicity has been marked. The parents of ethnically non-native speakers on the other hand, can be marked as being geographically non-native speakers of English.' [2]

This definition of nativeness is attractive since it provides (1) a tri-level distinction along the *native* —> *non-native* domain space; and (2) allows, whenever required, each sub-region to operate with separate phonological inventories (and underlying phonetic prototypes). Figure 1 shows the potential of the mentioned representation when applied to casually articulated continuous speech. On viewing this, it is apparent that phonetic features surfacing at the ethnically marked level of speaker-nativeness could easily be dismissed as mere production errors — or otherwise noted as 'stigmatised' pronunciations — when using a traditional (binary) framework. In some cases, such allophonic tendencies are simply moved to the socio-linguistic domain of 'style shifting' within an entirely native inventory [4].

Regardless of how small variations in pronunciation (and voice quality) may have been treated in the past by linguists, the power of their intrinsic group belonging should at least be explored, since the benefits to related machine learning and forensic disciplines are

potentially enormous. Additionally, if speakers are discretely classified in terms of both ethnicity (descent), and proximity (birthplace), prerequisite knowledge — implicit in earlier studies of bilingualism, for instance — could be put to further use [4, 6].

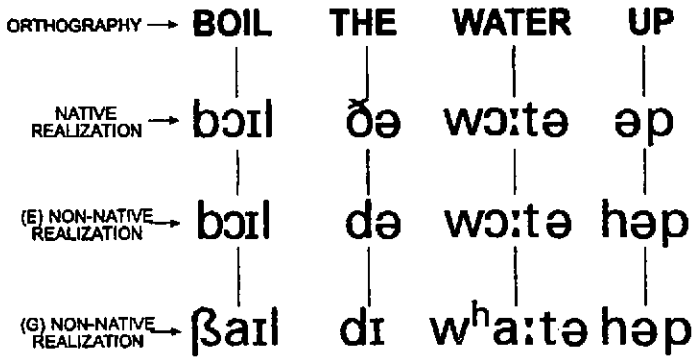

Figure 1. The explicit classification of speaker nativeness by ethnically (E), and geographically (G) derived measures reveals phonetic features that are often obscured in the presence of previously ill-defined terms. British-Caribbean and non-creolized Caribbean English speech are depicted at levels E and G, respectively.

2. THE ANALYSIS

Recent empirical data [2] show that when exposed to contextually reduced data — excerpted from a stream of continuous speech — listeners are able to discern phonetic properties that are in some way distinct from those of their mental prototypes for the target language. Moreover, it would seem that a listener's grouping capabilities on the grounds of speaker-ethnicity are likely to increase when in such a setting.

A monosyllabic focus word within one of the seven reduced contextual form (RCF) utterances used in [2] has been instrumentally examined in order to ascertain the differentiating qualities heard. Figure 2, below, illustrates the differences in signal intensity between one native and two geographically non-native samples.

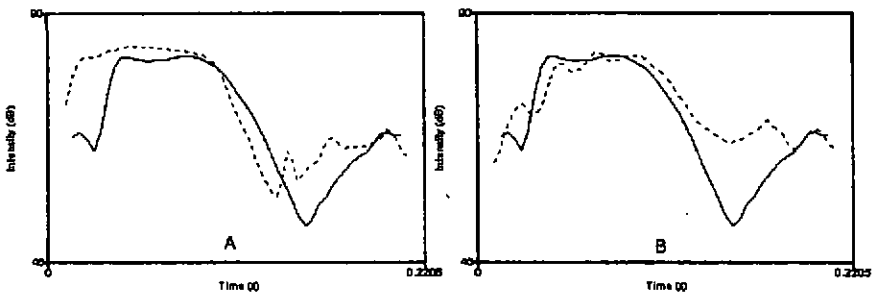

Figure 2. Intensity characteristics of the word 'drugs' produced in a connected environment. The native profile is shown with a continuous line. Asian (A) and Caribbean (B) tokens are depicted by dashed line.

The focus word, 'drugs', shown in this figure appears in a /C__V/ environment (/m/ and /ɑ/ respectively). Asian representations of the prevocalic region in the word favour [dɾ], whilst

the affricated and palatalised state of the Caribbean /dr/ gives rise to a [d͡ʒ] being realised. As far as the perception of experienced listeners is concerned, the Asian utterance was always correctly identified. Only one of the listeners erroneously identified the Caribbean utterance as being truly native. Even in this case, the affrication was clearly perceived (and subsequently documented) by the listener. It also should be noted however, that the related subject's Response Form from which these data are derived reveals that the confidence level when judging this token was low (marked 'guess').

In the intensity measurements we can also observe that the silence period is less marked for both the Asian (close to 57.5 dB) and Caribbean (about 66.5 dB) speakers (figures 1a and 1b respectively), where the level of the natively-produced closure is in the region of 52 dB. The overall result is that during listening, Caribbean and native representations of the word-final cluster /gz/ appear rather more elided than the slightly devoiced and more spectrally diffused [g̊z] heard in the Asian speech. The latter sounds were also preceded by a transient feature resembling a velar closure and resulted in a very brief plosive sensation (of less than 7ms). All of these auditory sensations were confirmed by spectrographic examination of the data (figure 3).

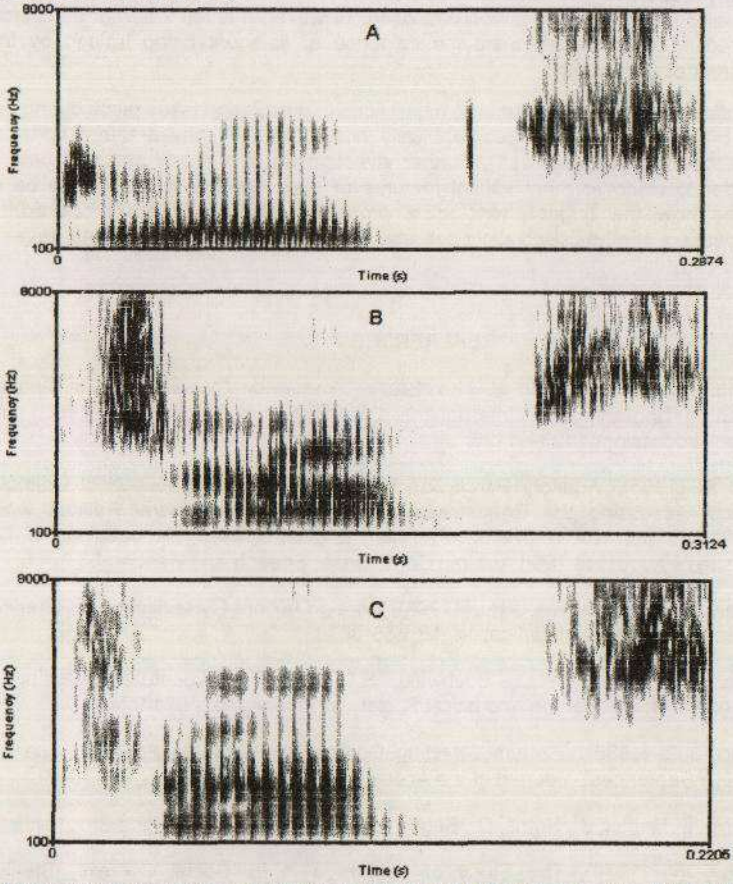

Figure 3. Examination of spectrograms relating to Asian (A), Caribbean (B) and native British English (C) utterances highlight marked differences in word-initial and -final consonant clusters /dr/ and /gz/ respectively.

The overall impression on listening to these non-native data, is a marked retroflexion (followed by an alveolar tap) for the Asian token, and an affrication in the Caribbean speech that phonologically renders the utterance similar in sound to the word 'jugs' /ʤugz/. It is interesting to note however, that the majority of listeners did not overly consider the /k/-like quality of the Asian token's word-final velar consonant a strong cue to grouping.

3. CONCLUSION

A monosyllabic focus word that featured in an earlier listening experiment has been examined with the intention of highlighting features of — auditorily perceived — non-native English speech to by instrumental means.

Associated data highlighted in [2] show that when consonant clusters within this and other focus words have been repeatedly presented to trained listeners, they have reported on several phonetic properties that have betrayed any notion of nativeness that may have, in more contextualized situations, remained unnoticed. However, it is felt that more empirical data is required in order to determine the influence of listeners being trained by the presentation order of stimuli, etc.

This work confirms that some (involuntary) differences in articulation takes place during the speech acts of proficient — but geographically non-native — speakers that consider English to be their first language (L1). Further investigation of similarly obtained data is needed in order to discover other salient features of such speech. It would also be of interest to see how the target phonological properties of English are produced by British-born speakers of ethnically distinct descent, in addition to their acoustic derivatives.

REFERENCES

[1] Arslan, L.M. and Hansen, J.H.L. (1997), 'Frequency Characteristic of Foreign Accented English', IEEE International Conference on Acoustics, Speech and Signal Processing, 1123—1126.

[2] 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 1998 Autumn Conference: Speech and Hearing 98.

[3] Arslan, L.M. and Hansen, J.H.L. (1996), 'Foreign Accent Classification in American English', Speech Communication, 18, 353-367.

[4] Elman, J.L., Deihl, R.L., Buchwald, S.E. (1977) 'Perceptual Switching in Bilinguals', Journal of the Acoustical Society of America, 62, 971-974.

[5] Flege, J.E. (1988), 'Factors Affecting Degree of Perceived Foreign Accent in English Sentences', Journal of the Acoustical Society of America, 84, 70-79.

[6] Cutler, A., Mehler, J., Norris, D., Segui, J. (1989), 'Limits on Bilingualism', Nature.

[7] Labov, W. (1970), The Study of Language in its Social Context', Stadium Generale, 23, 30-87.

[8] Raphael, L.J. (1972), 'Preceding Vowel Vuration as a Cue to the Perception of the Voicing Characteristic of Word-final Consonants in American English', Journal of the Acoustical Society of America, 51, 1296-1303.

[9] Flege, J. (1991), 'Age of Learning Affects the Authenticity of VOT in Stop Consonants Produced in a Second Language', Journal of the Acoustical Society of America, 89, 395-411.

[10] Darwin, C.J. (1975), "On the Dynamic use of Prosody in Speech Perception", in: A. Cohen and S.A. Nootebloom (eds.), 'Structure and Process in Speech Perception', Berlin: Springer, 187-194.