euronoise

Cue switching in the perception of approximants: Evidence from two English dialects

Rachael-Anne Knight[a], Christina Villafaña Dalcher[a] and Mark Jones[b]

[a]City University, Department of Language and Communication Science, Northampton Square, EC1V 0HB London, UK

[b]University of Cambridge, Phonetics Lab, Department of Linguistics, Raised Faculty Building, Sidgwick Avenue, CB3 9DA Cambridge, UK

knight@city.ac.uk

The present study examines sound change propagation from a phonetic standpoint, with particular attention to the salience of different formant frequencies as cues to perception of contrast in similar sounds. Thirty-three subjects (8 AE and 25 BE) performed an identification task in which they judged whether stimuli were more like /r/ or /w/. The stimuli comprised five sounds, copy-synthesised from a source [ɹ], where the values of F2 and F3 were adjusted to fall between the frequencies typical for [ɹ] and [w]. The only significant difference between the two dialect groups' performance occurred with a token in which F3 was kept at a frequency typical for [ɹ] and F2 was lowered to that of /w/. AE and BE speakers identified this token as /r/ 93% and 70% of the time, respectively. This is unexpected, as /r/ in both dialects is characterised by a low F3. However, the difference may be due to the existence of the 'labiodental' variant of /r/ in BE. As this variant does not have a low F3, BE speakers must tolerate a wider diversity of /r/-types than AE speakers. We suggest that the /r/ category in BE may be becoming increasingly defined by F2, rather than by F3, which has implications for future production of /r/ in this accent, and supports an acoustically motivated theory of sound change propagation.

1 Introduction

Approaches to the actuation and transmission of sound change in recent literature on the subject include both functional and formal accounts. Sound change may have its roots in the reduction of articulatory effort [1], listener failure to resolve ambiguities in speech signals [2], or preservation/simplification of phonological structure [3, 4]. Spread or transmission of change is generally taken to involve social factors such as differentiation [5], language contact [6, 7] or identity practices of speakers [8]. A recent theory identifies the mimetic behaviour of speakers as a factor [9], divorcing transmission of novel speech sounds from sociolinguistic motivations.

Here we examine a similar perspective on sound change and variation where the presence of a variant in a linguistic community alters the relative importance of acoustic cues for listeners who encounter that variant. To the extent that resetting of perceptual cue targets alters motor representations of speech, the listener-turned-speaker becomes an agent in the spread of the linguistic variant, albeit for different reasons than often assumed in the sociophonetic literature.

In support of an acoustic cue-related motivation for sound change spread, we present the results of a perceptual study testing the ability of speakers from two different linguistic communities to distinguish the /r/-/w/ contrast.

2 Background

This study examines the acoustic cues of two different approximants, /r/ and /w/, and a brief description of these sounds in terms of their acoustic characteristics and variation patterns is appropriate.

2.1 Phonetic qualities of /r/ and /w/

The standard description of /r/ in British and American English is a voiced postalveolar approximant, where the tongue tip is in wide approximation to the region of the palate behind the alveolar ridge. Although studies of American English /r/ show that speakers employ many different articulatory strategies [10], in production of this consonant there are two stable acoustic traits of /r/ in both dialects. The most salient is the low third formant (F3) [11, 12, 13, 14, 15]. Another is the proximity of the second and third formants (F2 and F3) (e.g., [16]). The labial-velar approximant /w/, in contrast, is generally characterised by a high F3 and low F2, resulting in a wide gap between these two formants [14, 17].

2.2 Sociophonetic variation of /r/

Many younger speakers in some regions of England now use a variant of /r/ that differs from the canonical form described in Section 2.1. Typically symbolised as [ʋ], it is characterised by the same acoustic qualities as developmental high-F3 /r/, and has been described in the literature as a labial or, more commonly, labiodental approximant [18] perhaps accompanied by velarisation [19, 20, 21]. Many earlier descriptions class this variant as a speech defect or as a feature of either immature speech (e.g. [22, 23]) or upper-class speech [19], but the labiodental realisation is becoming increasingly common in varieties across England. In Norwich in 1974 there were very few instances of the labiodental variant, but by 1983, 33% of speakers born between 1959 and 1973 used [ʋ] in their speech [24, 25].

In the United States, however, a labialised or labiodental [ʋ] is not attested in adult speech with regularity outside of Brooklyn, New York [19]. The assumption that speakers outside of metropolitan New York are not exposed to a labiodental variant is central to the current study and its results, although developmental forms in child speech may contain a high F3 in many varieties of AE and BE [18, 27, 28, 29].

2.3 Comparison of three approximants

Formant values of the three approximants in question, postalveolar [ɹ], labiodental [ʋ], and labial-velar [w] can be compared in the schematic spectrogram in Figure 1, below. The labiodental's second formant is similar to the mid-range formant frequency of [ɹ], while its third formant is similar to the high F3 of [w].

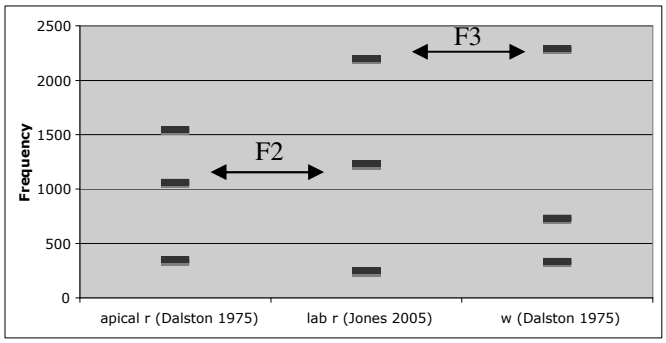

Figure 1 Formant values of three approximants

We may ask whether there are implications of the existence of 'labiodental' /r/ for the way in which listeners handle the /r/-/w/ contrast, specifically with respect to the utilisation of acoustic cues such as F2 and F3 frequency.

3 Perceptual study of /r/ variants

3.1 Overview

This study collected perception data from two groups of speakers who are assumed to differ in their exposure to adult 'labiodental' /r/ and tested for significant differences in perceptual cues between the speaker groups.

The subject pool comprised eight adult native speakers of American English from the Washington, DC area and 25 adult native speakers of British English. As the AE data was collected remotely, these subjects were not recorded, but all were judged to use a postalveolar /r/ based on auditory analysis. The BE speakers were recorded and found to use either postalveolar /r/ or 'labiodental' /r/, although some of these speakers varied their articulations by context. Based on sociolinguistic studies of /r/ variants in AE and BE, it was assumed that none of the AE speakers (having never lived in the New York City metropolitan area) are exposed to 'labiodental' /r/, while the BE speakers, regardless of individual productions, are regularly exposed to the adult variant.

3.2 Methodology

The perception experiment, built in PsyScope [26], consisted of two blocks – a forced choice identification task and a discrimination task. In the former, subjects were asked to judge whether stimuli in "a _ing" context were more like /r/ or more like /w/. In the latter, subjects decided whether pairs of stimuli in the same "a _ing" context were identical. The stimuli for both tasks comprised five copy-synthesised sounds from a source [ɹ] uttered by an adult male native speaker of BE, where the frequencies of F2 and F3 were manually adjusted. Table 1 shows the input formant frequencies of the synthetic stimuli and Figure 2 presents a schematic illustration of the five tokens. The formant frequencies output by the synthesiser differed from the input formants by an average of 8.4 Hz, with a maximum input-output variance of 24 Hz occurring in Stimulus A's F3.

Stimulus	F1	F2	F3	Description
A	355	1201	1682	/r/-like formants
B	355	963	1682	F2 at midpoint of /r/ and /w/, F3 /r/-like
C	355	1201	2541	F2 /r/-like, F3 /w/-like
D	355	725	1682	F2 /w/-like, F3 /r/-like
E	355	725	2541	/w/-like formants

Table 1 Input formant values of copy-synthesised stimuli

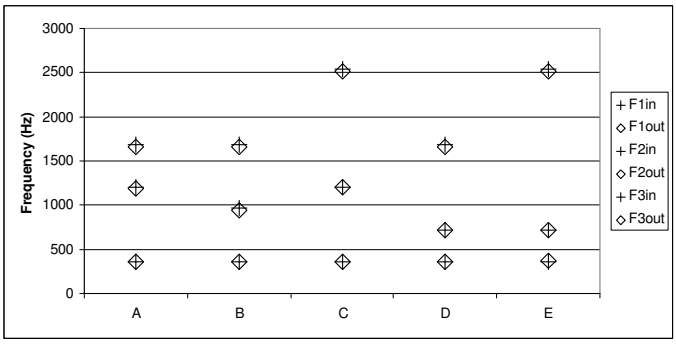

Figure 2 Frequencies F1-F3 in stimuli A-E showing relationships between the input to the synthesiser (crosses) and the output from the synthesiser (diamonds)

For the identification task, the total number of tokens equalled 50, with ten randomised repetitions of each stimulus. The discrimination task, used to assess subjects' perceptual sensitivity, comprised 105 tokens: five repetitions of each ordered pair of stimuli, with five instances of identical pairs used as controls.

Assuming low F3 to be the primary acoustic cue for /r/ identification, subjects' behaviour was predicted as follows: stimuli with a low, /r/-like third formant (Stimuli A, B, and D) would be heard as /r/, while Stimulus E, with /w/-like formants, would be heard as /w/. Stimulus C, however, synthesized to be most like 'labiodental' /r/, would be categorised differently depending on dialect: its relatively high third formant would prompt AE subjects to hear it as /w/, while its resemblance to the existing labiodental /r/ variant in SSBE would encourage BE subjects to classify it as /r/, despite the non-/r/-like third formant frequency.

3.3 Results

In the identification task most subjects in each of the dialect groups identified Stimuli A, B, C, and D (that is, all stimuli except for the token with /w/-like formants) as /r/ the majority of the time. Table 2 summarises the responses in terms of the percentage of /r/ responses to each of the stimuli, by dialect group.

	Stim A	Stim B	Stim C	Stim D	Stim E
AE	100	100	93	93	5
BE	99	97	88	70	2

Table 2: Percentage of /r/ responses to identification stimuli

The shaded cells in Table 2 highlight the response patterns that deviate from the predicted outcomes. Stimulus C, with a high, /w/-like F3, was identified as /r/ most of the time by both the AE and BE subjects, with no significant behavioural difference between the groups. We conclude that the high rate of /r/-identification of this stimulus by AE subjects may have resulted from the proximity of F2 and F3, the outweighing of high (non-/r/-like) F3 by high (/r/-like) F2, and confusion of the acoustic cues with those of /l/, which was not among the choices in the perception task.

We focus here on Stimulus D, the token with a typically [ɹ]-like low F3. This stimulus was predicted to be judged as /r/ in a majority of instances, and, averaging over all subjects, was in fact identified as /r/ three times as often as it was identified as /w/. However, when we split the subjects by dialect type, we find a robust difference in behaviour: the AE subjects judged the stimulus as /r/ 93% of the time; the BE subjects judged it as /r/ only 70% of the time. In fact, the only significant difference between the two dialect groups' identification of the five stimuli was found in the reaction to Stimulus D, based on independent sample T-tests ($t=3.146$, $p<.005$). No other statistically significant patterns were found with respect to identification or discrimination of stimuli, and there were no effects of /r/-type used on the BE speakers' judgments.

4 Discussion

We begin by presenting an overview of one theory of sound change, the unintentional and automatic imitation of sounds resulting in modifications to mental representations over time [9]. While a mimetic model may be capable of accounting for the spread of some phonetic variants throughout certain speech communities, we suggest that speakers switch the acoustic cues used in identification of a speech signal first, without necessarily copying ambient productions. A cyclical account of sound change propagation, rooted in contrast maintenance, is proposed.

It is normally assumed that, once actuated, a sound change is transmitted through a speech community along sociolinguistically motivated pathways (e.g., [1]). The mimesis model exemplified in [9] challenges this view by demonstrating that short-term changes may occur in production after brief exposure to a different realization, even when subjects have not noticed that a different variety is being spoken. Therefore the mimesis account suggests that sound changes may be transmitted without any sociolinguistic motivation. A similar conclusion is drawn indirectly from the results presented here. The results of the perceptual experiments reported here suggest that once a variant has reached a certain level of occurrence in a speech community – perhaps via an initial period of mainly sociolinguistic transmission – the variant acquires sociolinguistic autonomy. Its use may no longer be driven by sociolinguistic considerations, but the novel variant's characteristics contribute to the perceptual representation of a category for any speakers who encounter it. The characteristics of the new representation may then erode older speech production targets, and the novel variant becomes increasingly common, causing further erosion of the dominance of the older articulatory targets.

4.1 Cue-switching in the perception of approximants

The robust difference found between speakers of the two dialects in identification of a low-F3 stimulus may lie in a switching of acoustic cues on the part of the BE speakers, arising from the presence of an alternate form of /r/ in England and a continued pressure to differentiate /r/ and /w/.

As shown in Figure 1 above, the 'labiodental' variant on the rise in BE is acoustically characterised by a third formant quite similar to that of /w/ making it difficult for speakers to use this cue to distinguish /r/ and /w/ so a new differentiation strategy must be adopted. One distinguishing characteristic between 'labiodental' /r/ and /w/ is the frequency of the second formant. Figure 3 illustrates this alteration schematically.

The AE subjects, assumed to lack exposure to an adult high-F3 /r/ variant, experience no pressure to alter their acoustic cues in perception of the /r/-/w/ contrast.

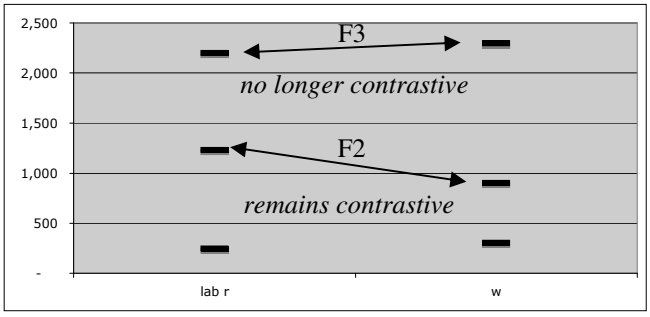

Figure 3: Formant contrasts: 'labiodental' /r/ and /w/ [21]

Given the role of F2 frequency as a likely contrastive cue for 'labiodental' /r/ and /w/, judgments of a stimulus where F3 is /r/-like and F2 is /w/-like as /r/ are predictable from the assumed exposure to 'labiodental' /r/ in a subject's linguistic environment. For speakers with this exposure, gradually weighting F2 more highly than F3 to contrast /r/ and /w/, results in Stimulus D being increasingly likely to be perceived as a /w/.

Subjects without any exposure to high-F3 'labiodental' /r/, presumably continue to rely on F3 as a contrastive cue and will therefore perceive Stimulus D as /r/, despite the fact that its F2 is identical to that of /w/:

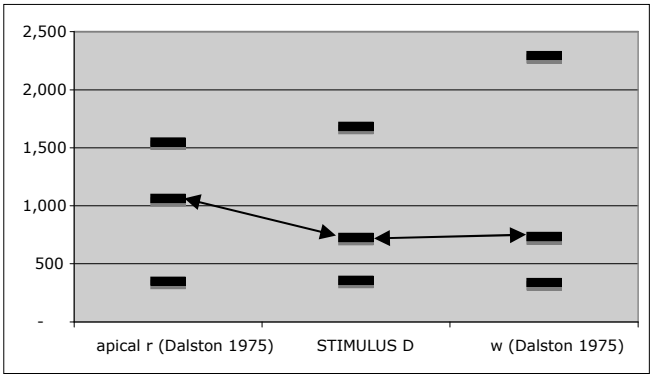

Figure 4: Comparison of F2 frequencies

Stimulus D (present study), /r/, and /w/ [29]

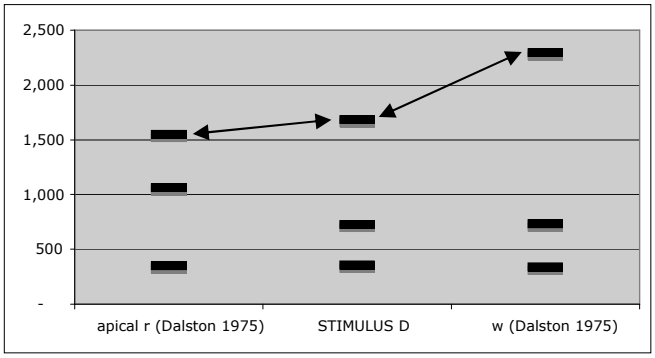

Figure 5: Comparison of F3 frequencies
Stimulus D (present study), /r/, and /w/ [29]

As a consequence of variation in the realisation of /r/ in BE, the approximant /r/ category in BE may be becoming increasingly defined by F2, rather than by F3. If this is the case, BE speakers will weight F2 more than F3 in their perceptual categorisation, and the F2 boundary between /w/ and /r/ will become sharper in BE relative to AE. BE speakers categorise Stimulus D as /r/ in 70% of the cases, but as /w/ 30% of the time. This trend in identification is evident throughout the BE subject pool: only three BE speakers showed a strong preference for /w/ when presented with Stimulus D. Of the remaining subjects, 13 preferred /r/ and eight exhibited little or no preference between /r/ and /w/.

4.2 A cyclical account of sound change

A crucial point we make here is the lack of significant effects of production type in /r/ identification on the part of the BE subjects. We confirm this both impressionistically and quantitatively.

In the impressionistic analysis of /r/ in initial, singleton context, the 25 BE subjects were divided into 11 labiodental /r/ users and 14 apical /r/ users. T-tests comparing the percentage of Stimulus D tokens by labiodental and apical /r/ users in the BE dialect pool yielded no significant results (t= -1.018, p=n.s.). An even more striking observation of the impressionistic analysis is that the labiodental /r/ users were less likely to identify Stimulus D as /r/ than the apical speakers (63% and 75% of /r/ responses to this stimulus were noted, respectively).

To confirm these results, and to accommodate the fact that many speakers classified as apical /r/ users actually had mixed apical and labiodental productions, we tested for correlations between percentage of /r/ responses to Stimulus D and three acoustic indicators of /r/-ness, absolute frequency of F3, distance between F3 and F2, and a derived coefficient of F3 times F3-F2 distance. None of the three tests yielded significant results, as shown in Table 3

	r	p
F3 frequency	-.298	n.s.
F3-F2 distance	-.151	n.s.
F3*F3-F2 distance	-.228	n.s.

Table 3: Results of correlation tests for acoustic /r/ indicators and responses to Stimulus D

The perceptual results presented here suggest a gradual demotion of F3 as a perceptual cue and a concomitant promotion of F2 as the main carrier of the /r/~/w/ contrast. Erosion of F3's perceptual status might explain the variability observed in F3 in the limited data on the subjects' own /r/ realisations. If F3 is less important as a cue to /r/, speakers need not be so concerned with attaining a particular F3 frequency. Some speakers may then allow F3 to vary, perhaps at random, perhaps by context. Continuing variability in F3 production feeds into the pool of /r/ representations, further undermining the status of F3 as an important cue. The increase in /r/ variability with respect to its third formant serves as a trigger to further acoustic cue-shifting, and subsequent changes in production – a cyclical shift in perception and production has been initiated. Thus a gradual erosion of low F3 instances of /r/, and a concomitant increase in 'labiodental' /r/ may be predicted across BE.

5 Conclusions and future research

This study has presented /r/-variant perception data from two distinct dialect groups differing in the types of rhotics existent in their linguistic environments. Examining formant frequencies of /[ɹ], [w], and three synthetic stimuli, we have attempted to explain why speakers of British English exhibit a different pattern in their categorisation of certain acoustic signals than speakers of American English. This dialect-dependent variation in perception appears to correlate with the presence or absence of variant forms in a speaker's linguistic environment, where such forms are sufficiently similar in acoustic characteristics to necessitate a shift in perceptual strategy. Furthermore, the shift in reliance on one acoustic cue to another has potential ramifications for speech production.

Rather than a purely descriptive account of /r/ variants in a linguistic community, the present study supplies us with a way to address sound change propagation from a phonetic point of view and to begin to answer questions about the ordered relationship between perception and production. The data herein suggests that the revision of mental representations in order to preserve contrast with existing sounds in a language's inventory can exist without concomitant production of such representations. Eventual copying of heard signals, therefore, may be explained not simply by means of random imitation of other speakers, but also as an after-effect of the resetting of acoustic cues used to distinguish similar sounds.

This study can be expanded in a number of directions. The result that those BE subjects who tended to use labiodental /r/ were slightly less likely to judge the ambiguous stimulus D as /r/ than subjects using low F3 /r/ needs further investigation. It suggests an indirect link between perception and production, but it may be that these subjects judged the stimuli as /r/ on other grounds, such as retained long-domain resonances for /r/ from the original source file (e.g. [30]), centre of gravity, or rate of formant transitions. A modified perception/production experiment incorporating priming would examine the potential effects of mimesis with respect to labiodental /r/. Weighting of different acoustic cues (absolute formant frequencies, F3-F2 ratios) and pass filtering may bring out further disparities between and within dialect groups. Perceptual testing of subjects

exposed to a 'labiodental' /r/ variant commonly found in certain non-rhotic AE dialects may support the claim of cue-shifting in contrasting /r/ and /w/. Articulatory data using ultrasound imaging on 'labiodental' /r/ and BE /r/ in general will allow comparisons of speech production across dialects – it may be the case that postalveolar [ɹ] in American and British English are not, in fact, as articulatorily similar as they are assumed to be.

References

[1] R. Kirchner, *An Effort-based Approach to Consonant Lenition*, University of California at Los Angeles Ph.D. Dissertation (1998)

[2] J. Ohala, "The listener as a source of sound change", in C.S. Masek, R.A. Hendrik, M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior: Chicago Linguistics Society*, Chicago: CLS, 178-203 (1981)

[3] P. Kiparsky, "The phonological basis of sound change", in J. Goldsmith (ed.) *Handbook of Phonological Theory*, Oxford: Blackwell (1995)

[4] B. Hayes, "Phonetically-Driven Phonology: The Role of Optimality Theory and Inductive Grounding", In M. Darnell, E. Moravscik, M. Noonan, F. Newmeyer, Ka. Wheatly (eds.), *Functionalism and Formalism in Linguistics, Volume I: General Papers*, Amsterdam: John Benjamins, 243-285 (1999)

[5] W. Labov, "The social stratification of English in New York City", Washington: Center for Applied Linguistics, (1972)

[6] N. Schilling-Estes, W. Wolfram, "Alternative models of dialect death: dissipation vs. concentration", *Language* 75, 3, 486-521 (1999)

[7] P. Trudgill, *Dialects in Contact*, Oxford: Blackwell, (1986)

[8] M. Bucholtz, "Why be normal?: Language and identity practices in a community of nerd girls", *Language and Society* 28, 203-223 (1999)

[9] V. Delvaux, A. Soquet, "The influence of ambient speech on adult speech productions through unintentional imitation", *Phonetica* 64, 145-173 (2007)

[10] J.R. Westbury, M. Hashi, M.J. Lindstrom, "Differences among speakers in lingual articulation for American English", *Speech Com-munication* 21, 203–226 (1998)

[11] P.C. Delattre and D. Freeman, "A dialect study of American r's by X-ray motion picture", *Linguistics* 44, 29–68 (1968)

[12] F. Nolan, *The Phonetic Bases of Speaker Recognition*, Cambridge: Cambridge University Press (1983)

[13] A. Alwan, S. Narayanan, K. Haker, "Towards articulatory-acoustic models for liquid approximants based on MRI and EPG data, Part II The rhotics", *Journal of the Acoustical Society of America* 101, 1078–1089 (1997)

[14] K. Stevens, *Acoustic Phonetics*, MIT Press (1998)

[15] C.Y. Espy-Wilson, S.E. Boyce, M. Jackson, S. Narayanan, A. Alwan "Acoustic modeling of American English /r/", *Journal of the Acoustical Society of America* 108, 343–356 (2000)

[16] F.H. Guenther, C.Y. Espy-Wilson, S.E. Boyce, M.L. Matthies, M. Zandipour, J.S. Perkell, "Articulatory tradeoffs reduce acoustic variability during American English /r/ production", *Journal of the Acoustic Society of America* 105, 2854–2865 (1999)

[17] C.Y. Espy-Wilson, "Acoustic measures for linguistic features distinguishing the semi-vowels /w j r l/ in American English", *Journal of the Acoustical Society of America* 92(2), 736–757 (1992)

[18] P. Kerswill, "Children, adolescents and language change", *Language Variation and Change* 8, 177–202 (1996)

[19] J. Wells, *Accents of English*, Cambridge: Cambridge University Press (1982)

[20] A. Cruttenden, *Gimson's Pronunciation of English* (6th edition), London: Edward Arnold (2001)

[21] M.J. Jones, "An acoustic study of labiodental /r/ in British English", Ms., University of Cambridge (2005)

[22] P. Foulkes, G.J. Docherty, "Variation and change in British English (r)", In H. van de Velde, R. van Hout (eds.) *r-atics: Sociolinguistic, Phonetic, and Phonological Characteristics of /r/ (Etudes et Travaux 4)*, 27–43, Brussels: ILVP (2001)

[23] A.C. Gimson, *An Introduction to the Pronunciation of English* (3rd edition), London: Edward Arnold (1980)

[24] P. Trudgill, *The Social Differentiation of English in Norwich*", Cambridge: Cambridge University Press (1974)

[25] P. Trudgill, "Norwich revisited. Recent linguistic changes in an English urban dialect", *English World-Wide* 9, 33–49 (1988)

[26] J.D. Cohen, B. MacWhinney, M. Flatt, J. Provost, "PsyScope: A new graphic interactive environment for designing psychology experiments", *Behavioral Research Methods, Instruments, and Computers* 25(2), 257–271 (1993)

[27] R.P. Klein, "Acoustic analysis of the acquisition of acceptable r in American English", *Child Development* 42, 543–550 (1971)

[28] R.M. Dalston, "A spectrographic analysis of the spectral and temporal characteristics of English semivowels spoken by three year old children and adults", Northwestern University (1972)

[29] R.M. Dalston, "Acoustic characteristics of English /w, r, l/ spoken correctly by children and adults", *Journal of the Acoustical Society of America* 57, 462–469 (1975)

[30] P. West, "Perception of distributed coarticulatory properties of English /l/ and /r/", *Journal of Phonetics* 27, 405-426 (1999)