York Papers in Linguistics

PARLAY proceedings Series 1

Issue 1:

Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York (**PARLAY 2013**) conference
6th September 2013

April 2014
ISSN 1758-0315

Editor:
Theodora Lee

Editorial Note

The editors of the this volume of the York Working Papers offered us – as invited speakers at the inaugural PARLAY conference – the opportunity to introduce the volume to you, and in so doing, to give you a flavour of the PARLAY conference which these papers represent.

PARLAY (Postgraduate Academic Researchers in Linguistics at York) was created to provide a forum in which early career researchers in linguistics can meet, exchange ideas and receive input on their work. It is not the first such event to be proposed or hosted in the UK, so what is the special flavour of the York postgraduate research community which PARLAY seeks to share beyond York itself?

A hallmark of linguistic research in the Department of Language and Linguistic Science at the University of York – and thus of PARLAY – is the desire to blend the theoretical and the empirical, and a reluctance to treat one without the other. The eight papers in this volume – just a subset of the 32 papers presented at PARLAY 2013 – reflect this desire. The theoretical positions espoused are diverse, but every paper is backed up by empirical data of one type of another, ranging from corpus data (small or large), through acoustic analysis of production data (scripted or spontaneous) to behavioural experiments (auditory perception or eye-tracking). Five of the authors come from the postgraduate research community at York, but we are delighted that three authors from sister universities in Europe are also represented.

It was our pleasure to accept an invitation to speak at the inaugural PARLAY conference in 2013, and we commend this volume, and future iterations of the PARLAY conference to you. We wish to thank the editors for their work on the volume, and congratulate them on the success of the 2013 conference series. We trust that PARLAY will continue to serve the needs of postgraduate researchers in linguistics at York and beyond, for many years to come.

Francis Nolan, Cambridge; Sam Hellmuth, York
March 2014
http://www.york.ac.uk/depts/lang/ypl
http://www.parlayconference.blogspot.co.uk

Contents

ENGLISH STRESS AND UNDERLYING REPRESENTATIONS

QUENTIN DABOUIS
Université François Rabelais

Abstract

This paper addresses the issue of underlying representations (URs) from a Guierrian perspective and the necessity of taking into account certain orthographic elements which are associated with phonological behaviours related to English stress. We propose that underlying vowels should be represented as abstract phonological objects and that the inclusion of orthographic consonant geminates and final mute <e> into URs should be considered. Additionally, we argue that the phonology should have access to morphosyntactic information, which implies that the input to the phonology should be polystratal. Eventually, after arguing that vowel reduction should occur after stress assignment, we will report the results of studies on vowel reduction and "stress preservation" showing that reduction is not systematic in unstressed syllables and that non-reduction can, in some cases, be attributed to the existence of a full vowel in a morphologically-related word or to a high frequency of the latter.

1. Introduction

In most rule-based or constraint-based approaches in phonology, we are faced with the problem of URs and their content. In the approach introduced by Guierre, this problem has never been treated extensively even though the large corpus studies which have been conducted should allow us to address that problem within the approach itself, but also to bring elements of reflection with a more general reach. First, a presentation of the Guierrian School will be given along with an essential distinction to be made between studying individual speakers and studying a language. After giving the arguments in favour of the choice of the latter, a few problems in relation to the role of orthography in English phonology will be addressed along with two additional questions closely linked with English stress: the interface with morphology and vowel reduction.

2. The Guierrian School

The "Guierrian School" is an approach which was introduced in the seventies by the French researcher Lionel Guierre. During the sixties, Guierre computerised Daniel Jones's pronouncing dictionary and put together one of the largest corpora on English pronunciation at the time (35,000 words). He then studied it in its entirety. His PhD thesis (1979) was presented as an answer to Chomsky and Halle's *Sound Pattern of English* (1968, hereafter *SPE*), and his main critique regarding *SPE* was the absence of any empirical data backing up the rules formulated. Therefore, using this corpus, he tested the efficiency of the rules of stress placement and vowel pronunciation proposed in *SPE* and found that many rules were not very efficient when tested empirically.

However, Guierre never held that the rules he proposed were to be found in any given native speaker of English's phonology since his object of study was the English language.

 ISSN 1758-0315

Unlike *SPE*'s authors who quickly abandoned it, Guierre made an extensive use of the written form of words and included strictly orthographic elements such as consonant geminates, final mute <e> and vowel diagraphs in the parameters conditioning stress placement and vowel value. If Guierre had a few ideas on which of these elements were to be included in a lexical (or underlying) representation, many of the researchers following his work (*e.g.* Deschamps, Fournier, Trevian) did not all agree on these. However, Deschamps (1983) argued that English orthography is phonological. However, all agree that English phonology is strongly influenced by morphology, and we should consider how to include morphological information in URs as well. To sum up, the Guierrian approach is mainly characterised by the study of pronouncing dictionaries' data and the use of orthography when necessary.

This paper is an attempt to present some of these elements and the arguments that may or may not lead to including them in URs. The notation used in this paper is the one used by all authors working within the approach introduced by Guierre, which is the following:

- Angle brackets are used for orthography (*e.g.* <original>)
- Square brackets are used for phonetics (*e.g.* [əˈrɪdʒɪnəl])
- Slashes are used for phonology (*e.g.* / ? /)

Before investigating the contents of URs, it is necessary to specify the object of study, and more precisely "whose" URs are treated here.

3. Speaker VS Language

When one studies a phenomenon like English stress, it seems appropriate to define what is being studied exactly and if the aim is to study the language or individual speakers, as the means of investigation will be defined accordingly. Both options present some problems, a few of which are laid down below:

Speaker:

- How can we pick an "ideal", "average" or "representative" speaker?
- What tests should be used in order to get quality data?
- How can we get enough data so that it can be considered representative of a given speaker's phonology?

Language:

- How can we define the English language and its boundaries?
- What data should be used?
- Should it take into account the relative stratification of the language (*e.g.* items' frequencies, specialised lexical fields, learned vocabulary)?

Although we can assume that some rules found in a given speaker's phonology may be found in the language, it may not necessarily be the case, so it does not seem reasonable to say that what we find to be true for a given speaker will hold for the language as well, and vice versa.

One of the main characteristics of the Guierrian School is, as mentioned above, to have chosen to study the language[1]. Even though defining the language might be difficult, we will follow Saussure's definition:

[1] Even though the study of individual speakers is not excluded, it is just seen as distinct. Some Guierrian studies have already done tests on speakers (Martin, 2011) and more will most likely come.

> "It is both a social product of the faculty of speech and a collection of necessary conventions that have been adopted by a social body to permit individuals to exercise that faculty." (1916: 25)

In order to try and study the language, the choice has been made to analyse pronouncing dictionaries like Jones's *Cambridge English Pronouncing Dictionary* (hereafter CEPD) or Wells's *Longman Pronunciation Dictionary* (LPD). These dictionaries, even though not without some mistakes, allow us to access vast amounts of words that could be judged representative of the language[2]. The varieties described in these dictionaries are the standard ones (Received Pronunciation and General American), and the approach presented here does not claim that what is true for these varieties of English is true for all varieties. However, Martin (2011) conducted an intervarietal study of over 3,500 words known to be unstable in RP (*e.g.* with stress variants or belonging to classes which have numerous exceptions) and found that in over 90% of cases, the position of stress was the same in RP, GA and SAusE (Standard Australian English). Therefore, it seems that English, at least in Kachru's "Inner Circle" (1992), is very stable in its stress system.

Recent studies like Abasq et al. (2012) or Dabouis (2012) have also used frequency information from the *Corpus of Contemporary American English* to address the problem of rare or specialised vocabulary by excluding low-frequency items.

Thus, when one is using this methodology, it is possible to study a given class of words by analysing all the items of that class which can be found in the dictionaries. Then, it is possible to give the *productivity* of a class, the *efficiency* of the rule applying to it (if any) and the *list of exceptions* to the rule at stake. With such an inductive approach, formulating rules becomes a question of statistics as well as conventional analysis: a rule can only be called a rule if its efficiency is around 90%. If a phenomenon occurs in 70-90% of cases, we will be talking about a *tendency*, and under that figure we will be talking about chance. For example, Descloux et al (2010) looked at over 2,500 dissyllabic verbs and found the following stress patterns:

	Early Stress		**Late stress**		***Total***
	Nb	%	Nb	%	
Suffixed	177	74%	63	26%	*240*
Compounds	245	85%	44	15%	*289*
Prefixed	92	**7%**	1170	**93%**	*1262*
Bases	673	89%	85	11%	*758*
Total	*1187*	***47%***	*1362*	***53%***	*2549*

Table 1. Stress in dissyllabic verbs (from Descloux et al, 2010)

The figures in this table show that the usually assumed rule that dissyllabic verbs are stressed on their second syllable is not a rule at all as they are almost equally divided between early stress and late stress. However, the criterion which seems to be crucial here is prefixation, and we can formulate the rule "prefixed verbs are late-stressed" as it is the case in 93% of cases. Additionally, we could say there is a tendency for suffixed dissyllabic verbs to be early-stressed but, as English suffixes tend to have idiosyncratic properties, the latter should be studied individually and not only within this inventory.

[2] Collie (2008) calls the language described in CEPD "an artificial idiolect of English".

When we choose to study the language, we have access to two forms: the phonetic and the orthographic forms. To determine the contents of URs, we can draw from both accessible forms and them only, as doing otherwise would imply including elements in the representation which cannot be justified but by the theory itself. This reproach was indeed often made to *SPE* in the case of words like *right* or *nightingale* which would present an underlying velar fricative /x/ never attested on the surface explaining the pronunciation of their stressed vowels.

Moreover, psycholinguistic concerns such as the balance between computation and lexical storage, sometimes called "the division of labour" (Bermúdez-Otero & McMahon, 2006), which are essential when one is trying to model individual speakers' phonologies, are irrelevant here. Obviously, the chosen form of representation is determined by the nature of the rules or constraints which will be applied in order to derive the surface representation. For that matter, the rules to keep in mind here are the ones described by Fournier (2007, 2010b), which could be adapted into rule-based or constraint-based models.

4. Sketching Empirically-Based Underlying Representations

In this section, we will start with a presentation of the proposed representation of vowels, with three main points:

- vowels can be grouped in series which relate to the same orthographic vowel;
- there can be more phonological vowels than there are phonetic vowels (in a word);
- we cannot ignore the behaviour of words with final mute <e>.

Then we will tackle the main problem between orthography, phonology and phonetics with regards to consonants: written consonant geminates (*e.g.* <bb>, <rr>, <mm>).

Eventually, we will deal with the question of the interface with morphosyntax, as several morphosyntactic elements are needed in the representation to derive stress.

One principle to keep in mind during the following paragraphs and that we will adopt here is what Guierre (1979: 33) called the "uniqueness of lexical forms" (in French: "unicité des formes lexicales"). According to that principle, having two possible pronunciations for a same lexical unit does not entail that we must have one lexical form for each of them. For Guierre, *controversy* ([ˈkɒntrəvɜːsi][3] ~ [kənˈtrɒvəsi]) or *albino* ([ælˈbiːnəʊ] (GB) ~ [ælˈbaɪnoʊ] (US)) have only a single lexical form, and the "least bad" known representation of that form is the orthography (*Ibid.*).

4.1.Vowels

4.1.1. Series and their Representation

If we have a look at the following table[4] taken from Fournier (2010b: 98), we can see that each written vowel can have:

- four[5] different *values* when it is a monograph (*e.g.* r-coloured, checked, free and r-coloured free);

[3] All the phonemic transcriptions given in this paper are British pronunciations taken from LPD, unless indicated otherwise.
[4] This table concerns only stressed vowels, so it does not include reduced vowels.

- two different *values* when it is a diagraph (*e.g.* free and r-coloured free).

V^{r} r-coloured vowel	**V̆** checked vowel	**Monographs <V>**	**V̄** free vowel	**V̄^{r}** r-coloured free vowel	**Diagraphs <V̄V>**
[ɑː]	[æ]	<a>	[eɪ]	[eə]	<ai, ay / ei, ey>
[ɜː]	[e]	<e>	[iː]	[ɪə]	<ea, ee / ie**>
[ɜː]	[ɪ]	<i>	[aɪ]	[aɪə]	<ie*, ye>
[ɔː]	[ɒ]	<o>	[əʊ]	[ɔː]	<oa**, oe*>
[ɜː]	[ʌ (ʊ)]	<u>	[(j)uː]	[(j)ʊə]	<e(a)u, ew/ ue*>
			[ɔː]	[ɔː]	<au, aw>
			[uː]	[uː]	<oo>
			[ɔɪ]		<oi, oy>
			[aʊ]	[ɔː (aʊə)]	<ou, ow>

* : final ** : non-final

Table 2. Correspondences between orthography and phonetics for stressed vowels (after Fournier, 2010b: 98)

In Fournier's terminology, the vowels [ɑː], [æ], [eɪ], [eə] all have the same *quality*, *e.g.* they all derive from the same written vowel, they only differ in *value*. A set of very efficient rules[6] (summed up in Fournier 2010b: 141) determine which value a written vowel is going to have in a given context. If we take the orthography into account, it is easy to see that these vowels are related, but it does not mean that such a relationship has a phonological status[7].

However, morphologically-related pairs with vowel alternations seem to go in that direction:

(1)

V̄~ V̆	<divine> [dɪˈvaɪn]	~	<divinity> [dɪˈvɪnəti]
	<profane> [prəˈfeɪn]	~	<profanity> [prəˈfænəti]
	<serene> [səˈriːn]	~	<serenity> [səˈrenəti]
V^{r} ~ V̆	<isobar> [ˈaɪsəʊbɑː]	~	<isobaric> [ˌaɪsəʊˈbærɪk]
	<fluor> [ˈfluːɔː]	~	<fluoric> [flʊˈɒrɪk][8]
	<scar> [ˈskɑː]	~	<scarify> [ˈskærɪfaɪ]
V̄^{r} ~ V̆	<barbarian> [bɑːˈbeəri‿ən]	~	<barbaric> [bɑːˈbærɪk]
	<empire> [ˈempaɪə]	~	<empiric> [ɪmˈpɪrɪk]
	<compare> [kəmˈpeə]	~	<comparative> [kəmˈpærətɪv]

It seems that these alternations have a status in the language independently of the orthography. Another type of alternation needs to be taken into account before we can formulate a proposition for the representation of vowels: the alternation between "foreign vowels" and "indigenous vowels".

---

[5] That does not include foreign vowels, see §4.1.2.

[6] For example, the rule given in (10) or the rule associating words in *-ic* with a short vowel both have an efficiency of over 95%.

[7] As I am concerned here with the language, I will not argue for or against the reality of such vowels for English speakers. For a review on the study of vowel alternations, see Eddington (2001).

[8] This is the British pronunciation according to LPD. According to that same dictionary, American English has two possible pronunciations: [flʊˈɔːrɪk] ~ [flʊˈɑːrɪk], the first one respecting isomorphism with the base.

4.1.2. Foreign Vowels

Another of Guierre's arguments to postulate underlying vowels equivalent to orthographic vowels is the alternation between foreign[9] free vowels ($\bar{V}^e$) and indigenous[10] free vowels ($\bar{V}$). We can find three main foreign vowels:

	$\bar{V}$	$\bar{V}^e$
<a>	[eɪ]	[ɑː]
<e>	[iː]	[eɪ]
<i>	[aɪ]	[iː]

Table 3. Indigenous and foreign free vowels

Approaches which do not take into account the orthography usually do not have a concept of "foreign vowels", for phonetically the latter belong to the set of English vowels. What actually makes them "foreign" is precisely their relationship with the orthography, and following our proposition here, with their UR. When words containing these vowels get "naturalised" (*i.e.* adapted to the indigenous phonological system), the changes in pronunciation are highly predictable, especially the indigenous vowel which is going to be chosen, as in:

(2) <albino> [ælˈbiːn əʊ] (GB) ~ [ælˈbaɪnoʊ] (US)
<tomato> [təˈmɑːtə ʊ] (GB) ~ [təˈmeɪt̬oʊ] (US)

In these examples, we can see that British English has preserved a "foreign" pronunciation whereas American English has adapted these words to its phonological system by adapting the vowels: [iː] → [aɪ] ; [ɑː] → [eɪ]. Of all the values they could take, the vowels change to the indigenous vowel corresponding to the same orthographic vowel. This could be seen as a simple influence of orthography, but Guierre argued that it was because the two values found in these words (foreign and indigenous) relate to the same underlying vowels, in the case of the words in (2), /i/ and /a/.

We will not go as far as claiming that it is so, for it would imply to attribute features to these vowels, and we would require synchronic rules for the Great Vowel Shift, such as the Vowel Shift Rule (discussed in McMahon, 2007). The idea is not to say that all values derive from the same underlying vowel through various alterations, but only to say that they are related, and that they share a common *quality*. Therefore, the representation for vowels proposed here is one that refers to abstract objects representing a series of surface vowels, and these abstract objects do not have any phonological content. It is only the context in which these objects occur which is going to determine which *value* is going to surface. We can find series of minimal pairs with vowels having the same value but only differing by their quality (*e.g. pat, pet, pit, pot, put* for checked vowels; *mate, mete, mite, mote, mute* for free vowels). As far as notation is concerned, the option favoured here is that of using the orthographic symbols for these abstract objects (*e.g.* /e/ would represent the series [ɜː], [e], [iː], [ɪə] (+ foreign [eɪ])) but we could very well adopt other notations (*e.g.* V_e). The choice of the representation is potentially debatable, but is of secondary importance here.

[9] "Foreign vowels" are the ones which imitate the pronunciation of these vowels in the language from which the words usually containing these vowels are borrowed.
[10] "Indigenous vowels" are regular English vowels as shown in Table 2.

4.1.3. Diagraphs

There are two environments in which vowels are usually short, unless they are represented by a diagraph:

- Before a consonant cluster or before the final consonant of a word:

(3) <auction> [ˈɔːkʃən] ~ <action> [ˈækʃən]
<seat> [ˈsiːt] ~ <sit> [ˈsɪt]

- When they are found in the antepenultimate syllable[11]:

(4) <speechify> [ˈspiːtʃɪfaɪ] ~ <specify> [ˈspesəfaɪ]

Diagraphs could be represented by their orthographic form as well to refer to the reduced series that they represent (free vowels and r-coloured free vowels; *e.g.* /ee/ would represent the series [iː], [ɪə]). Another to formalise their behaviour could be to say that they function as diacritics of length, as the vowels they represent are usually long.

4.1.4. Synaeresis

There are words for which the count of vowels (and thus syllables) is different between the orthographic and phonetic forms, as in *ocean* and *partial*, which both have three orthographic vowels but only two phonetic vowels, as two vowels got historically compressed and reduced into one under synaeresis. However, these words exhibit an interesting behaviour when derived:

(5) <ocean> [ˈəʊʃən] → <oceanic> [ˌəʊʃiˈænɪk]
<partial> [ˈpɑːʃəl] → <partiality> [ˌpɑːʃiˈæləti]

In their derivatives, *oceanic* and *partiality*, the two vowels that were compressed into one are distinct as the strong suffixes *-ic* and *-ity* entail the placement of primary stress on the second syllable of the two. This leads us to think that there are not two but in fact three phonological vowels in *ocean* and *partial*, and that it is a reduction phenomenon which causes the last two to be compressed into one. In general, we consider that there are as many syllables as there are orthographic vowels[12] except in two cases:

- Vowel diagraphs, as they represent only a single vowel (*e.g.* <ea>: *sea* ≠ *react*);
- Final mute <e>, which will be examined below.

4.1.5. Final Mute <e>

Final mute <e> is well-known by English native speakers, for they learn about its special behaviour with regards to the vowel preceding it, so that it is sometimes called "magic e" by teachers who teach reading and writing. It constitutes what Guierre called a "tensing context", *i.e.* a context in which vowels are "tense"[13] when they are stressed or unreduced[14]. We could write down the rule for final mute <e> as follows.

---

[11] This phenomenon is usually known as Trisyllabic Shortening/Laxing, but Guierrian authors usually talk about "Luick's rule" Luick (1898), who first described the phenomenon.

[12] However, syllabic consonants (as in words in <Cm#>, <Cle#>, <Cre#>) are never counted as a phonological syllable, even though they can constitute a phonetic syllable.

[13] Guierre considered all the values given in Table 2 except V̆ to be "tense". However, in the case of final mute <e>, it concerns only V̄, V̄^{e} and V̄^{r}.

(6) $V \rightarrow \bar{V} / \_\_C^1e$

Examples in (7) illustrate this behaviour under stress, those in (8) when the vowel is unstressed and unreduced, and those in (9) when reduction is optional.

(7) *fate* (≠ *fat*) ; *bite* (≠ *bit*) ; *cote* (≠ *cot*) ;

(8) ˈ*demonstrate ;* ˈ*anecdote ;* ˈ*dynamite ;*

(9) ˈ*composite* [aɪ ~ ɪ] *;* ˈ*advocate* [eɪ ~ ə]

Orthographically, this rule can be extended to all vowels in this context, and can be formulated as in (10). Examples illustrating that rule are listed in (11).

(10) $V \rightarrow \bar{V} / \_\_C^1V\#$

(11) *a*ˈ*roma,* ˈ*baby, bi*ˈ*kini,* ˈ*coma, ko*ˈ*ala,* ˈ*photo, tor*ˈ*pedo, vol*ˈ*cano, …*

In this context, final mute <e> functions as a vowel, which is not very surprising for it was once pronounced (Bermúdez-Otero, 1998, Duffell, 2008, Minkova, 1982, 1991) and the rule in (10) was not an orthographic rule at the time. However, it functions mostly as a diacritic.

It should be noted that it never counts as a syllable when it comes to stress assignment. This is shown by the stress patterns in non-derived words of three syllables or more[15], which are subject to the Normal Stress Rule (NSR), when they are not prefixed non-substantives[16].

(12) **Normal Stress Rule**[17]
Words of three syllables or more are stressed on their penultimate syllable.

As illustrated by the following examples, final mute <e> never counts as a syllable:

(13) ˈ*envelope ;* ˈ*paradise ;* (and not **en*ˈ*velope ; *pa*ˈ*radise*)

Thus, final mute <e> seems to have a status in the language both through the phonological behaviours associated with its presence and simply through its existence in one of the two forms we have access to (in this case, the orthography). The question of its representation at the underlying level leaves two possible options. Either we can represent it:

- As the vowel /e/, which would have the same properties than other vowels overall but would have the specificity of being erased before stress assignment when found in final position. This could either be done by a deletion rule in a rule-based model or by a constraint forbidding its presence on the surface in a constraint-based model;
- As a diacritic, which has the advantage of not including an element in the representation that will need to be erased later in the derivation. However, in doing this we lose the parallelism between the behaviour of final mute <e> and that of other vowels.

As we leave this question open for now, let us turn to another element which presents similar problems: consonant geminates.

---

[14] Not all authors agree on the possibility for vowels to be unstressed and unreduced as they consider that vowel length is always associated to some level of stress (secondary or tertiary). For Guierre, this need not be the case, and he refuted the existence of stresses after primary stress (this view could be supported by the phonetic study by Plag & Kunter, 2011, who found that pre-tonic secondary stress was phonetically indistinguishable from primary stress, whereas it is not the case for post-tonic secondary stress).

[15] Here the syllable count does not include <e>.

[16] Indeed, prefixed non-substantives have a different behaviour: their prefixes are ignored when it comes to stress assignment (Fournier, 2007). An example of the behaviour can be seen in the study on verbs reported in Table 1.

[17] For the exact context of application of the NSR, see Fournier (2007, 2010a).

4.2. Consonant Geminantes

Consonant geminates are the main point of divergence between orthography, phonetics and phonology: they are represented orthographically by a consonant cluster; they behave like one phonologically but always surface as a single phonetic consonant. They are very efficient for the prediction of stress placement and vowel pronunciation:

(14) *Coˈlŏssus, vaˈnĭlla, ˌcigaˈrĕtte, ˈnŏbble* (≠ *ˈnōble*), *ˈtĭttle* (≠ *ˈtītle*), *ˈmĭrror*

As pointed out by Deschamps (1982), they also explain the behaviour of <u>, usually long when stressed unless it is before a consonant cluster or in a final closed syllable, in words like those in (15) which parallel the behaviour before a consonant cluster, as in (16).

(15) *butter, mutter, rubber*

(16) *function, pustule, vulture*

However, like final mute <e#>, all we can observe is a phonological behaviour, not its cause, as they are never pronounced as geminates. Once again, we have two options:

- Represent them as underlying geminates and simplify them later in the derivation;
- Represent them with a diacritic indicating their specific behaviour.

4.3. Interface with Morphosyntax

In his thesis (1979), Guierre argued against the idea that syllable weight was central in the phonological system of English[18], especially with regards to vowel quality. Vowels are extremely variable according to the context in which they appear, and it seems unreasonable to attribute an underlying value to them[19]. He also warned against potential circular logic which could be derived from that concept, such as *SPE*'s argument that vowels are long because they are stressed and that they are stressed because they are long[20].

Guierre argued that if the relationship between vowel length and stress is to be oriented, stress should come first. In fact, he argued that vowel values are determined mostly by the context in which they appear, and that stress allows for the expression of these values whereas reduction – when it occurs – obscures them. Thus, for him vowel value and stress assignment could be processed in parallel, and independently[21]. Only reduction needs to occur afterwards.

However, stress has been shown to be morphologically-influenced, and according to Fournier (1998), more than what is usually assumed. A very well known phenomenon is that of "strong suffixes", more often called "Class I suffixes" (Spencer, 1991), as opposed to stress-neutral suffixes, "Class II suffixes"[22]. Here are a few examples:

(17) Class I: *-ion, -ity, -ic, -ate,…*
Class II: *-ness, -less, -hood, -ful, -ly,…*

[18] So did Fournier, 2010b.
[19] Except maybe in the case of vowels which are represented by diagraphs, which are almost always long.
[20] This logic is still found in more recent articles, as in Duanmu (2010), where we can read: "Stressed syllables are heavy/Heavy syllables are stressed".
[21] However, this view is not shared by all authors who worked after him. For Fournier, one can predict the value of a vowel efficiently only when that vowel is bearing stress. In other terms, vowel value comes only once stress assignment has taken place.
[22] Even though the definition Fournier gives of strong suffixes is not exactly equivalent to the common assumptions about Class I suffixes, for more details, see Fournier (1998).

This difference could be formally represented the way it is represented in Lexical Phonology (Kaisse & Shaw, 1985, Kiparsky, 1982) or, more recently, in Bermúdez-Otero's version of Stratal OT (2012), with Class I rules applying at the "stem-level" and Class II rules applying at the "word-level"[23], even though it would require some adaptations[24].

In the Guierrian approach, the main parameter conditioning English phonology is thought to be morphology, and semantically-defined morphological domains. Different morphological units behave differently phonologically, thus the notion of morphologically-defined phonological domain is definitely relevant.

URs should then include information relative to these morphologically-defined phonological domains, which we could represent with labelled brackets corresponding to the different strata, such as that which can be found in many generative works, like the following examples for *regarding* and *classifiable*:

(18)

$$[_{w}\,[_{s}\,\text{re+gard}]_{s}\,\text{iŋ}]_{w}$$
$$[_{w}\,[_{s}\,[_{s}\,\text{class}]_{s}\,\text{ifi}]_{s}\,\text{able}]_{w}$$

(W: word level; S: stem level)

Additionally, as illustrated by the rule concerning prefixed non-substantives, information about syntactical categories is necessary for stress assignment. Thus, we could use another labelled bracketing in order to display that information:

(19)

$$[_{V}\,[_{V}\,[_{P}\,\text{re}]_{P}\,[_{R}\,\text{gard}]_{R}\,]_{V}\,\text{iŋ}]_{V}$$
$$[_{A}\,[_{V}\,[_{N}\,\text{class}]_{N}\,\text{ifi}]_{V}\,\text{able}]_{A}$$

(P: prefix; V: verb; A: adjective; N: noun; R: root)

To sum up, when it comes to English phonology, we need information from other linguistic levels (or modules)[25], and that requires that the input to the phonology should contain all that information. Therefore, the input would have to be polystratal. The details concerning the contents or the number of these strata are, however, well beyond the scope of this paper.

5. Vowel Reduction

If we consider vowel values to be predictable by the context in which vowels appear, we need to mention cases in which that value is not expressed, *e.g.* when reduction occurs[26]. Reduction is usually seen as the "norm" in unstressed syllables, especially those adjacent to primary

[23] The assignment of a given suffix to a given level is seen as an idiosyncratic property of that suffix.

[24] For example, it seems problematic to include both non-derived words and words containing Class I suffixes in the stem-level when only the former are subject to the rule described in note 16. Additionally, the concept of "fake cyclicity" proposed by Collie (2008) after Bermúdez-Otero & McMahon (2006) and Bermúdez-Otero (in preparation) seems adapted to capture the relationships of isomorphism between morphologically-related words, especially the idea that the whole form of an embedded word is accessible to the phonology to derive the surface form of a related embedding word, which is an idea that is usually assumed by Guierrian authors.

[25] Which can itself be conditioned by yet other linguistic levels, as when semantics condition morphological boundaries (*e.g. reform ~ re-form* → one unit versus two units, semantic and morphological).

[26] Reduction is seen here as resulting from stress.

stress. A recent PhD dissertation by Dahak (2011) showed that this was not the case, as the percentage of unreduced vowels varied considerably according to the syllabic position under scrutiny.

Syllabic rank	% of non-reduced vowels
1**0**0	3.4%
1**0**00	5.1%
-2**0**1-	22.2%
-1**0**	26.6%
1**0**	33.9%
-10**0**0	53.6%

Table 4. Percentage of non-reduction in unstressed syllables according to the syllabic rank[27]

As we can see in this table, the highest proportion of reduced vowels is found in post-tonic syllables which are adjacent to the primary-stressed syllable in three or four syllable words. However, the lowest proportion of reduction is found in the middle syllable of a final string of three unstressed syllables.

Some other recent studies (Collie 2007, Kraska-Szlenk, 2007) have shown what are often called "stress preservation"[28] effects. Indeed, they have showed that the relative frequency of a base and its derivative had an influence on stress preservation. Indeed, if the derivative is more frequent than the base, then reduction is more likely to occur, and vice versa, as evidenced in Table 5 below.

			(x per 10^6 words in spoken section of COCA)		
			base		derivative
a.	*cyclic stress*				
	cond[ɛ́] mn	còndˌ[ɛ] mn-ation	7.09	>	2.57
	imp[ɔ́]rt	ìmp[ɔ̀]rt-ation	5.15	>	0.62
b.	*variable stress*				
	cond[ɛ] nse	còndˌ[ɛ ~ ə]ns-ation	0.28	≈	0.22
c.	*noncyclic stress*				
	cons[ɝ́]rve	còns[ə]rv-ation	1.65	<	9.11
	trà␣nsp[ɔ́]rt	trà␣nsp[ə]rt-ation	7.23	<	23.54

Table 5. Stress preservation and frequency (from Bermúdez-Otero, 2012: 32, after Kraska-Szlenk 2007: §8.1.2)

[27] The figures 1, 2 and 0 between slashes are commonly used by Guierrian authors to represent stress patterns: /1/ stands for primary stress, /2/ for secondary stress and /0/ for unstressed syllables. Therefore, the stress pattern of *ˌciga'rette* can be represented as /201/.

[28] However, within the framework presented here this phenomenon would be described in terms of non-reduction of unstressed vowels in related derivatives.

We found a similar effect in Abasq et al (2012) and showed that, in dissyllabic prefixed noun/verb pairs (*e.g. record, concern, process*), vowel reduction is less likely to occur if there exists a stress variant in which that vowel bears stress, *i.e.* there are more chances of having the noun *record* pronounced [ˈrekɔːd] than [ˈrekəd] because of the existence of the verb pronounced [riˈkɔːd]. Additionally, it was found that, for these words, reduction was more widespread in prefixes than in roots, hinting that different morphological units may have a different "resistance" to reduction, for instance that prefixes tend to reduce more than roots. This second finding is in line with a phenomenon Guierre described in his thesis (1979: 253): initial pretonic vowels followed by a consonant cluster tend to remain unreduced, except when they are monosyllabic prefixes, in which case they massively undergo reduction.

Eventually, one could ask what should be done about the representation of vowels which are always reduced. Precisely because they are always reduced, their quality can never be accessed phonetically. Guierre argued that if one should form a new word like **cymbalic*, it would most likely be pronounced [sɪmˈbælɪk], even though in *cymbal*, the second vowel is completely reduced. In that case, *-ic* predicts the position of stress and the value of the stressed vowel (checked), but it is the orthography which provides the vowel quality (<a>). Without the orthography, any checked vowel could be used, but the orthography determines which one is to be chosen[29].

6. Conclusion

We have argued in favour of an abstract UR of English vowels to account for a variety of surface alternations. The choice of such a representation raises the question of the abstractedness of URs. We believe that abstract forms can be justified if they can be supported by phonological or orthographic evidence and that they need not have any phonological content as long as they are distinctive.

Vowel diagraphs, consonant geminates and final mute <e> could all be formalised as diacritics or in a form close to orthography. However, the choice between these two options seems futile as both imply added lexical information. We would personally favour the second option, as it would be closer to one of the two accessible forms.

Furthermore, we have proposed that the input to the phonology should contain information from other linguistic levels and should therefore be polystratal.

Eventually, we discussed the issue of vowel reduction, the view that the latter obscures the quality of vowels and the idea that this quality could potentially be maintained by the existence of morphologically-related word in which it is not obscured.

To conclude, it seems difficult to study English phonology with no reference to orthography and the latter should probably be taken into account more often. Some arguments were presented to demonstrate an influence of orthography in the language, such an influence can also be found when studying individual speakers, as in Taft's psycholinguistic experiments (2006) seem to show.

[29] For this example, phonological priming would of course exclude [sɪmˈbɒlɪk] (i.e. *symbolic*), but not [sɪmˈbelɪk] nor even [sɪmˈbʌlɪk].

Acknowledgements

We are grateful to J.-M. Fournier, M. Martin, M. Cotton-Kinch and two anonymous reviewers for their corrections and constructive comments. Responsibility for any mistakes or omission is mine alone.

References

ABASQ, V., DABOUIS, Q., DESCLOUX, E., FOURNIER, J-M., FOURNIER, P., GIRARD, I., MARTIN, M., VANHOUTTE, S. 2012. "Stress in Dissyllabic Verb/Noun Pairs", Poster presented during the 20th MFM (Manchester Phonology Meeting) at the University of Manchester, United Kingdom (25-27th May).

BERMÚDEZ-OTERO, R. In preparation. *Stratal Optimality Theory: Synchronic and diachronic applications*. Cambridge: Cambridge University Press.

BERMÚDEZ-OTERO, R. 2012. "The Architecture of Grammar and the Division of Labour in Exponence", Trommer, Jochen (ed.). *The morphology and phonology of exponence (Oxford* Studies in Theoretical Linguistics 41). Oxford: Oxford University Press, 8-83.

BERMÚDEZ-OTERO, R. 2011. "Cyclicity" Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice (eds). *The Blackwell companion to phonology*. Malden, MA: Wiley-Blackwell.

BERMÚDEZ-OTERO, R. 2003. "The acquisition of opacity", in Jennifer Spenader, Anders Eriksson, and Östen Dahl (eds), *Variation within Optimality Theory: Proceedings of the Stockholm Workshop on 'Variation within Optimality Theory'. Stockholm: Department of Linguistics, Stockholm* University, 25-36.

BERMÚDEZ-OTERO, R. 1998. "Prosodic optimization: the Middle English length adjustment". English *Language and Linguistics* 2: 169-197.

BERMUDEZ-OTERO, R. & MCMAHON, A. 2006. "English Phonology and Morphology" Aarts, Bas, & McMahon, April (2006). *The handbook of English linguistics*. Oxford: Blackwell. 382-410.

CHOMSKY, N. & HALLE, M. 1968. *The Sound Pattern of English*, Cambridge, Massachusetts, London, England: The MIT Press.

COLLIE, S. 2008. 'English stress preservation: the case for "fake cyclicity"', *English Language and Linguistics* 12 (3): 505-32.

COLLIE, S. 2007. *English stress-preservation and Stratal Optimality Theory.* PhD thesis, University of Edinburgh, Edinburgh. Available as ROA-965-0408, Rutgers Optimality Archive, http://roa.rutgers.edu.

DABOUIS, Q. 2012. "Types of Derivation and Stress Placement: the Case of Suffixed Adjectives in *-al*", communication pronounced at the international conference "Interfaces in English Linguistics" (IEL) 2012 at Károli Gáspár University in Budapest, Hungary (12^{th}-13^{rd} October).

DAHAK, A. 2011. *Étude diachronique, phonologique et morphologique des syllabes inaccentuées en anglais contemporain* [*Diachronic, phonological and morphological study of unstressed syllables in contemporary English*]. PhD Thesis, Université de Paris Diderot.

DAVIES, M. 2008- . *The Corpus of Contemporary American English*: 425 million words, 1990-present. Online: http://corpus.byu.edu/coca

DESCHAMPS, A. 1982. "L'orthographe de l'anglais est-elle phonologique ?" ["Is English orthography phonological?"], *Colloque d'Avril sur l'anglais oral, 68-96, Villetaneuse : Université de Paris-Nord, CELDA,* diffusion APLV.

DESCHAMPS, A., O'NEIL, M. 2007. "Obituary: Lionel Guierre". *Language Sciences* 29, 492–495

DESCLOUX, E., GIRARD, I., FOURNIER, J.-M., FOURNIER, P., MARTIN, M. 2010. "Structure, Variation, Usage & Corpora: The Case of Word Stress Assignment in Disyllabic Verbs". Communication pronounced at the 2010 PAC Workshop in Montpellier, France (September 2010).

DUANMU, S. 2010. "Onset and the Weight-Stress Principle in English" dans *Linguistic essays in honor of Professor Tsu-Lin Mei on his 80th birthday*, ed. HONG Bo, WU Fuxiang, and Chaofen SUN. Beijing: Commercial Press.

DUFFELL, M. J. 2008. *A New History of the English Metre*. Volume 5 of Studies in Linguistics, MHRA.

EDDINGTON, D. 2001 "Surface analogy and spelling rules in English vowel alternations", *Southwest Journal of Linguistics*, 20. 85-105.

FOURNIER, J.-M. 2010a. "Accentuation lexicale et poids syllabique en anglais : l'analyse erronée de Chomsky et Halle" ["Stress and syllable weight in English: Chomsky and Halle's erroneous analysis"], Communication pronounced at the 8th Rencontres Internationales du Réseau Français de Phonologie at the University of Orléans, France (1-3rd July).

FOURNIER, J.-M. 2010b. *Manuel d'anglais oral* [*Oral English manual*]. Paris: Ophrys.

FOURNIER, J.-M. 2007. "From a Latin syllable-driven stress system to a Romance vs Germanic morphology-driven dynamics", *Language Sciences* 29, 218-236, London: Elsevier.

FOURNIER, J.-M. 1998. "Que contraignent les terminaisons contraignantes ?" ["What are strong endings?"].Topiques, *Nouvelles recherches en linguistique anglaise*, Travaux XCIII du CIEREC, Publications de l'Université de Saint-Etienne.

FOURNIER, J.-M. 1996. "La reconnaissance morphologique" ["Morphological recognition"], 8ème Colloque d'Avril sur l'anglais oral, 45-75, Villetaneuse : Université Paris-Nord, CELDA, diffusion APLV.

GUIERRE, L. 1982. "Grammaire et lexique en phonologie : Les oppositions accentuelles catégorielles" ["Grammar and Lexicon in Phonology : Categorial Stress Oppositions"], 1er colloque d'avril sur l'anglais oral [1st April conference on oral English], J.-L. Duchet, J.-M. Fournier, J. Humbley & P. Larreya, éd., Université de Paris-Nord : CELDA, diffusion APLV.

GUIERRE, L. 1979. *Essai sur l'accentuation en anglais contemporain – Eléments pour une synthèse* [*Essay on stress in contemporary English – Elements for a synthesis*], Paris, Université Paris-VII.

JONES, D. 2006. *Cambridge English Pronouncing Dictionary*, 17th edition, edited by P. Roach, J. Setter & J. Hartman, Cambridge: CUP.

LUICK, K. 1898. Beiträge zur englischen Grammatik III. Die Quantitätsveränderungen im Laufe der englischen Sprachentwicklung. Anglia 20: 335±62.

KACHRU, B. B. 1992. *The Other Tongue: English across Cultures*. Urbana and Chicago: University of Illinois Press.

KAISSE, E. & SHAW, P. 1985. "On the Theory of Lexical Phonology", Phonology Yearbook 2, 1-30.

KIPARSKY, P. 1982 "From Cyclic to Lexical Phonology" in H. van der Hulst and N. Smith, eds., *The Structure of Phonological Representations*. Part I, Foris, Dordrecht, 131-176.

KRASKA-SZLENK, I. 2007. Analogy: the relation between lexicon and grammar (LINCOM Studies in Theoretical Linguistics 38). Munich: LINCOM Europa.

MARTIN, M. (2011), *De l'accentuation lexicale en anglais australien standard contemporain* [*Of Lexical Stress in Contemporary Standard Australian English*], PhD Thesis, Université de Tours.

MCMAHON, A. 2007. "Who's afraid of the vowel shift rule?", *Language Sciences* 29, 341–359. London: Elsevier.
MINKOVA, D. 1991. *The History of Final Vowels in English. The Sound of Muting*. Topics in English Linguistics 4; Berlin, New York: Mouton de Gruyter.
MINKOVA, D. 1982. "The environment for open syllable lengthening in Middle English". *Folia Linguistica Historica* 3.2.29-58.
PLAG, I., KUNTER, G. 2007. "The phonetics of primary vs. secondary stress in English". Online at http://www2.uni-siegen.de/~engspra/DFG-Project/Plag-Kunter-2007.pdf (accessed 3/09/2013).
SAUSSURE, F. DE. 1916. *Cours de linguistique générale* [*Course in General Linguistics*]. Payot. (2005 edition).
SPENCER, A. 1991. *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Cambridge: Cambridge University Press.
TAFT, M. 2006. "Orthographically influenced abstract phonological representation: Evidence from non-rhotic speakers." *Journal of Psycholinguistic Research*, 35(1), 67-78.
TREVIAN, I. 2003. *Morphoaccentologie et processus d'affixation de l'anglais* [*English morphoaccentology and affixation processes*], Bern: Peter Lang.
WELLS, J.C. 2008. *Longman Pronunciation Dictionary*, 3rd edition, London: Longman.

Quentin Dabouis
Laboratoire Ligérien de Linguistique (lll@univ-tours.fr)
Université François Rabelais
Tours
France
Email : quentin.dabouis@gmail.com

INTRODUCING CONTEMPORARY PALATALISATION

OLIVIER GLAIN
Université Jean Monnet de Saint-Etienne
Université Jean Moulin – Lyon 3

Abstract

New instances of palato-alveolars have been reported in various regions and countries of the English-speaking world over the past few decades. Those new instances of /ʃ, tʃ, ʒ, dʒ/, which I propose to bring together under the common name of *Instances of Contemporary Palatalisation* (henceforth, ICPs), are the results of processes of palatalisation. I argue that ICPs can be seen as the products of a phonological process that finds its roots in the context of significant post-World War Two social and political changes, both within and beyond the British Isles. What is common to all four ICPs is not the input but the output of the process, that is to say systematic palato-alveolars. The phonological view adopted is therefore that Contemporary Palatalisation expresses product-oriented, rather than source-oriented generalisations (cf. Bybee, 2001: 126). I argue that ICPs are the continuation of a historic pattern endemic to English, which has systematically yielded palato-alveolars throughout the history of the language. This paper presents selected results of a corpus study based on a PhD project undertaken at the University of Lyon.

1. Introduction

Palatalisation is a particularly productive process that lies at the heart of linguistic change in Indo-European languages. While some instances of word palatalisation are considered to be standard in a great many varieties of English (e.g. *issue*, *nature*), the status of others, mostly associated with the latter part of the 20th century and the beginning of the 21st century, is more controversial from a prescriptive point of view. I have labelled such phenomena *Instances of Contemporary Palatalisation* (ICPs). As to the phonological process which leads to ICPs, I have called it *Contemporary Palatalisation* so as to distinguish it from previous processes in the history of the English language. ICPs are variants mostly associated with younger speakers.

First, I will focus on the process of palatalisation that underlies ICPs and explain how it operates in four phonetic environments. In order to see how ICPs fit within the larger context of sound change at the end of the 20th century, the reader will be reminded of the correlation that can be found between World-War two social and linguistic changes in Britain and in the USA.

After briefly surveying the different varieties of English in which ICPs may occur, I will explain how the phenomenon mirrors diachronic patterns that have led to palatalisation throughout the history of English. Then, I will present some of the results taken from a corpus study undertaken for my PhD thesis. Finally, I will propose an explanation for the development of ICPs within the framework defined by Smith (2007), whose cognitive model of change posits that sound change typically goes hand in hand with *social* change. Smith's

 ISSN 1758-0315

theory will be used to explain the emergence of ICPs in the context of significant post-World War Two social and political changes, both in Britain and in the USA.

2. Instances of Contemporary palatalisation (ICPs)

2.1 Yod coalescence after /t, d/ *in stressed syllables*

This is a type of assimilation where the approximant /j/ (yod) fuses, or coalesces, with preceding /t, d/, resulting in affricates /tʃ, dʒ/, e.g. *tune* /ˈtjuːn/ → /ˈtʃuːn/ ; *dune* /ˈdjuːn/ → /ˈdʒuːn/. The yod is the assimilator.

In an article on Received Pronunciation (RP), Wells (1997: 22-23) writes that 'English has long had a tendency to convert /tj/ into /tʃ/ and /dj/ into /dʒ/' (e.g. *nature*). Indeed, in Early Modern English, borrowings from French were gradually palatalised in items like *nature* and *fortune* (Crystal 2003). Wells observes that the process spread to new words in the mid-twentieth century to include words like *actual, perpetual, gradual, graduate*, whose everyday forms contain the affricate /tʃ/ or /dʒ/, their variants with /j/ being 'mannered' or 'artificial'. Wells finally notes that a new change occurred in the late 20th century, whereby coalescence continued to 'widen its scope', to reach stressed syllables in words like *Tuesday, dune*. Following the results of my corpus study (cf. section 5.1), I suggest that this is the period of history when all four ICPs really started to diffuse into the community. Therefore, ICPs are originally non-standard variants whose diffusion has become noticeable since the last few decades of the 20th century.

First, there was considerable resistance to accept coalescence in stressed syllables, as is shown in Ramsaran 1990, Wells 1997 and in various editions of the *Longman Pronunciation Dictionary* (henceforth, LPD), and the *English Pronouncing Dictionary* (henceforth, EPD). Linguists were reluctant to consider it as part of the standard accent and to include it in the description of RP. However, a study by Hannisdal (2006) turned the tables. As a direct consequence of her work, Wells decided to include coalescence in stressed syllables into descriptions of RP (LPD 2008). On his blog, he even explains that he had been wrong not to do it before:

> In LPD I labelled these variants "non-RP". Clearly I was wrong to do so (even if it's true for people of my own advanced age).
>
> (Wells 2007)

Wells implies that these variants constitute a change in progress, mostly associated with younger speakers, which is confirmed by various pronunciation preference polls in LPD 2008. Cruttenden (2008: 81) also considers that yod coalescence in stressed syllables as a change that is "well-established" in RP.

Of course, most accents of North America and other varieties in which yod elision is particularly prominent do not display this ICP.

2.2. Yod palatalisation after /s, z/ *in stressed syllables*

Words like *presume* and *assume* have traditional forms /prɪˈzjuːm, əˈsjuːm/ that can be palatalised into /prɪˈʒuːm, əˈʃuːm/. The assimilator is /j/, which retracts the articulation of both /s/ and /z/. Therefore, underlying /sj/ is palatalised into /ʃ/ and underlying /zj/ is

palatalised into /ʒ/. The phenomenon is widely accepted in unstressed syllables but it is less so in stressed syllables. For instance, the palatalised variants are listed as non-standard variants in LPD 2008 and they are not even listed at all in EPD (the remarks made in this article concern all editions of EPD. This is the reason why no date is mentioned.). The phenomenon is not as common as coalescence in /tju, dju/ sequences, partly because yod dropping is more common in these environments and partly because fewer items contain /sju, zju/ than /tju, dju/ sequences in the English lexicon. Nevertheless, these variants appear to be progressing, even in the standard accent:

> Coalesced forms in the onset of accented syllables, e.g. in *assume*, *presume* are increasingly heard in RP
>
> (Cruttenden 2008: 227)

As with the first ICP, most accents of North America and other varieties in which yod dropping is particularly prominent do not display this ICP.

2.3. Palatalisation of /s/ *in initial* /st, str, stj/ *clusters*

Our third ICP concerns words like *stop, start, stress, street, student, stew,* which display the palatalised variants /ˈʃtɒp, ˈʃtɑːt, ˈʃtres, ˈʃtriːt, ˈʃt(j)uːdənt, ˈʃt(j)uː/. In items like *stress*, *street* and *student*, *stew*, it is the /r/ and the /j/ which respectively retract the articulation of the alveolar fricative. In the case of /st/, the cluster that is the least likely to yield palatalisation, identification of a particular assimilator is much less obvious, as /s/ and /t/ are both alveolars. Such instances of palatalisation may well be the result of a *paradigmatic* type of assimilation[1].

EPD does not list any palato-alveolars in these environments while the existence of palatalised variants of /str, stj/ is only mentioned in passing in LPD 2008 (52).

In a study about the palatalisation of /str/ clusters, Rutter (2011) uses acoustic measurements to compare ten English speakers' realisations of the onsets /ʃ/, /ʃr/, /ʃtr/, and /s/. He finds out that most of the occurrences of /str/ clusters produced by these speakers fall within their normal range for /ʃ/, as opposed to various intermediate phonetic realisations falling somewhere between a typical /ʃ/ and /s/. The results indicate that the change towards /ʃ/ is complete for those speakers.

2.4. Palatalisation of /s/ *by* /r/

This fourth- and last- ICP can be found in items like *anniversary, classroom, estuary, grocery, nursery*, which have palatalised variants /ˌænɪˈvɜːʃri, ˈklɑːʃruːm, ˈeʃtjʊri, ˈgrəʊʃri, ˈnɜːʃri/ (LPD 2008). It is again the /r/ that triggers the assimilation process by retracting the articulation of /s/. These palatalised variants are listed in LPD, for both British and American English, but not in EPD.

[1] Paradigmatic assimilations occur when sounds interact on a paradigmatic axis (Pavlík 2009: 5). The high frequency of palatalised variants of /str, stj/ patterns may contribute to the palatalisation of the /st/ cluster.

3. The diachronic aspect of English palatalisation

Overall, linguists have shown reservations as to the inclusion of ICPs within standard English accents. It therefore seems perfectly reasonable to assume that they are originally non-standard variants. From a diachronic point of view, ICPs seem to be the continuation of a historic process whereby palato-alveolar fricatives and affricates have gradually diffused into English. It is assumed that the only proto-Germanic palatal was /j/ (Stévanovitch 2008: 21). Throughout the history of the English language, new palato-alveolars have repeatedly been created as allophones of pre-existing phonemes. The phonetic innovations thus produced were eventually fossilised (phonologised).

(1) In Old English, [k] → [tʃ] in certain environments (e.g. Old English *Cinn* [k] → Contemporary English *chin* [tʃ]; Old English *tæcan* [k] → Contemporary English *teach* [tʃ]).

(2) In Middle English, then Modern English (15th century: rare, 16th end 17th centuries: more common), [sj] → [ʃ] (e.g. *nation* [nasjõ] → [neɪʃən]; *sure* [syr] → [sjuːr] → [ʃʊə]).

(3) In the 17th century [zj] → [ʒ] (e.g. *measure* [məzyr] → [mezjuːr] → [meʒə]).

(4) In the 17th century [tj] → [tʃ] (e.g. *nature* [natyr] → [naːtjuːr] → [neɪtʃə]; *fortune* [fortyn] → [fortjuːn] → [fɔːtʃən]).

(5) In the 17th century [dj] → [dʒ] (e.g. *soldier* [soldjər] → [səʊldʒə]).

(Stévanovitch 2008: 24).

4. Post-World War Two social and linguistic changes in Britain and the USA

In Britain, RP began to lose ground in the second part of the 20th century. At the same time, media's interest in non-standard pronunciations arose. Regional accents appeared on the BBC[2] and have been on the increase in all media ever since. That phenomenon has had tremendous psychological repercussions. Prescriptivism in pronunciation had been particularly well-developed since the end of the 18th century. As a result of that long prescriptive tendency, non-standard speakers of British English felt linguistically insecure. *Linguistic insecurity* can be defined as the lack of confidence experienced by speakers when they believe that the way they speak does not conform to - and is inferior to - the standard variety (Calvet 2011: 47). Linguistic insecurity started to decline with the increase of exposure of non-standard varieties in the media.

The decades following World War Two were characterised by significant social changes in Britain. Hannisdal (2006) makes a link between this socio-historical context and the decline of RP:

> Up until the middle of the 20th century RP reigned supreme as the unrivalled English pronunciation standard. But in the decades after the Second World War Britain underwent radical social changes which also left their marks on the linguistic development and on the attitudes towards accent. Along with the general social changes, the role of RP also changed considerably. Between 1944 and 1966 the number of universities in Britain doubled and higher education became available to people from diverse social backgrounds. The increased democratisation in the

[2] In the 1920s, the BBC had decided that all its presenters had to be RP speakers (hence the term *BBC English*). That decision would help give special importance to RP. 'By using only RP speakers as announcers and newsreaders, the BBC underlined the social importance of the accent, and in the public mind RP became even closer linked with high status and intellectual competence' (Hannisdal 2006: 13).

> educational system extended into the occupational and public life. Professional and academic careers became open to people from the lower social strata, who of course were non-RP speakers. Regional accent features "massively invaded the realms of the social elite" (Wotschke 1996: 221) and the hegemony of RP was broken. An educated speaker was no longer synonymous with an RP speaker, and RP was no longer the exclusive property of a narrow social class.
>
> (Hannisdal 2006: 15)

In 1970, Gimson wrote:

> The acceptance of the BBC accent, *i.e.* some form of RP, as a standard can no longer be said to be common amongst the younger people. The social structure of the country is much less rigid than it was forty years ago, and the young are particularly apt to reject authority of any kind. This general rejection includes the accent of the "Establishment", *i.e.* RP.
>
> (Gimson 1970: 18-19)

Thus, the decline of RP coincided with major social changes that went hand in hand with an increasingly equalitarian ideology in Britain. Kerswill (2007: 38) also notes that the loss of RP's privileged status accompanied the strong social mobility in post-World War Two Britain. In addition, the emergence of non-standard pronunciations in new contexts should be seen 'in the context of the ideology, first emerging in the 1960s, of gender and racial equality and the legalisation of contraception, abortion and homosexuality – coupled with a generally greater access to education' (Kerswill 2007: 51).

The decline of RP as a sort of international standard after World War Two extended far beyond Europe at a time when Britain lost its Empire. In the United States, some sort of international English based on RP served as a model in high society. Thus:

> r-less pronunciation, as a characteristic of British Received Pronunciation, was also taught as a model of correct, international English by schools of speech, acting, and elocution in the United States up to the end of World War II. It was the standard model for most radio announcers and used as a high prestige form by Franklin Roosevelt[3].
>
> (Labov, Ash & Boberg 2006: 46)

Furthermore, the post-World War Two decades also witnessed the decline of a certain form of formal speech in public contexts in the USA.

The art of oratory has long been part and parcel of American culture. One easily recalls having heard important speeches from the 20th and the beginning of the 21st century (e.g. John Fitzgerald Kennedy, Martin Luther King, Barack Obama). Some other speeches have been passed down from one generation to the next even in the absence of recordings (e.g. Abraham Lincoln). McWhorter (2012: 109) explains that listening to speeches used to be a form of entertainment in the USA. For example, 'before Abraham Lincoln delivered the Gettysburg Address, a professional orator named Edward Everett delivered a two-hour formal speech to entertain the crowd'. McWhorter (2012: 109-110) observes that the 20th century witnessed a gradual shift from formal and written-based types of public speaking to speeches that became more informal and much closer to real spoken English. The shift was part of a more general change. Indeed:

> the difference between spoken and written language has been key in a general transformation in American language culture over the past several decades from one focused on written forms to one focused on spoken ones. This has been influenced in part by the spread of recording technology and in part by late 20th-century countercultural movements that rejected traditional forms of oratory.
>
> (McWhorter 2012: 109)

[3] A passage from a speech by F.D. Roosevelt can be heard at the address below. This speech exhibits a great many similarities with RP.
http://www.youtube.com/watch?v=4Wo9Q3WJHjA

Let us leave aside the question of technology and focus on the countercultural movements. McWhorter (2012: 111) explains that the major changes came after the 1960s 'and the general trend toward questioning the establishment'. One of the linguistic consequences of the ideology of the time was that a more natural, much less ceremonial form of language became the preferred style, which had longer-term repercussions:

> By 1981, at which point countercultural America had settled back down into something more conservative again, American rhetoric, even in the most formal settings, had changed. Modern speechmaking was more like talking, and orators took pride in sounding more like the common man.
>
> (McWhorter 2012: 111)

5. The spatial and temporal dimension of ICPs

ICPs have perhaps mistakenly been associated with the south-east of England and/or with Estuary English (Altendorf 2003: 69, Altendorf and Watt 2008: 213, Cruttenden 2008: 87, Coggle 1993: 51-52, LPD 2008: xix, Wells 1982: 331). However, Contemporary Palatalisation has been noted in many varieties of English in England (including RP), in Scotland, Canada, Australia, New Zealand, South Africa, as well as in several varieties of US English (see Glain 2013: 141-147 for detailed references).

As it is often implied in the literature that the variants that I call ICPs are associated with younger speakers, I wanted to check whether they constituted a change in progress. Such changes should be observed through the apparent time method:

> The first and most straightforward approach to studying linguistic change in progress is to trace change in apparent time: that is, the distribution of linguistic variables across age levels.
>
> (Labov 1994: 45-46)

5.1. A corpus study

In order to compare recordings of speakers of all ages from various parts of the English-speaking world, I turned to the IDEA[4] public website (*International Dialects of English Archives*), created by Paul Meier and hosted by the University of Kansas. I studied recordings of people from England, Scotland, the USA, Australia and New Zealand.

On this website, informants are asked to read a text and are then interviewed (in most cases). I will now present a selection of the results that I obtained for British English. The study was based on 216 recordings of the text entitled "Comma gets a Cure", that contains a number of potential ICPs[5], and on 315 recordings of interviews of speakers of both sexes and of all genders. Therefore, the IDEA corpus allowed me to study both scripted and unscripted speech.

Let us first turn to some of the results for England and Scotland[6]. The corpus clearly shows that there have been more and more speakers with some degree of contemporary palatalisation over the years. Speakers who do not palatalise at all are becoming the minority. In a number

[4] Special thanks to IDEA (International Dialects of English Archive).

[5] The words that may contain ICPs are the following: *story*, *district*, *duke*, *street*, *stressed*, *strut*, *strong*, *stroking*, *tune*.

[6] Welsh speakers are not included in my study of Britain. I did not find a single Welsh speaker whose speech displayed contemporary palatalisation in the IDEA British corpus.

of cases, the period in time when the change appears to gather momentum coincides with the speakers who were born in the late 1960s/early 1970s. This pattern seems to repeat itself with different ICPs, as illustrated with the graphs below (the point in time when the change becomes more apparent has been circled in figures 2 and 3). It corresponds to the end of the 20th century and it therefore matches the evolution noted by Wells about yod coalescence (cf. section 2.1). The analysis of the American corpus has yielded similar results (Glain 2013: 298-303).

In the USA, palatalised variants become more and more common as we move from one age group to the next. The pattern described about Britain applies there too: the change becomes more noticeable with the speakers born in the late 1960s/early 1970s.

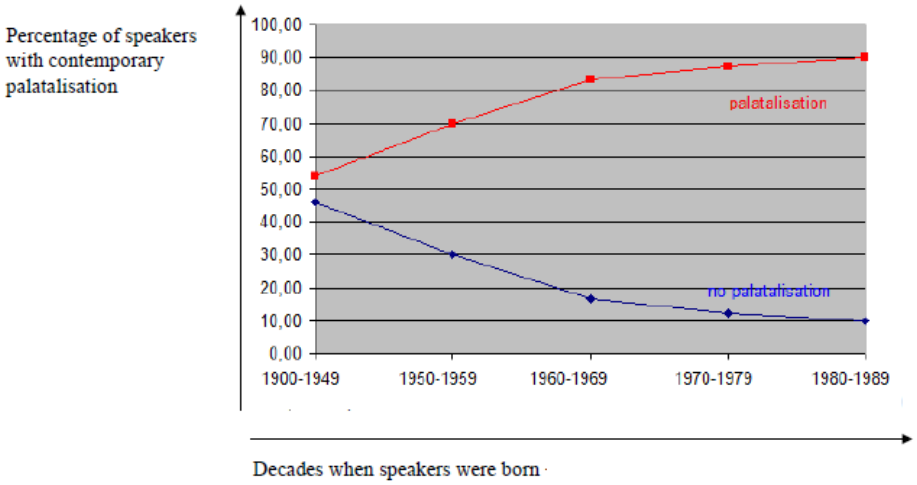


Figure 1: Overall palatalisation, England and Scotland, scripted speech

Figure 1 shows people who have some degree of contemporary palatalisation[7] in their speech vs. those who display no contemporary palatalisation at all.

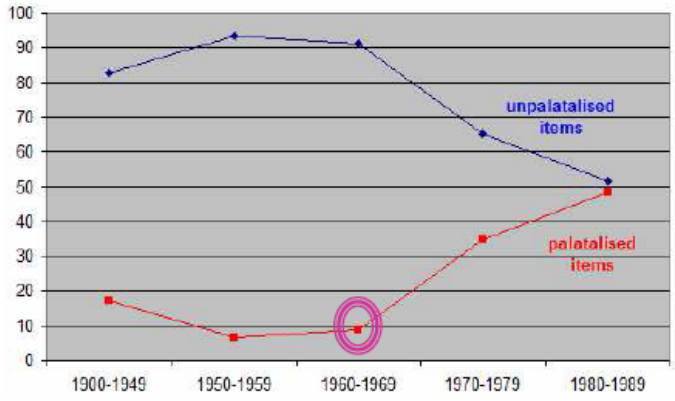


Figure 2: /str/ palatalisation, England and Scotland, scripted speech

For each age group, figure 2 shows the percentage of the items with /str/ that have been palatalised vs. the percentage of those that have not.

7 No speaker from the IDEA has exclusive contemporary palatalisation.

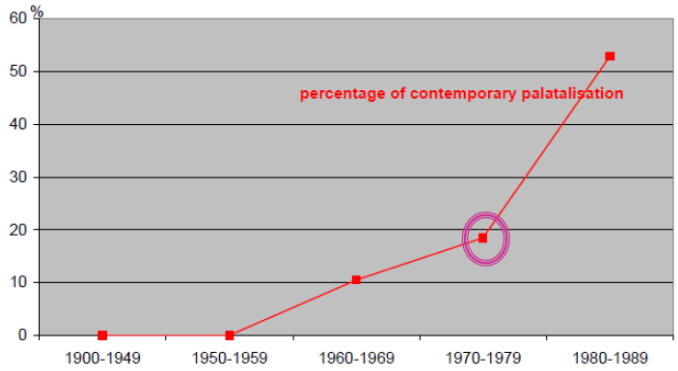


Figure 3: Overall palatalisation, England and Scotland, unscripted speech

Figure 3 only concentrates on the percentage of speakers who display some degree of contemporary palatalisation. The overall pattern is quite clear and the change gathers momentum in the 1970s.

The figures will be interpreted in the next section.

5.2. Some sociolinguistic observations

Let us consider the figures for scripted speech. In Britain, Scottish speakers display a higher rate of overall contemporary palatalisation than English speakers (93% vs. 75%). Within England, the only speakers who do not display any contemporary palatalisation in their speech are those from the north-east. Indeed, I could not find a single ICP in the speech of speakers from the following counties/regions: Northumberland, Tyne and Wear, County Durham, Yorkshire. London and south-eastern speakers do not seem to palatalise any more than their counterparts from other regions. Men palatalise a little more than women (84% vs. 74%).

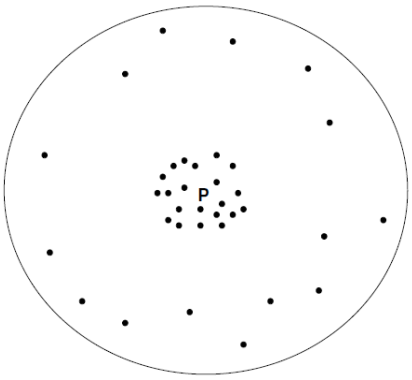
In the USA, the speakers who display the highest rate of contemporary palatalisation are those associated with the varieties known as *Southern American English* and *African American Vernacular English* (AAVE). As regards Southern speakers, 63% of them display some degree of contemporary palatalisation (vs. 53% for the national average). A massive 82% of AAVE speakers display contemporary palatalisation (vs. 53% for speakers of other varieties). That those two varieties should exhibit similar patterns is not really surprising as it is well-known that AAVE shares a number of features with Southern varieties of US English (Edwards 2008: 182). As was the case with Britain, men palatalise more than women (64% of men display overall palatalisation vs. 45% of women). This might seem a little surprising as women are typically viewed as the leaders of linguistic change when we are dealing with supra-regional innovations (Labov 2001: 516).

The figures obtained from the study of unscripted speech (Glain 2013: 307-320) reflect the same sociolinguistic differences. There appears to be no significant differences between scripted and unscripted speech in relation to the effective production of ICPs. That is surprising insofar as the opposition scripted/unscripted speech has sometimes been considered as an important principle of variation when reduction processes are concerned, unscripted speech being much more likely to yield reduced forms such as assimilations (e.g. Shockey 2003: 17). The principle of variation that best characterises ICPs is that of lexical frequency,

whereby high-frequency items are much more likely to undergo reduction. This has been corroborated by a survey of the phonetic transcriptions of the lexical items that may potentially undergo contemporary palatalisation, both in EPD and in LPD (Glain 2013: 267-279). In the IDEA corpus, the items which are most often palatalised are *during*, *straight*, *strong*, *street*, *strange*, *restaurant*, *grocery*/*groceries*. They are all fairly common words.

6. Explaining the change: Smith's cognitive model

Smith (2007) proposes a model that includes elements from theories of change based on the coarticulatory nature of speech (Lindblom 1986, 1990; Ohala e.g. 1981, 1989, 1993a, 1993b, 1994, 2003 ; Blevins 2004). At the same time, Smith includes the social dimension of change in his model. The traditional dichotomy between internal and external change is not relevant as Smith (2007: 74) works within the framework of cognitive linguistics. Cognitivists 'posit an intimate, dialectic relationship between the structure and function of language on the one hand, and non-linguistic skills on the other' (Taylor 1996: ix). In other words, within a cognitive framework, there is no clear-cut boundary between the mental processes associated with human language and those related to the rest of human experience.
Smith (2007: 19) writes that, for any given phone, speakers have a repertoire of variants that they can choose from according to the situation. Among those variants, there exists a prototypical realisation of the phone that corresponds to its phonological representation. Smith reminds the reader that Jones's definition of the notion of phoneme is 'a family of related sounds' (Jones 1956: 172). These related sounds are organised around the prototypical value, as illustrated below.



. = actual realisations
P = prototypical realisation
Illustration 5: A family of sounds (Smith, 2007: 20)

The prototypical value can vary from a speaker to another. It follows that the listener may change his/her pronunciation, through a process of identification with and adoption of the prototypical value of his/her interlocutor. Smith (2007: 11) maintains that a true phonological change occurs at the level of the individual if the adoption of the new value modifies the listener's phonological system. The more frequent the contact with speakers whose phonological system is different, the more significant the change in the listener's system.

Yod coalescence after /t, d/ can clearly illustrate the model proposed by Smith. When a speaker (A) who only has palatalised forms in unstressed syllables (e.g. *actually, fortune, duality, durability*) interacts with another speaker (B) who has palatalised forms in both

unstressed *and* stressed syllables (e.g. *actually, fortune, duality, durability, tune, tutor, dune, reduce)*, adoption of the prototypical values of B may lead to a modification of the consonantal system of A. Indeed, /tj, dj/ may merge with /tʃ, dʒ/, modifying A's system from A1 to A2.

A1 (initial system of A)

/tʃ/	*actually, fortune*
/dʒ/	*duality, durability*
/tj/	*tune, tutor*
/dj/	*dune, reduce*

→

A2 (A's modified system, following contact with B)

/tʃ/	*actually, fortune, tune, tutor*
/dʒ/	*duality, durability, dune, reduce*

A's consonantal system is partly modified following a process of imitation.

The reason why the ongoing evolution of sounds does not lead to the complete breakdown of communication is that the vast majority of the innovations that come from the interaction between two speakers are not diffused into the speech community (Smith 2007: 12-13). If change is always potential within variation, there has to be an interaction between extra- and intralinguistic factors at a particular time in order for a particular change to occur and then to diffuse into the community (Smith 2007: 10). This raises the question of the *actuation* of change on a large scale. Why is a given change actuated at a particular time? Why not earlier? Why not later? Smith (2007) argues that the reason why some innovations catch on in the community is often related to social considerations. The evolution may even originate in major historical events or ideological changes (Labov 2010: 44).

If we consider Britain, the emergence of ICPs in the latter part of the 20th century clearly coincides with the period when RP began to lose ground on account of a particular socio-historical context (cf. section 4). Such a context is likely to stimulate linguistic change within a cognitive model of change such as that proposed by Smith. It is therefore theoretically sound to posit that the development of ICPs, non-standard variants, was indeed triggered by an interaction between intra- and extralinguistic factors that led to the decrease of the standard accent.

American English does not have yod coalescence after /t, d/ or yod palatalisation after /s, z/. However, the loss of an international prestige model and of traditional forms of oratory (cf. section 4) certainly contributed to the rise of the other two ICPs in the USA. It is entirely possible that, in shifting away from written, overarticulated speeches, modern speechmaking has participated in an overall change favourable to processes of phonetic reduction based on coarticulation and hypoarticulation, such as ICPs. In the USA, like in Britain, the 1945-1970 period witnessed a radical change in speaking standards. Those innovations seem to have been triggered by social changes.

7. Are ICPs phonetic or phonological phenomena?

Like previous instances of palatalisation in the history of English (cf. section 3), it is clear that ICPs originate in purely phonetic assimilations. However, as those previous palatalised variants were gradually phonologised, it is worth wondering if ICPs remain strictly as phonetic phenomena today. Can we consider that they might be phonological variants? In order to try and answer those questions, I carried out an experiment with 30 speakers (15 were English and 15 were American).

I came up with a list of words which I asked my 30 informants to read slowly and syllable by syllable. They also had to read a text that contained the same words. The point of the experiment was to determine whether the ICPs that were actually produced were connected speech phenomena or rather corresponded to the speakers' citation forms of the lexical items considered. The words were *tube*, *astute*, *dune*, *reduce*, *assume*, *presume*, *resume*, *student*, *street*, *stop*, *start*, *Australia*, *grocery*, *classroom*.

Some speakers palatalised certain items when they read the text, but not the list of words, which is not surprising considering the particularities of connected speech phenomena. On the other hand, some speakers palatalised the same items both in the text and in the list of words, which seems to be an indication that there might be more than connected speech phenomena at work. The really surprising part of the experiment was that other speakers occasionally palatalised an item when they read it syllabically, but not when it was part of the text. It does not seem far-fetched to suggest that the apparent variation in the citation forms of the items considered reflects a true underlying variation within the group considered. There appears to be *variable phonological representations* of the items considered. Indeed, the speech is slower and less variable when the informants read the words syllable by syllable, which is more likely to bring the true underlying representations to the fore. Such representations are more easily lost in more rapid connected speech.

For English speakers, the most productive ICP was yod coalescence after /t, d/ in stressed position, which yielded 47% of palatalised forms in the syllabic reading. The second most productive ICP was yod palatalisation after /s, z/ in stressed syllables (37% of palatalised forms), followed by /stj/ palatalisation (25%), palatalisation of /s/ by /r/ (19%) and /str/ palatalisation (12%). As far as American speakers were concerned, palatalisation of /s/ by /r/ was the most frequent ICP (with 25% of palatalised forms), followed by /str/ palatalisation (19%).

Of course, there is no absolute certainty that the palatalised variants are phonological for those speakers, but it still seems to be an indication that some lexical items are stored with palato-alveolars for certain speakers. ICPs might be a little more than mere surface phenomena.

8. Conclusion

In this paper, I have introduced the phonological phenomenon which I have labelled *Contemporary Palatalisation*, as well as its lexical manifestations, *Instances of Contemporary Palatalisation* (ICPs).

While being phenomena mostly associated with the last fifty years, ICPs are the continuation of a long historical process which has systematically led to palatalisation in English. Contemporary Palatalisation is therefore an example of a synchronic process that is in fact the manifestation of systematic, diachronic ones. Ohala (1994: 375) maintains that the coarticulation of phonemes in synchrony have the exact same effect as other co-occurrences, which have been identified at the diachronic level. Blevins (2004: 18) shares the same view,

arguing that phonetic innovations often mirror diachronic processes. Therefore, Contemporary Palatalisation may well be triggering a real change within the community, as there seems to be variation at the underlying level, as some speakers appear to have phonological representations with fossilised palatalisation.

I have tried to give an answer (at least in the specific case of ICPs) to what Weinreich, Labov and Herzog (1968) call *the actuation problem* and regard as 'the heart of the matter', i.e. why a change occurs at a particular time. Everything indicates that the particular sound changes listed in this paper are part of a larger linguistic trend whereby spoken English underwent radical changes in the second part of the 20th century. Such changes operated within the overall context of a certain democratisation of society in Britain and America, which went hand in hand with the development of more informal forms of language.

References

ALTENDORF, ULRIKE. 2003. *Estuary English: Levelling at the interface of RP and south-eastern British English.* Berlin: Gunter Narr Verlag Tübingen.

ALTENDORF, ULRIKE. & WATT, DOMINIC. 2008. In: Bernd Kortmann & Clive Upton (eds) *Varieties of English 1: The British Isles*. Berlin: Mouton de Gruyter. pp. 194-222.

BLEVINS, JULIETTE. 2004. *Evolutionary Phonology.* New York: Cambridge University Press.

BYBEE, JOAN. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

CALVET, LOUIS-JEAN. 2011. *La sociolinguistique.* Paris: Presses Universitaires de France.

COGGLE, PAUL. 1993. *Do You Speak Estuary? The new Standard English.* London: Bloomsbury.

CRUTTENDEN, ALAN. 2008. *Gimson's Pronunciation of English.* London: Hodder Education.

EDWARDS, WALTER. 2008. *African American Vernacular English: phonology.* In: Edgar Shneider (ed) *Varietes of English 2: the Americas and the Caribbean.* Berlin: Mouton de Gruyter. pp. 181-191.

GIMSON, ALFRED. 1970. *An Introduction to the Pronunciation of English*. London: Edward Arnold.

GLAIN, OLIVIER. 2013. *Les Cas de Palatalisation Contemporaine (CPC) dans le monde anglophone.* Lyon: unpublished PhD thesis of the University of Lyon.

HANNISDAL, BENTE. 2006. *Variability and change in Received Pronunciation, A study of six phonological variables in the speech of television newsreaders.* Bergen: unpublished PhD thesis of the University of Bergen.

JONES, DANIEL. 1956. *Pronunciation of English*. Cambridge: Cambridge University Press.

KERSWILL, PAUL. 2007. "Standard and non-standard English". In: David Britain (ed) *Language in the British Isles.* Cambridge: Cambridge University Press.

LABOV, WILLIAM. 1994. *Principles of Linguistic Change, Volume 1: Internal Factors.* Chichester: Wiley-Blackwell.

LABOV, WILLIAM. 2001. *Principles of Linguistic Change, Volume 2: Social Factors.* Oxford: Blackwell.

LABOV, WILLIAM. 2010. *Principles of Linguistic Change, Volume 3: Cognitive and Cultural Factors*. Chichester: Wiley-Blackwell.

LABOV, WILLIAM, ASH, SHARON & BOBERG, CHARLES. 2006. *Atlas of North American English: Phonetics, Phonology and Sound Change.* Berlin: Mouton de Gruyter.

LINDBLOM, BJÖRN. 1986. "On the origin and purpose of discreteness and invariance in sound patterns". In: Joseph Perkell & Dennis Klatt (eds), *Invariance and Variability in Speech Processes.* Hillsdale: Lawrence Erlbaum Associates. pp. 493-523.

LINDBLOM, BJÖRN. 1990. "Explaining phonetic variation: a sketch of the H&H theory". In: William Hardcastle & Alain Marchal (eds) *Speech Production and Speech Modelling*. Amsterdam: Kluver. pp. 403-439.

MCWHORTER, JOHN. 2012. *Myths, Lies and Half-Truths of Language Usage.* Chantilly: The Great Courses.

MEIER, PAUL. 1997. *The International Dialects of English Archive (IDEA).* http://www.dialectsarchive.com

OHALA, JOHN. 1981. "The listener as a source of sound change". In: *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society. pp. 178-203.

OHALA, JOHN. 1989. "Sound change is drawn from a pool of synchronic variation". In: Leiv Breivik *et al.* (eds). *Language Change: Contributions to the study of its causes*. Berlin: Mouton de Gruyter. pp. 173-198.

OHALA, JOHN. 1993a. "Coarticulation and Phonology". In: *Language & Speech* 36, SAGE Journals. pp. 155-170.

OHALA, JOHN 1993b. "The phonetics of sound change". In: Charles Jones (ed) *Historical Linguistics: Problems and Perspectives*. London: Longman. pp. 237-278.
OHALA, JOHN. 1994. "Hierarchies of environments for sound variation". In: *Acta Linguistica Hafniensia* 27.37. http://linguistics.berkeley.edu/~ohala/papers/hierarchies.pdf. pp. 371-382.
OHALA, JOHN. 2003. "Phonetics and Historical Phonology". In: Brian Joseph & Richard Janda (eds) *The Handbook of Historical Linguistics*. Oxford: Blackwell. pp. 669-686.
PAVLÍK, RADOSLAV. 2009. "A Typology of Assimilations". In: *SKASE Journal of Theoretical Linguistics*, vol. 6, n°1. pp. 2-26
RAMSARAN, SUZAN. 1990. "RP: fact and fiction". In: Susan Ramsaran (ed), *Studies in the pronunciation of English. A commemorative volume in honour of A.C. Gimson*. London: Routledge. pp 178-190
ROACH, PETER, HARTMAN, JAMES & SETTER, JANE (eds). 2006. *Cambridge English Pronouncing Dictionary.* Cambridge: Cambridge University Press.
RUTTER, BEN. 2011. "The consonant cluster /stɹ/ in English". In: *Journal of the International Phonetic Association*, Volume 41 / Issue 01. pp 27-40
SHOCKEY, LINDA. 2003. *Sound Patterns of Spoken English.* Oxford: Blackwell.
SMITH, JEREMY. 2007. *Sound Change and the History of English.* Oxford: Oxford University Press.
STEVANOVITCH, COLETTE. 2008. *Manuel d'histoire de la langue anglaise des origines à nos Jours.* Paris: Ellipses.
TAYLOR, JOHN. 1996. *Linguistic Categorization.* Oxford: Oxford University Press.
WEINREICH, URIEL, LABOV WILLIAM & HERZOG MARVIN. 1968. *Empirical Foundations for a Theory of Language Change*. Austin: University of Texas Press
WELLS, JOHN. 1982. *Accents of English.* Cambridge: Cambridge University Press.
WELLS, JOHN. 1997. "Whatever Happened to Received Pronunciation?". In: Carmelo Medina & Concepción Soto (eds) *II Jornadas de Estudios Ingleses*. Jaén: Universidad de Jaén. pp 19-28
WELLS, JOHN. 2007. *John Wells's phonetic blog.* http://www.phon.ucl.ac.uk/home/wells/blog0704.htm
WELLS, JOHN. 2008. *Longman Pronunciation Dictionary*. London: Longman.
WOTSCHKE, INGRID. 1996. "Socio-regional speech versus 'Oxford accent' – trends and fashions in the pronunciation of English English". In: *Hermes, Journal of Linguistics* 17: 213-235.

Olivier Glain
Département Techniques de Commercialisation
IUT de Roanne
Université Jean Monnet de Saint-Etienne
Roanne, France
professional email: olivier.glain@univ-st-etienne.fr
personal email: oglain@club.fr

THE ROLE OF GENDERED SOCIOLINGUISTIC VARIABLES AS PERCEPTUAL CUES

ANIA KUBISZ
University of York

Abstract

This paper investigates perceptions of speaker-indexical information from gender-specific phonetic variables in the absence of speakers' fundamental frequencies. The results revealed that listeners not familiar with a dialect under investigation were not sensitive to speaker-indexical information embedded in the phonetic variants. The results also showed that in their evaluations, male and female listeners often did not differentiate between localised and supra-local variants. Finally, the perceptual differences noticed between male and female listeners were not statistically significant.

1. Introduction

Previous socioperceptual studies focus on identifying speaker indexical information such as ethnicity (Purnell et al. 1999, Wolfram 2000), geographic origin (Bezooijen & Gooskens 1999, Clopper et al. 2005) or personality traits (Lambert et al. 1960, Ball & Giles 1988, Bezooijen 1988). Researchers have also investigated female and male voice identification (Biemans 2000, Munson & Babel 2007). Even though it has been established that listeners are quite accurate at identifying adult female and male voices, it is still unclear how listeners identify gender in the speech signal (Munson & Babel 2007). Literature provides evidence that fundamental frequency impacts femininity and masculinity judgments (Munson & Babel 2007, Foulkes et al. 2010). However, fundamental frequency is not always a decisive factor. First of all, there is an overlap of female and male pitch ranges, such that a lower-pitched female voice might be erroneously taken for a higher-pitched male voice and vice versa (Foulkes et al. 1999, Biemans 2000). Furthermore, Johnson et al. (1999) showed in their study that a voice judged as most stereotypically female had lower mean fundamental frequency than the non-stereotypical female voice.

Finally, it has been reported that listeners are able to distinguish male and female speakers in the absence of acoustic information present in speakers' fundamental frequency (Coleman 1971, Lass et al. 1975, Assmann & Nearey 2007, Hubbard & Assmann 2013). These findings imply that parameters of the vocal tract are not the main factors deciding whether a speaker sounds feminine or masculine, which further means that gender-specific acoustic information does not rely heavily on fundamental frequency.

Because fundamental frequency is not the main cue to speakers' gender identification, it is hypothesised that when speaker-social information embedded in fundamental frequency is not accessible to the listener, this type of information can be identified from gender-specific phonetic variants.

Therefore, this paper examines the role of sociolinguistic variants as cues to the identification of the speaker-social information when listeners are exposed to speech that sounds gender-ambiguous.

Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York (**PARLAY 2013**) conference
 ISSN 1758-0315 Issue 1

This study has the same aim as Foulkes et al.'s (2010) study. It is hypothesised that listeners familiar with the dialect and particular variant realisations should be sensitive to speaker indexical information embedded in gender-correlated phonetic variables. However, listeners with no previous exposure to the dialect are not expected to be able to access this information.

A set of gender-correlated phonetic variables identified in the Newcastle dialect were selected for the purpose of this study. Variables are sociolinguistically marked in terms of speaker gender, age and social class. It was decided to use Newcastle English phonetic variant in the study for the following reasons (Milroy et al. 1994a, 1994b, Docherty & Foulkes 1999, Watt & Milroy 1999, Watt 2000, Watt 2002, Watt & Allen 2003, Foulkes et al. 2005, Beal et al. 2012). Firstly, Tyneside English is characterised by rich realisations of vowel variants, stop realisations, and others. Because Newcastle is considered to be the hub of the North East region, its dialect has been extensively researched and described (Milroy et al. 1994a, 1994b, Docherty & Foulkes 1999, Watt & Milroy 1999, Watt 2000, Watt 2002, Watt & Allen 2003, Foulkes et al. 2005, Beal et al. 2012). Furthermore, Tyneside English is stereotypically perceived as the variety spoken in all of the North East.

While further research will investigate and compare perceptions of speaker-social information provided by Tyneside listeners, North East listeners and listeners from outside of these two regions, the present paper focuses on listeners from outside of the North East or Tyneside area, who are hence unfamiliar with the dialect.

Perceptions of Newcastle-localised variants were compared and contrasted with perceptions of other localised variants or non-marked supra-local variants.

2. The Newcastle dialect

Great Britain is characterised by an abundance of regional dialects. The North East of England, with the Newcastle dialect being one of many spoken in the region, is no different. However, outsiders tend to have a distorted view of the North East. They seem to neglect a number of distinct dialects, such as Sunderland or Middlesbrough dialects, present in the region and consider the Geordie dialect to be spoken anywhere up north (Pearce 2009, Beal et al. 2012). However, each of the dialects in the region is characterised by distinctive phonetic features.

Variation in the use of some vowels and consonants is one of the main phonetic cues revealing social and regional characteristics of the speakers within the North East (Beal et al. 2012: 26). However, there is also rich variation within the Newcastle dialect in terms of the use of phonetic variants. In fact, the Newcastle dialect is characterised by an array of localised phonetic variants, which are marked sociolinguistically, as they are not only gender- but also age- and class-specific (Watt & Allen 2003: 269). It is these features that distinguish Tyneside speakers from speakers south of the River Wear or Teesside speakers (Beal et al. 2012). It is also these features that distinguish speakers within the Newcastle dialect.

The section below provides an account of variation and possible realisations of the FACE, GOAT and NURSE vowels investigated in the study.

Two perceptually prominent vowels in Tyneside English are the FACE and GOAT vowels (Watt 2000, Beal et al. 2012). Not only is there a significant variety of realisations of these vowels but also different variants are used by older and younger speakers (Watt 2000).

Watt (2000, 2002) lists three types of realisations of the FACE and GOAT vowels and groups them into monophthongs, centering diphthongs and closing diphthongs. The most commonly occurring and unmarked variants of FACE and GOAT in Tyneside English are the monophthongal realisations, [eː] and [oː]. These realisations are also found in other varieties of North East English, and as such, are supra-local (Beal at al. 2012: 31).

Monophthongal [eː] and [oː] are found across male and female speakers of different ages and social backgrounds in Newcastle English. The only exceptions are older working-class male speakers who, instead, use the centering diphthong [ɪə] as a realisation of the FACE vowel. The GOAT vowel is realised as monophthongal [oː], the centring diphthongal [ʊə] or the fronted monophthongal [ɵː] in this group of speakers (Watt & Milroy 1999, Watt 2000, 2002, Beal at al. 2012).

While the diphthongal FACE and GOAT variants [ɪə] and [ʊə] are found in all of the North East, they are, in fact, associated with Tyneside English and considered to be traditional and old-fashioned, and as such are characteristic of older working-class males (Watt 2000, 2002, Beal at al. 2012). [ɪə] can be also found in the speech of younger working-class males, although much less frequently than in older working-class males (Watt & Milroy 1999). [ʊə] is less frequently used by other groups of male speakers than older working-class. For example, older middle-class or younger working-class speakers use it less frequently, and younger middle-class speakers use it very rarely (Watt & Milroy 1999).

The closing diphthongs are [eɪ], which is a realisation of the FACE vowel, and [oʊ] and [ɵː], which are realisations of the GOAT vowel. Overall, [eɪ] is not a common variant in Tyneside English, yet it is becoming more popular among younger middle-class speakers. It is used most often by young female middle-class speakers, followed by young middle-class male speakers (Watt 2000).

The closing diphthong [oʊ] is also widely used in other parts of the country. In Newcastle, this realisation is used by young middle-class speakers (Watt 2000, Beal et al. 2012). The fronted monophthongal [ɵː], on the other hand, is associated with male speakers and is used most frequently by younger middle-class males but also older and younger working class males. However, the variant is becoming less common in general and female speakers refrain from using it (Watt & Milroy 1999, Watt 2000).

Finally, Watt & Allen (2003: 269) and Viereck (1968: 69, 70) provide more examples of the realisation of the GOAT vowel which make the vowel contrast in Tyneside English even more varied. For example, [ɪə] can be found in words like [stɪən] *stone,* [hɪəm] *home,* [bɪən] and *bone,* and [aː] in words like *snow* [snaː]. These pronunciations occur in older working-class male speakers and are considered to be old-fashioned even by Viereck (Viereck 1968).

Another vowel associated with significant variability in the region is the NURSE vowel, which can be realised as the localised retracted [ɔː], fronted [øː] and centralised [ɜː] (Watt 1998, Watt & Milroy 1999, Beal et al. 2012).

While the first variant is now rare and associated with older working-class male speakers, the two other variants are more commonly used in Tyneside English than [ɔː]. The centralised [ɜː] is most common and also supra-local. Watt (1998) and Watt & Milroy (1999) point out that the fronted variant [øː] is marked for age and gender, as it is associated with female speakers, and especially younger middle- and working-class females who use it more frequently than [ɜː].

In general, localised vowel variants seem to be associated with older and usually male speakers. Younger speakers, especially females, tend to prefer supra-local variants, widely used across the region and the country (Beal et al. 2012).

Overall, a decrease in the use of localised, traditional forms can be observed in Tyneside English (Watt 2000). In their place, new, non-regional forms are adopted. The process results in a reduction of the number of vowel variants in use and implies language levelling, which results in formation of a more uniform repertoire of phonetic variation, one that is closer to other varieties of British English (Watt 2000, Watt 2002). At the same time, the supra-local forms new to the region seem to be less socially and geographically marked.

3. Method

For the purpose of this study talker pitch was shifted to obtain the effect of gender-ambiguous-sounding voice.

This study uses single-word stimuli. The advantage of using single words over connected speech is that listeners can focus with greater ease on the specific type of information present in the acoustic signal (Munson 2007). At the same time, this approach allows the researcher to control for more parameters and therefore draw more reliable conclusions from the data when analysing which phonetic cues listeners rely on.

3.1. Stimuli

Stimuli selected for this study occur in three phonological contexts: word-finally in open syllables, preceding a nasal, and preceding a fricative in one instance.

A total of four voices were used in this study. Two phoneticians recorded target stimuli using Newcastle variants and two other speakers recorded fillers used in the study.

Preliminary tests revealed that in terms of the range of possible pitch manipulation and the final outcome in terms of voice naturalness, male voices gave better results than female voices. Therefore, only male voices were used in this study.

The tokens were recorded in a recording studio to .wav sound files at a sampling rate of 44.1 kHz and 16 bit mono resolution.

Speakers were in their forties and mid-twenties. All tokens were manipulated in Adobe Audition 3.0 (Adobe, 2007) using the Pitch Shifter function to raise pitch and obtain the effect of gender-ambiguous-sounding voice. In addition to preserving the tempo of the samples, high precision and default appropriate settings were selected. Pitch Shifter allows changes in fundamental frequency by semitones and cents, where 1 semitone is equal to 100 cents. Each token was manipulated individually between 1.0 and 4.0 semitones.

The algorithm implemented by the Pitch Shifter allows the speech tempo to be preserved and the formant values to be adjusted to changes in pitch (Adobe, 2007). Because this study investigates perception of gendered phonetic variables in the absence of gender-specific fundamental frequency, the aim was to manipulate only one of the phonetic cues, that is, fundamental frequency. Preserving tempo and adjusting formant values to changes in pitch sustained other acoustic features of the recordings. Furthermore, this approach allowed the researcher to control for pitch and draw more specific conclusions about the acoustic cues responsible for perceptions of speaker-indexical information.

All tokens were normalised for volume in Adobe Audition CS5.5 (Adobe, 2012) using the Match Volume function. A single token was pre-selected and the remaining tokens were matched in volume to the pre-selected token using the file total root mean square power (RMS) function and limiting settings to ensure the output files were not clipped or overly loud.

Finally, the naturalness and gender ambiguity of the stimuli and fillers were judged by a male and a female sociophonetician familiar with the dialects of North East England.

3.2. Procedure

The experiment was conducted in laboratory conditions and administered in E-Prime 2.0 (Psychology Software Tools, Inc., 2012). At the beginning of the experiment there was a training session, after which participants were given time to ask questions. A total of 396 single-word stimuli and fillers were presented over Sennheiser HD 280 Pro headphones at a comfortable hearing level, one at a time. Each stimulus was played once only. The entire session was estimated to take about 40 minutes and there were two breaks in between. Because the study was rather long, participants were instructed to time the breaks themselves.

Visual representations of stimuli were simultaneously projected onto a computer screen. In order to avoid visual priming, except for two instances referring to filler words, visual word representations excluded images of men or women. The role of visual stimuli was to help listeners not familiar with the Newcastle dialect understand the recordings. The images also served as an additional element in the study, which alleviated a possible feeling of boredom.

Listeners were instructed to listen to each stimulus and evaluate it using a Visual Analogue Scale (VAS) slider with a 0 to 100 point scale, incrementing by 1 point and logging participant choices on the *x* axis (Groot, 2013). Listeners were also asked to go with their first impressions and to not overthink their choices. Furthermore, the pace at which the stimuli were presented and the fact that listeners heard each stimulus once only gave listeners just enough time to reach a decision.

Stimuli were presented in a fixed order and the slider was reset to a midpoint position on the scale after each evaluation. Additionally, the slider did not allow for stimuli to be left unrated and so, in order to proceed, participants had to move it.

Each speaker was evaluated three times along three dimensions: how male or female they sounded, how old or young they sounded and how middle class or lower class they sounded. These alternatives were presented in a mixed order for each block in such a way that every stimulus was rated along only one dimension per block and on all three of them in total. Each participant rated all three blocks.

3.3. Participants

Listeners were volunteers recruited from the undergraduate and graduate student bodies at the University of York. Four male and four female listeners participated in the study. Listeners were ages 19 to 24. Additionally, a Newcastle male listener, aged 26, took part in the study.

All listeners considered themselves to be middle class and, except for the Newcastle listener, all participants were speakers of varieties other than the Newcastle English or a North East English variety. Except for one female and two male listeners, all participants declared

speaking one or more foreign languages. None of the participants suffered from flu or reported a hearing impairment.

3.3. Data analysis

As has been already mentioned, male and female groups were comprised of four participants. Each participant evaluated between two to three words in each of the conditions. This means that data were clustered and they could be correlated rather than independent. First, in order to analyse data, median values for each of the participants were determined. Applying this type of measure ensured that data were independent which, in turn, allowed statistical analysis. Because the data distribution was skewed, a non-parametric Mann-Whitney test was applied. This type of test compares differences between two medians. In this case, median values in male and female groups were compared. Additionally, because of the small size of the total sample, the exact probability option of the test was selected.

It was hypothesised that there might be differences in perception of the variants between male and female listeners. For this reason, results in the two groups are presented separately. However, because one of the participants was a Newcastle listener, it was decided that he should be included in the analysis for the sake of providing a brief comparison of the results obtained in the male and female groups of listeners from outside of the North East with the results provided by the user of the variety under investigation. However, data provided by the Tyneside listener were excluded from statistical analysis.

Because there were only four listeners in the male group and four more in the female group, it was decided that bar plots, rather than box plots, should be used when visualising experimental results.

The slider appeared at the midpoint on the scale after each audio stimulus was played. Additionally, the continuum evaluation scale itself was quite long, which made it relatively easy for the participants to drag the slider back to the centre of the scale if they wished to rate a token at 50 per cent. However, when interpreting the results, it was assumed that ratings in the range between about 45 and 55 per cent on the scale report mid-evaluations and refer to gender-ambiguous-sounding voice. As has been mentioned, the age differences between participants were not significant as all listeners were in their late teens and early to mid-twenties. Therefore, as far as age evaluation is concerned, it could be assumed that perceptions of speaker age should not differ to a considerable extent between the participants. Thus age evaluations between 45 and 55 per cent on the scale were analysed as referring to young but mature-sounding voices. Finally, midpoint evaluations of speaker social class could mean that the speaker is somehow bringing features of the two classes together, yet not sounding definitely middle- or lower-class.

4. Results

The following section focuses on the evaluation of speaker-indexical social information of localised and supra-local vowel variants of FACE, GOAT and NURSE occurring in Tyneside English.

The localised FACE [ɪə] variant is characteristic of older working-class male speakers and [eː] is a supra-local variant (Watt & Milroy 1999, Watt 2000, Beal et al. 2012). A closer look at the results presented in Figure 1 provides some interesting observations.

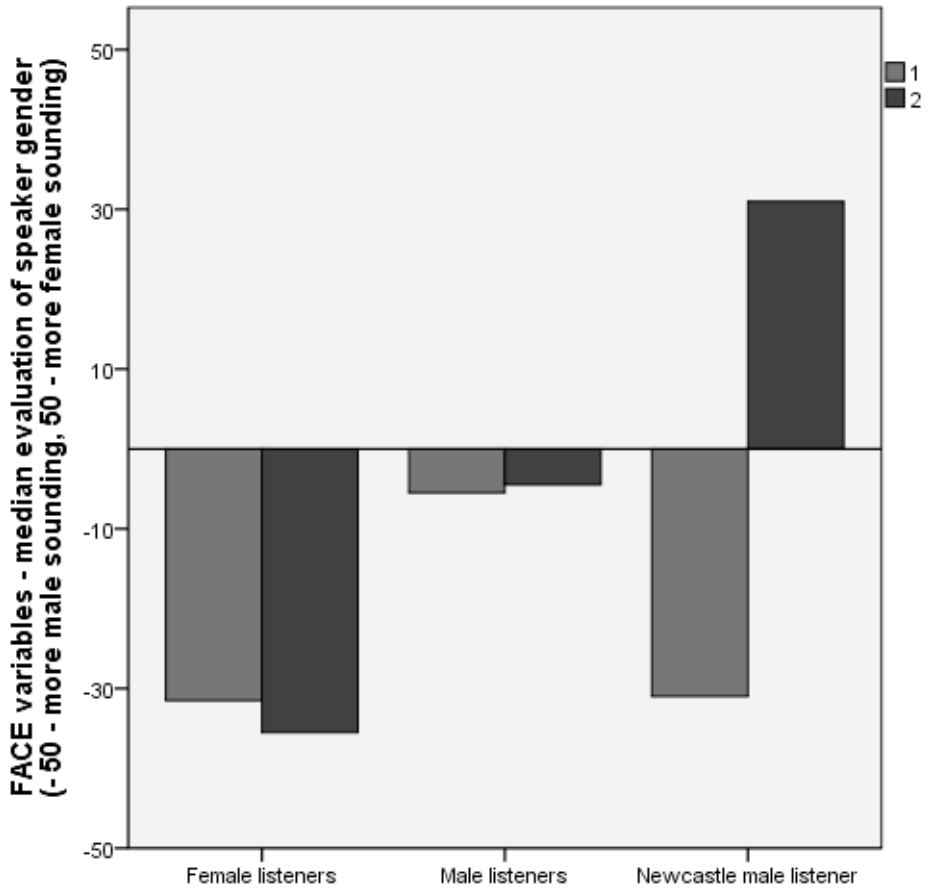


Figure 1. FACE localised [ɪə] (1) and supra-local [e:] (2) variants -- evaluation of speaker gender.

Overall, the localised variant was evaluated as male-sounding by all groups of listeners (Fig.1.). As can be seen, the female group and the Newcastle listener in particular strongly perceived the variant as male-sounding. However, male listeners identified it as much less male-sounding than female listeners or the Newcastle listener.

Even though [ɪə] is, in fact, found in speech of older male speakers, male and female listeners were not expected to be sensitive to the localised Newcastle variant. Thus the results need to be accounted for differently. One possible explanation is that upon hearing an unfamiliar variant people tend to perceive it as lower class and male-sounding (Beal et al. 2012). At the same time however, the supra-local variant [e:] was evaluated by the male and female groups of listeners almost identically to the localised [ɪə] variant. This would suggest that to these listeners, the two variants did not differ to any considerable extent.

Nevertheless, a difference in evaluations of the two variants provided by the Newcastle listener can be noticed. Interestingly, while the localised variant [ɪə] was judged as male-sounding, the supra-local variant [e:] was perceived as female-sounding. Thus it seems that the Newcastle listener was sensitive, at least to some extent, to the two phonetic realisations of the FACE vowel present in Tyneside English.

Figure 2 presents speaker age evaluation of the FACE vowel variants under investigation.

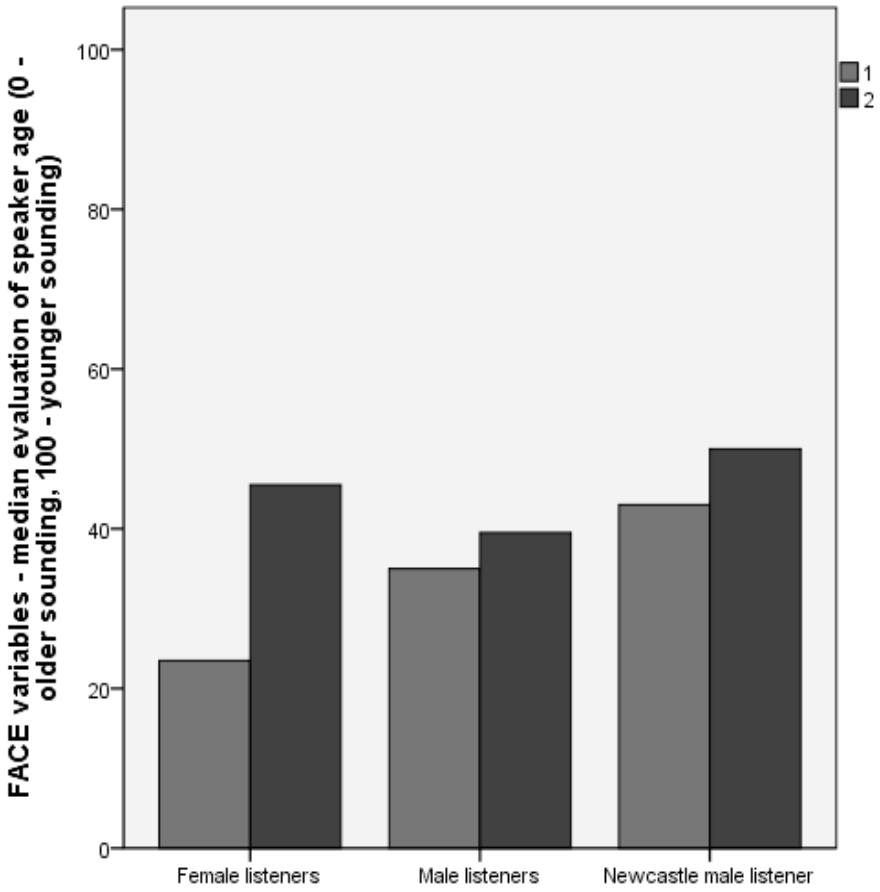


Figure 2. FACE localised [ɪə] (■1) and supra-local [eː] (■2) variants -- evaluation of speaker age.

As far as speaker age evaluation is concerned, it can be easily observed that both male and especially female listeners judged the localised variants as older-sounding.

Even though female listeners found also the supra-local variant to be overall rather older-sounding, they did find it as mature-sounding and considerably younger than the localised variant. Male listeners, on the other hand, did not perceive the two variants to be significantly different.

The results produced by the Newcastle listener in terms of age identification confirm the expected for the most part. Both variants were perceived as mature but young-sounding, yet, the localised variant was identified as slightly older-sounding.

Figure 3 illustrates evaluations of speaker social class on the basis of the FACE vowel variants.

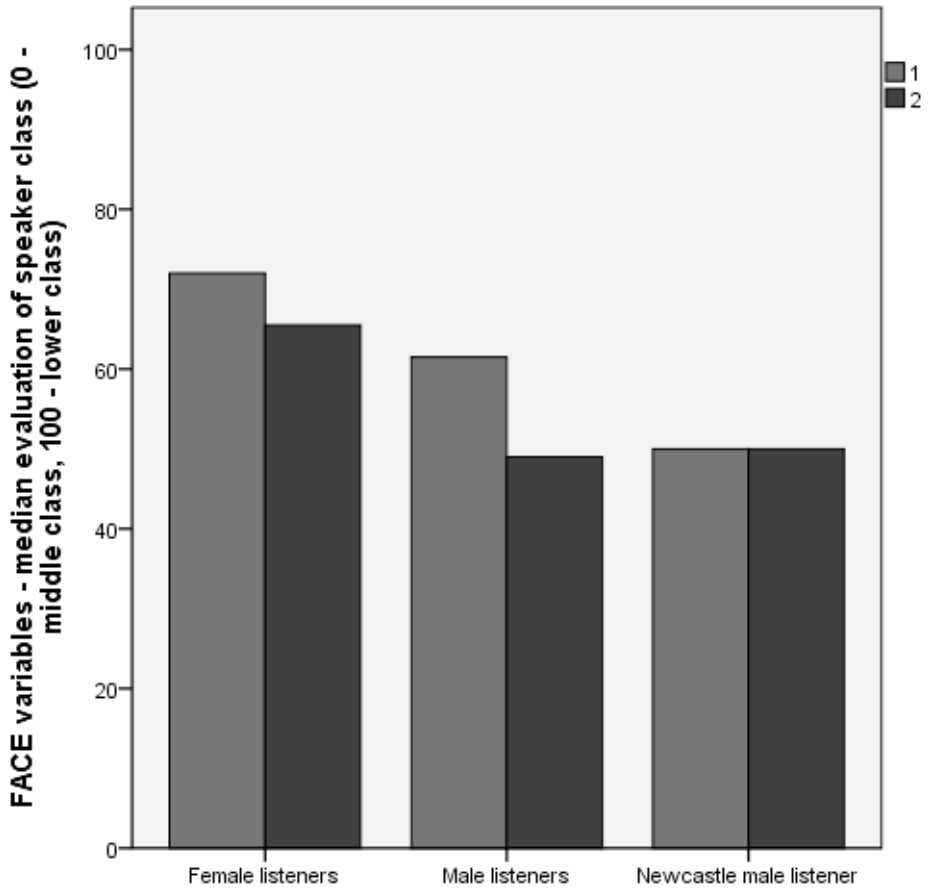


Figure 3. FACE localised [ɪə] (■1) and supra-local [e:] (■2) variants -- evaluation of speaker social class.

Overall, speaker social class of the FACE localised variant was evaluated as lower class by both female and male listeners. The supra-local variant, on the other hand, was categorised as lower class sounding by female listeners and as combining middle and lower class features by male listeners.

Furthermore, a difference between male and female evaluations can be noticed. Female listeners perceived the recordings to be generally more lower-class-sounding than male listeners, who by contrast, found them to be slightly less lower-class-sounding.

The Newcastle male listener categorised both variants at midpoint. It could be that the variants were, in fact, identified as being used by speakers combining middle- and lower-class features. The other possibility is that the listener did not feel comfortable evaluating the class parameter. Whether or not there is a pattern will be investigated in the following sections of the paper.

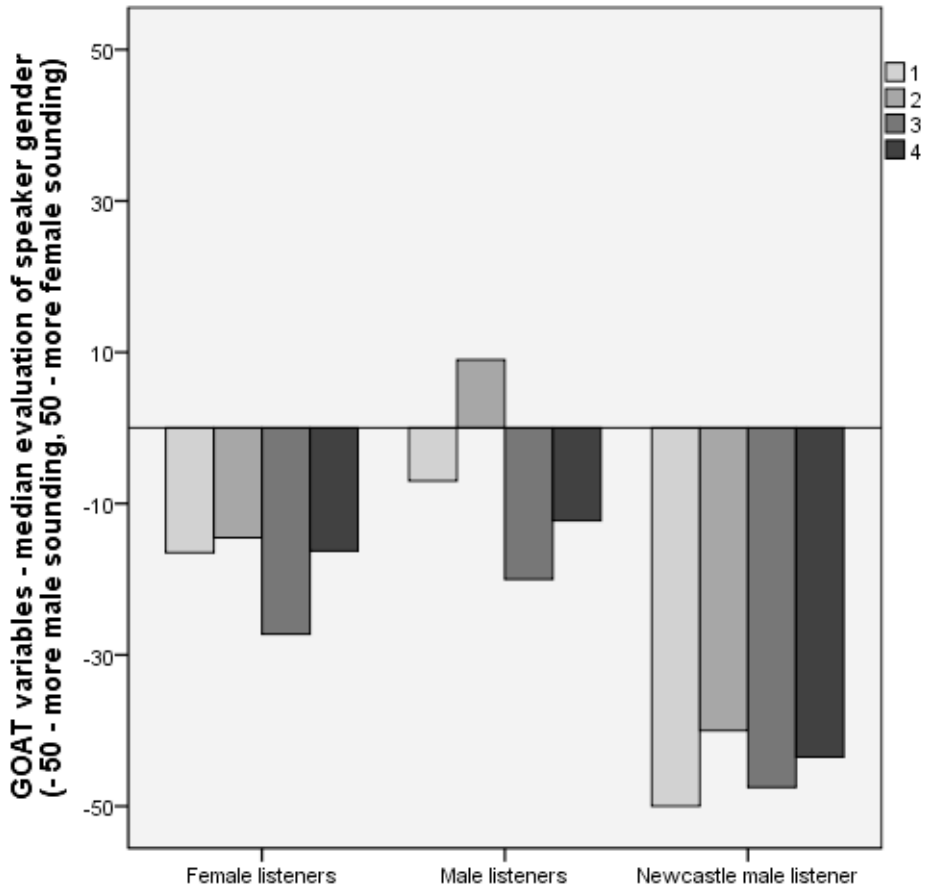


Figure 4. GOAT localised [ʊə] (1), [ɵː] (2) and [iːə] (3) and supra-local [oː] (4) variants -- evaluation of speaker gender.

Figure 4 presents evaluations of speaker gender of the three localised GOAT variants, the centring diphthong [ʊə] fronted monophthongal [ɵː], archaic [iːə] and the supra-local monophthongal variant [oː]. All variants in the GOAT group are associated with male speakers in Tyneside English (Viereck 1968, Watt 1998, Watt 2000, Beal et al. 2012).

The centring diphthong [ʊə] is characteristic of older working-class males, [ɵː] is used most often by younger middle-class males but also older and younger working-class males and [iːə] is found in the speech of older working-class males.

Both localised and the supra-local variants were categorised as male-sounding by all groups of listeners. The only exception was the fronted monophthongal [ɵː] variant, which was perceived as slightly female-sounding by the male listeners. It is worth pointing out however, that across male and female groups of listeners the archaic variant [iːə] was evaluated as the most male-sounding of all the variants.

Interestingly, the Newcastle listener perceived the supra-local variant along with the localised variants as definitely male-sounding.

The shift in perceptions between male and female speakers observed earlier in gender evaluations of the variants of the FACE vowel can be noticed here as well. In general, the male group found the voices to be less definitely male-sounding than the female group.

Figure 5 shows judgements of speaker age of the GOAT vowel variants.

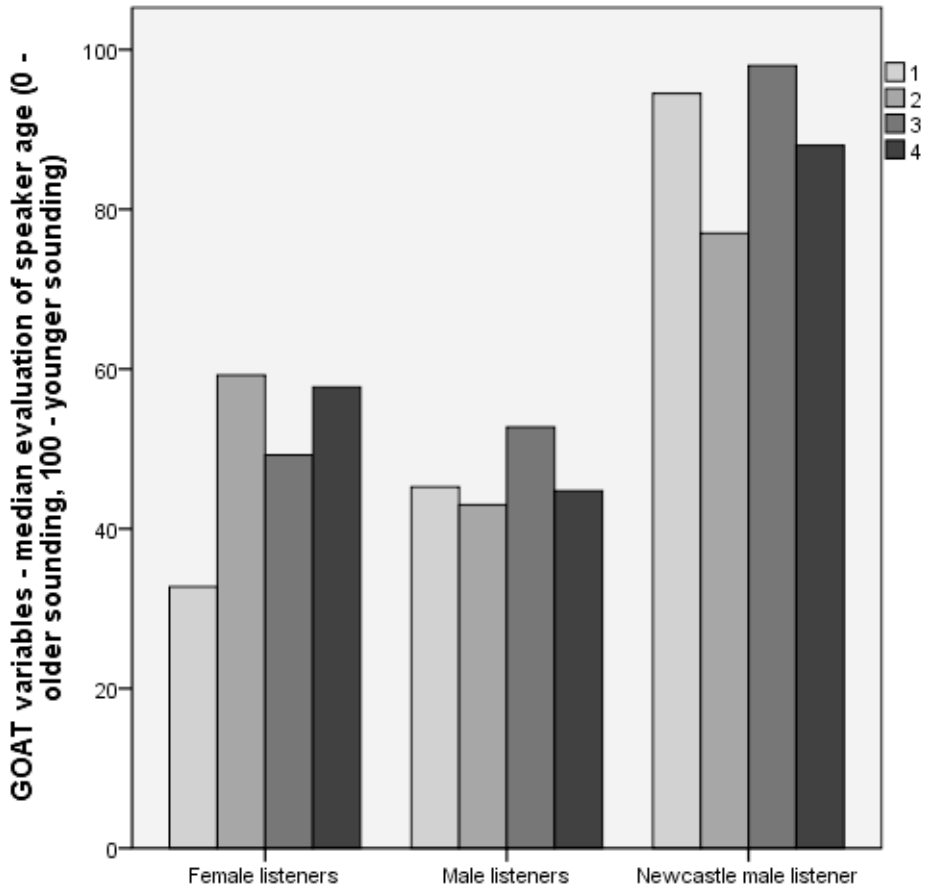


Figure 5. GOAT localised [ʊə] (1), [ɵː] (2) and [iːə] (3) and supra-local [oː] (4) variants -- evaluation of speaker age.

Evaluations of speaker age show that female listeners found the [ʊə] diphthong to be older-sounding and the monophthongal [ɵː] along with the supra-local monophthongal [oː] to be slightly younger-sounding (Fig. 5).

The archaic diphthongal [iːə] was categorised as used by mature but young speakers, which might suggest that the listeners were uncertain as to how to evaluate it. Male listeners, by contrast, rated all variants around the midpoint, finding them to be mature but young-sounding. [ʊə], [ɵː] and [oː] were perceived to be only slightly older-sounding than [iːə] in this group of listeners.

The Newcastle listener, on the other hand, perceived all variants as definitely young-sounding.

Figure 6 illustrates speaker class evaluations of the GOAT vowel variants.

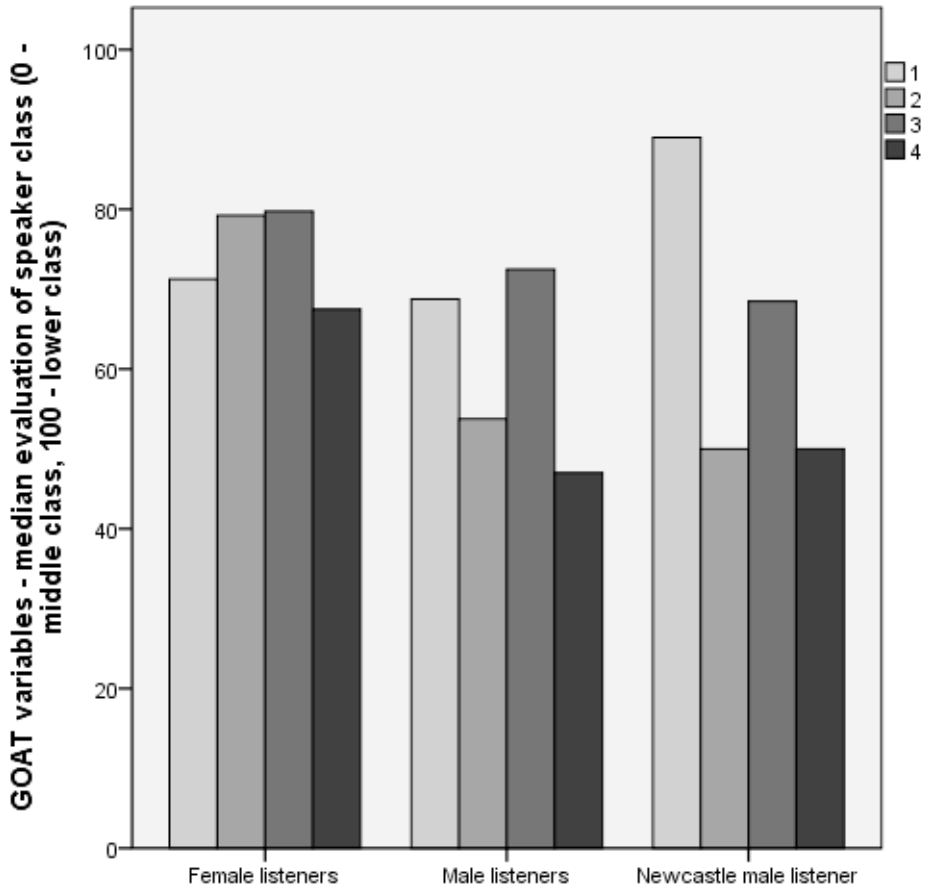


Figure 6. GOAT localised [ʊə] (1), [ɵː] (2) and [iːə] (3) and supra-local [oː] (4) variants – evaluation of speaker social class.

Class evaluation results reveal that female listeners found all variants to be lower-class-sounding. However, there was more variation is class perception in the male group. While the centering diphthong [ʊə] and archaic [iːə] were also found to be overall lower-class-sounding, the fronted monophthong [ɵː] was rated as only slightly lower-class-sounding and the supra-local monophthong [oː] was perceived as slightly more middle-class-sounding. This might suggest that the two variants were found to be used by speakers combining features characteristic of the middle and lower classes.

It is interesting to see that the pattern of evaluations found in the male group is somewhat similar to the one seen in the Newcastle listener.

The evaluations provided by the Newcastle listener correspond with the social classes of speakers who use the variants under investigation.

Interestingly, the results suggest that the Newcastle listener and, somehow, male listeners were possibly sensitive to indexical information such as speaker class carried by the Newcastle GOAT variants.

Figure 7 presents evaluations of speaker gender of the NURSE vowel variants.

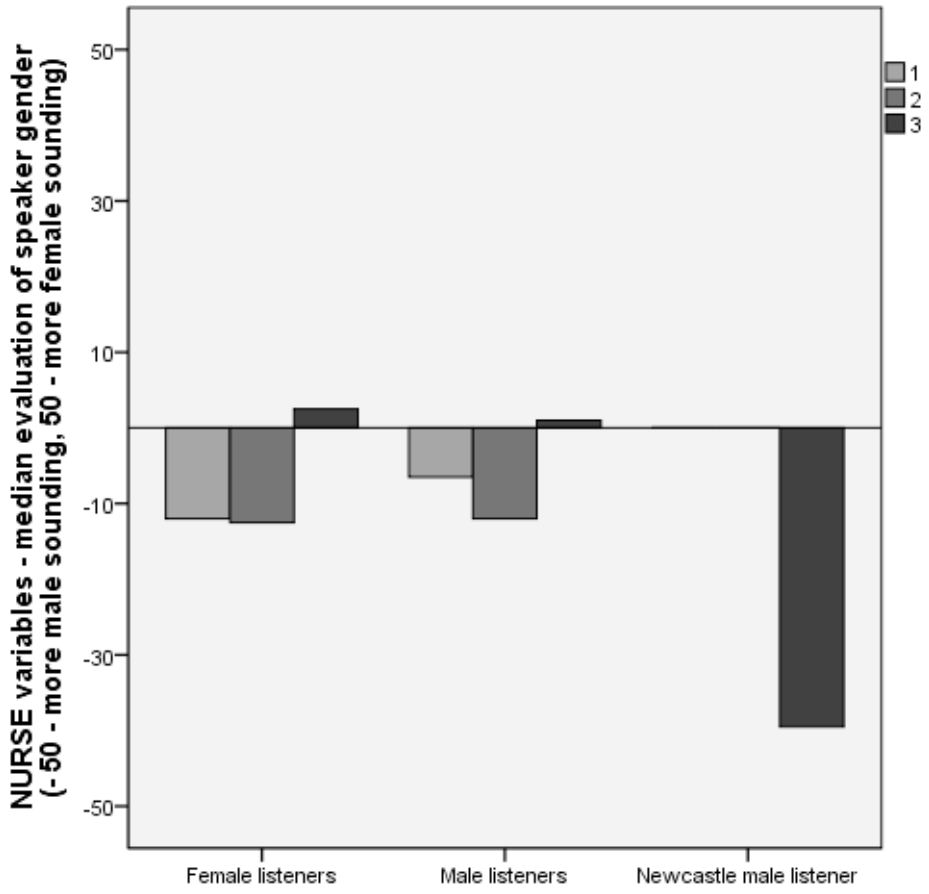


Figure 7. NURSE localised [øː] (1), [ɔː] (2) and supra-local [ɜː] (3) variants -- evaluation of speaker gender.

The final vowel investigated in the study is the NURSE vowel (Fig. 7). Fronted [øː] is typically found in young middle- and working-class females but also in older working-class females. While the retracted [ɔː] is used by older working-class males, the centralised [ɜː] is a supra-local NURSE variant (Watt 1998, Watt & Milroy 1999, Beal et al. 2012).

Both groups of listeners categorised the local variants almost unanimously as overall male-sounding. However, male listeners found the retracted variant to be more female-sounding. Furthermore, the evaluations were in the upper regions of the midscale which means that the listeners did not find the variants to be strongly male-sounding.

The supra-local variant was perceived as non-gender-specific by the two groups of listeners.

Interestingly enough, the Newcastle listener categorised the localised variants as gender-ambiguous, yet he categorised the supra-local variant as definitely male-sounding.

Figure 8 illustrates evaluations of speaker age of the NURSE vowel variants.

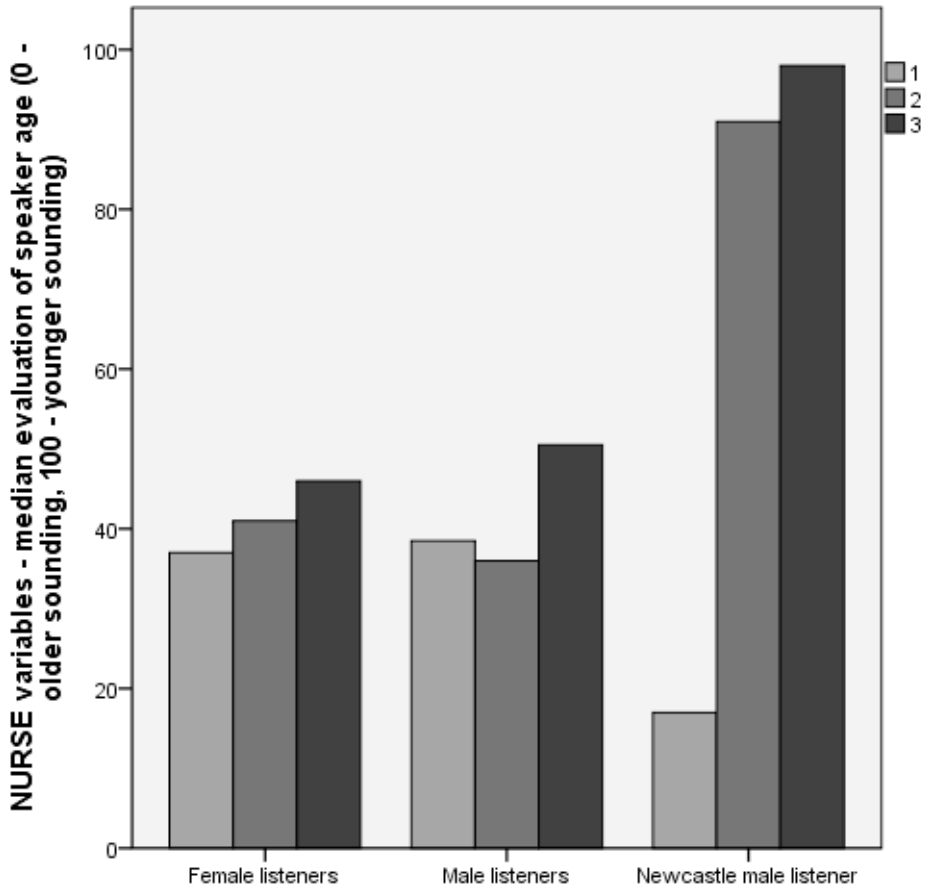


Figure 8. NURSE localised [øː] (1), [ɔː] (2) and supra-local [ɜː] (3) variants -- evaluation of speaker age.

Age categorisation results reveal that male and female groups of listeners found the localised variants to be in general slightly older-sounding (Fig. 8).

The supra-local variant, on the other hand, was perceived as only slightly younger-sounding than the localised variants. Both male and female listeners evaluated it as mature but young-sounding.

The Newcastle listener found the female variant [øː] to be definitely old-sounding and the following variants, male [ɔː] and supra-local [ɜː], as definitely young-sounding.

Figure 9 presents evaluations of speaker social class of the NURSE vowel variants under investigation.

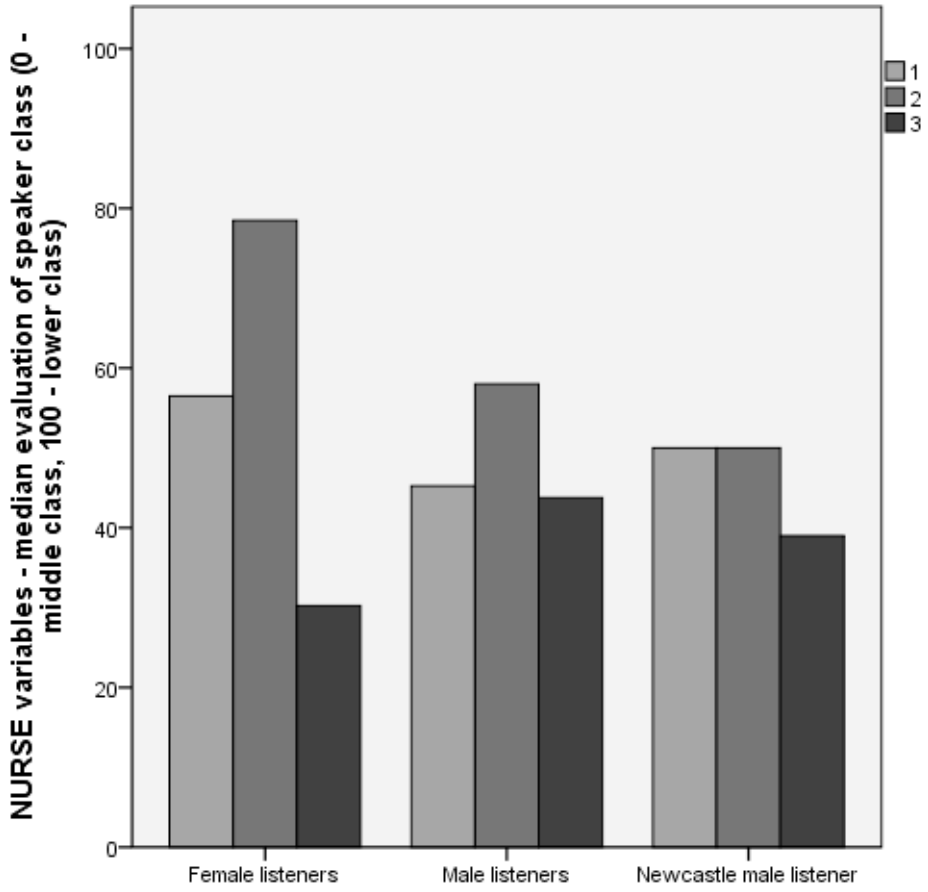


Figure 9. NURSE localised [øː] (1), [ɔː] (2) and supra-local [ɜː] (3) variants -- evaluation of speaker class.

The class evaluation results show that the female variants [øː],was categorised as combining middle- and lower-class features by the female group (Fig. 9). Male evaluations seem to be heading in the same direction yet, they are not as strong, since the variant was categorised as slightly more middle-class-sounding.

The traditional male variant [ɔː], on the other hand, was judged as slightly lower-class-sounding by the male group, and definitely as lower-class by female listeners.

The supra-local variant was evaluated as overall middle-class-sounding across all groups of listeners, with male listeners finding it only slightly middle-class-sounding.

Similarly, as was the case with class evaluation of the FACE variants, also here the Newcastle listener categorised the two lower class localised variants at midpoint.

5. Conclusions

Preliminary results show that listeners from outside of the North East did not seem to be sensitive to indexical information present in localised Newcastle variants or even supra-local North East variants, relating to gender, age or class. As a result, listeners could not extract this information form the acoustic signal with any significant accuracy. In fact, in a number of cases, localised and supra-local or even national variants were perceived as almost identical or quite similar. Therefore, the main question of the role of gendered phonetic variables in evaluating speaker-social information seems to have been answered only partially. It will be further investigated in the main study whether familiarity with the dialect under investigation provides more promising results.

Two patterns of male-female evaluations of speaker gender and class were recorded. It was observed that males tended to evaluate speaker gender as overall more female-sounding than female listeners, who found the stimuli to be more male-sounding in comparison.

The second pattern refers to evaluations of speaker class, where female listeners tended to evaluate speaker class as lower than male listeners.

Despite these patterns visible in the plots, statistical analyses did not reveal any significant differences between male and female listeners when evaluating localised or supra-local variants. However, it might be that the results of statistical tests did not indicate statistical significance simply because the sample was not large enough. Overall, there was not enough evidence to reject the null hypothesis and assume there were perceptual differences between male and female listeners.

Finally, as far as class evaluations are concerned, ratings around the midpoint on the scale suggest that perhaps listeners failed to identify any specific speaker class information, or they did not feel comfortable judging the parameter. This could be the case especially when listeners recognise a variant as familiar-sounding.

References

ADOBE AUDITION 3.0 User Guide. (2007). *Adobe Systems Incorporated.*

ASSMANN, P.F. & NEAREY, T.M. 2007. Relationship between fundamental and formant frequencies in voice preference. *Acoustical Society of America* 122 (2), 35-43.

BALL, P. & GILES, H. 1988. "Speech Style and Employment Selection: The Matched Guise Technique." In: *Doing Social Psychology: Laboratory and Field Exercises*, ed. Glynis M. Breakwell, Hugh Foot, and Robin Gilmour, 121–49. Cambridge: Cambridge Univ. Press.

BEAL, J., BURBANO ELIZONDO, L. & LLAMAS, C. 2012. *Urban North-Eastern English: Tyneside To Teesside.* Edinburgh : Edinburgh University Press.

BEZOOIJEN, R. VAN. 1988. "The Relative Importance of Pronunciation, Prosody and Voice Quality for the Attribution of Social Status and Personality Characteristics." In: *Language Attitudes in the Dutch Language Area*, ed. Roeland van Hout and Uus Knops, 85–103. Dordrecht: Foris.

BEZOOIJEN, R. VAN & GOOSKENS, C. 1999. Identification of Language Varieties : The – Contribution of Different Linguistic Levels. *Journal of Language and Social Psychology,* 18(1). 31-48.

BIEMANS, M. 2000. *Gender Variation in Voice Quality.* PhD dissertation, University of Utrecht. Utrecht: LOT.

CLOPPER, C., CONREY B., PISONI D.B. 2005. Effects of Talker Gender on Dialect Categorization. *Journal of Language and Social Psychology*. 24: 182.

COLEMAN, R.O. 1971. Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research* 14, 565-577.

E-PRIME. (2012). Psychology Software Tools, Inc.

DOCHERTY, G. J. & FOULKES, P. 1999. Derby and Newcastle: instrumental phonetics and variationist studies. In P. Foulkes & G. Docherty eds. *Urban Voices: Accent Studies in the British Isles*. London: Arnold, 46–71.

FOULKES, P., DOCHERTY, G.J., KHATTAB, G. & YAEGER-DROR, M. 2010. Sound judgements: perception of indexical features in children's speech. In Preston, D. & Niedzielski, N. (ed.) *A Reader in Sociophonetics*. Berlin: de Gruyter, 327-356.

GROOT, P. 2013. *Pauls' Pages*. [Online] Available at: http://www.pfcgroot.nl/ [Accessed 14 May 2013].

HUBBARD, D.J. & ASSMANN, P.F. 2013. Perceptual adaptation to gender and expressive properties in speech: The role of fundamental frequency. *Acoustical Society of America* 133 (4), 2367–2376.

JOHNSON, K.,ELIZABETH A. STRAND, E.A. & D'IMPERIO, M. 1999. Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27, 359-384.

LAMBERT, W. E., HODGSEN, R. C., GARDNER, R. D. & FILLENBAUM, S. 1960. Evaluational Reaction to Spoken Language. *Journal of Abnormal and Social Psychology,* 60, 44–51.

LASS, N. J., HUGHES, K. R., BOWYER, M. D., WATERS, L. T. & BOURNE, V. T. 1975. Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Acoustical Society of America* 59 (3), 675-678.

MILROY, J., MILROY, L., HARTLEY, S. & WALSHAW, D. 1994a. Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English. Language Variation and Change, 6, 327-357.

MILROY, J., MILROY, L. & HARTLEY, S. 1994b. Local and supra-local change in British English: the case of glottalisation. *English World-Wide* (15), 1-33.MUNSON, B. 2007. The Acoustic Correlates of Perceived Masculinity, Perceived Femininity, and Perceived Sexual Orientation. *Language and Speech* 50 (1), 125 –142.

MUNSON, B. & BABEL, M. 2007. Loose Lips and Silver Tongues, or, Projecting Sexual Orientation Through Speech. *Language and Linguistics Compass* 1/5, 416–449.

PEARCE, M. 2009. A Perceptual Dialect Map of North East England. *Journal of English Linguistics* 20 (10) , 1-31.

PURNELL, T., IDSARDI, W. & BAUGH, J. 1999. Perceptual and phonetic experiments in American English dialect identification. *Journal of Language and Social Psychology,* 18, 10–30.

VIERECK, W. 1968. A diachronic-structural analysis of a northern English urban dialect. *Leeds Studies in English*, 65-79. [Online] Available at: [http://www.leeds.ac.uk/lse/lse.html [Accessed 27 Sept 2013].

WATT, D. 1998. *Variation and Change in the Vowel System of Tyneside English*. PhD thesis, the University of Newcastle upon Tyne.

WATT, D. 2000. Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Language Variation and Change*, 12, 69–101.

WATT, D. 2002. 'I don't speak with a Geordie accent, I speak, like, the Northern accent': Contact-induced levelling in the Tyneside vowel system. *Journal of Sociolinguistics* 6(1), 44-63.

WATT, D. & ALLEN, W. 2003. Tyneside English. *Journal of the International Phonetic Association*, 33, 267 - 271.

WOLFRAM, W. 2000. On Constructing Vernacular Dialect Norms. *Chicago Linguistic Society* 36, 335–58.

Ania Kubisz
PhD Student
Department of Language and Linguistic Science
University of York
Heslington
York
email: ania.kubisz@york.ac.uk

PHONOLOGICAL 'WILDNESS' IN EARLY LANGUAGE DEVELOPMENT: EXPLORING THE ROLE OF ONOMATOPOEIA

CATHERINE E. LAING
University of York

Abstract

This study uses eye-tracking to single out the role of 'wild' onomatopoeia in language development, as described by Rhodes (1994). Wildness–whereby extra-phonetic features are used in order to reproduce non-human sounds–is thought here to facilitate infants' understanding of onomatopoeic word forms, providing a salient cue for segmentation and understanding in the input. Infants heard onomatopoeic forms produced in familiar and unfamiliar languages, presented in a phonologically 'wild' (W) or 'tame' (T) manner. W forms in both familiar and unfamiliar languages were hypothesised to elicit longer looking times than T forms in both familiar and unfamiliar languages. Results reflect the role that onomatopoeia play in early language development: wildness was not found to be a factor in infants' understanding of word forms, while reduplication and production knowledge of specific stimuli generated consistent responses across participants.

1. Introduction

Onomatopoeia appear amongst the early words of infants acquiring many languages, yet it is not uncommon for studies of child language development to disregard these word forms when analysing infant speech (e.g., Behrens, 2006, Genesee et al., 2008). While onomatopoeia could be considered as marginal to the adult language, they often constitute a considerable portion of an infant's first word forms (Kauschke & Klann-Delius, 2007, Menn & Vihman, 2011 (see appendix)), and a focus on this early vocabulary may explain some of the developments that follow as an infant's lexicon progresses towards the adult model.

Onomatopoeia are derivative of sound symbolism, which is a fully integrated feature of many languages, including Korean and Japanese (Ivanova, 2006). Sound symbolism, or 'mimetics', draws from the phonetic properties of a word to represent the synesthetic features of the object or state that it describes (Rhodes, 1994), resulting in a highly expressive parallel lexicon which is fully established as part of the language (e.g., Japanese 'pika' *a flash of light*, 'goro' *a heavy object rolling* (Kita, 1997)). Onomatopoeic word forms differ in that they constitute phonetic imitations of sounds in the environment, produced within the limits of the vocal tract. Unlike mimetics, onomatopoeia do not express physical features through the word's phonetic or phonological properties, but rather they serve to phonetically reproduce non-human sounds (e.g., 'thud', 'vroom').

The use of mimetics in language development has been well-documented in the literature, found to facilitate the learning of Japanese novel verbs amongst Japanese and English-speaking adults and infants (Imai *et al*., 2008, Kantarzis *et al*., 2011). Mimetic forms appear at the very onset of Japanese infants' word production, where they are used with a high level of accuracy (Tsujimura, 2005), and become increasingly complex over time as the infant acquires a full lexicon of both mimetic and non-mimetic words (Iwasaki *et al*., 2007). This

Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York (**PARLAY 2013**) conference
 ISSN 1758-0315 Issue 1

evidence towards a role for mimetics in early word learning suggests that onomatopoeic words may similarly support the learning of new word forms in the early output.

Whether the infant is acquiring a language that is rich in sound symbolism, such as Japanese, or a language which contains little (if any) sound symbolism, such as English, it appears that sound symbolic words–both mimetic and onomatopoeic–could be perceptually salient to infants acquiring their first word forms. Rhodes (1994) describes a model of 'wild' and 'tame' onomatopoeia, which explains the extent of phonetic imitation found in the features of the word form. 'Tame' forms are produced within the phonetic norms of the ambient language, adhering to normalised phonological structures that are familiar to the speaker, while 'wild' forms make use of the vocal tract's full capacity in order to approximate as closely as possible to the sound that the speaker is imitating. Wild forms draw upon vocal gestures that are not ordinarily used in the adult language, raising the question as to whether it is precisely these phonetic 'special-effects' that render onomatopoeic forms more salient in the speech stream, thus facilitating perception, memory, and eventually production of infants' earliest word forms. Wild onomatopoeia do not correspond to the typical segments and syllable-structures of the adult language, and may provide a perceptual attractor for infants as they attend to the speech stream.

This study uses eye-tracking to address infants' perception of wildness in onomatopoeic forms, which is presumed to provide a highly salient linguistic 'hook' in the input, facilitating lexical memory and the formation of word representations in language development. Wild features are assumed here to provide prosodic cues in the input, while being easier to recall than the typical native language phonology, to which tame forms adhere. It is hypothesised that infants will respond most systematically to the wild forms, thought to be easily recognisable due to their idiosyncratic 'special effect' features. This would indicate that infants respond most readily to the linguistically atypical features of the speech stream when acquiring language, underlining those features which are essential to the earliest stages of language development, but which do not necessarily match the words or phonemes that will eventually form the adult output.

2. Method

2.1. Participants

Nineteen Swedish infants (10 male, 9 female) between the ages of 14 and 16 months were tested (mean age 461.5 days). Infants were all full-term, and acquiring Swedish as their first language. A further five infants participated in the experiment but were excluded from the analysis due to fussiness during the eye-tracking procedure (4) or experimenter error (1).

2.2. Stimuli

Six onomatopoeic words (OWs)–all animal sounds[1]–which consistently appeared on English, Swedish, German and French adaptations of the MacArthur-Bates Communicative Development Inventory (CDI, Fenson *et al*., 1994) were selected for use in the experiment to ensure that participants were likely to have had prior experience of the stimuli. Two different photographic images of each of the corresponding animals were selected: the animals were all stood facing in the same direction, looking towards the infant from the right hand side, and

[1] The OW equivalents of COW, SHEEP, DOG, CAT, DUCK and ROOSTER were used in the experiment.

presented on a grey background. OWs were recorded in Swedish (the familiar language, L_F), and in languages unfamiliar to the infants (L_U)–Chinese, Arabic and Urdu.

Audio stimuli were recorded by native speakers, all female postgraduate students in the Linguistics departments at the universities of York or Stockholm. Each student was first asked to produce the OW as they would produce it when speaking to a toddler, as if imitating the animal in question ('wild' W). The students were then asked to produce the words with no added prosodic features, keeping to the natural phonology and stress pattern of their native language ('tame' T). Each word was produced once in each recording, adhering to the conventional full form of the word; words which would normally undergo reduplication were reduplicated (e.g. *quack quack*), while those which the speakers deemed as having no reduplication in production were recorded without reduplication (e.g. *cock-a-doodle-doo*).

Four adults, none of them speakers of any of the L_U languages, were then tested on their recognition of the W_U and T_U stimuli prior to the analysis. Only one of the 24 stimuli was found to be unrecognisable by all of the adults, which was removed from the analysis. Of the seven stimuli that were judged incorrectly by at least one of the adults, all but one were produced in a T manner. These results confirmed the suitability of the stimuli used in the infant experiment, as well as supporting the hypothesis that wildness facilitates word recognition.

2.3. Procedure

The experiment was controlled using E-Prime, with the visual stimuli presented using a 17" Tobii Studio 1750 eye-tracking monitor. Caregivers held the infant on their laps in a chair placed in front of the screen, and a five-point infant calibration was taken for each participant before the experiment began. The experimental procedure lasted approximately four minutes, during which time the caregiver was asked to wear headphones playing music from a Swedish radio station.

The experiment consisted of a salience phase and a test phase: during the salience phase pairs of images were displayed on the screen for 4000ms, before a centralising image of a baby appeared in the middle of the screen which served to 'reset' the infants' eye-gaze prior to the test phase. The image disappeared automatically upon fixation (or after 4000ms if the infant did not fixate), and the OW was heard through speakers on either side of the screen immediately after offset of the fixation image; the test phase lasted for 3000ms. After the experiment infants were rewarded with a certificate and parents were asked to complete a Swedish CDI questionnaire.

Each infant heard a total of 24 OWs: each of the four conditions (W_F, W_U, T_F and T_U) for each of the six animals, with a distribution of all three unfamiliar languages across the stimuli. The order of data output and the target's location on screen was randomised using E-Prime. Selection of the distractor image was partially randomised in E-prime according to the size of animal in the target image: to ensure against confusion between the images (e.g. sheep and dog, duck and rooster), animals were grouped into two categories–'small' and 'large'–and for each trial the distractor image was chosen from the opposite category to avoid ambiguity.

3. Results

Fixations during test phase were analysed from a window of 300-1800ms after onset of the stimulus. The proportion of looking towards the target was calculated for each trial as a percentage of the total fixation time for both target and distractor, and a mean looking time was calculated for each infant in each condition.

3.1. Wildness and Familiarity

A two-way repeated measures ANOVA was carried out with two factors: wildness (W vs. T) and familiarity (L_F vs. L_U), with proportion of looking towards the target image as the dependant variable (n = 19). This revealed no significant effect for wildness ($F(1, 18) = 3.428$, $p = .081$) or familiarity ($F(1, 18) = .486$, $p = .495$). The interaction was not significant either ($F(1, 18) = 1.617$, $p = .220$), and as is evident in Figure 1, results were around chance (0.50) for all conditions.

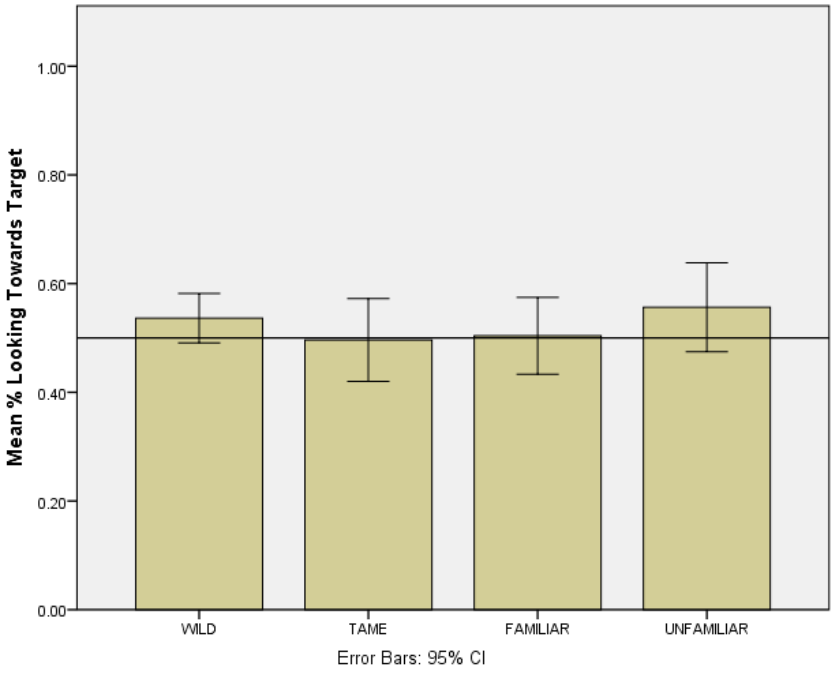


Figure 1: Results for all infants across conditions

These results raise the question of the interpretation of 'familiarity' in this experiment: were the stimuli really familiar to the infants, and if so, how familiar? In order to address this question, it is necessary to investigate the infants' knowledge of the individual OW forms used in the experiment. Results from the CDI questionnaires were used to determine infants' knowledge of individual word forms, both in terms of the OW, and the conventional word (CW) equivalent (for example, *woof* versus *dog*).

3.2. Knowledge of Stimuli

Breaking down the findings in this manner made it possible to explore the results in more depth. Infants were given two knowledge scores for each of the six target words–one for the OW and one for the CW–based on whether or not they were able to produce the word form. Results were then separated into knowledge groups in accordance with these scores. This

approach was based on the assumption that being able to produce the CW would only strengthen an infant's representation of its OW counterpart, while also suggesting that an infant is more advanced in his language development when he has started to produce CWs. The scoring conventions are presented in Table 1.

	Produces word	Doesn't produce word[2]
OW	1	0
CW	1	0

Table 1: Knowledge Scoring for OWs and CWs

Scores were allocated for individual stimuli, meaning that an infant may be in OW1 and CW0 group for DOG if he is able to produce *woof* but not *dog*. The average looking time towards target for infants in each knowledge group was calculated.

3.2.1. Knowledge 0

The CW0 group (n = 17) was analysed with a two-way repeated measures ANOVA using the same two factors, and a significant interaction between wildness and familiarity was found ($F(1, 16) = 8.557$, $p = .01$). As shown in Figure 2, infants who were not able to produce the CW looked longer than chance at L_U words only. This is illustrated more clearly in Figure 3, where familiarity can be seen to interact with wildness.

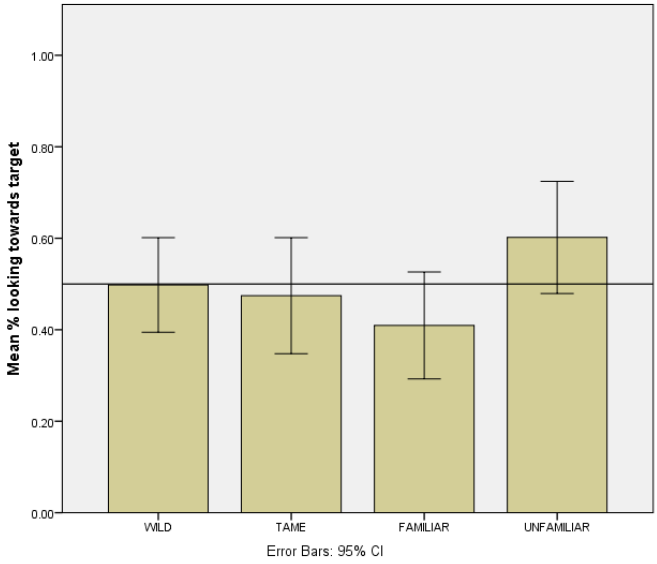


Figure 2: Results for CW0 infants

[2] Infant was scored as 0 if they had not yet produced the word form, whether or not they were reported to understand the form by the parent. Initially three scores were given ('produces', 'understands', 'doesn't understand') but as no difference was found between 'understands' and 'doesn't understand', these categories were merged.

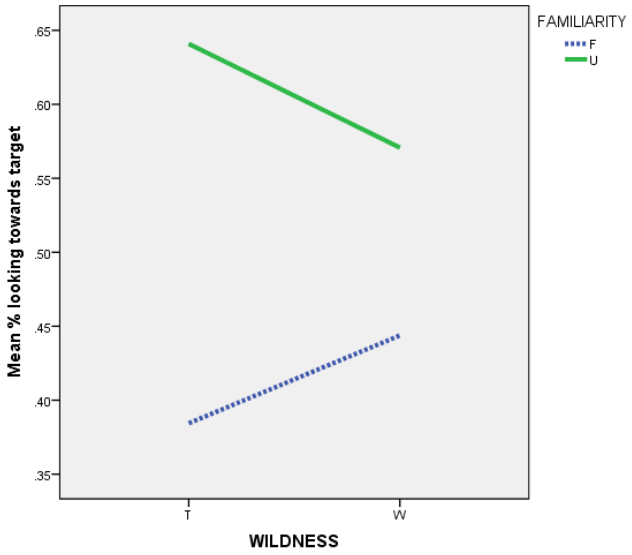


Figure 3: Effects of wildness and familiarity for CW0 infants

The OW0 group (n = 17) was then subjected to the same analysis, and a significant effect was found for familiarity ($F(1, 16) = 5.346$, $p = .034$), while wildness yielded a marginally significant effect ($F(1, 16) = 4.125$, $p = .059$). No effect was found for the interaction of wildness x familiarity. Again, Figures 4 and 5 show a bias towards L_U words, while infants in this group tended to respond above chance to T but not W stimuli.

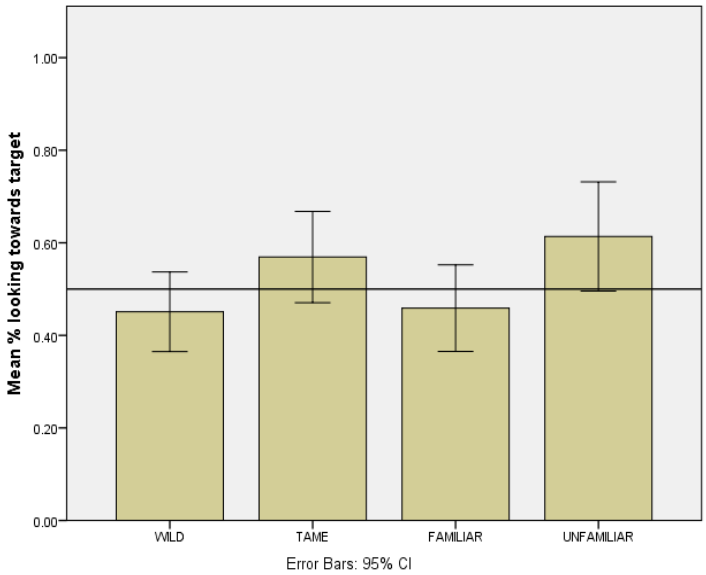


Figure 4: Results for OW0 infants

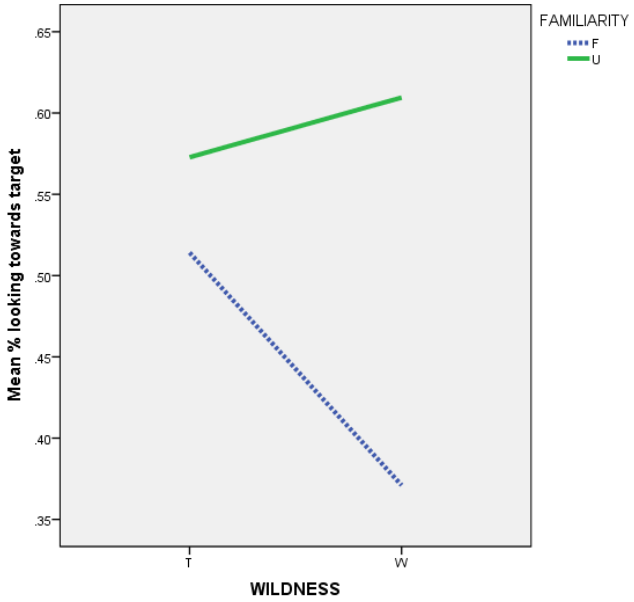


Figure 5: Effects of wildness and familiarity for OW0 infants

Finally, infants in knowledge group 0 for both CWs and OWs were analysed (n = 13), and the interaction between wildness and familiarity was found to be significant ($F(1, 12) = 6.037$, $p = .03$). The bias towards L_U stimuli can be observed in Figure 6, which was most pronounced in the W condition (Figure 7).

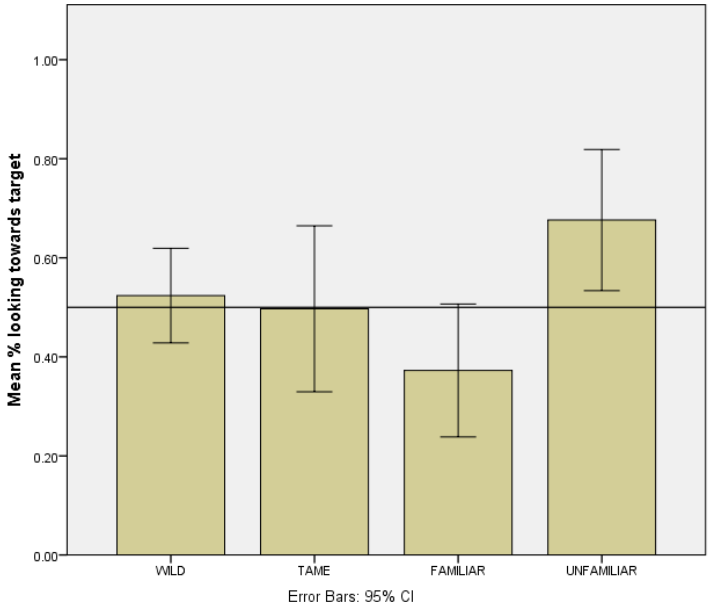


Figure 6: Results for OW0+CW0 infants

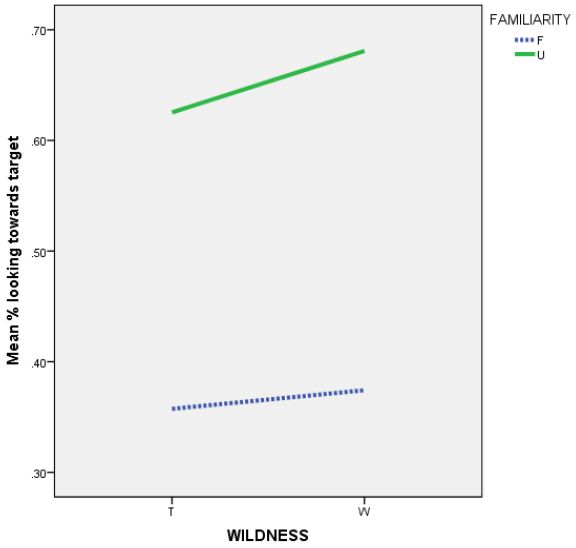


Figure 7: Effects of wildness and familiarity for OW0+CW0 infants

Results from this group show a consistent trend towards the L_U condition in all three parts of the analysis, demonstrating a bias towards this condition amongst infants in the earlier stages of language development. No trend relating to wildness was found in this condition, however, which indicates that wildness did not play a role in these infants' perception of the stimuli. This goes against the original hypothesis that infants will draw from the wild features of OWs in early language development in order to facilitate understanding.

3.2.2. Knowledge 1

Results for CW1 and OW1 groups were analysed using the same model. No significant effect was found for either group (CW1: n = 9, familiarity: $F(1, 8) = .789$, $p = .4$, wildness: $F(1, 8) = .244$, $p = .635$; OW1: n = 13, familiarity: $F(1, 12) = .537$, $p = .478$, wildness: $F(1, 12) = .034$, $p = .857$). In contrast to the results from knowledge group 0, it can be seen in Figure 8 that infants in the CW1 group tended to look to target above chance in all conditions except L_U. A similar effect can be found for infants in the OW1 group, who looked to target above chance in all conditions (Figure 9).

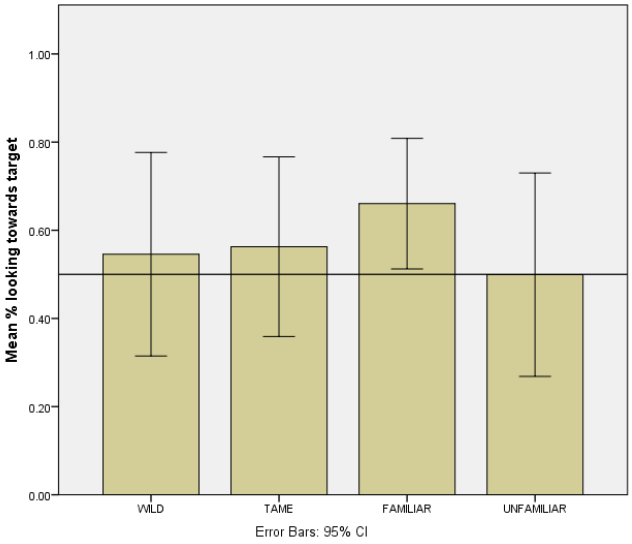


Figure 8: Results for CW1 infants

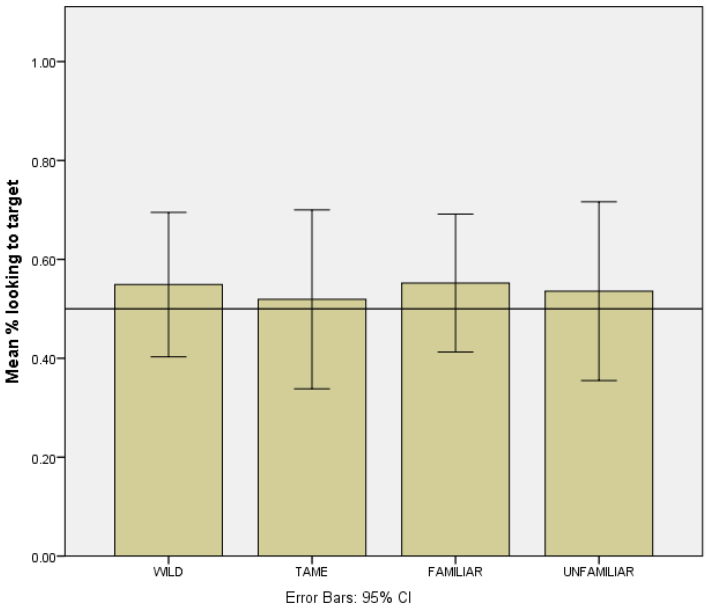


Figure 9: Results for OW1 infants

Infants with a knowledge score of 1 for both OWs and CWs were analysed (n= 7). No significant effect was found (Wildness: $F(1, 6) = 1.764$, $p = .232$, Familiarity: $F(1, 6) = 3.995$, $p = .093$), but Figure 10 shows biases towards both W and L_F stimuli. The lack of effect in these three tests could be due to low sample size, since the results show much longer looking times than those for knowledge group 0.

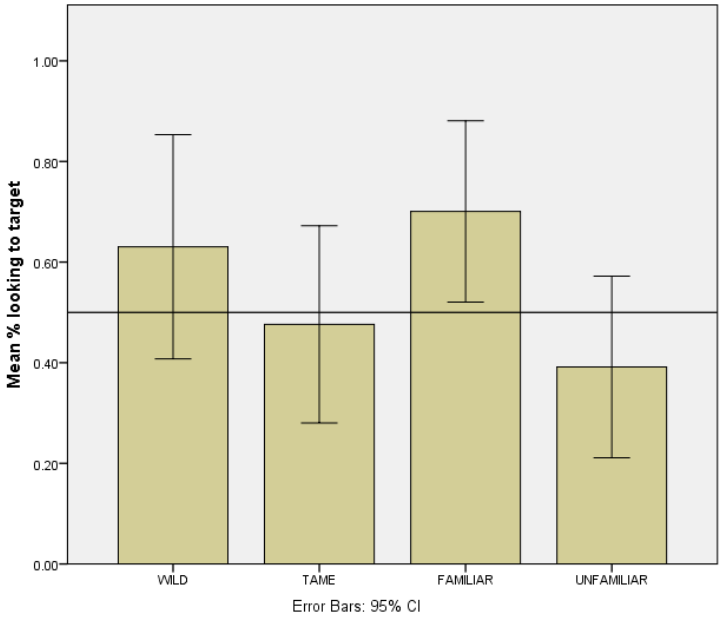


Figure 10: Results for OW1+CW1 infants

Infants in this group have been found to show a very different response to those in knowledge group 0. Figure 11 shows looking times for both OW+CW stimuli, where the difference between the groups can be seen more clearly. Results for responses to familiarity are almost inverted across the two groups, while wildness appears to affect responses only for infants in knowledge group 1.

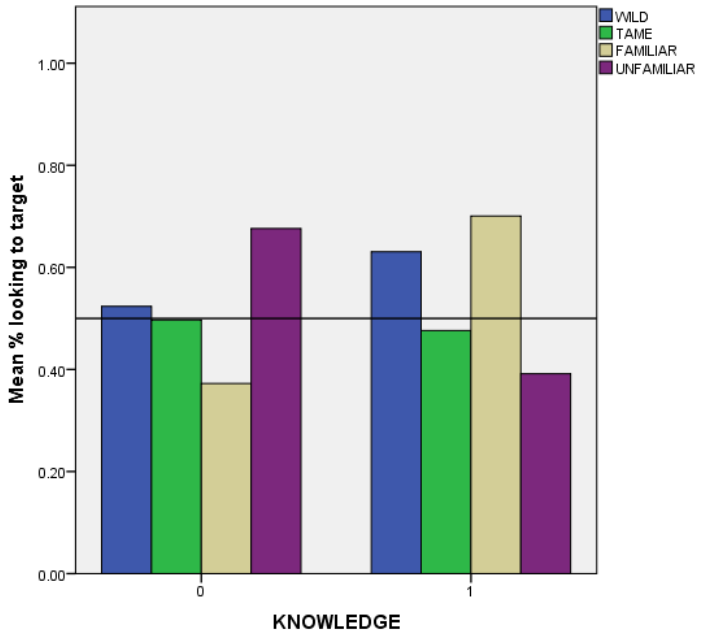


Figure 11: Results for both OW+CW knowledge groups across all conditions

3.3 *Reduplication*

Finally, the L_F and L_U stimuli were compared in order to discern whether any differences between forms in individual languages could be causing the discrepancy in response between the two groups. It was found that 100% of the L_U stimuli contained reduplication, of which

all but one form were fully reduplicated (e.g. Chinese DOG [wʌŋwʌŋ], Arabic ROOSTER [kukuku:ku.kukuku:ku]), while only two of the six L_F OWs contain reduplication. These two stimuli (both L_F and L_U forms) were analysed separately in order to single out reduplication as a feature in this analysis. No preference was observed when the data were considered as a whole, while trends across the two knowledge groups remained consistent with previous findings (Figures 12 and 13): a significant effect was found for familiarity in knowledge group 0 ($F(1, 10) = 5.248$, $p = .045$ ($n = 11$)), and no effect was identified amongst the knowledge group 1 participants, possibly due to a low sample size for this group ($n = 5$).

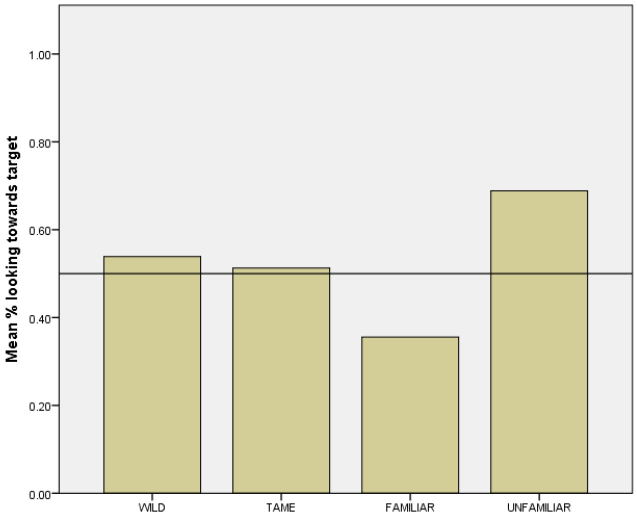


Figure 12: Results for reduplicated stimuli across OW0+CW0 infants

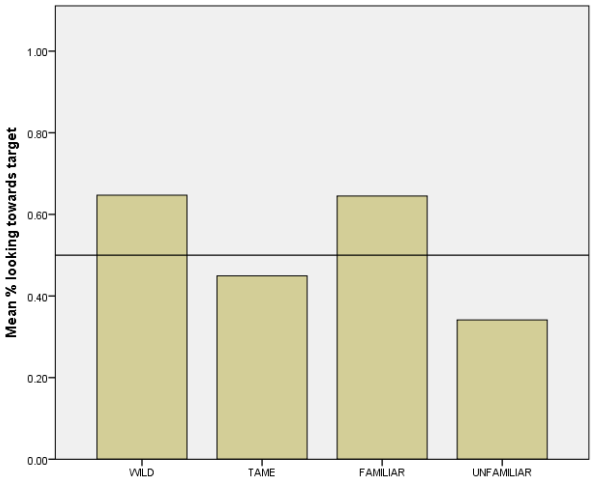


Figure 13: Results for reduplicated stimuli across OW1+CW1 infants

Duration of the individual stimuli was then measured using Praat, and a four-way repeated measures ANOVA showed L_U forms to be significantly longer then L_F forms ($F(3, 8) = 14.61$, $p = .001$). Furthermore, a two-way independent measures ANOVA found that W forms were significantly longer than T forms ($F(1, 22) = 4.803$, $p = .039$), as shown in Figure 14.

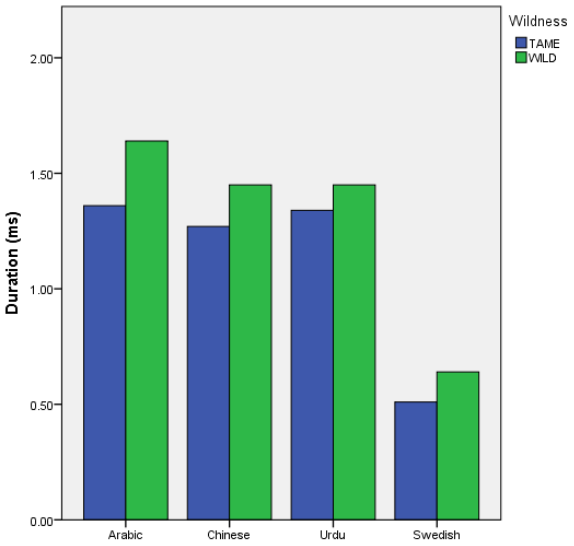


Figure 14: Average duration of W and T stimuli across languages

Results were then divided up according to duration of the various stimuli. Those stimuli that were shorter than 1.25s (the midpoint of the range of all results) were classed as 'short', and included all of the L_F and 11 of the L_U stimuli, and those which were 1.25s and longer were classed as 'long', and included only L_U stimuli. A two-way repeated-measures ANOVA showed that infants looked significantly longer at 'long' stimuli: $F(1, 18) = 9.33$, $p = .007$ ($n = 19$) (see Figure 15).

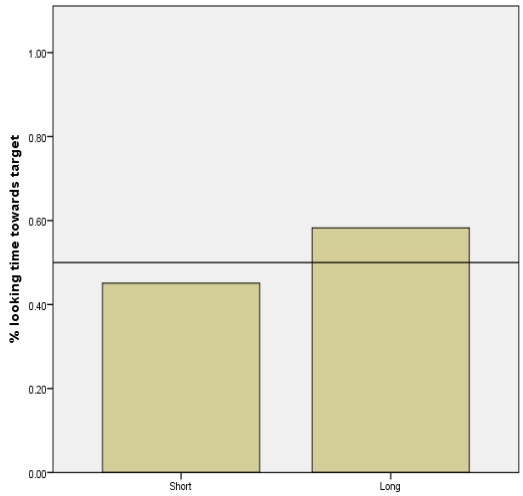


Figure 15: Average looking time towards 'short' and 'long' stimuli

While reduplication was not found to be a factor in its own right, it appears to bear some importance in terms of these results, which show that longer words elicit longer looking times. In the case of most of the individual stimuli used in this experiment, reduplication essentially doubled the infants' input, repeating the OWs in full to not only increase the amount of input received in each trial, but also to reiterate the information that the infants were receiving for each.

In sum, the results did not stand up to the hypothesis, and a closer analysis has shown that wildness does not appear to facilitate perception amongst infants who are unable to produce the word forms. Infants have been found to respond above chance to both W and T stimuli,

but the results have shown the strongest biases towards L_U and L_F stimuli. It seems that wildness is a superfluous feature of OWs, which does not facilitate comprehension. Furthermore, reduplication has been found to be a prominent feature of the L_U stimuli, which may be an indirect contributing factor in the recognition of these word forms. Overall, the findings from this experiment suggest that infants who were able to produce the words perceived in the experiment attended to the speech stream differently to those who cannot yet produce the words, but that the infants as a group were biased towards those words which they perceived for the longest period. These responses will be discussed further below.

4. Discussion

These findings go against the hypothesis that infants would show the strongest response to W forms in both L_F and L_U conditions. No result was found when the data was considered as a whole, but patterns began to emerge in the results when these were broken down further to consider the infants' knowledge of each of the target words. Infants who were not yet able to produce the target word form– either OW, CW or both–showed a significant preference for the L_U condition, while at no point did wildness appear to play a role for this group of participants. The opposite result was found for participants in knowledge group 1, who showed a preference for L_F stimuli throughout the data, but which was not substantiated by a significant result in the analysis.

The opposite effects observed in the results of knowledge groups 0 and 1 can explain why no effect can be seen in Figure 1: together, the two sets of results work to 'cancel out' one another, which reflects the extent of the discrepancy between the two groups. The obvious question to ask here is why infants at different stages of lexical development are producing these opposite effects, and what it might be that changes between the stages at which an infant comprehends a word form and produces a word form that leads them to this stark difference in perception.

Infants in knowledge group 0 responded most strongly to the L_U forms, indicating a strong reliance on the one thing that most of these forms had in common: extended word duration. The results seen here could reflect a lag in response time amongst this group, who may not have begun perceiving the word form as quickly as those who were able to produce the target word. Reduplication may also have contributed to these results, not only in adding duration to the infants' experience of individual stimuli, but also by reiterating the phonological information for each of the L_U stimuli, and thereby facilitating perception. The use of reduplication in the stimuli used in this experiment is consistent with its common occurrence in IDS (Sundberg, 1998), which is thought to increase the salience of word forms in the early input.

While infants in knowledge group 1 do not show any significant trends in their results, emerging patterns in response to the L_F stimuli (both W and T) may relate to the influence of production on perception: infants in the early stages of word production are more likely to be drawn to those words in the input that they can produce themselves, and thus responses may be stronger towards these stimuli. This 'articulatory filter', as discussed by Vihman (1993), reflects the role of auditory feedback from an infant's own output in terms of his perception of the input. As an infant's phonological output develops, the articulatory filter prompts him to attend to those features of the input which are active in the output, a sort of 'phonetic matching' which supports the infant in the development of phonological memory (Vihman & DePaolis, 2000; Vihman, 2014). As stated by Vihman (2014), 'the child's first word production should *facilitate recognition of* (*and support attention to*)…words that resemble

the word forms that are in the child's productive repertoire' (ch.9, emphasis added).This explanation can account for the increased attention to L_F forms in knowledge group 1's results, as phonetic representations are bolstered by the infants' production of these word forms. Again we see a familiarity response here, but this time this has been 'reset' to the least complex form–that is, the form that best fits the infants' output. Finally, the lack of consistent response to W or T stimuli suggests that these features may be arbitrary in relation to the phonological structure of the word form. It could also be that phonological wildness is specific to particular target words or even individual infants; more results are needed for this group if these options are to be investigated.

5. Conclusion

These results provide no concrete evidence for the role of wildness in word learning; wild features in the input appear to be arbitrary when paired with words the infant recognises and, more importantly, words he can already produce. However, the results do suggest that the reduplicative features of many OWs may prompt perception and, later, production in language development. It is likely that the wild vs. tame paradigm is specific to individual infants' experiences of individual word forms, and thus perception of OWs cannot be measured across-the-board in such a way. These findings highlight the essential interplay between perception and production in early language development, demonstrating the breadth of an infant's early lexical categories which account for both degeneracy and variability in the input, before the onset of production brings about perceptual narrowing, providing feedback specific to the phonological categories of the ambient language.

References

BEHRENS, H. 2007. "The input–output relationship in first language acquisition". *Language and Cognitive Processes*, 21(1-3), pp. 2-24.

BLOOM, P. 2000. *How Children Learn the Meanings of Words*. Cambridge,MA:MIT Press.

FENSON, L., DALE, P. S., REZNICK, J. S., BATES, E., THAL, D. J., & PETHICK, S. J. 1994. "Variability in early communicative development". *Monographs of the Society for Research in Child Development*, 59(5), Serial Number 242.

GENESEE, F., NICOLADIS, E. & PARADIS, J. 1995. "Language differentiation in early bilingual development". *Journal of Child Language*, 22, pp. 611-631.

IMAI, M., KITA, S., NAGUMO, M. & OKADA, H. 2008. "Sound symbolism facilitates early verb learning". *Cognition***,** 109(1), pp. 54-65.

IVANOVA, G. 2006. "Sound-symbolic approach to Japanese mimetic words". *Toronto Working Papers in Linguistics***,** 26, pp. 103-114.

IWASAKI, N., VINSON, D. P., & VIGLIOCCO, G. 2007. "What do English Speakers Know about *gera-gera* and *yota-yota*? A Cross-linguistic Investigation of Mimetic Words for Laughing and Walking". *Japanese-Language Education around the Globe*, 17, pp. 53-78.

KANTARTZIS, K., IMAI, M. & KITA, S. 2011. "Japanese Sound-Symbolism Facilitates Word Learning in English-Speaking Children". *Cognitive Science,* 35**,** pp. 575-586.

KAUSCHKE, C. & KLANN-DELIUS, G. 2007. "Characteristics of maternal input in relation to vocabulary development in children learning German". In I. Gülzow & N. Gagarina (eds.) *Frequency Effects in Language Acquisition: Defining the Limits of Frequency as an Explanatory Concept.* Berlin: Walter de Gruyter GmbH.

KITA, S. 1997. "Two-dimensional semantic analysis of Japanese mimetics". *Linguistics***,** 35,

pp. 379-415.
MENN, L. & VIHMAN, M. 2011. "Features in child phonology: inherent, emergent, or artefacts of analysis?". In N. Clements & R. Ridouane (eds.), *Where Do Phonological Features Come From? The nature and sources of phonological primitives*. (Language Faculty and Beyond 6.) Amsterdam: John Benjamins.
RHODES, R. 1994. "Aural Images". In J. Ohala, L. Hinton & J. Nichols (eds.) *Sound Symbolism.* Cambridge, UK: Cambridge University Press.
SUNDBERG, U. 1998. *Mother tongue – Phonetic Aspects of Infant-Directed Speech.* Stockholm: PERILUS.
TSUJIMURA, N. 2005. "Mimetic verbs and innovative verbs in the acquisition of Japanese". *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 31(1), pp. 371-382.
VIHMAN, M. M. 2014. *Phonological Development: The first two years*. (2nd ed.). Oxford: Blackwell.
VIHMAN, M. M. 1993. "Variable Paths to Early Word Production". *Journal of Phonetics*, 21, pp. 61-82.
VIHMAN,, M. M. & DEPAOLIS, R. A. 2000, "The role of mimesis in infant language development: Evidence for phylogeny?". In C. Knight, M. Studdert-Kennedy & J. Hurford (eds.) *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge, UK: Cambridge University Press.

Acknowledgements

I would like to thank Lena Renner, Ellen Marklund, Elisabet Cortes for their invaluable help in carrying out the eye-tracking study, as well as Professor Francisco Lacerda for allowing me to work at the University of Stockholm's Babylab. This research could not have taken place without funding from the ESRC's OIV scheme.

Catherine E. Laing
Department of Language and Linguistic Science
University of York
Heslington, York
YO10 5DD
United Kingdom
Email: catherine.laing@york.ac.uk

AGREEMENT PATTERNS IN *IT*-CLEFTS: A MINIMALIST ACCOUNT

EWELINA MOKROSZ
John Paul II Catholic University of Lublin

Abstract

The paper proposes a structural analysis of English *it*-clefts which accounts for different agreement patterns with clefted personal pronouns corresponding to the subject gap in the cleft clause. The agreement patterns under investigation are patterns already reported in the literature (Akmajian 1970, Sornicola 1988) as well as those that were found in a questionnaire filled in by a group of native speakers in 2012. An attempt will be made to account for the agreement variations that appear in *it*-clefts within the theoretical framework of the Minimalist Program (e.g., Chomsky 2008) incorporating Kratzer's (2009) view on bound pronouns. Specifically, possessive pronouns will be argued to enter the derivation with valued φ-features. Reflexives, on the other hand, will require feature valuation. Case variations on the clefted pronoun will be accounted for with the three observations originally made by Quinn (2005).

1. Introduction

A significant part of the literature on *it*-clefts revolves around the accounts of their information structure and syntactic derivation. The former assign *it*-clefts a marked informative interpretation (e.g., Prince 1978). The latter have divided linguists between the so-called extraposition (e.g., Akmajian 1970) and expletive analysis (e.g., Chomsky 1977). Less attention, however, has been devoted to an explanation of the different agreement patterns *it*-clefts exhibit. Among *it*-clefts, the most interesting ones from the point of view of agreement are those with a clefted pronoun corresponding to the subject position in the cleft clause.[1] Example sentences showing the relevant variations are presented in (1), (2) and (3).

(1) a. It is *me* who *is* responsible.
b. It is *I* who *am* responsible.

Akmajian (1970: 153)

(2) It is *me* who has to protect *himself/myself*.

Akmajian (1970: 157)

(3) It is *me* that hit *his/*my* own father.

Akmajian (1970: 160)

The apparent subject function of a clefted pronoun in each cleft clause above in (1-3) suggests agreement with the cleft verb as observed between subjects and verbs in other types of clauses in English. (1a), however, shows that a lack of such agreement has no bearing on the

[1] In the paper the following terms will be used to refer the relevant parts of an *it*-cleft sentence: a clefted pronoun (a focused element in the immediate postcopular position of the main clause, i.e., *me* in (i)), a cleft clause (a subordinate clause with a gap corresponding to the clefted phrase, i.e., *who […] likes flowers* in (i)) and a cleft verb (the main verb in the cleft clause, i.e., *likes* in (i)).

(i) It is me who likes flowers.

Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York (**PARLAY 2013**) conference
 ISSN 1758-0315 Issue 1

grammaticality of the relevant sentence. Under Chomsky's (1981) binding Principle A, a reflexive pronoun as an anaphor has to be bound (c-commanded) by the antecedent located in the same minimal clause (CP). This, in turn, entails full agreement, i.e., in person and number, between the c-commanding phrase and the anaphor. Yet, the cleft clause in (2) seems to violate the well-known binding restriction. Another equally unexpected fact is the ungrammaticality of the first person possessive pronoun in (3). It could be that the father was hit by his own child. The following questions immediately arise: Are the violations just noted new pieces of evidence against well known linguistic generalizations? Is each pattern generated by a different syntactic structure or are some other explanations due?

The first part of the discussion below (section 2) is devoted to the presentation of agreement patterns observed in the literature from the 70s and 80s. The collected data will be supplemented with those provided by the questionnaire completed in 2012. Section 3 attempts to account for the reported agreement variations within the Minimalist framework.

2. Agreement patterns in it-*clefts*

This section juxtaposes the data on *it*-clefts from two works, namely, Akmajian (1970) and Sornicola (1988), with the data from the questionnaire.[2] The attention is given to the relations of the clefted pronoun to the cleft verb, a reflexive pronoun and a possessive pronoun. Akmajian (1970) examines the data from three dialects which he refers to as Dialect I, Dialect II and Dialect III. In Dialect I, the clefted pronoun always bears Accusative case and agrees with the cleft verb solely in number. The clefted pronoun in Dialect II bears Nominative case when it corresponds to the subject gap. The agreement with the cleft verb is the same as in Dialect I. Dialect III exhibits dependency between the case of the clefted pronoun and full/partial agreement between the clefted pronoun and the cleft verb. In particular, full agreement, i.e., in person and number, is judged to be grammatical only when the clefted pronoun bears Nominative case. Accusative case on the clefted pronoun is accompanied by partial agreement in number between the clefted pronoun and the cleft verb. The agreement variations just presented are exemplified below.[3]

(4) Dialect I

It is *me* who *is* responsible.

(5) Dialect II

a. It is *me* who(m) John *is* after.
b. It is *I* who *is* sick.

Akmajian (1970: 152)

[2] The questionnaire was distributed by the author of the paper to a group of five native speakers of English at the age range from the early 40's to the late 60's. Each speaker received a list of *it*-clefts, to which they were asked to assign one of the three judgements provided below.

(i) ungrammatical (ungr)
(ii) acceptable but non-standard (acc)
(iii) grammatical and standard (gr)

It has to be admitted that no context introducing *it*-clefts was included in the questionnaire.

[3] Sornicola (1988) reports the same three patterns as Akmajian (1970).

(6) Dialect III

a. It is *I* who *am/*is* responsible.
b. It is *me* who **am/is* responsible.

What the questionnaire shows and what Akmajian (1970) and Sornicola (1988) fail to observe is the fact that *the clefted pronoun can appear in Accusative case and agree in a person and number with the cleft verb* (see (7) below). Interestingly, some speakers found the agreement pattern in (8) acceptable.[4] This is the only pattern in which no number agreement is noted. The patterns not recorded by Akmajian (1970) and Sornicola (1988) constitute 31% of all agreement patterns found in the questionnaire.

(7) It is *me* who *like* flowers. 3-ungr, 1-acc, 1-gr

(8) It is *them* that *likes* flowers. 2-ungr, 2-acc, 1-gr

Additionally, Table 1 presents grammaticality judgements of individual speakers.

Examples with particular agreement patterns	**Central Wales (border with England)**	**Newcastle**	**London**	**Dublin**	**Norwich**
	I	II	III	IV	V
a. It is *me* who *likes* flowers.	gr	acc/gr	acc/gr	acc	gr
b. It is *I* who *like* dogs.	ungr/acc	ungr/gr	ungr/acc	ungr/acc/gr	ungr
c. It is *I* who *likes* dogs.	acc	gr	ungr/acc	acc/gr	ungr
d. It is *me* who *like* flowers.	ungr/acc	ungr	ungr	ungr	ungr

Table 1: Agreement patterns by speakers

It turns out that individual respondents show great disparities accepting almost all variations. Speaker V is the only one who consistently rejected constructions with Nominative case on the clefted pronoun. Almost all speakers judged pattern (d) in Table 1 as ungrammatical.

While describing the agreement on reflexives located in cleft clauses of *it*-clefts, Akmajian distinguishes three patterns. Under the most popular one, the reflexive exhibits an invariant 3rd person feature regardless of the person feature of the clefted pronoun. The agreement between the reflexive and the clefted pronoun is manifested only in number as in (9). In the second pattern, there is full agreement (see 10 below). Sornicola (1988) remarks that this pattern can be encountered in the most widespread variety of English.[5] A variation within one dialect as in (11) is also possible as noticed by Akmajian (1970) but not Sornicola (1988).

[4] Peter Sells (p.c.) has remarked that *them that likes flowers* could be interpreted as a relative clause in some English dialects. In Mokrosz (in prep) it is shown that the reported agreement variations in *it*-clefts can be observed also in relative clauses. Thus, the remark concerning relative clauses should not affect the analysis presented in this paper.

[5] Sornicola (1988) does not explain what she means by the most widespread variety of English.

(9) It is *me* that cut *himself* so badly.

(10) It's *me* that cut *myself*.

(11) It's *me* who has to protect *myself*/*himself*.

Akmajian (1970: 155-156)

Like Akmajian (1970), the questionnaire results show a slight preference for partial agreement, i.e., only in number, between the clefted pronoun and the reflexive in the cleft clause, though full agreement is still acceptable. Sornicola (1988: 352), however, maintains that only full agreement between the clefted pronoun and the cleft verb guarantees full agreement between the clefted pronoun and the reflexive. The grammatical examples presented by Akmajian (1970) and the corresponding ones in the questionnaire contradict Sornicola's generalisation. Specifically, they show that partial agreement between the clefted pronoun and the cleft verb may co-occur with full agreement between the clefted pronoun and the reflexive. The questionnaire also aimed to check whether the change of a case from Accusative to Nominative on the clefted pronoun would have any bearing on the grammaticality reports originally provided by Akmajian and Sornicola. The following results were collected.

(12) It is *I* who *like myself*. 2-ungr, 3-gr

(13) It is *I* who *likes myself*. 3 ungr, 2-gr

(14) *It is *I* who *like himself*. 5-ungr

(12) and (13) show that Sornicola's assumption is not valid for Nominative clefted pronouns either. The ungrammaticality of (14) remains to be explained.

Surprisingly, the person feature of the reflexive does not always have to overlap with the value of the person feature of the clefted pronoun. The relevant sentence is presented below.

(15) It is *me* who *likes yourself*. 1-ungr, 2-acc

Akmajian (1970) also draws attention to constructions with an obligatory identity between the subject and some possessive pronoun. These are as follows: constructions with certain idioms, reflexive possessives and certain verbs of perception. The relevant examples are presented in (16a, 17, 18). According to Akmajian (1970), possessive pronouns show 3rd person feature regardless of the person feature of the clefted pronoun. This observation concerns only Dialect I and no information is given on Dialect II and III.

(16) a. Was it *you* that held *his*/**your* breath for five minutes?
b. *Or was it *John* that held *your* breath for five minutes.

(17) It's *me* that hit *his*/**my* own father.

(18) It was *me* who felt a spider crawl up *his*/**my* leg.

Akmajian (1970: 159-160)

According to Akmajian, *your* in (16a) would be judged anomalous in Dialect I as it entails a contrast presented in (16b). The explanation concerning contrast may seem unclear as clefts are inherently burdened with a contrastive focus. Importantly, Akmajian (1970) draws his conclusions only on the basis of constructions with certain idioms, reflexive possessives and some verbs of perception. The negation introduced by Akmajian in *it*-clefts with such phrases shows why only the 3rd person feature on the possessive is grammatical.

(19) a. It was *me* who felt a spider crawl up *his* leg.
b. It was *me* who felt a spider crawl up **my* leg.

(20) a. *It wasn't *me* who felt a spider crawl up *my* leg.
b. It wasn't *me* who felt a spider crawl up *his* leg.

Akmajian (1970: 160)

The ungrammaticality of (20a) supports Akmajian's (1970) assumption on the 3rd person feature. A different behaviour of possessives in cleft clauses can be noted in constructions other than those studied by Akmajian (1970).

(21) a. It was *me* who lost *his* brush.
b. It was *me* who lost *my* brush.

(22) a. It wasn't *me* who lost *my* brush.
b. It wasn't *me* who lost *his* brush.

In (21) the possessive pronouns point to the possessors of the brush. The test with negation presented in (22a) and (22b) shows that both of them, i.e., *his* and *my*, are grammatical. This observation does not overlap with the conclusion made by Akmajian (1970) about examples (16-18). Specifically, *his* in (21a) points only to an external referent while *his* in (16a, 17-18) refers to the same person as the personal pronoun preceding it. *Me* in (21b) as opposed to (19b) can be contrasted with the 3rd person pronoun *him* because it is possible that a person other than the owner of the brush could lose the brush. Thus, it is grammatical to have a possessive pronoun agreeing with the clefted pronoun in both person and number or only in number.

3. A Minimalist account of agreement patterns in it-*clefts*

The following section starts with an outline of Kratzer's (2009) analysis. The explanation proposed in section 3.2 for the lack of uniformity in person feature to a large extent takes advantage of Kratzer's main assumptions. Subsequently, some doubts are voiced about Kratzer's proposals and a final analysis of *it*-clefts with clefted subject pronouns is outlined. The section ends with a possible answer to case variations observable on a clefted pronoun.

3.1. Kratzer (2009)

According to Kratzer (2009), bound pronouns such as reflexives, possessive pronouns and relative pronouns originate with a defective set of features complemented by a binder via a functional head such as v or C.[6] Thus, in a sentence such as (23) the reflexive as a bound pronoun receives the value for a person and number feature from the nominal phrase *John* via v.

(23) John likes *himself/*myself*.

Bound reading depends on the φ-feature compatibility between a functional head and a bound pronoun guaranteed by Feature Transmission under Binding and Predication.

(24) *Feature Transmission under Binding*

The φ-feature set of a bound DP unifies with the φ-feature set of the verbal functional head that hosts its binder.

Kratzer (2009: 195)

[6] According to Kratzer (2009: 187-188), bound pronouns are 'interpreted by assignment functions' while referential pronouns 'refer to salient individuals'.

(25) *Predication (Specifier-Head Agreement under Binding)*

When a DP occupies the specifier position of a head that carries a λ-operator, their φ-feature sets unify.

Kratzer (2009: 196)

The abovementioned feature compatibility is not as apparent in the case of possessive pronouns as it is in the case of reflexives. According to Kratzer (2009), the two possessive pronouns in the sentences below allow bound reading even though their person features differ from the person features of functional heads hosting their binders.

(26) a. We are *the only people* who take care of *our* children.

Kratzer (2009: 201)

b. I am *the only one* who is brushing *my* teeth.

Kratzer (2009: 208)

Under Kratzer's analysis, in the case of clauses headed by a relative pronoun as in (26), it is only the possessive pronoun that originates underspecified (unvalued person, unvalued number) while the verb is born with a valued person and number. Subsequent unification of features between the possessive pronoun and the verb guarantees the bound reading of the possessive pronoun. The possessive pronoun in (26a) bears the set [1st, pl] and in (26b) the set [1st, sg]. Relative pronouns as minimal pronouns enter the derivation with unspecified number and person features. As bound pronouns they have two sources of φ-features they eventually receive: (a) embedded little *v* which passes the features under *Predication* and (b) the head of a relative clause coreferential with a relative pronoun. As a result, a feature clash arises on the relative pronouns in (26a) and (26b) above. The particular feature sets ultimately making up the feature set on the relative are presented in (27) below.

(27) a. who = *the only people* [3rd, pl] + *v* [1st, pl]
b. who = *the only one* [3rd, sg] + *v* [1st, sg]

The bound reading of a given item is possible only under feature compatibility between this item and a functional head carrying its binder. Yet, the examples in (27) show the opposite. Both verbs, namely *take care* and *is brushing* carry a 3rd person feature whereas it is the 1st person feature which is expected.[7] In order to explain the deviation from the rule of the bound reading under feature compatibility, Kratzer (2009) refers to a language specific markedness of features obeyed by a morphophonological spell-out. Specifically, in English a person feature dominates a gender feature on nominal expressions. On verbs, on the other hand, a person feature is marked, which is why a gender feature is chosen more often. Importantly, Kratzer (2009) follows a view widely held in the linguistic literature on the 3rd person as a non-person (e.g., Benveniste 1966). In Kratzer's (2009) analysis, however, not only is the 3rd person considered a non-person but it also acquires a new status of a gender feature. As evidence for the feature markedness in English just outlined, Kratzer (2009) provides two examples, one in (28) and the other in (29). The ungrammaticality of (29) with the 3rd person on the possessive pronoun indicates that it is a 1st person and not a 3rd person that is unmarked on pronouns.

(28) The teacher and I have done *our* best to fix the problem.

(29) *The teacher and I have done *their* best to fix the problem.

Kratzer (2009: 210)

[7] Since Kratzer (2009) provides the two examples in (26) as cases in need of explanation, it will be assumed that the verb carries the 3rd person feature rather than the 1st person.

Thus, it should come as no surprise that the possessive pronouns in (26) bear the 1st person feature while it is the 3rd person under the markedness of features in English which is spelled out on the verb.

One of the most conspicuous problems in Kratzer's analysis concerns the features of the verb. Specifically, its valued/unvalued status depends on the type of a binder it hosts. In the presence of a nominal subject expression with valued features, it is born unvalued. The opposite is observed in the case of a relative pronoun in the binder position. Since both a nominal subject expression and a relative pronoun enter the derivation after the verb, the valued/unvalued status of the features on the verb faces a look-ahead problem.

The following two examples also require some explanation.

(30) It is *me* who likes *myself/himself*.

To explain the grammaticality of the reflexive *himself* in (30), it has to be assumed after Kratzer (2009) that the verb enters the derivation with a third person feature. The grammaticality of *myself* in the same sentence would have to be accounted for by a reference to the first person on the clefted pronoun. Let us consider a variety of English in which the person feature on the reflexive is always compatible with the person feature of the clefted pronoun while the verb shows the 3rd person.

(31) It is *me* who likes *myself*.

In this case, under Kratzer's analysis, the verb would have to be born with the 3rd person feature. The choice of the 3rd person feature on the verb would be incidental and unmotivated.

3.2. Derivation of it-*clefts*

After Kratzer (2009) it can be argued that reflexives, being born with an unvalued person and number, require feature valuation. There are some points, however, which mainly from the Minimalist standpoint require some modifications.

Firstly, it is likely that the relative pronoun originates with an unvalued number feature but with a valued person feature, namely the 3rd person feature. This conclusion is drawn on the basis of an analogy that can be made between relative pronouns and other *wh*-pronouns such as interrogative pronouns, which are assumed to carry the 3rd person feature. The 3rd person feature on the verb in the example below shows that the item it agrees with can carry the 3rd person feature.

(32) Who wants some ice-cream?

Secondly, in the Minimalist Program (e.g., Chomsky 2008) the verb is argued to enter the derivation with unvalued person and number features. No optionality with regard to the presence of features is permitted.

Finally, the nature of possessive pronouns is worth considering. According to Kratzer (2009), possessive pronouns can be analysed as either referential or bound pronouns. For example, *my* as a referential pronoun in example (33) refers to a prominent individual in a given communication. The reading of *my* in (33) as a bound pronoun indicates that nobody else around there can take care of his/her own children.

(33) I'm the only one around here who can take care of *my* children.

Kratzer (2009: 188)

It also has to be checked whether the same interpretation, i.e., bound and/or referential, can be assigned to a possessive pronoun in the cleft clause. The bound and referential readings of *my* in (34) corresponding to (33) are presented in (35a) and (35b), respectively.

(34) It is only *me* around here who can take care of *my* children.

(35) a. (?)The bound reading of *my* indicates that nobody else around here can take care of his/her own children; and
b. Under the referential reading, *my* refers to a prominent individual in a given communication; thus, the person uttering (34) is the only who can take care of their children and not anybody else.

It appears that the referential reading sounds more acceptable than the bound reading. Thus, in (34), the phrase *my children* is better interpreted as a prominent individual in a context. The possible argument for the unique referential interpretation of *my* in (34) could be the semantic complexity represented by *it*-clefts. *It*-clefts are argued to carry a presupposition of existence and exhaustivity (e.g., Percus 1997). Thus, in the *it*-cleft in (34) it is presupposed that there is someone who can take care of *my children*, which already points to *my children* being a prominent individual. The exhaustivity that *it*-clefts carry indicates that *me* is the only person who can do that. Thus, possessive pronouns in cleft clauses will be analysed as referential items entering the derivation with already valued person and number features.

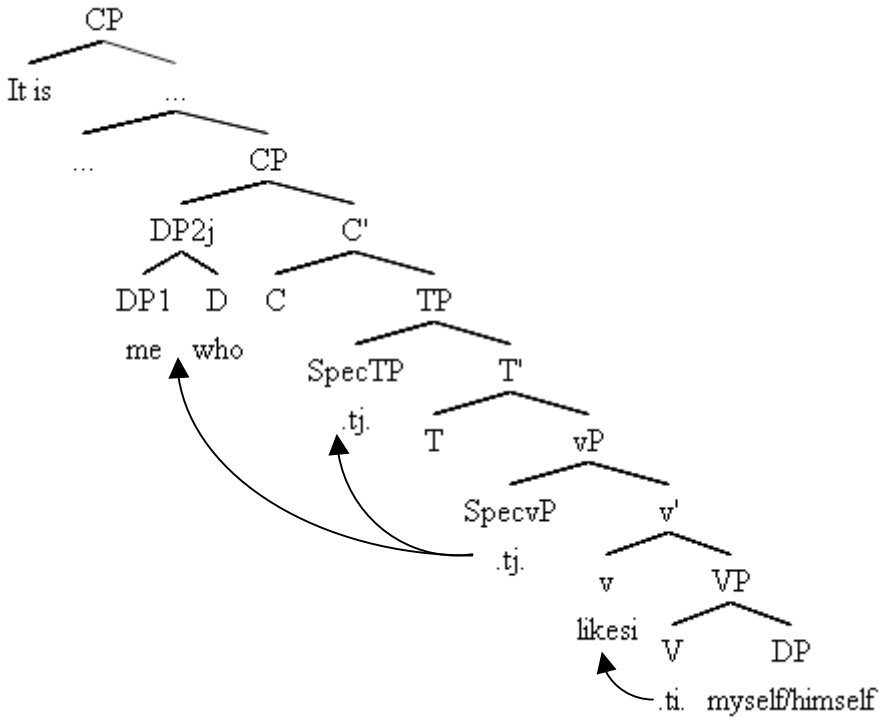
3.2.1. The origin of the clefted pronoun

The aim of this section is a structural account of *it*-clefts with a clefted subject pronoun and a co-referential reflexive in the cleft clause.

(37) is a structural analysis of an example sentence in (36).

(36) It is me who likes myself/himself.

(37)



The verb carries [uPers, uNum, Acc] while a reflexive is born unspecified for person and number features; thus, no valuation is possible. A relative pronoun and a personal pronoun originate in the same DP, i.e., DP2. As a result of their Merge, the *wh*-pronoun has its unvalued feature valued eventually carrying the following set [3rd/1st Pers, sg Num, uCase]. The same set is passed to the reflexive via Kratzer's (2009) operation of Transmission. Which person value will dominate on the reflexive as well as on the verb after Agree is resolved later in the derivation. The relative pronoun and the personal pronoun have an unvalued case which makes them still active in the derivation. C with the edge feature and T with an unvalued person and number *probe in parallel* triggering movement to Spec, CP and Spec, TP, respectively (Chomsky 2008: 147).

By following Kratzer's (2009) idea on bound pronouns, it is possible to explain the varied agreement patterns referring to the markedness of 1st and 2nd person feature on the verb and the markedness of the 3rd person feature on the reflexive. The ungrammaticality of the example repeated below can be connected with a semantic clash between the marked 1st person on the verb and the marked 3rd person on the pronoun.

(14) *It is *I* who *like himself*.

Another example that appears to be problematic for our analysis is reproduced below.

(8) It is *them* that *likes* flowers.

It is possible to argue that *them* in example (8) functions as a collective noun phrase like the noun *team*. Thus, depending on the context *them* may receive a singular or plural interpretation.[8]

Some explanation is due with regard to the nature of the complex DP2 proposed in (37). In our analysis the clefted pronoun originates in the specifier of DP2 analogously to possessive DPs in which, according to Abney (1987), a nominal phrase is born in the specifier while the possessive marker in the head D. With this assumption it is possible to explain why it is grammatical to cleft proper names and pronouns but ungrammatical to use them in relative clauses. A phrase such as a complex DP2 allows proper names and pronouns, argued to occupy D heads, to originate in the same phrase as the relative *who*. Finally, the analysis in (37) clearly shows why in the case of *it*-clefts with two subordinate clauses like the one in (38), it is always the last one which is interpreted as a cleft clause while the first as a relative clause.

(38) It is Mark who likes violence who hit John.

3.2.2. Case mismatch

As already noted, the choice of the case is not always dependent on the function the clefted pronoun corresponds to in the cleft clause. A similar study, yet, covering more structures and

[8] The reviewer suggested considering the contrast between the two examples below.

(i) you who are/*is so beautiful
(ii) I who am/is? so grateful to be here…
The research of the Corpus of Global Web-Based English (Davies 2013) showed that the 3rd person feature on the copula in (i) can be grammatical.

(iii) it is you who is being patronising (and insulting) […].
http://www.bbc.co.uk/blogs/paulhudson/2012/01/will-we-see-the-northern-light.shtml

interviewing more respondents was conducted by Quinn (2005). The results of her research to a great extent overlap with ours. Firstly, she notices that focus position of an *it*-cleft is basically an *Accusative* case position in subject and non-subject clefts. Secondly, Nominative case on the pronoun is less popular than Accusative but occurs in both subject and non-subject clefts. Out of all Nominative personal pronouns *he* is more popular than *she*, *they*, *I*, *we*. Quinn (2005) lists three factors that have an influence on the choice of a case on the clefted pronoun. One of them concerns competition between three types of cases, namely Argument, Positional and Default case.[9] Quinn reports that in Present Day English Positional case takes precedence over Argument and Default case. Since the clefted pronoun corresponds to the subject position in the cleft clause, i.e., Spec, TP, it will be spelled out with Nominative case as this case is associated with the functional head T (Quinn 2005: 58). The second factor refers to a c-command relation; a pronoun in a c-commanding position will appear in its gracile form, namely *me, he, she, they, we*, as opposed to a c-commanded pronoun which features a robust form, i.e., *I, him, her, us, them* (Quinn 2006: 151-153). The third factor concerns a tendency observed by Quinn in the results from the questionnaires she distributed. Since many speakers interviewed by Quinn consistently chose the objective form of the pronoun in given structures including *it*-clefts, Quinn (2005: 171) concludes that there must be a general tendency towards invariant forms of pronouns, i.e., *me, him, her, us, them*, in all contexts. The three factors listed by Quinn appear to account for the case variations on the clefted pronoun in the most satisfactory way.

4. Conclusion

The unprecedented degree of reported variations in the agreement patterns in *it*-clefts constitutes a challenge for the assumptions made within the Minimalist Program. What has been proposed in fact derives the answers not only from the minimalist analysis but also from semantics. The former guided the derivational steps proposed in (37) while the latter provided answers in the spirit of Kratzer (2009) on the person variations on reflexive pronouns. Quinn's explanations of case on the clefted pronoun seemed to be well-founded as they relate not only to purely syntactic phenomena but also to some general observations on the form of pronouns in Present Day English. The acceptability of the following sentence still awaits its explanation.

(15) It is me who likes yourself.

[9] Quinn (2005: 57-59) provides the following definitions of each type of case mentioned in the text.

(i) Argument Case (Arg-Case)
The overt case form of any structural argument of a predicate must comply with the structural linking between cases and arguments in the θ-structure.

(ii) Positional Case (Pos-Case)
The overt case form of an argument noun phrase appearing as the specifier of an agreement-related functional head at Spell-Out must match the case/agreement features of this functional head, iff the position of the noun phrase at Spell-Out differs from its θ-position.

(iii) (Positional) Default Case (Def-Case)
The overt case form of any noun phrase not influenced by Pos-Case must match the default case of a language. In Modern English, the default case is the objective case.

Acknowledgements

I would like to offer my special thanks to the reviewers of this paper for very useful comments.

References

ABNEY, STEVEN, PAUL. 1987. "*The English Noun Phrase in Its Sentential Aspects*". Doctoral dissertation, MIT, Cambridge, Mass.

AKMAJIAN, ADRIAN. 1970. "On deriving cleft sentences from pseudo-cleft sentences". *Linguistic Inquiry 1*. pp. 149-168.

BENVENISTE, ÉMILE. 1966. "*Structure des relations de personne dans le verbe*. *Problèmes de linguistique générale"* [Structure of person relations on the verb. Problems for general linguistics]. Paris: Gallimard. pp. 225-236.

CHOMSKY, NOAM. 1977. "On Wh-Movement". In: Culicover, P., T. Wasow & Adrian Akmajian (eds) *Formal Syntax*. New York: Academic Press. pp. 71-132.

CHOMSKY, NOAM. 1981. "*Lectures on Government and Binding*". Dordrecht: Foris.

CHOMSKY, NOAM. 2008. "On Phases". In: Freidin, R., C. P. Otero & M. L. Zubizarreta (eds) *Foundational Issues in Linguistic Theory*. Cambridge, Mass.: MIT Press. pp. 133-166.

DAVIES, MARK. 2013. "*Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries"*. Available online at http://corpus2.byu.edu/glowbe/.

KRATZER, ANGELIKA. 2009. "Making a Pronoun. Fake Indexicals as Windows into the Properties of Pronouns". *Linguistic Inquiry, 40. 2*. pp. 187-237.

MOKROSZ, EWELINA. (in prep). "*How relative are cleft clauses in English* it-*clefts?"*

QUINN, HEIDI. 2005. *"The Distribution of Pronoun Case Forms in English"*. Amsterdam: John Benjamins.

PERCUS, ORIN. 1997. "Prying open the cleft". In: Kusumoto K. (ed) *Proceedings of NELS* 27. Amherst, Mass.: GLSA. pp. 337-351.

PRINCE, ELLEN, F. 1978. "A comparison of wh-clefts and it-clefts in discourse". *Language 54*. pp. 883-906.

SORNICOLA, ROSANNA. 1988. "*It*-clefts and *Wh*-clefts: two awkward sentence types". *Linguistics 24*. pp. 343-379.

Ewelina Mokrosz
Department of Theoretical Linguistics
John Paul II Catholic University of Lublin
Lublin, Poland
email: ewelinamokrosz26@tlen.pl

CLICKS IN CHILEAN SPANISH CONVERSATION

VERÓNICA GONZÁLEZ TEMER
University of York

Abstract

Chilean Spanish speakers were audio and video recorded engaging in conversation to find out whether clicks in Chilean Spanish occur in sequences similar to the ones reported for English. Some other research questions are whether there are any tokens that function alongside clicks in which case it is relevant to see what their articulatory properties and interactional functions are. Finally, if there are gestures accompanying clicks, it is important to devise their functions. A methodological approach that combines the sequential techniques of Conversation Analysis (CA), and the phonetic techniques of impressionistic observation and instrumental analysis is employed with naturally-occurring conversation. The results show that clicks do have a regular distribution and are part of bigger meaning-bearing prosodic constructions in Chilean Spanish, which entails that they are indeed linguistic. Moreover, the similarities with English in the way a click helps to show "gearing up to speak", to index new sequences that are disjunctive with the prior, to signal trouble in finding words; and the differences such as the particular use of clicks used to display affect found for Chilean Spanish support this argument.

1. Introduction

Clicks are non-pulmonic, velaric, ingressive sounds that are part of the consonant systems of some southern African languages such as Xhosa and Zulu and Khoisan languages of the Kalahari region, but are mostly unknown in other places of the world. (Clark et al, 2007:120).

They occur more extensively in languages when considered as communicative functions such as the expression of 'yes' and/or 'no' (Gil, 2011), approval (Ladefoged, 2005:170), disapproval and regret (Clark et al, 2007:18), as well as exasperation (Laver, 1994). They are subsumed by Ward (2006:27) as a sign of personal dissatisfaction, and are also associated with communication with babies or with animals (Gimson, 1970:31; Clark et al, 2007:18; Güldemann & Stoneking, 2008:95). Eklund (2008:238) supports the statement that ingressive clicks are used for paralinguistic purposes in French by quoting Havet (1875) who points out that "ingressive t is used to express doubt; and ingressive palatal t expresses surprise, but can also be used to call horses"

Gil (2011) lists a variety of languages around the world where phonemic and non phonemic clicks are used. He provides a typological distribution map (see figure 1).

Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York (**PARLAY 2013**) conference
 ISSN 1758-0315 Issue 1

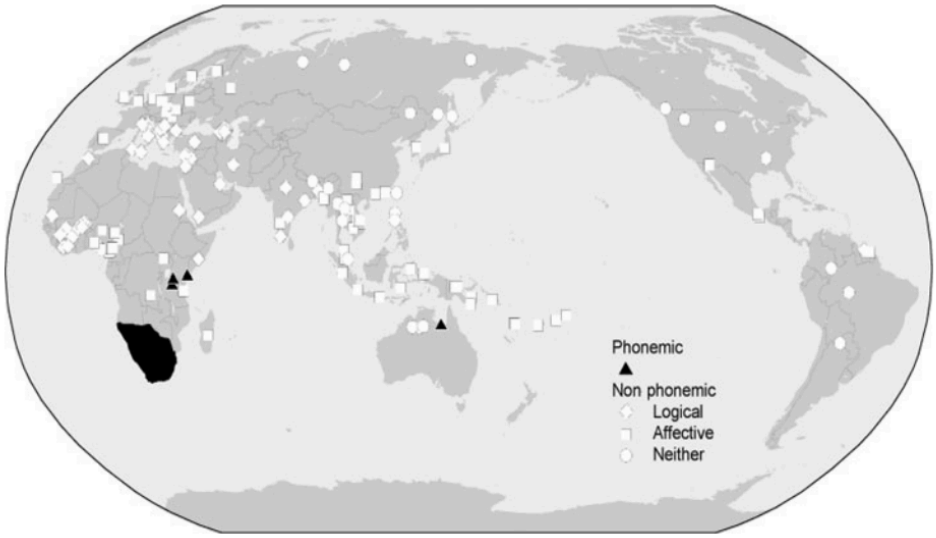


Figure 1. The global distribution of clicks (Gil 2011, Güldemann & Stoneking, 2008).

Güldemann & Stoneking (2008:95) concludes from this map that clicks should not be considered unusual speech sounds concerning their production and use, as they are widespread both "geographically and genealogically" among human languages.

1.1 Phonetic overview

Clicks have traditionally been characterised as ingressive, velarically intiatied suction stops (Catford, 2001:53). Nasal airflow, voicing and different phonation types can be coarticulatory accompaniments of clicks (Ladefoged, 2005:170). These accompaniments are common and phonologically meaningful for those languages that have clicks as parts of the repertoire of speech sounds.

Thus, in English they are not very common although there is occasional aspiration according to Ogden (2013:5) who also claims clicks in English are sequentially followed by "one of a large number of tokens such as *oh, ah, aw, oo*," which might be articulated in different ways with regard to articulatory properties such as amplitude, duration, voice and vowel quality, and glottal stop initiation. These tokens and their particular articulatory properties are essential in the display of affect and these are the ones that require a close acoustic analysis.

Ogden (2013:5) simplifies the classification of clicks in English to two, a central and a lateral one because as Ladefoged & Maddieson (1996:256-257) claim, the articulation of clicks is complex and the rate of movement of the articulatory release, tongue shape, apical and laminal distinction as well as the contact surface of these articulators, are aspects of the production of clicks that are not paid attention to in their description. These features are likely to be captured when using other methods such as ultrasound. This is what presupposes the main complexity, but it also possible that the features previously described do not have an impact on the different functions clicks can have. Therefore, a simple classification is what will be done for this study as well.

1.2 Functions of clicks

Recent studies on English (Wright, 2005, 2007, 2011a, 2011b; Ogden, 2013) demonstrate that clicks have an orderly, sequential distribution that can convey different functions according to their embedded contexts of production. Some of the findings are listed below.

1.2.1. Sequence markers

As clicks are relatively loud transient sounds, Ogden (2013:11) suggests clicks mark incipient speakership, acting as an audible signal in conversation ensuring others acknowledge this. The physical principle behind these clicks is that when people are about to speak, they open their vocal tract and this can produce a velarically initiated ingressive sound, in other words a click (Ogden, 2013:23). When there is contact and separation between the articulators generally with inhalation initiated pulmonically, then a percussive is produced. Clicks and percussives are often the culmination of other physical activities; lip closure, swallow, and release.

Following the CA terminology proposed by Sacks et al (1974), these clicks are generally found turn-initial, or Turn Constructional Unit (TCU) initial. TCUs are considered to be the basic units of talk, these are identifiable because there is a Transition Relevant Place TRP) at the end of them signalling the interlocutor in the conversation can start producing talk or more talk could be produced by the same speaker, this means a turn can be composed on one TCU or multiple TCUs. Similarly, Wright (2005, 2007, 2011a, 2011b) shows how clicks manage aspects of sequence organisation. They occur at boundaries between sequences, for instance at the end of a phone call. This type of click is what Wright (2007) calls New Sequence Indexing (NSI) clicks as they shift the direction of the talk from one topic to another. She proposes a particular sequential organisation and articulatory properties that differentiate one sequence as disjunctive with the prior which accounts for the function of clicks in such sequences.

1.2.2. Clicks in word searches

Clicks are often produced during word searches, and are used to signal trouble in finding words in conjunction with particles such as "*uh, uhm, mmm*, uttered on a mid-level pitch perhaps tagged on to (full forms of) words; in-breaths and swallowing; gaps in talk at points where longer syntactic units are projected." (Ogden, 2013:23).

The use of clicks in word searches seems to be more linguistically structured and particular than those that mark incipient speakership. Ogden (2013:23) suggests they may be an evolved form of more organic sounds such as clicks and percussives that mark incipient speakership."

1.2.3. Stance-taking clicks

Ogden (2013:3) states that the lay interpretation of clicks in English-speaking cultures associates them with a negative stance and is often described as ‘tutting’.This makes reference to both the sound and the negative stance conveyed .

Clicks that display some kind of stance or affect are designed to be heard, with a loud and deliberate action. They may be accompanied by body movements such as head turns and facial expressions. Ogden (2013: 23) writes

> "they are embedded in sequences of talk which can project the stance displayed in the turn with the click, and/or elaborated on after the click is produced, e.g. through overt linguistic material."
>
> (Ogden, 2013: 23)

Because clicks are often accompanied by response tokens such as *oh, ah, aw, oo* in the display of stance or by gestures, it is very difficult to demonstrate they can display stance by themselves. Ogden (2013:24) asserts that it is essential to look at longer sequences and non-verbal behaviour so as to better understand what is involved in stance-taking from a linguistic point of view.

All things considered, one of the motivations for this study lies in the fact that there is very little information about clicks in Spanish except for accounts of their onomatopoeic use in American Spanish (Zilio, 1986:139,143). There is a need for a crosslinguistic study of clicks to determine whether they work differently in different languages in both form and function, or if there are sequential differences. This would provide solid arguments to say they are linguistic or at least part of the speakers' cultural knowledge. This research addresses the following questions:

1. Do clicks in Chilean Spanish occur in sequences similar to the ones reported for English?
2. Are there any tokens that function alongside clicks? If so, what are they and what are their articulatory properties and interactional functions?
3. Are there gestures accompanying clicks? If so, what are they and what are their functions?

The rest of the paper is divided as follow: first the methodology for the research is explained in terms of data collection and different types of analyses in section 2. Then the results are presented, categorised and analysed in section 3. Finally, some concluding remarks are derived in section 4.

2. Methodology

For this investigation, a methodological approach that combines the sequential techniques of CA, the phonetic techniques of impressionistic observation and instrumental analysis were employed. To make sense of phonetics in conversation, a proper model of the way conversation works is needed, i.e. CA, as it is in conversation and through sequential analysis that we find a regular distribution for clicks whose function is not arbitrary, but learnt. This method is therefore important for making new discoveries about the nature of language.

2.1. Data

The data for the present study was collected from five pairs of Chilean participants, five female and five male speakers who ranged from 18 to 49 years of age. Participants knew each other from before, they were acquaintances, friends or married couples and the researcher knew them all. They were video recorded for thirty minutes per pair in a studio at the University of York. The couples sat next to each other so the camera could capture facial gestures and hand movements clearly. The cameras used were a JVC HY-GM100E full HD and Sanyo Xacti HD both recording onto SDHC cards at 1440 x 720. The audio was recorded at 44.1KHZ 16bit with a Zoom H4n audio recorder using a Beyerdynamic Opus 55 headset microphone, a DPA 4066 headset microphone and a Sennheiser EW100 radio microphone.

The audio was recorded on two channels to make the data more manageable for overlapping talk.

As previously mentioned, clicks occur at sequence boundaries; therefore there was the need to generate instances where participants would change the topic. Therefore, one of the participants in every couple was asked to go through a list of topics to cover. Generating some kind of emotional response was also relevant for this study, hence a mixture of positive, evocative and sensitive topics were included in the list: an experience at the hairdressers, life in York, the 2010 earthquake in Chile, etc.

2.2. Interactional and Phonetic Analysis

The data for this research is drawn from naturally-occurring conversation because as Wright (2005:2) explains "meaning is context-bound, emergent and contextually determined". Therefore, the method of analysis to be used in the study of clicks in this study is based on the orientations that participants display in naturally-occurring talk and not on an intuitive interpretation of what a particular utterance is thought to mean without considering its context. This methodology is what allows for conversation analysis to have the rigour of scientific analysis and to be the means to understand social communication.

The phonetic analysis of the data is carried out together with the sequential analysis. Impressionistic and acoustic analyses are employed to account for the parameters (pitch range, pitch movement, loudness, duration and articulatory properties) that function alongside the sequences studied. Parametric listening techniques proposed such as the ones described by Kelly & Local (1989) were used in this study, they propose such techniques to understand noises produced by the vocal tract as a range of co-working components that vary and have movement in order to avoid falling into categories of description unsuitable for naturally-occurring speech. Where possible the observations made have been corroborated through the use of PRAAT (Boersma & Weenink, 2013) by means of acoustic imaging to measure frequency, pitch, intensity and timing among other uses.

3. Results

This study comprises 2.5 hours of data from which 38 clicks could be identified. This means on average people click once every 4 minutes. In terms of classification by place of articulation, waveform and spectrogram observation helped to distinguish clicks from percussives when in doubt as from the waveform it is possible to see that clicks show two transients while percussives show only one in general. Video data helped in the recognition of lateral clicks as lips are twitched (generally to one side) as well as bilabial clicks as both lips are pressed together without protrusion. This is when impressionistic records play an important role. Figure 2 shows the classification of clicks by place of articulation.

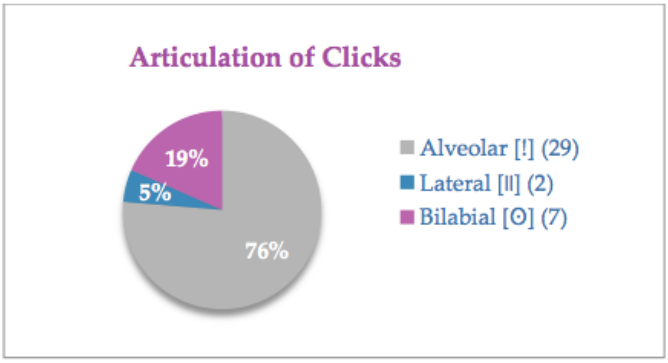


Figure 2. Number and percentages of clicks by place of articulation

Considering the turn as the basic unit of organisation in conversation (Sacks et al, 1974), the subsequent unit is the TCU. Figure 3 shows that if clicks are grouped by position in the turn, the majority are TCU-initial (28, 74%), 5 are TCU-medial (13%). These results are consistent with the findings for English (Ogden, 2013:9). Lastly, 5 clicks are TCU-final (13%) and most of these are of a special kind which will be discussed in 3.3.2.

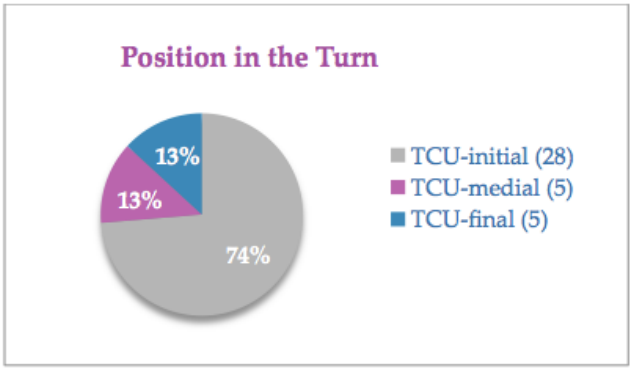


Figure 3. Number and percentages of clicks regarding position in the turn

Considering the different functions of clicks as proposed for the findings in English presented in the previous section, figure 4 shows that of the 38 clicks, 11 mark incipient speakership, 10 are in word searches and 17 display stance or affect. These different functions and contexts of clicks will be considered in detail by looking at different examples in the rest of this section.

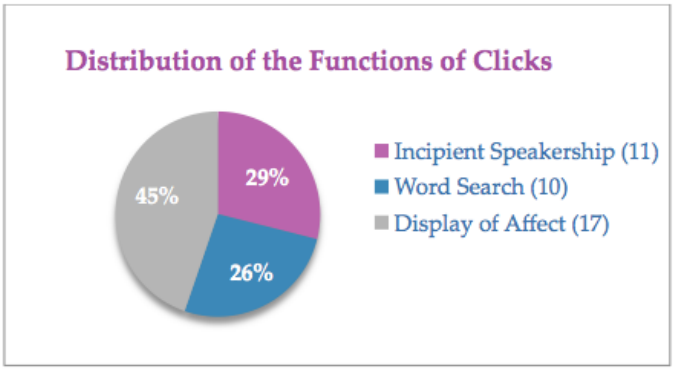


Figure 4. Number and percentage of clicks by function

3.1. Incipient speakership

In this study, there were cases where clicks were both turn-initial and TCU-initial. They are clearly signalling that the speaker is about to speak or wants to keep on talking. These clicks can also be as a result of other physical activity such as lip closure or swallowing. This is what was considered as plain incipient speakership. In line 5 of Example 1 Lía has started a new TCU that is cut off apparently because of the need to swallow which is what she does, after this she produces a click and restarts the TCU (line 7) using the same word she had used in line 5. (For transcription conventions see appendix).

Example 1
P1_Disney

```
01 Lía:  [to hasta los] DIÁlogos son medios (0.3) cantaitos;=
         everything even the  dialogues   are  kind of       sung
02       =y son como el (.) °hh (0.3) las co[neXIOnes]
          and they are like the           the  connections
03 Rosa:                                    [YA:     ].
                                            right
04 Lía:  entre una canción y la otra,=
         between one  song      and the other
05 Lía:  =pero,
          but
06 Rosa: YA:;
         right
07 Lía:→((swallowing))! <<f>pero no HAY otra conversación>.
                          but   there is no other  conversation
```

In other cases, clicks not only mark incipient speakership, but also introduce a new sequence that is disjunctive with the prior in terms of the topic of the conversation. Wright (2007) calls them **new sequence indexing** (NSI) clicks. In the following example there are several characteristics both in the sequential organization and articulatory properties that mark out the following sequence as being disjunctive with the prior; in plain terms, there is a change of topic.

Although there are few cases of these in the current data, it is interesting to discuss these in more detail as many of the features found for Chilean Spanish and discussed below are in accordance with Wright's (2007, 2011b) findings for this type of sequence in English. First, in terms of sequential properties, the type of sequence that precedes the NSI click turn in Example 2 below is an assessment, one of the categories proposed by Wright, and the 'prefatory discontinuity markers' that occur in English such as *anyway*, *okay* and *well*, also occur in Spanish, *bueno* 'well' in Example 2. Second, the pitch characteristics regarding the closing down of the sequence previous to the NSI click (low in the speaker's range and with a narrow pitch span) and the characteristics of the click-initiated new sequence (high in the speaker's range and with a wider pitch span) reflect what Wright (2007:1071)) proposes and what was found in this study. Third, Wright (2007:1070) suggests the clicks are often released with the simultaneous initiation of an in-breath, which can have a high amplitude and a relatively long duration and that is the case of Example 2. However, some other articulatory properties presented by Wright (2007:1071) such as a complete closure in the final syllable of the sequence preceding the click or glottalisation in the onset of the first lexical item in the click-initiated new sequences were not observed in the current data. The following example 1 shows many coincidences with what has been found for English.

In Example 2, León's closing down of the sequence in line 9 is breathy and low in intensity (42dB) with a pitch of 129Hz that is low in the speaker's range (average = 144Hz) with a narrow pitch span. León produces a long in-breath and another pause in line 11 after which he produces an alveolar click (see figure 5) followed by a TCU that is high in intensity all along (62dB) when compared to the speaker's average intensity (50dB) and the pitch is high (216Hz) on the speaker's range. The combination of an in-breath and a click is perceived by the interlocutor as marking incipient speakership as she makes no attempt to overlap in talk. This incipience is reinforced in this TCU by its loudness and high pitch. The speakers have gone off track in the conversation when León reintroduces the question asked by Rita earlier in the conversation with the discontinuity marker *bueno* 'well', getting back on the track of the conversation. The new sequence receives a fitting response (line 14) and Rita does not make an attempt to return to the previous topic.

Example 2

P3_MividaenYork

01 Rita: <<all>es que yo creo que aCÁ también es frio;>=
it's just that I think that it is also cold here
02 =pero es que uno <<whispery>está acostumbrado a:> al
but one is used to the
03 [FR]Ío y al vientito,=
cold and the wind
04 León: [<<p>SÍ.>]
yes
05 Rita: =este DÍA de verano por ejemplo de hoy. (0.7)
this summer day for example
06 León: si igual han haBIdo algunos días
yes there's been some
07 <<breathy>súper fríos acá:.=
really cold days here
08 =con VIENto súper helado.=
with very cold wind
09 =<<p>que yo CREO que puede ser como Punta Arenas.>>
I think it could be like Punta Arenas (a Chilean southern city)
10 León:→°hhh (0.4) ! <<f><<h>bueno y la mi VIda en York
well and the my life in York
11 es be es bastante amena.>>=
is is quite nice
12 =porque vivo con mi faMIlia?(0.5)
because I live with my family
13 Rita: ha ha haha[ha ha ha]
14 León: [<< :-)>de la que TÚ eres parte>.]
which you are part of
15 (0.3)

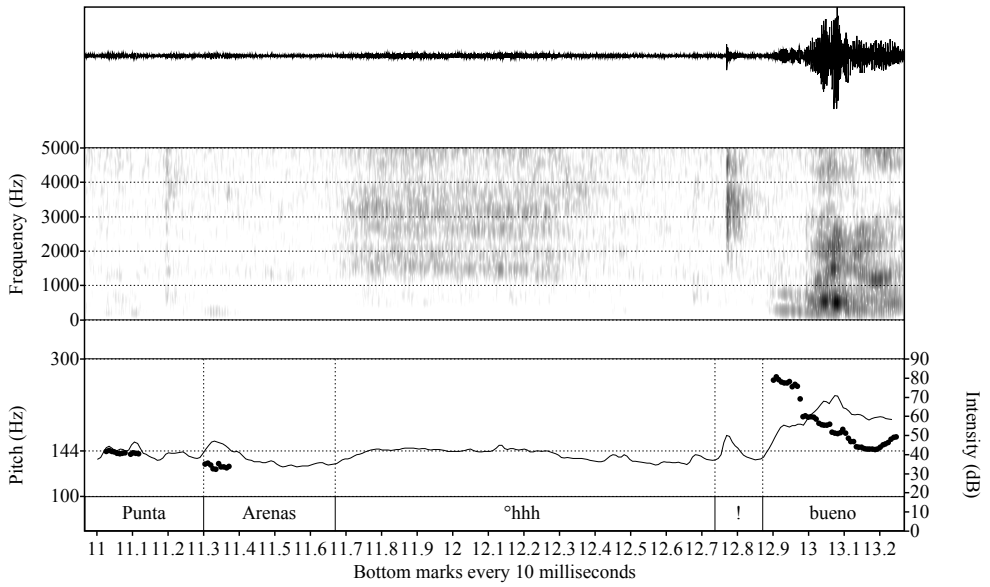


Figure 5. P3_MividaenYork - Waveform, spectrogram, pitch and intensity of closing down sequence previous to NSI click + in-breath + click-initiated new sequence.

3.2. Word Searches

Similarly to English, clicks are used in Chilean Spanish to signal trouble in finding words. They occur in conjunction with particles such as *eh, ehm*, in-breaths and swallowing or gaps at talk. Example 3 is a typical word search example. Rita is talking about how energy is displayed in massive events. In line 16 the problematic sequence begins with the lengthening of *mismo* 'same', resulting in *mismo:m:* and a long pause (1.6 seconds). In line 17 she swallows and produces a bilabial click. This is followed by the particle *ehm* (see Figure 6). The following TCU has a high pitch (259Hz) and a high amplitude (60dB). From line 18 through 22 Rita tries to explain the meaning of the word she is looking for. León displays understanding in line 23 by means of the particle *mm*. In lines 24 and 25 Rita identifies the word, *catarsis* 'catharsis' and repeats it in the following line reassuring herself it is the correct word with whispery voice quality and low amplitude. This is followed by a new TCU, line 27 where Rita ends this sequence by saying *esa es la palabra* 'that's the word' which does the same as the previous line, but also proves this was indeed a word search. As shown in Example 3, the TCU after the click is high in pitch and amplitude even though the word search is not resolved at that point but much later in line 25. This displays some similarity between what follows a click in new sequence indexing and word searches.

Example 3

P3_Catarsis

01 Rita: NO poh;
not really
02 allá es DONde quería llevar el a-=
that's where I wanted to take the
03 =el aSUNto de la energía.=
the energy matter
04 =esa MISma pasión que antes estaba <<len>confrontada
that same passion that before was confronted
05 en la batalla>? (0.6)

```
        in      battle
06      <<all>eso se traduce así;>
              that  translates     like that
07      en el f en el dePORte en el futbol
        in       in     sport     in     football
08      <<p>esta súper claro>.
             it is  very   clear
09      (0.5)
10 León:    MM.
             yes
11      (0.2)
12 Rita: es la misma enerGÍA pero- (0.2)
        it is the same   energy     but
13      pero <<len>PUESta: en una competencia: deportiva>;
        but          put       in  a    sports      competition
14      (0.6)
15      yo cacho que la MÚsica es <<p><<h>igual>>, (1.2)
        I believe    that it's  the  same  with  music
16      tiene el MISmo:m:, (1.6)
        it has   the  same
17  →   ((swallows)) ⊙  eh:: (1.4)
                        hmm
18      <<f>no ES eh>-=
            it's not hmm
19      =ES como ah eh- (.)
         it's like    hmm
20      todo es a faVOR de (así m),(.)
        everything is in favour of
21      no hay en CONtra o a favor,=
        there is nothing against or  in  favour
22      =no es una LUcha;
         it's not  a    struggle
23 León: [MM,    ]
          yes
24 Rita: [es como] SIMplemente <<creaky>una:>, (0.4)
          it's like      simply                a
25      <<len>caTARsis>. (0.2)
              catharsis
26      <<whispery><<p>caTARsis (xxx xxx). (0.3)
                       catharsis
27      Esa es la palabra.>>
        that's    the  word
28      (1.0)
```

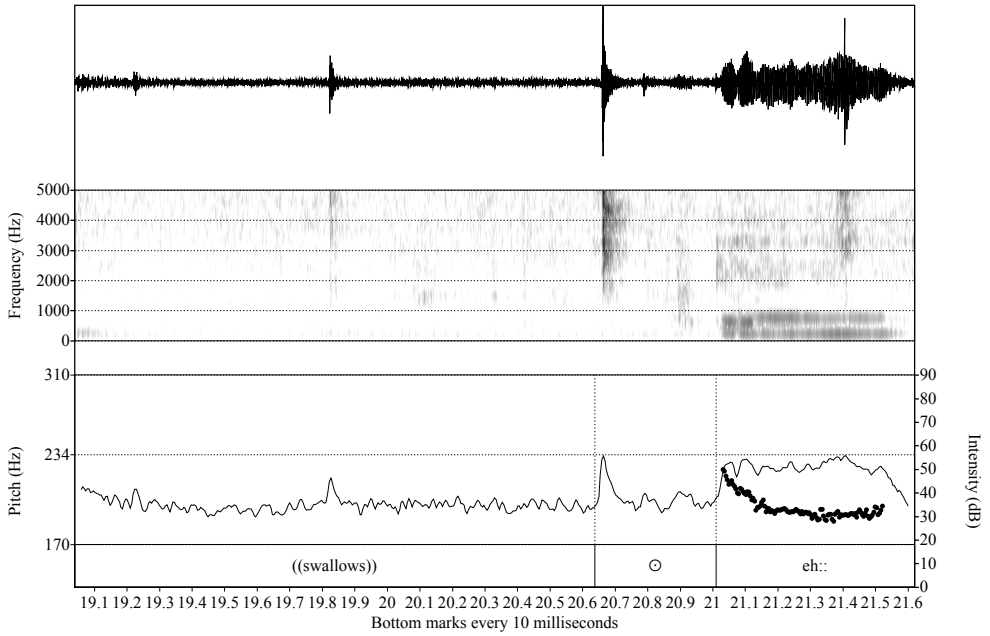


Figure 6. P3_Catarsis - Waveform, spectrogram, pitch and intensity of swallowing + click + particle *eh*

3.3. Display of stance

As described in the introduction, clicks "are part of a family of verbal and non-verbal practices" (Ogden, 2013: 23-24) to display stance at the beginning of turn, and this would not be the case of percussives. This is also true for Chilean Spanish, although there many of the cases involving clicks and gestures where clicks are turn-final.

Clicks that display some kind of stance or affect are high in amplitude and are often a deliberate action and they are designed for the interlocutor to hear (Ogden, 2013: 23). They may be accompanied by response tokens such as *ah* in Chilean Spanish and gestures such as eye blinks and hand movements. The following sections will try to shed some light on what the phonetic properties of the tokens in the environment of clicks are and see what bodily movements are associated with them (in the current data). In this respect, proposals regarding the construction created by the association of clicks and gestures will be given.

3.3.1. Clicks + ah *as change-of-state token*

The following is a comparison of examples from the same speaker where she displays understanding by means of a sequence that begins either with a click followed by the change-of-state token *ah* or just with this particle. In both cases the particle *ah* is turn initial and prefaces words that further display understanding on what so far has been problematic in the conversation, hence the denomination "change-of-state token".

Regarding what is being done in these cases, we could consider what Heritage (1984:321) states about the particle *oh* in English. He claims "there are occasions in talk where recipients may wish to show that prior talk has been adequately descriptive and/or that they have competently understood its import." In this sense, the repositioning of *oh* to turn initial is a resource used by speakers to ensure they display their understanding to their interlocutors.

Example 4 comes from a group of cases where a speaker initiates the turn with a click followed by the change-of-state token *ah* 'oh'. Rocío asks Lucas a question in line 2, from there onwards they start negotiating their memories regarding what he has done, particularly what the last film he saw at the cinema was. As he fails to remember, in line 25 he changes the direction of the conversation and refers to a film he would like to see instead. After a 0.8 seconds pause, Rocío does a click (line 28) followed by the particle *ah* 'oh' which is 0.8 seconds long, breathy and has a falling intonation contour (see figure 7). This construction displays she knows the reference Lucas has mentioned.

Example 4

P4_BradPitt

01 Rocío:pero: SÚper chistosa;=
but really funny
02 =y tú cuál fue la última que VISte? (0.8) en el CIne?
and you what was the last (film) you saw at the cinema
03 Lucas:<<len>en el CIne.> (0.6) uh: <<len>no he Ido hace::,>
at the cinema wow I haven't been since
04 (0.9)
05 [desde que lleGAmos a york]que no he ido al cine.
since we arrived in York I haven't been to the cinema
06 Rocío:[<<h>ah tú no has Ido acá en york.?]
oh you haven't gone here in York
07 Lucas:(0.7) <<creaky>un a::.
one
08 (1.1)
09 un año: y TRES meses que no voy al cine.
one year and three months that I haven't gone to the cinema
10 (0.5)
11 Rocío:<<creaky>O:[::h>
wow
12 Lucas: [y en chile la última que fuimos a ver al
and in Chile the last one we went to see to the
13 cine ni me aCUERdo cual es
cinema I don't even remember which one it is
14 (1.0)
15 QUÉ podría haber sido?
what could have been
16 (0.6)
17 Rocío:no SÉ.
I don't know
18 (2.7)
19 no me acuerdo como (.) SPIderman o[:,
I don't remember something like Spiderman or
20 Lucas: [NO::-
no
21 (1.0)
22 que mala meMOria,
such a bad memory
23 (0.9)
24 Rocío:((laughs)) (0.8)
25 Lucas:pero me gustaría ir a VER la de:: (0.2) brad pitt eh
but I would like to see the one with Brad Pitt hmm
26 world war zeta.

```
          World   War   Z
27        (0.8)
28 Rocío:→[ !            ]Ah::::.
                           oh
            [((eye blink))]
29        (1.4)
30        pero al CIne <<creaky>o:::, (.) te gustaría VERla: por
          but at the cinema            or          would you like    to see it  on
31        internet.>
          the internet
32        (0.6)
33 Lucas:<<h>no: <<creaky>me gustaría verla> en el CIne yo creo.>
               no                I would like to see it      at the cinema   I think
34        (1.0)
35        aunque AHOra igual hay hartas que son tres de,
          although   now           there are lots    that  are    3D
36        (1.0)
37        que uno podría verlas en el CIne.
          that   one  could     watch      at the cinema
```

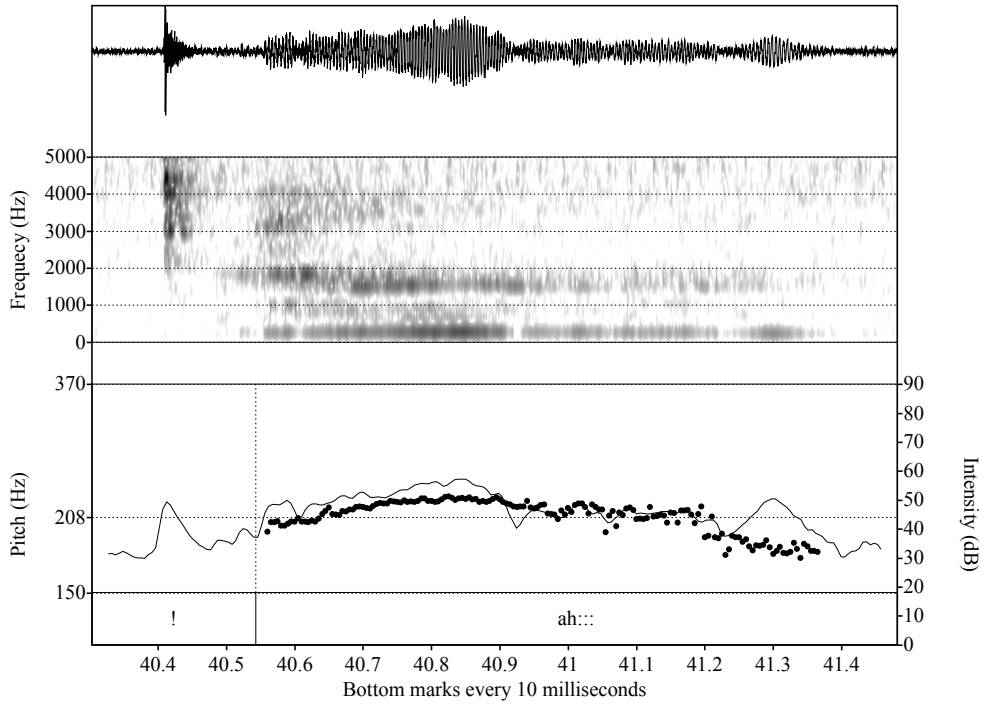


Figure 7. P4_BradPitt - Waveform, spectrogram, pitch and intensity of click + ah:::

The speaker in question also blinks at the same time she clicks as can be seen in in figure 8 where there are three examples where Rocío has her eyes closed in the production of the click and opens them afterwards. This could be a further sign that this click is displaying recognition as it is important for the maintenance of these kinds of conversations that speakers understand one another and that they display this understanding by means of gestures and continuers. The click and the blink in these cases are a recognisable practice, this is, once a speaker has done it, the trajectory of the current action can go where it should be and the conversation progresses.

Figure 8. Blinks produced simultaneously with clicks + subsequent eye openings for three cases in the collection

It is interesting to see that there are similar cases with and without clicks in Chilean Spanish. Example 5 is one from a group of cases where the same speaker produces sequences of realisation, this is, the particle *ah* 'oh' plus an utterance that displays understanding, where no clicks are uttered. Lucas states a fact in lines 17 and 18, but Rocío in line 20 disputes this. In line 22 Lucas insists on his assumption. Rocío in line 23 asks *sí* 'yes' with a rising intonation to confirm, then Lucas confirms it in line 24 and adds *acuérdate* 'remember' in line 26 which is in overlap with Rocío's realisation in line 27. The change of state is marked by the particle *ah* ‘oh’ that comes in fast together with the rest of the turn and in overlap with the previous turn (see figure 9). Lucas responds to this realisation with *sí* ‘yes’ in line 30.

Example 5

P4_Reloj

01 Lucas:oye este reloj ya está bien maLIto,
hey this watch is quite bad now
02 (2.4)
03 Rocío:por QUÉ?
why
04 Lucas:(0.3) porque MIra está todo ne:gro aquí;
because look it's all black here
05 Rocío:(.) <<h>DÓNde;> (1.4)
where
06 porque es por el meTAL poh. (0.7)
it's because of the metal
07 Lucas:vamos a tener que camBIARlo, (1.0)hh (0.8)
we're gonna have to change it

```
08        [hhh°
09 Rocío:[he visto Unos por internet;(0.5)
          I've seen      some  on the internet
10        que son baRA<<laughing>tos;> hehehe
          that  are   cheap
11 Lucas:(0.4) ?<<creaky>ah> (0.4) <<whispery>CUÁle:s,
                              oh                    which ones
12 Rocío:(0.9) <<h>como así de PLÁStico?>=
                     like          plastic ones
13        =<<h>bien: deseCHAbles.>=
               quite    disposable
14 Lucas:=no: po si la (.) la idea es comPRAR uno que dure
          no      if  the      the  idea   is   to buy     one  that  lasts
15        po;=
16        =<<creaky>ESte: la correa es de cuero-> (0.2)
                     this one  the  strap     is made of leather
17        <<creaky> y ya te ha durado CUÁNto,> (0.2)
                     and it has already lasted for how lomg
18        todo lo que estamos JUNtos. (0.2)
          all the time we've been        together
19        CAsi;
          almost
20 Rocío:(0.5)<<all>n:<creaky>o no me lo regaLASte> al
                     no              you didn't give it to me       at the
21        principio> (.)
          beginning
22 Lucas:<<h><<all>te o regalé para> tu primer cumPLEaños;>
                     I gave it to you   for      your first       birthday
23 Rocío:(0.2) SÍ:?
                yes?
24 Lucas:(0.3) SÍ:-
                yes
25        (2.0)
26 Lucas:          [aCUÉRdate;       ]
                    remember
27 Rocío:→<<all>[ah me regalaste]> el reLOJ,
                  oh  you gave me         the  watch
28        y un día en el sPA.
          and a day    at   the  spa
29        (3.5)
30 Lucas:SÍ-
          yes
31 Rocío:(.)o sea hace CUÁNto tiempo? (0.9)
             that means how  long       ago?
32        lleVAmos?
          we've been together?
33        (2.0)
34 Lucas:hace [TRES años;]
          three  years  ago
35 Rocío:     [llevamos  ]CUAtro años juntos.
                we've been together four  years
36        (1.3)
37 Lucas:cuatro Años. (0.2) <<creaky>tcho::h>
          four years                          wow
38 Rocío:(.) tch (0.4) <<all>pégate con una PIEdra en el
```

```
                hey                    hit yourself with  a      rock       on the
39          [pecho oye.]>
            chest
40 Lucas:[ha  ha     ]haha
41          (0.5)
42 Rocío:tch ha
43          (1.0)
```

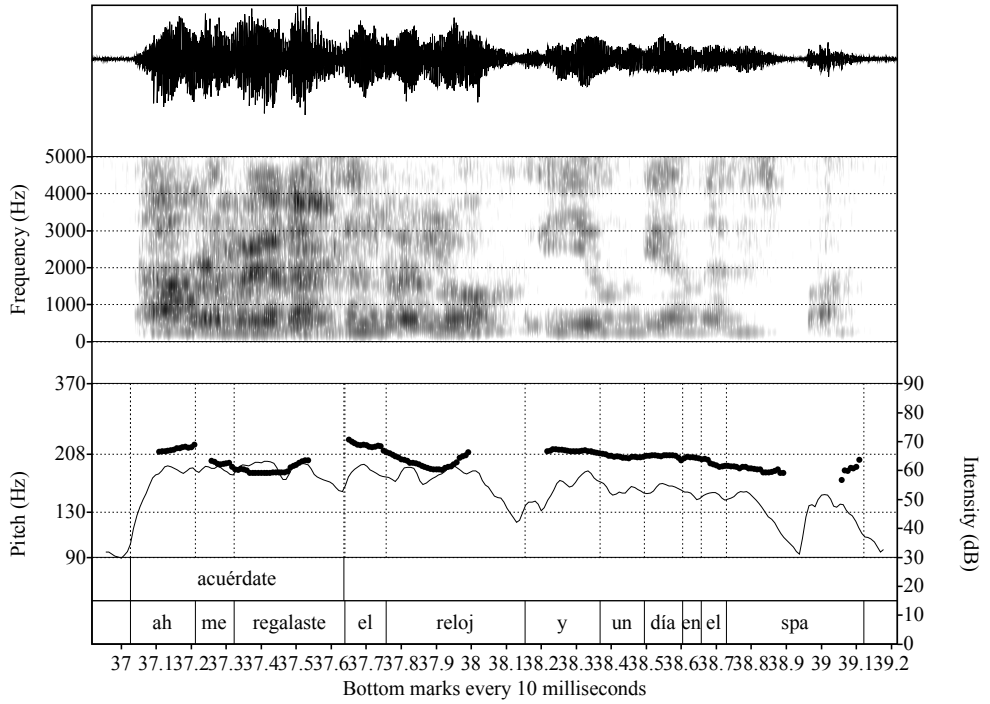


Figure 9. P4_Reloj - Waveform, spectrogram, pitch and intensity of *ah*

In all these examples, there are some important differences in length. In examples where clicks are uttered, the particle *ah* ranges from 508 to 838 milliseconds, as opposed to a range from 110 to 145 milliseconds in the cases without clicks. Furthermore, in examples with clicks there are long pauses (0.4 to 0.8 seconds) before the clicks are produced. This suggests those cases with the structure click + *ah* are a way of acknowledging the length of the pause which could be assumed as problematic and the click is clearly a sign of incipient speakership. Examples without clicks on the other hand are quite different. In all of them the *ah*-prefaced turn comes in overlap, this shows they are not dealing with a problem, but just help to create intersubjectivity, the relation between one's subjectivity and that of another.

Regardless whether there is a click or not, in some of the examples, both participants are negotiating their memories about certain events as they have different perspectives and recall different details about them. The display of understanding comes when one of the participants is able to recognize such events. In other examples one participant is recounting events and the displays of understanding seem to be inferences that the listener has drawn in relation to such events. Interestingly, in these examples the storyteller always confirms what the listener has said to display her understanding.

All things considered, there seems to be certain orderliness concerning the display of understanding with and without the accompanying production of clicks. Both in the location relative to the prior turn and in the articulatory rate in their production. When referring to the

particle *oh* in English - equivalent to *ah* in Chilean Spanish as already mentioned - Schegloff (1991:157) adds

> "*oh* can claim a change in the speaker's state, but its utterance enacts an interactional stance and does not necessarily reflect a cognitive event."
>
> (Schegloff, 1991:157)

However, in all of these cases, *ah* does in fact seem to bear a cognitive load often signalling a shared understanding between speakers, or an inference by one speaker is made and then confirmed.

3.3.2. Clicks and gestures

The following cases from the present study show sequences in which clicks were used in a display of affect and are also accompanied by gestures. These clicks are of a very specific nature forming part of a bigger construction that includes body movements and particular articulatory properties such as lip protrusion. These accompany the production of the click alongside distinctive lexical choices such as repetition. All of this is constrained by contextual cues.

The first case (Example 6) shows how the repetition of clicks is used to express negation in Chilean Spanish. What is interesting about this example is the strength of the formulations throughout the sequence and how that relates to the high number of negative particles and repetition of lexical items. There is the probability that the speaker who does most of the talking orients to this negative stance so she gets a strong response, but as it is possible to see in the transcription that such is not the case. There is repetition of the word *nada* 'nothing' in line 7 where the speaker goes into creak with a decreasing pitch and intensity as can be seen from figure 11. Line 23 is very similar to line 7, this time the word *no* is repeated with lip protrusion and head shaking as can be seen in figure 12. This is followed by three alveolar clicks that also have lip protrusion and are accompanied by head shaking. This evidence makes clear that the clicks here are another way of saying *no* as can be seen in figure 10 where Lía's head movements change as she says *no* and produces the clicks. In lines 29 and 30, there is more repetition of lexical items, the word *cantado* 'sung' is repeated six times as shown in figure 13. The repetition of the click seems to function as a metronome (Ogden, 2013: 314) as these occur in regular time intervals. In a sequence that is so loaded with repetition, it is easy to understand that the click is repeated following the same fashion, perhaps as a way of reiterating the point, especially because all of these repetitions are self-initiated and do not convey repair.

Example 6

P1_LesMiserables

01 Rosa: es un mu[siCAL]?
is it a musical?
02 Lía: [es un musi]cal enTEro. °h
it is a musical entirely
03 enTEro.=
entirely
04 =o sea no hay ninGUna <<all> parte de la película
I mean there is no part of the movie
05 donde es conversada>.
where they talk

```
06       (0.2)
07 Lía:  <<dim>N[Ada nada <<creaky>nada nada>>].
                 nothing nothing          nothing nothing
08 Rosa:         [Ah:: no te puedo creer ya    ],
                  Oh    I   can't    believe you  ok
09       (.)
10 Rosa: [YA::         ]
         right
11 Lía:  [to hasta los] DIÁlogos son medios (0.3) cantaitos;=
         everything even the  dialogues    are   kind of        sung
12       =y son como el (.) °hh (0.3) las co[neXIOnes]
          and they are like the                 the   connections
13 Rosa:                                          [YA:     ].
                                                   right
14 Lía:  entre una canción y la otra,=
         between one  song       and the other
15 Lía:  =pero,
          but
16 Rosa: YA:;
         right
17 Lía:  ((swallowing))! <<f>pero no HAY otra conversación>.
                                 but    there is no other  conversation
18       (0.3) no es como los de DISney que:=
               it's not   like   the Disney ones      where
19       =acTÚAN un rato;
          they act for a while
20       [CANtan un rato,]
          they sing for a while
21 Rosa: [SÍ y cantan ya.]
          yes and they sing right
22 Lía:  <<creaky>y después acTÚANn otro rato>;
                    and then      they act      another while
23     → <<creaky><<all><<dim>NO [no] no no> ¦w ¦w ¦w (0.2) °h (.)
                                 no   no    no  no
24 Rosa:                          [YA.]
                                  right
25 Lía:  NO.
         no
26       (0.5)
27       <<breathy>ahhh [ok>
                   oh    ok
28 Lía:                 [esto ES:- (0.2) e:hm hh° (0.6)
                          this   is            hmm
29       <<creaky><<dim><<acc>canTAdo cantado cantado
                              sung       sung       sung
30       cantado cantado cantado>>>.
         sung     sung      sung
31       TOdo el rato; (.)
         all   the  time
32       <<creaky>TOdo el rato>.
                   all   the  time
32 Rosa: [°h <<f>MIra y <<all>pero y de qué,]
                   look  and       but    what about
33 Lía:  [o sea la película emPIEza   ]con el can[tando y:]-
          I mean  the  film       starts        with him  singing      and
```

```
34 Rosa:                                                [ha ha ]ha ha
35 Lía:  (0.3) eh:: <<h>como pa darte el conTEXto>?
                hmm        like    to   give you the  context
36       así hasta COmo, (0.3)
         and even when they say
37 Rosa: [<<h>como una introducción>?]
               like     an     introduction
38 Lía:  <<h>en FRANcia> del <<creaky>no sé que>.
              in   France              I  don't   know   when
```

Figure 10. P1_LesMiserables - no followed by three clicks with lip protrusion, sequential images show headshaking.

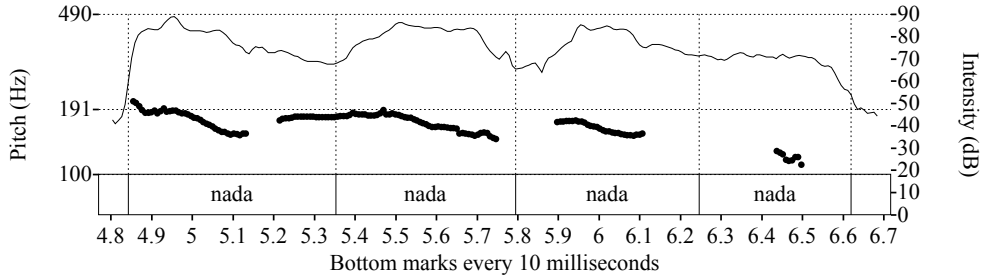


Figure 11. P1_LesMiserables - Pitch and intensity of the repetition of *nada* 'nothing'

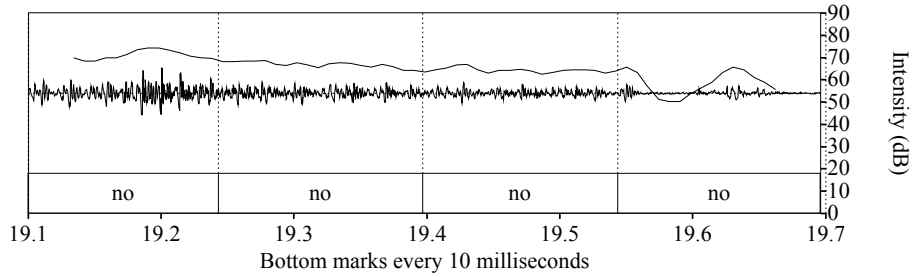


Figure 12. P1_LesMiserables - Waveform and intensity of the repetition of *no*

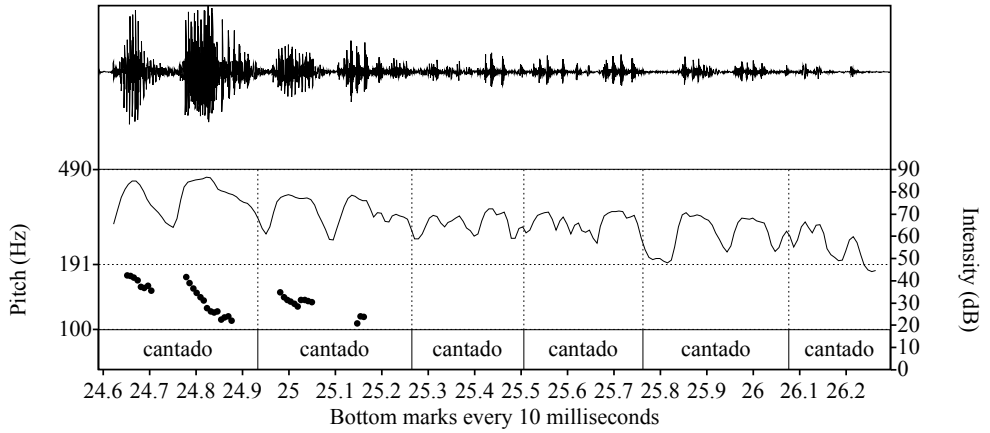


Figure 13. P1_LesMiserables - Waveform, pitch and intensity of the repetition of *cantado* 'sung'

For the analysis of example 7, some concepts that Liddel & Metzger (1998) have proposed for the study of sign language will be introduced. The first one is the idea of blended mental spaces where speakers can imagine that real entities are present anywhere around them (Liddel & Metzger, 1998:663). Another important concept is that of constructed actions which refers to the narrator's construction of another's actions and these provide a visual message that needs to be accurate to be understood (Liddel and Metzger, 1998:672). These ideas help to explain the clicks in the following examples which are part of a constructed action.

In example 7, Lía is talking about how, when her daughter plays, her eyes shine. Rosa laughs in overlap with Lía's line 13 where she proceeds to imitate the ways her daughter's eyes shine by means of gestures and an alveolar click that has lip protrusion, the gesture and protrusion can be seen in figure 14. From the video, it is possible to see that she intended to click more than once because she keeps the lip posture for longer than the click duration. The waveform in figure 15 shows there is activity and a peak in intensity similar to that of the click, perhaps indicating a percussive. The pitch trace is also included to support this argument.

Lía has real space elements to use in this construction as well, her own eyes in this case and elements that are part of her recollection of how her daughter's eyes shine. The shine, *per se*, is something that cannot be imitated with a body movement or subject to deixis. What Lía uses instead is the wiggling of her fingers to resemble the flickering of the eyes shine and the click is aligned with the hand movement. In line 14 Rosa displays her understanding with the word *ya* 'right'.

Example 7

P1b_Ojitos

01 Rosa: A:Y [que entrete.]
oh that's fun
02 Lía: [°hh](.)y AHÍ yo la voy a ver poh. (0.2)
and there I go to see her
03 Rosa: [Y:A.]
ok
04 Lía: [jugar](0.3) pero en verDAD a mí me encanta
play but the truth is I love
05 verla jugar;=

to see her play
06 =no porque TANto que me guste el voleibol,=
not because I like volleyball that much
07 =<<all>sino que> °hh (0.2)
but
08 <<creaky>si alguna vez> PUEdes ir a verla,(.)
if you ever have the chance of seeing her play
09 a Ella le cambia la cara:; (.)
her face changes
10 o sea ella se MEte a la cancha, (.)
I mean she goes into the field
11 <<len>y le BRIllan> <<creaky>los ojitos>, (.)
and her little eyes shine
12 Rosa: ha[hahaha]
13 Lía: → [así como] ǃ͡ʷ
like
14 Rosa: YA.
right

Figure 14. P1b_Ojitos - click with lip protrusion and wiggling of fingers

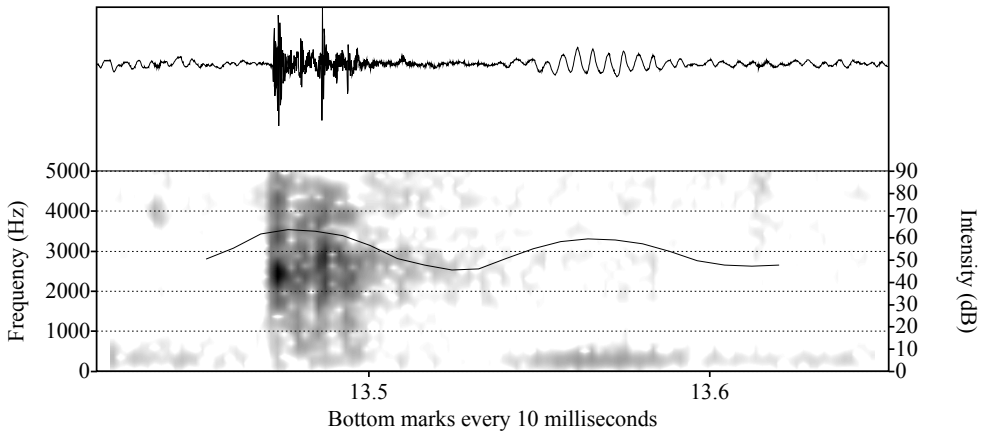


Figure 15. P1b_Ojitos - Waveform, spectrogram, pitch and intensity of click [ǃ͡ʷ] + possible percussive

4. Conclusion

Regarding the research questions previously posed, clicks in Chilean Spanish do occur in sequences similar to the ones reported for English. The findings for the distribution of clicks in terms of their function for Chilean Spanish show that for incipient speakership, NSI clicks behave similarly to English. The sequences which precede them are of the same kind, they have equivalent prefatory discontinuity markers, they are accompanied by in-breaths and they have the same pitch properties regarding the sequence previous and posterior to the NSI click. In the case of plain incipient speakership, clicks are turn- or TCU-initial. They are generally the culmination of other physical activity such as lip closure or swallow and function as an audible signal that may indicate one is about to talk and wants the interlocutor to acknowledge this.

Clicks in word searches were also found to function as they do in English; in order to signal trouble in finding words. They occur in conjunction with particles such as *eh*, *ehm*, which are equivalent to the hesitation markers found for English (Ogden, 2013:23), along with in-breaths and swallowing or gaps in talk. These clicks present similar properties to incipient speakership, but are more linguistically stable and structured.

Ogden (2013:24) had suggested that plain incipient speakership was likely to be similar in other languages and subject to individual differences that Scobbie et al (2011) identifies as "amount and texture of saliva", and individual features such as "'neutral' mouth postures". This short study may not be able to shed light on any of these individual differences, but it has shown that clicks marking incipient speakership do exist in Chilean Spanish and, most importantly, they behave similarly both sequentially and in terms of their phonetic environment to those in English. Ogden (2013:24) also suggested that since word searches and NSI involve more overtly linguistic elements, there was the likelihood to find important formal and functional differences between languages. As proved by this research, that is not the case because clicks in words searches and NSI clicks are very similar and so are their sequential environment and the articulatory properties that rule them.

For clicks that display some kind of stance or affect, the ones that are accompanied by a change-of-state token resemble similar cases in English. This research also proved that these cases with clicks differ from those displays of understanding without clicks, but still have a change-of-state token particularly in the duration of such token although the function is essentially the same for both cases. Eye blinks were also found to be simultaneously produced with clicks in the examples studied.

Finally, clicks were found to display stance when accompanied by gestures. They seem to be part of a bigger construction that includes body movements and particular articulatory properties (lip protrusion) as well as distinctive lexical choices (e.g. repetition) constrained by contextual cues. Ogden (2013:24) had also suggested this type of click to prove different in different languages and this is true for Chilean Spanish.

The findings of this research contribute to our understanding of the phonetic and interactional organisation of everyday social exchange in a language that had not been subject to this kind of study. Clicks are clearly not phonemic in Chilean Spanish, but the fact that they have a regular distribution and are part of prosodic constructions that convey meaning entails that they are indeed linguistic. Moreover, the similarities with English in the way a click helps to show 'gearing up to speak', to index new sequences that are disjunctive with the prior, to signal trouble in finding words; and the differences such as the particular use of clicks used to display affect found for Chilean Spanish support this argument.

Further research is needed to see if these findings are consistent and stable and to give a detailed account of those sequences in which clicks are said to occur but are not part of the data in this study. The need for conversations in less controlled settings is essential for such task.

Acknowledgements

I am grateful to Prof. Sonia Montero for her support and Dr. Richard Ogden for his supervision and valuable feedback on the content of this article.

References

BOERSMA, PAUL AND WEENINK, DAVID. 2013. Praat: doing phonetics by computer [Computer program]. Version 5.3.48, retrieved 1 March 2013 from http://www.praat.org/

CATFORD, JOHN CUNNISON. 2001. *A practical introduction to phonetics*. Oxford: OUP.

CLARK, JOHN, COLIN YALLOP AND JANET FLETCHER. 2007. *An introduction to Phonetics and Phonology*. Oxford: Blackwell Publishing.

EKLUND, ROBERT. 2008. Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association* 38. 3. 235-324.

GIL, DAVID. 2011. Para-Linguistic Usages of Clicks. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library.

GIMSON, A. C. 1970. *An introduction to the pronunciation of English*, 2nd edn. Bristol: Edward Arnold.

GÜLDEMANN, TOM AND MARK STONEKING. 2008. A historical appraisal of clicks: a linguistic and genetic population perspective. *Annual Review of Anthropology.* 37. 93-109.

HAVET, LOUIS. 1875. Observations phonétiques d'un professeur aveugle. *Mémoirs de la Société de Linguistique de Paris 2*, 218-221.

HERITAGE, JOHN. 1984. A change-of-state token and aspects of its sequential placement. *Structures of social action: Studies in conversation analysis*. 299-345.

KELLY, JOHN AND JOHN LOCAL. 1989. *Doing Phonology*. Manchester: Manchester University Press

LADEFOGED, PETER AND IAN MADDIESON. 1996. *The Sounds of the World's Languages.* Oxford: Blackwell Publishing.

LADEFOGED, PETER. 2005. *Vowels and Consonants: An Introduction to the Sounds of Languages*. 2nd edn. Oxford: Blackwell.

LAVER, JOHN. 1994. Principles of Phonetics. Cambridge: Cambridge University Press.

LIDDELL, SCOTT K., AND MELANIE METZGER. 1998. Gesture in sign language discourse. *Journal of Pragmatics* 30. 6. 657-697.

LOCAL, JOHN AND GARETH WALKER, 2005, 'Mind the gap': further resources in the production of multi-unit, multi-action turns. *York Papers in Linguistics Series 2*, 1, 133-143

OGDEN, RICHARD. 2013. Clicks and percussives in English conversation. *Journal of the International Phonetic Association*. 43. 3. In press

SACKS, H., SCHEGLOFF, E.A., JEFFERSON, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50. 4. 696-735.

SCHEGLOFF, E.A. 1991. Conversation analysis and socially shared cognition. *Perspectives on Socially Shared Cognition.* Washington, DC: American Psychological Association. 150-171.

SELTING, MARGRET, PETER AUER, DAGMAR BARTH-WEINGARTEN, JÖRG BERGMANN, PIA BERGMANN, KARIN BIRKNER, ELIZABETH COUPER-KUHLEN, ARNULF DEPPERMANN, PETER GILLES, SUSANNE GÜNTHNER, MARTIN HARTUNG, FRIEDERIKE KERN, CHRISTINE MERTZLUFFT, CHRISTIAN MEYER, MIRIAM MOREK, FRANK OBERZAUCHER, JÖRG PETERS, UTA QUASTHOFF, WILFRIED SCHÜTTE, ANJA STUKENBROCK, SUSANNE UHMANN1. 2011. A system for transcribing talk-in-interaction: GAT 2. *Gesprächsforschung* 12.
Online: www.gespraechsforschung-ozs.de/heft2011/heft2011.htm.

WARD, NIGEL. 2006. Non-lexical conversational sounds in American English. *Pragmatics and Cognition.* 14. 1. 129-182.

WRIGHT, MELISSA. 2005. Studies of the phonetics-interaction interface: clicks and interactional structures in English conversation. PhD Thesis, University of York, UK.

WRIGHT, MELISSA. 2007. Clicks as markers of new sequences in English conversation, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 1069-1072.

WRIGHT, MELISSA. 2011a. On clicks in English talk-in-interaction. *Journal of the International Phonetic Association*. 41. 02. 207-229.

WRIGHT, MELISSA. 2011b. The phonetics–interaction interface in the initiation of closings in everyday English telephone calls. *Journal of Pragmatics*, 43. 4. 1080-1099.

ZILIO, GIOVANNI M. 1986. Expresiones extralingüísticas concomitantes con expresiones gestuales en el español de América. Actas de los congresos de la Asociación Internacional de Hispanistas IX. 139-151.

Appendix

Transcription conventions

GAT 2 (Selting et al, 2011) with some modifications.

http://www.gespraechsforschung-ozs.de/heft2011/px-gat2-englisch.pdf

Sequential structure
[ ] overlap and simultaneous talk
[ ]
= latching
→ refers to a line of transcript relevant in the argument

Inbreaths and outbreaths
°h / h° in- / outbreaths of appr. 0.2-0.5 sec. duration
°hh / hh° in- / outbreaths of appr. 0.5-0.8 sec. duration
°hhh / hhh° in- / outbreaths of appr. 0.8-1.0 sec. duration

Pauses
(.) micropause, up to approximately 0.2 sec duration
(0.2)- (2.0) measured pause (notation with one digit after the dot)

Other segmental conventions
: lengthening, by about 0.2-0.5 sec.
:: lengthening, by about 0.5-0.8 sec.
::: lengthening, by about 0.8-1.0 sec.
ʔ cut-off by glottal closure
eh, ehm, hesitation signals, so-called 'filled pauses'

Laughter and coughing
haha syllabic laughter
hehe syllabic laughter
hihi syllabic laughter
((laughs)) description of laughter
<<laughing> > laughter particles accompanying speech with scope
((coughs)) non-verbal vocal actions and events
<<coughing> > …with indication of scope

Accentuation
SYLlable focus accent

Final pitch movements of intonation phrases
? rising to high
, rising to mid
– level
; falling to mid
. falling to low

Changes in pitch register, with scope

<<l> > lower pitch register
<<h> > higher pitch register

Loudness and tempo changes, with scope

```
<<f> >        forte, loud
<<ff> >       fortissimo, very loud
<<p> >        piano, soft
<<pp> >       pianissimo, very soft
<<all> >      allegro, fast
<<len> >      lento, slow
<<cresc> >    crescendo, becoming louder
<<dim> >      diminuendo, becoming softer
<<acc> >      accelerando, becoming faster
<<rall> >     rallentando, becoming slower
```

Changes in voice quality and manner of articulation, with scope

```
<<creaky> >       glottalized, "vocal fry"
<<whispery> >     change in voice quality as stated
<<breathy> >      change in voice quality as stated
```

Other phonetic phenomena

```
[⊙] bilabial click
[!] alveolar click
[‖] lateral click
```

Gestures

```
((eye blink))     gesture as stated
```

Verónica González Temer
Department of Language and Linguistic Science
University of York
Heslington
York
email: vmg503@york.ac.uk

PRODUCTION AND PERCEPTION OF SMILING VOICE

ILARIA TORRE

University of York

Abstract

This study investigates phonetic aspects in the production and perception of Smiling Voice, i.e. speech accompanied by smiling. A new corpus of spontaneous conversations is recorded to compare the formant frequencies of Smiling Voice (SV) and Non-Smiling Voice (NSV); the hypothesis that smiling raises formant frequencies is proven to be valid also on spontaneous speech, after previous research found this hypothesis to be true for scripted speech. Then, two perception experiments are carried out to test the hypothesis that listeners can recognize instances of SV extracted from the corpus of spontaneous data and instances of NSV obtained from read speech. Once the hypothesis is confirmed, a second perception experiment is performed to attempt to locate the point, in an artificial continuum from SV to NSV, where such perception happens.

1. Introduction

1.1. Smiling

The co-occurrence of speech and smile has been given various names in the literature, such as "speech-smile" (Kohler 2007), "smiled speech" (Émond and Laforest 2013) or "Smiling Voice" (Pickering et al. 2009). Here, the name Smiling Voice will be adopted, as opposed to Non-Smiling Voice (which describes neutral speech produced with no accompanying facial gesture); the abbreviations SV and NSV will sometimes be used.

Smiling is a universal phenomenon, common in humans and other animals (Mehu and Dunbar 2007: 271; van Hooff 1975: 231), which involves many movements in the facial area, especially in the region of the eyes, brows and mouth (Mehu and Dunbar 2007: 270). Ultimately, smiling shortens the vocal tract (Shor, 1978: 89), which affects the properties of the sound produced simultaneously.

1.2. Smiles and emotions

Smiles are mostly considered expressions of positive emotions such as happiness or joy (Kohler 2007: 21), although it is possible to fake smiles (Ekman and Friesen 1982: 244). Whatever emotion smiles can convey, humans seem to be quite good at recognizing it. This has been demonstrated with a number of experiments using only audio stimuli (Laukka 2005), only visual stimuli (Eisenbarth and Alpers 2011), non-verbal audio stimuli (Hawk et al. 2012) or crossed audio-visual stimuli (de Gelder and Vroomen 2000), in an application of the McGurk Effect, i.e. a phenomenon caused by a mismatched audio-video set of stimuli, which leads to the perception of an alien sound (McGurk and MacDonald 1976).

The present study focuses on some phonetic aspects of the production and perception of Smiling Voice. Even though the question of how Smiling Voice is linked to emotions was not addressed in any phase of the research, it is not excluded that emotions might have played a role in some of the experiments, especially the perception ones; after all, most of the study concentrates on naturally-occurring conversations, where emotions are put on display constantly.

Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York (**PARLAY 2013**) conference

 ISSN 1758-0315 Issue 1

2. *Production of Smiling Voice*

2.1. *Previous findings on the acoustics of Smiling Voice*

Some of the research on Smiling Voice includes the work by Kohler (2007: 21), who found that smiling increases the formant frequencies of the sound. Tartter (1980) found that smiling increases the formant frequencies, fundamental frequencies and, at least for some speakers, duration and amplitude, of the speech portion. These findings were confirmed by Tartter and Braun (1994), Drahota et al. (2008) and Fagel (2009). However, works on Smiling Voice are still scarce, and their methodology has yet to be proven valid. For example, most of the literature on the topic has used scripted speech, i.e. instructing actors or naïve speakers to artificially produce Smiling and Non-Smiling Voice. One of the few works employing natural conversations (Drahota et al. 2008) analysed the effect of smiling on vowels, without taking into account whole utterances, but it is clear that the acoustics of smiling affect whole portions of speech. Only Émond and Laforest (2013), in their very recent work on the prosodic correlates of Smiling Voice, used whole smiled words coming from a corpus of spontaneous conversations, but – since it was impossible for them to find a non-smiled counterpart for each word in that corpus – they chose other non-smiled words that had similar characteristics (e.g., same number of syllables and same duration). It is evident, however, that it would be pointless to make a comparison of the formant frequencies between such words, because the segments constituting the original words will be different.

The present research explores the hypothesis that changes in formant frequency from neutral voice to Smiling Voice affect data coming from naturally-occurring conversations as well.

2.2. *Methodology*

Five pairs of friends (four females and six males), all native English speakers, participated in two recording sessions. They were initially audio and video recorded while same-sex pairs conversed in a sound-proof booth of the University of York, using two Sanyo Xacti video recorders, one Zoom H4n audio recorder and two Beyer Opus 55 headset Microphones, at a sampling frequency of 44.1 kHz. In this first recording session, participants could choose to play a number of games that were designed to elicit smiles (e.g. picking objects from a box and recall experiences related to them (as used by Beskow et al. 2009: 190), playing "Taboo" (a card game that involves describing words), completing children's crossword puzzles, and playing rhymes).

Picture 1: The recording booth[1]

[1] Written permission to use audio-visual material from the recordings was obtained by all participants prior to the recordings, in accordance to the guidelines specified by the Ethics Committee at the University of York.

After the session, all the videos were watched in order to identify the utterances that contained smiles all throughout. Only single words or short intonational phrases (identified as chunks of speech associated with one intonation pattern, as described in Wells 2006: 6) were selected. Words or phrases which overlapped with the other participant's speech, contained instances of laughing voice or laughter, or were accompanied by background noise were discarded. When selecting the utterances that contained Smiling Voice, a smile was defined as a facial gesture involving a rising of the cheek muscles, a widening of the mouth and an upwards curving of the lips, as described by Shor (1978: 88).

The selected words and intonational phrases were then isolated in the corresponding audio files and extracted using the software Praat (Boersma and Weenink 2012), thus obtaining one short audio file for each instance of Smiling Voice.

The same subjects were called shortly afterwards to participate in a second recording session, where they were asked to read the list of words or phrases that they had originally uttered with Smiling Voice. They were instructed to read in a manner that was "as neutral as possible", without smiling. In this way, a significant number of pairs of SV and NSV were obtained, and the mean frequencies of the first three formants of each token were computed running a script in the software Praat. Even though it is clear that these pairs differed in more elements than just smiling (amplitude and f0, to name a few), since they were uttered in completely different contexts, the focus of this research was on the differences among the first three formants in the target words. Therefore, the reading modality was chosen as one of the ways to obtain exactly the same word that had originally been uttered in the spontaneous conversations.

2.3. Results

The final data set was composed of 3462 frequency values (1154 SV/NSV tokens * 3 formant values). These were computed both in Hertz and in Bark scales (using the formula provided by Traünmuller 1990: 99), in order to investigate both the production and perception aspects of Smiling Voice. A 2-tailed t-test was performed on the data to obtain the mean difference between frequencies in SV and NSV, then for each formant, then only for tokens containing rounded segments, and then for gender-specific differences. A discussion of all the results is given in paragraph 2.4.

Figure 1 summarizes the results of the first tests. The top charts show that, in general, the formant frequencies in Smiling Voice are higher than in Non-Smiling Voice. The frequencies of the first formant were found to be generally higher in Smiling Voice, while Non-Smiling Voice resulted in higher second formant frequencies. Finally, the frequencies of the third formant turned out to be much higher in Smiling Voice, with a difference of 53 Hz and 0.111 Bark.

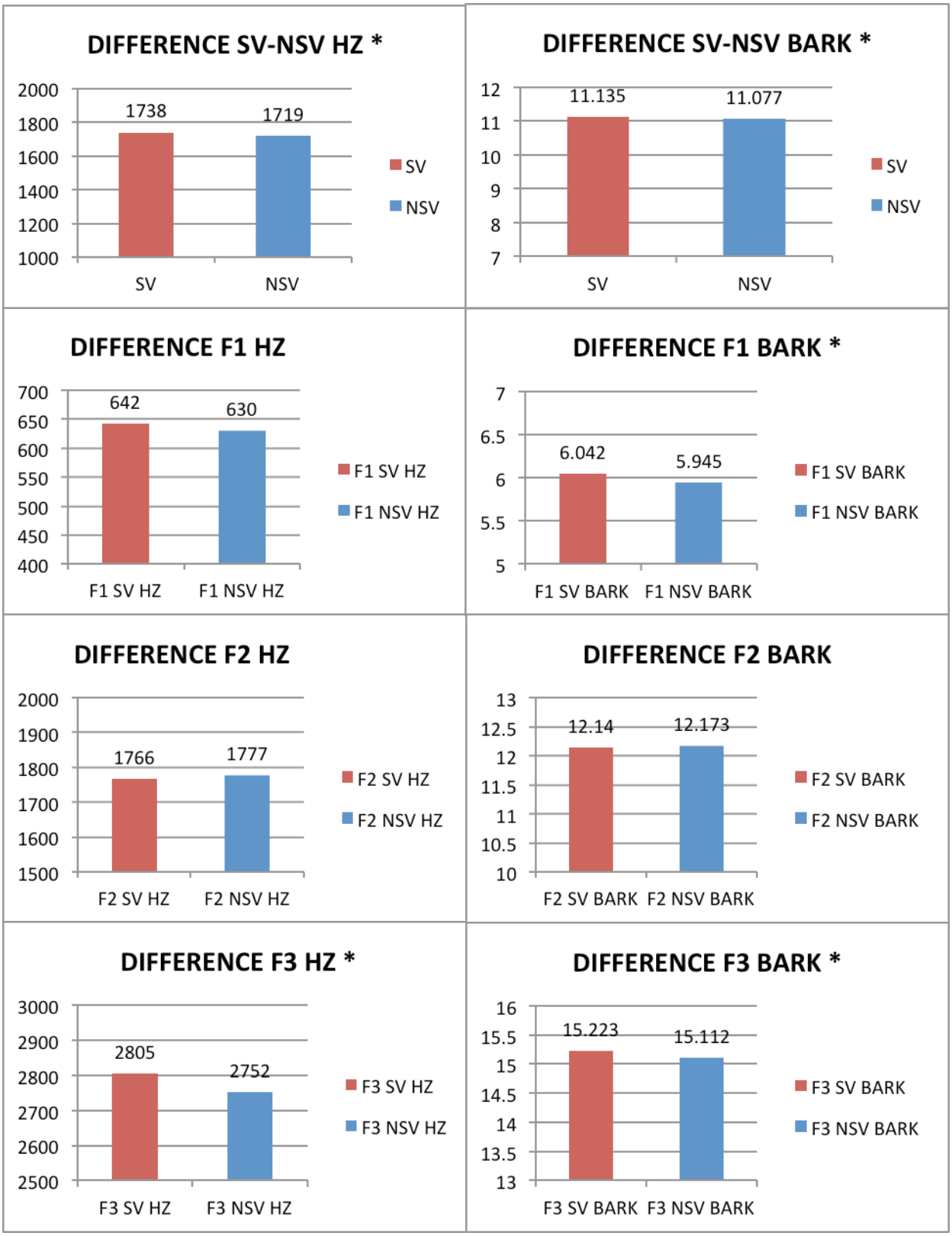


Figure 1: Differences in the average formant frequencies between Smiling and Non-Smiling Voice in Hertz (top left) and Bark (top right), and in f1, f2 and f3 frequencies.

Figure 2 represents the results of the test on the words that contained rounded vowels and/or consonants. The result was a mean difference of 73.7 Hz (top charts in figure 2.4 below), which is approximately seven times the mean difference calculated for all the words that did not contain rounding.

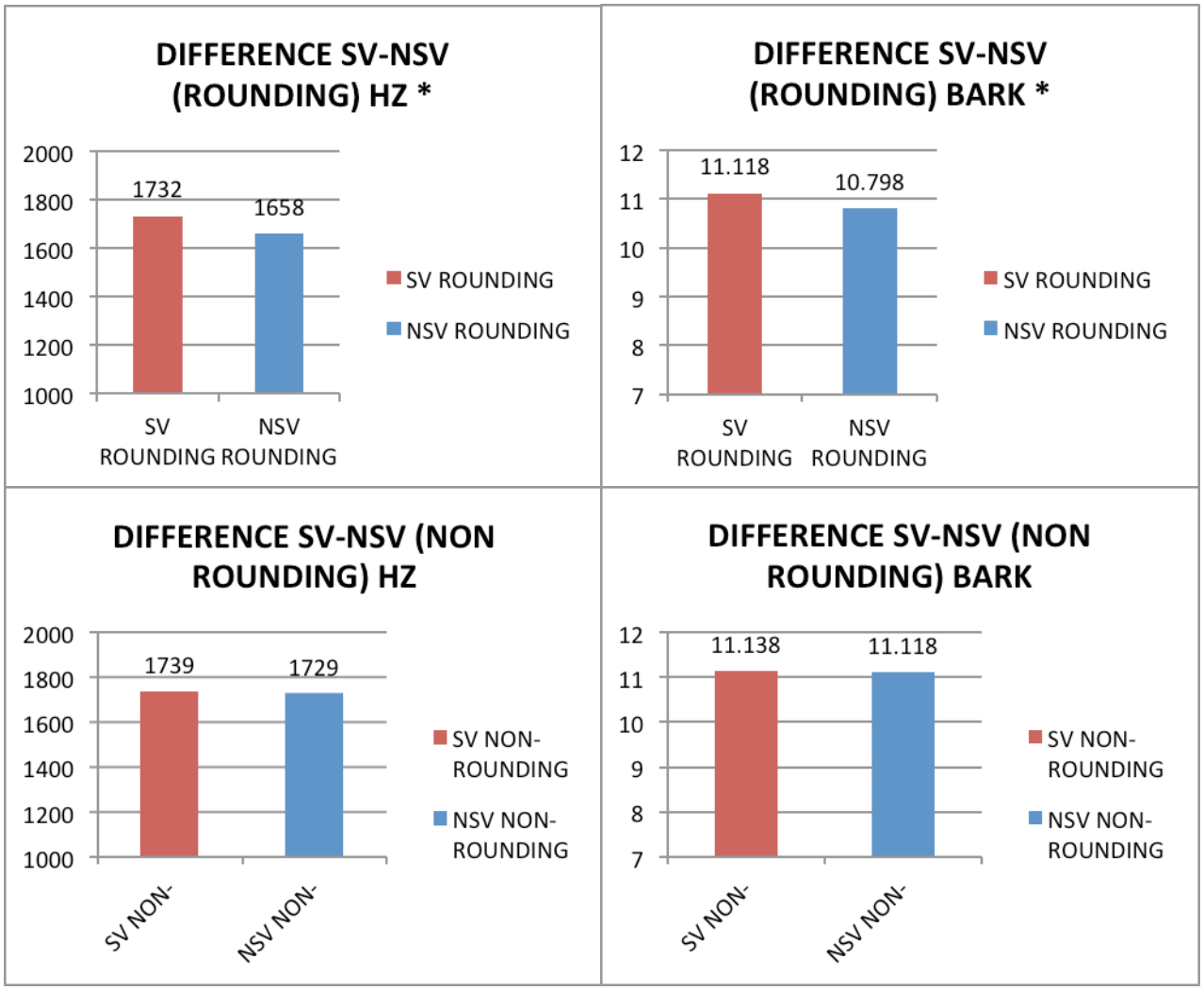


Figure 2: Differences in the average formant frequencies between Smiling and Non-Smiling Voice, in words containing rounding (top row) and words not containing rounding (bottom row).

Finally, Figure 3 shows differences in the data uttered by male speakers in comparison with the data collected from female speakers. Apart from showing that females' formant frequencies are higher than males', as expected (Clark et al. 2007: 269), it shows that, while the results for the male speakers are consistent with the average findings of Figure 1, female speakers present the reverse pattern, with frequencies in Non-Smiling Voice being higher than in Smiling Voice.

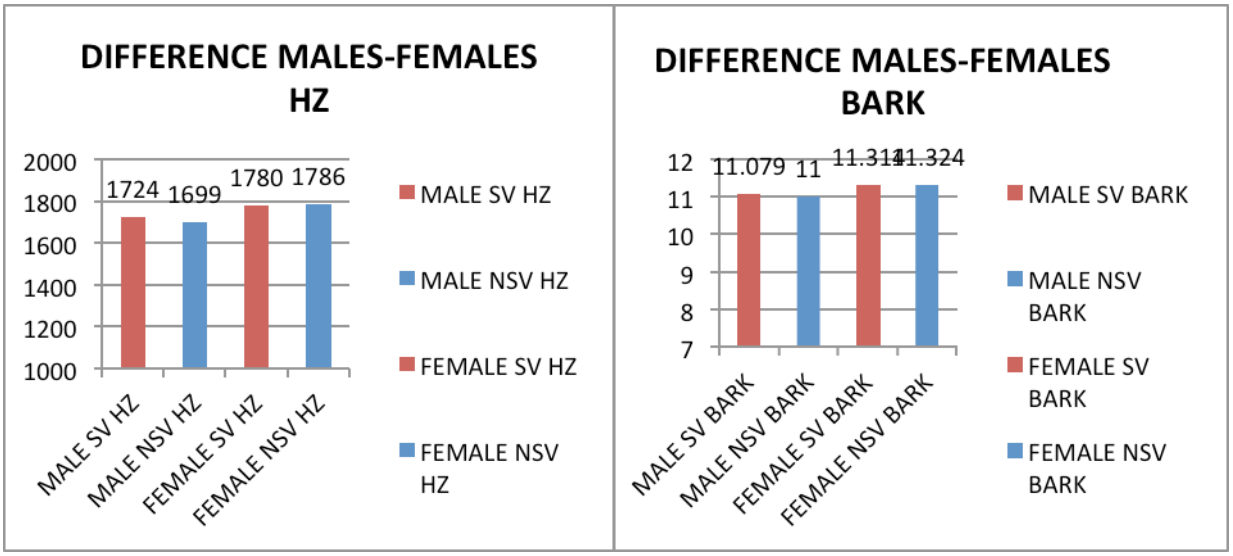


Figure 3: Differences in the average formant frequencies between Smiling and Non-Smiling Voice in male and female speakers.

2.4. Discussion

The general finding of this study is consistent with the hypothesis that smiling raises the formant frequencies of speech. This result appears to be statistically significant on the whole ($p<0.005$) and for words containing rounding ($p<0.005$). This second finding is not surprising, as lip rounding – which is usually accompanied by lip protrusion, and which lengthens the vocal tract (Fant 1960:116) – is the phenomenon that is affected the most by smiling – which, on the other hand, as said before shortens the vocal tract.

Examination of the individual formant frequencies, however, shows significant differences from previous experiments. In particular, Tartter (1980: 26) found that all three formant frequencies were higher in the smiling register, and considered f2 to be the most significant. Therefore, in a following study, Tartter and Braun (1994: 2105) reported only f2. In this study, however, f2 turned out to be the least significant value.

Differences in f1 appear not to be significant if the values are considered in Hertz ($p = 0.008$), but they start to be if the values are considered in Bark ($p = 0.005$), which means that, in general, differences in f1 in the passage from smiling to non-smiling are perceivable by a human ear. On the other hand, and contrary to previous findings, differences in f2 are negative, which means that, in general, Smiling Voice in this corpus has lower f2 frequencies than Non-Smiling Voice. However, these differences appear not to be statistically significant in both Hertz, and Bark ($p = 0.119$ and $p = 0.213$ respectively). What is striking is the comparison with Tartter (1980: 26), who found that, for 67% of his speakers, differences in f2 were statistically higher for smiling modality ($p < 0.002$), and with Lasarcyk and Trouvain (2008: 346), who found that lip spreading raises mainly the frequencies of f2. Since the second formant is the resonance associated with the length of the vocal tract, it would have been predictable that a shortening in the vocal tract would have raised the formant frequencies. However, in Tartter's research only scripted speech was used, which may have therefore biased the acoustics of the final data. Furthermore, neither the present study nor Tartter's employed a large enough number of speakers to make statistically relevant claims.

Instead, differences in f3 are statistically significant both in Hertz and in Bark ($p < 0.005$), and they represent the highest difference in the corpus (53 Hz and 0.111 Bark). This suggests that, of the three formants, f3 is the one that changes the most from smiling to non-smiling. This finding confirms previous results: Tartter (1980: 26) found that the difference from the

scripted Smiling and Non-Smiling Voice was higher in f3 as well. However, she did not provide an explanation for it, and, in another study she chose to concentrate on f2, because the similar effect on the three formants was considered "redundant" (Tartter and Braun 1994: 2104).

Finally, there is a significant difference between the male and female speakers used in the recordings for this study. Even though, as expected (Huber et al. 1999: 1540; Clark et al. 2007: 269), females' formant frequencies are generally higher than males', what is striking is the relative difference between SV and NSV. Males have generally higher formant frequencies for Smiling Voice than Non-Smiling Voice, whereas females have generally lower formant frequencies for Smiling Voice than Non-Smiling Voice. This result might suggest that the male speakers employed in the present recordings tend to smile "more broadly" than the female speakers, but no such tendency was observed in the videos. Also, as only a small number of speakers were used in this study, it would be unreliable to account for this claim. Furthermore, a broad smile is not produced only by a change in the vocal tract (and formant frequencies changes refer exclusively to changes in the vocal tract), but also by using many different facial muscles, therefore requiring different amounts of energy for individual speakers. Such gestures, as already explained, are not considered here.

3. Perception of Smiling Voice

3.1. Previous research

Although, as mentioned above, smiling does not necessarily correspond to a particular emotion, it is true that some emotions can be expressed by smiling, and most of the studies carried out so far have concentrated on the perception of smiles as linked to an emotion. In particular, it has been found that factors such as a listener's age (Lambrecht et al. 2012: 535; Paulmann et al. 2008: 265) or gender (Van Strien and Van Beek 2000: 650; Paulmann et al. 2008: 267) can influence said perception.

In this research, importance has been given to gender differences in the perception of Smiling Voice, which has been studied in two experiments. In the first one, subjects listened to a set of stimuli and were asked to recognize which stimuli were uttered while smiling and which ones were not. The second experiment constitutes an attempt to find the moment - in an artificial continuum from Smiling Voice to Non-Smiling Voice - where smile perception starts or ends.

3.2. Experiment 1

As mentioned above, the first perception experiment of this research seeks to confirm the hypothesis that listeners can recognize words uttered in Smiling Voice coming from spontaneous conversation, as opposed to words that were read without smiling.

3.2.1. Methodology

The stimuli for this experiment were taken from the recordings done initially to study the acoustics of Smiling Voice (section 2.2). They were selected from the words and phrases that had been already extracted from the first recordings, in pairs (i.e. for each word in Smiling Voice there was a corresponding word in Non-Smiling Voice). The final set was composed of

41 stimuli, 19 of which were filler items. 16 listeners, 8 males and 8 females, all native English speakers, volunteered to do the experiment, which took place in a quiet room, one person at a time. The audio was played on a laptop, and the listeners wore a noise-reducing Sennheiser HD 280 Pro set of headphones. They were given an answer sheet with 41 boxes, and were asked to fill each box with an S (if they thought that the corresponding word had been uttered with a smile) or with an N (if they thought that the corresponding word had been uttered without smiling). The listeners' response times were not recorded, but none of them asked to listen to a stimulus twice, and all of them managed to write each answer in the time slot provided (i.e. the 6 second pause between one stimulus and the next). No participant withdrew from the experiment, and no boxes on any answer sheet were left blank.

3.2.2. Results and discussion

As Figure 4 shows, the number of correct answers surpasses by far the number of incorrect ones: the participants answered correctly 77.9% of the time. This figure shows that listeners actively recognized Smiling and Non-Smiling Voice, without simply guessing (a normal distribution test confirmed that their performance was significantly better than chance, $z > 1.645$).

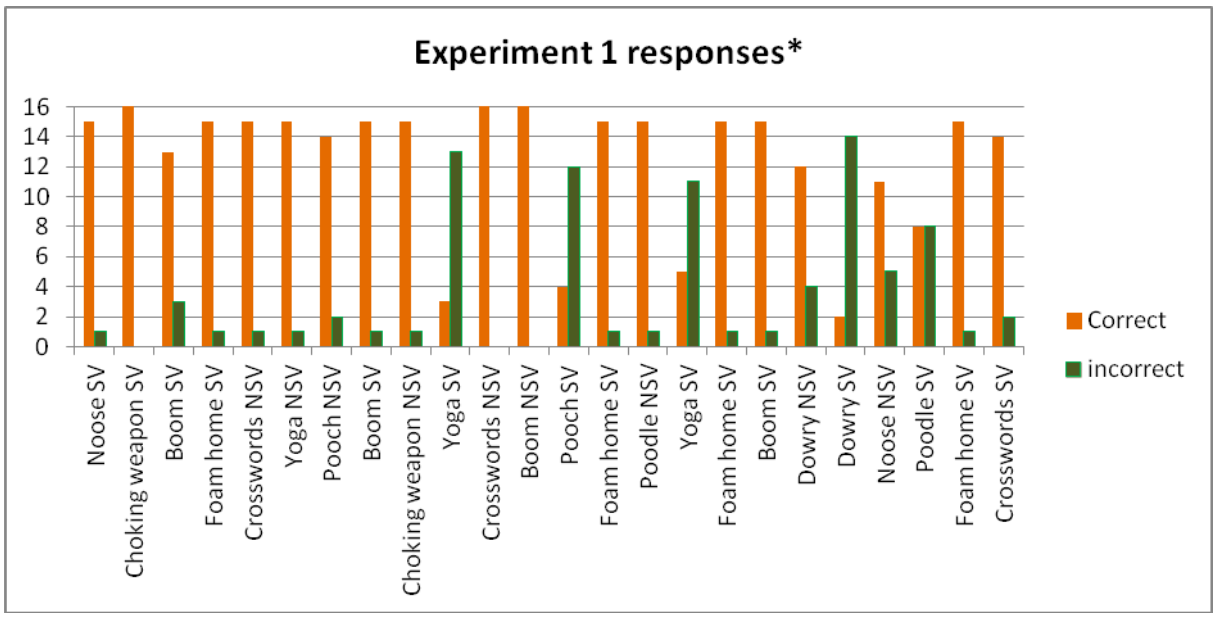


Figure 4: Responses to all the stimuli of experiment 1

It is possible to see that four stimuli in particular seemed to be difficult to classify for the participants. These are *Yoga* (SV), *Pooch* (SV), and *Dowry* (SV). These three words have in common the fact that they were uttered by a male, and that they present overall different prosodic features from the rest of the target words. In particular, all of them are marked by a low-falling pitch contour, and one of them (*Dowry*) is uttered with creaky voice. The speaker's gender can be excluded from being an influence on the listeners' perception, because other smiled target words uttered by males were recognized as such by the majority of listeners. What seems to change in these four words is that their prosodic characteristics resemble some of the features of read speech, which is what the non-smiling data set is made of. Figure 5 shows the pitch contours of the smiled and neutral versions of the three problematic words, in comparison with the smiled and neutral versions of three other target words that seemed not to cause any problems in the listeners (shown in Figure 6).

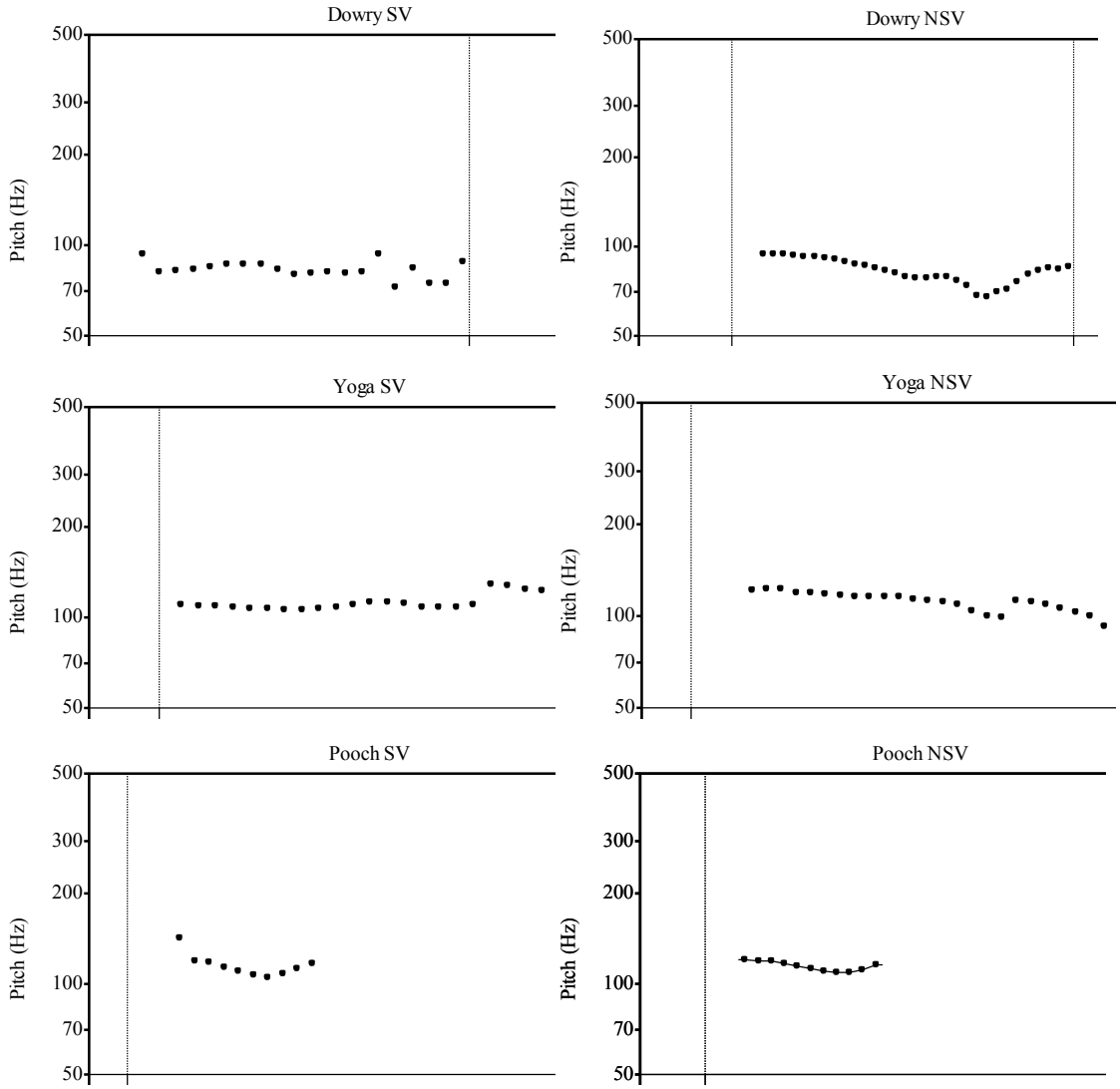


Figure 5: Pitch contours of three problematic words in experiment 1.

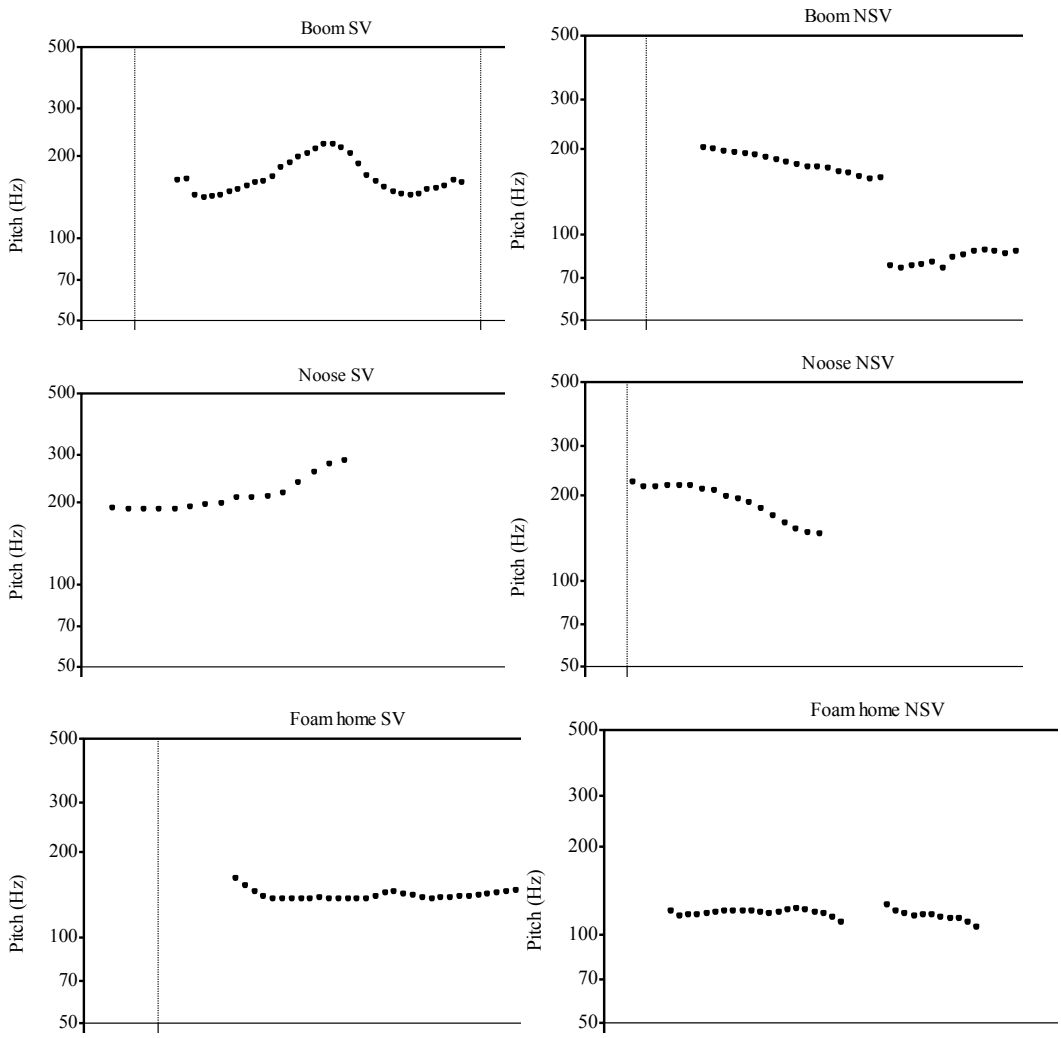


Figure 6: Pitch contours of three unproblematic words in experiment 1.

There is a growing amount of literature on the prosodic differences between spontaneous and read speech. For example, Laan and van Bergem (1993: 572) found that prosody plays a vital role in the differentiation between read and spontaneous speech: in their experiment, they artificially manipulated the f0 frequencies of read and unscripted utterances, so that one contained the frequency of the other, and found that listeners' performances significantly diminished. Similarly, the listeners' answers in the present experiment might have been biased by the fact that all the other words that they had classified as N presented similar prosodic characteristics, whereas the words that they classified as S tended to show a higher amount of variation in their suprasegmental features. This variation, according to Vaissière (2005: 252), seems to be associated with emotions such as happiness and pleasantness, which, in turn, tend to be associated with smiling. A low, monotonous fundamental frequency and slower speaking rate, on the other hand, tend to be associated with sadness and boredom (Vaissière 2005: 252; Lasarcyk and Trouvain 2008: 345). These results, however, partly contrast with Aubergé and Cathiard (2003: 95). In their data, what changed the most from spontaneous to (in their case) acted speech was amplitude, rather than f0. Their spontaneous data, however, was strictly controlled, and some of their speakers were professional actors. Also, their data was entirely in French, and there might be some language-specific differences in the use of intensity in the carrying of spoken emotions. Instead, Lasarcyk and Trouvain (2008: 347) found that the main cue in the recognition of lip spreading on single

vowels was f0; but they were aware that using single vowels would not provide complete results, and pointed out that future research should use longer utterances. They did not exclude the possibility that their listeners associated each stimulus with an emotion, but neither in their research nor in the present one were emotions mentioned to listeners in the perception experiments. They suggested that future research should use smaller f0 manipulations in the perception experiments, and, in a way, that is what has been done in the present research: f0 has not been manipulated at all, but it was originally different in the SV-NSV pairs.

All these considerations, however, are made on a small number of occurrences, and should be only intended as an attempt to explain the anomalies in the particular perception experiment carried out here.

Apart from these particular cases, the listeners' responses (summarized in Figure 7) show that they were better, on average, at recognizing Non-Smiling Voice than Smiling Voice. An independent samples t-test showed that this result is statistically significant (p=0.000).

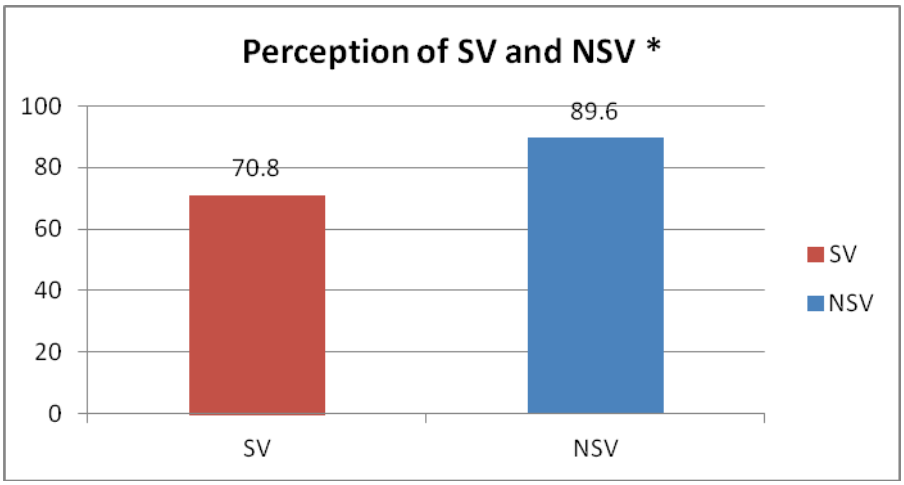


Figure 7: averages of correct identification of Smiling and Non-Smiling Voice.

As for possible gender differences, Figure 8 shows that females were generally better than males at recognizing both Smiling and Non-Smiling Voice. An independent sample t-test for equality of means, however, showed that these differences were not statistically significant ($p > 0.1$): there is no evidence that one gender performed significantly better than the other.

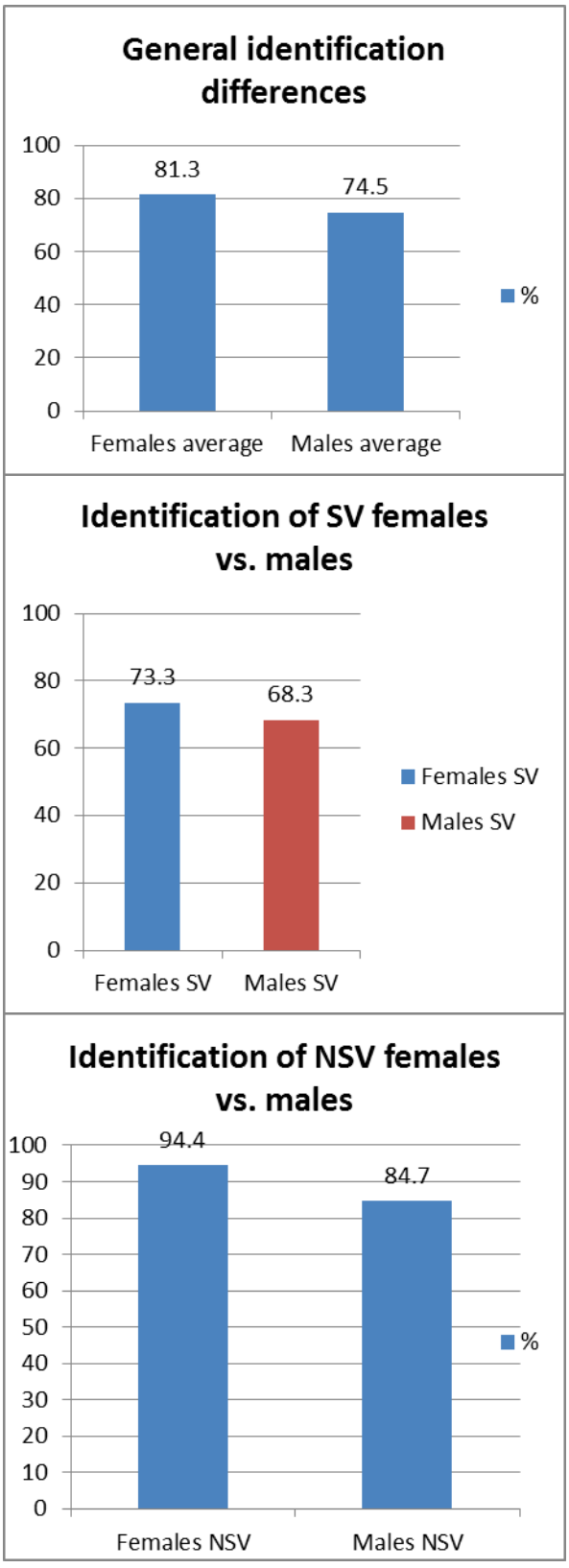


Figure 8: Gender differences in recognizing SV and NSV.

These results seem to be consistent with the findings of some previous studies on gender differences in perception. For example, Schirmer et al. (2002) found that female listeners performed better than males in perception experiments, even though in a second study (Schirmer et al 2005; Paulmann et al. 2008: 267) they found that gender disparities are eliminated if listeners are told to take into account prosody when doing the task. Similarly, an experiment involving nonsense words whose different prosodic characteristics carried different emotions (Pell et al. 2009: 430) proved that prosody is an essential cue for distinguishing emotions in speech.

In the present experiment, participants were only instructed to make a distinction between Smiling Voice and Non-Smiling Voice, and it is therefore possible that one of the cues that they were listening to was prosody. This would confirm the impression that the anomalies in the four words that were mistaken by the majority of listeners were due to a specific prosodic pattern. Listeners were asked to recognize smiles, and it is possible that they associated smiles with happiness, leaving the purely acoustic characteristics of the speech that they were hearing in the background.

3.3. Experiment 2

The second perception experiment, once confirmed that listeners can recognize Smiling and Non-Smiling Voice, seeks to determine at which point, in a continuum from smiling to neutral modality, such recognition happens.

3.3.1. Methodology

A native English speaker (female, aged 24) was audio recorded while reading six words. She was instructed to read each word twice, once with a straight face and once with stretched lips, but keeping the same intonation and, as far as possible, the same duration. This was necessary because of the configuration of Akustyk[2] , the programme that was used to synthetize a 5-step sound continuum from the "smiled" word to the neutral word. After many trials, it became evident that Akustyk could not create sound continua of files containing obstruent or voiceless sounds, or if the original starting and ending words had different intonation contours or very different durations. For these reasons, the words to be used had to contain voiced, non-obstruent sounds, and a similar prosodic pattern. The words that were chosen are: *Arrow, Loan, Mole* and *Wool*. Two more words (*Pool* and *Spoon*) were used as fillers. For this experiment, only words containing rounding were used because, since the formant frequencies show a wider variation from smiling to non-smiling modality in rounded words than in unrounded ones, it would have been easier for Akustyk to create intermediate steps that were as far as possible from each other.

Figure 9 shows the variations in the trajectories of the first three formants in the continuum of one of the target words. The three horizontal lines represent the formant trajectories, and the vertical columns of red dots represent the steps that Akustyk created from the formant frequencies of the start word to the formant frequency of the end word.

[2] Akustyk is an extension for Praat that allows to synthetize speech in a number of different ways, including creating a speech continuum. It was downloaded from http://bartus.org/akustyk in April 2013. The software is no longer available to download from the author's website, but it is free and open source, and it is possible to copy a version of it from the computer of someone who already has it.

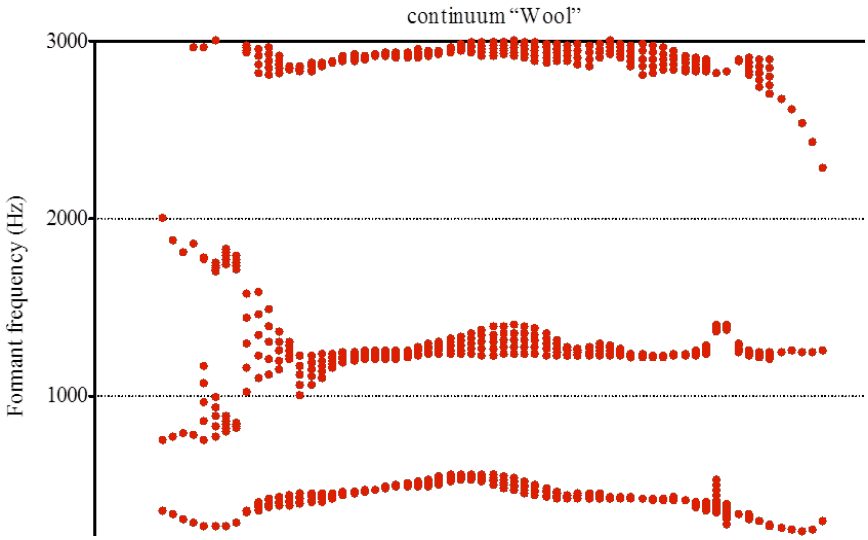


Figure 9: Continuum SV-NSV for the word W*ool*

16 native English speakers (8 males and 8 females), different from the participants in experiment 1, volunteered to take part in the experiment, which followed the same procedure as experiment 1.

3.3.2. Results and discussion

As it is possible to see from Figure 10, the experiment responses were far from uniform. In three out of four cases, the majority of listeners could at least identify the first step in the continuum as being smiled; but, for the case of *Mole*, only 3 listeners out of 16 (18.75%) correctly identified the first step as being smiled. As for the rest, the general tendency was that the majority of stimuli were identified as Non-Smiling. During the de-briefing, many subjects reported to the researcher that they had found it very difficult to distinguish any difference at all.

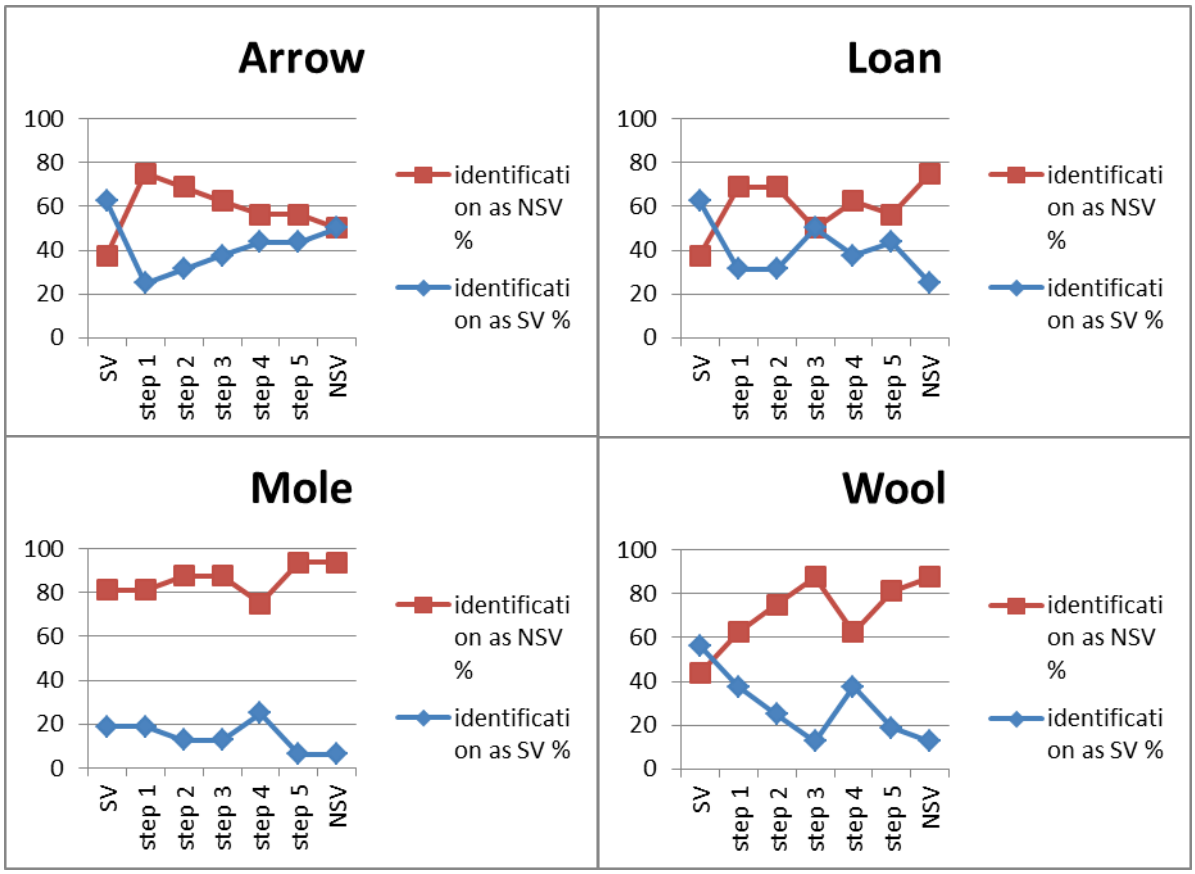


Figure 10: Trend answers for each of the target words of experiment 2.

These results could be due to the fact that the intermediate steps, which were artificially produced, were not realistic enough, or to the fact that the quality of the Akustyk's outputs was still too poor. Although the quality of the manipulation might be a good explanation for the difficulty of this experiment, it is also possible that listeners relied on prosodic cues to operate a distinction between SV and NSV. Since Akustyk was unable to create a continuum from two words with different intonations, in the final outputs only the formant frequencies were changed. Therefore, the intonation remained the same, since the speaker who was recorded uttered all her words with the same, low-falling intonation.

However, the stimuli used in this experiment were too few, and the quality of the synthesis was too poor, to allow to draw such a conclusion. More experiments are needed, with better synthesizer tools and more stimuli, before drawing a general conclusion. The present results contrast with Robson and MacKenzieBeck (1999), who found that listeners could recognize the effects of lip spreading on speech, and associate it with smiling. In their experiment, however, the stimuli were presented to the listener in an orderly succession, and they had to choose between pairs of stimuli, so that they were progressively trained in recognizing the differences. In the present experiment, instead, the stimuli order was randomized, and there were no cases of close steps being presented one after the other.

3.4. General discussion

Experiment 1 confirms that speakers can recognize Smiling Voice without being presented with visual stimuli, and supports the hypothesis that this recognition happens even when the audio stimuli comes from unscripted speech.

Experiment 2 does not present any significant result on the location of the recognition of Smiling Voice in an artificial continuum from Smiling Voice to Non-Smiling Voice. However, another interesting question is posed: is prosody essential in the perception of Smiling Voice? Future research will have to find an answer.

4. Conclusion

The present study examined some phonetic characteristics of the phenomenon of Smiling Voice.

It was found that, as predicted by the model of the vocal tract and by past literature, instances of Smiling Voice taken from spontaneous speech increase the frequencies of the first three formants, and that this increase is statistically significant. Differences were found among the individual formants: the most noticeable difference turned out to be in f3, and the least noticeable difference turned out to be in f2. Differences were also found in relation to the gender of the speaker, with the difference in the formant frequencies between Smiling and Non-Smiling Voice being higher in male speakers than in females.

To investigate the auditory correlations of Smiling Voice, two original perception experiments were carried out. Experiment 1 confirmed that listeners can distinguish words uttered in smiled and read modalities without many difficulties, and that, if difficulties arise, they are probably due to a lack of prosodic cues to differentiate the two modalities. The importance of prosody is stressed again in Experiment 2, where the responses indicated that the majority of stimuli were identified as non-smiled, even when they had originally been uttered with a smiling gesture. This suggests that prosody plays a vital role in the recognition of Smiling Voice, and that more accurate synthesizers (or experiments involving a manual synthesis of speech) should be used to investigate the matter further.

It is possible that the importance of prosody is due to the fact that prosody contributes to carrying emotional meaning in speech, and that listeners tend to associate smiling with an emotion. For the same reason, future research should also make sure that listeners do not make use of semantic cues for the recognition of Smiling Voice, e.g. by linking a particular stimulus to an emotion. This problem could be solved by using stimuli in smiling and neutral modalities with different degrees of prosodic variation coming from languages that the listener does not know. Another path to follow that the present research did not pursue was analysing data coming exclusively from spontaneous speech. In fact, although in this study it was possible to use spontaneous instances of Smiling Voice, the corresponding instances of Non-Smiling Voice were not spontaneous, as the speakers were reading from a list.

References

AUBERGÉ, V. AND CATHIARD, M. 2003. Can we hear the prosody of a smile? *Speech Communication 40*, pp. 87–97.

BESKOW, J., EDLUND, J., ELENIUS, K., HELLMER, K., HOUSE, D. AND STRÖMBERGSSON, S. 2009. Project presentation: Spontal - multimodal database of spontaneous dialog. *Proceedings of Fonetik 2009, the 12th Swedish Phonetics Conference. University of Stockholm, Sweden, 10-12 June 2009, Department of Linguistics, Stockholm University,* pp. 190-193.

BOERSMA, P. AND WEENINK, D. 2012. Praat: doing phonetics by computer [on line]. Version 5.3.53, available from http://www.praat.org/ [October 2012].

CLARK, J., YALLOP, C. AND FLETCHER, J. 2007. *An introduction to Phonetics and Phonology. Third edition.* Oxford: Blackwell.

DE GELDER, B. AND VROOMEN, J. 2000. The perception of emotions by ear and by eye. *Cognition and emotion 14 (3)*, pp. 289-311.

DRAHOTA, A., COSTAL, A. AND REDDY, V. 2008. The vocal communication of different kinds of smile. *Speech Communication 50*, pp. 278–287.

EISENBARTH, E. AND ALPERS, G. W. 2011. Happy Mouth and Sad Eyes: Scanning Emotional Facial Expressions. *Emotion 11 (4)*, pp. 860-865.

EKMAN, P. AND FRIESEN, W. V. 1982. Felt, False, and Miserable Smiles. *Journal of Nonverbal Behavior 6 (4)*, pp. 238-252.

ÉMOND, C. AND LAFOREST, M. 2013. Prosodic correlates of smiled speech. *Proceedings of Meetings on Acoustics, ICA 2013 Montreal, Montreal, Canada, 2 - 7 June 2013.*

FAGEL, S. 2009. Effects of Smiled Speech on Lips, Larynx and Acoustics. *Proceedings of AVSP 2009, Norwich, Sept 10th-13th*, pp. 18-21.

FANT, G. 1960. *Acoustic Theory of Speech Production*. Mouton, The Hague.

HAWK, S. T., FISCHER, A. H. AND VAN KLEEF, G. A. 2012. Face the Noise: Embodied Responses to Nonverbal Vocalizations of Discrete Emotions. *Journal of Personality and Social Psychology 102 (4)*, pp. 796-814.

HUBER, J. E., STATHOPOULOS, E. T., CURIONE, G. M., ASH, T. A. AND JOHNSON, K. 1999. Formants of children, women, and men: The effects of vocal intensity variation. *Journal of the Acoustical Society of America 106 (3)*, pp. 1532-1542.

KOHLER, K. J. 2007. 'Speech-smile', 'speech-laugh', 'laughter' and their sequencing in dialogic interaction. *Proceedings of the Interdisciplinary Workshop on the Phonetics of Laughter, Saarbrücken, 4-5 August 2007*, pp. 21-26.

LAAN, G.P.M. AND VAN BERGEM, D. R. 1993. The contribution of pitch contour, phoneme durations and spectral features to the character of spontaneous and read aloud speech. *Proceedings of Eurospeech '93, Berlin*, pp. 569-572.

LAMBRECHT, L., KREIFELTS, B. AND WILDGRUBER, D. 2012. Age-related decrease in recognition of emotional facial and prosodic expressions. *Emotion 12 (3),* pp. 529-539.

LASARCYK, E. AND TROUVAIN, J. 2008. Spread Lips + Raised Larynx + Higher F0 = Smiled Speech? - An Articulatory Synthesis Approach. *Proceedings of the 8th International Seminar on Speech Production, Strasbourg, France, 8-12 December 2008*, pp. 345-348.

LAUKKA, P. 2005. Categorical Perception of Vocal Emotion Expressions. *Emotion 5 (3),* pp. 277-295.

MCGURK, H. AND MACDONALD, J. 1976. Hearing lips and seeing voices. *Nature 264*, pp. 746-748.

MEHU, M. AND DUNBAR, R. I. M. 2007. Relationship between Smiling and Laughter in Humans (Homo sapiens): Testing the Power Asymmetry Hypothesis. *Folia Primatol 79*, pp. 269-280.

PAULMANN, S., PELL, M. D. AND KOTZ, S. A. 2008. How aging affects the recognition of emotional speech. *Brain and Language 104*, pp. 262-269.

PELL, M. D., PAULMANN, S., DARA, C., ALASSERI, A. AND KOTZ, S. A. 2009. Factors in the recognition of vocally expressed emotions: a comparison of four languages. *Journal of Phonetics 37*, pp. 417-435.

ROBSON, J. AND MACKENZIEBECK, J. 1999. Hearing smiles - perceptual, acoustic and production aspects of labial spreading. *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco, 1-7 August 1999*, pp. 219-222.

SCHIRMER, A., KOTZ, S. A. AND FRIEDERICI, A. D. 2002. Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research 14*, pp. 228-233.

SCHIRMER, A., KOTZ, S. A. AND FRIEDERICI, A. D. 2005. On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research 24*, pp. 442-452.

SHOR, R. E. 1978. The production and judgment of smile magnitude. *Journal of General Psychology 98*, pp. 79-96.

TARTTER, V. C. 1980. Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics 27 (1),* pp. 24-27.

TARTTER, V. C. AND BRAUN, D. 1994. Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America 96 (4),* pp. 2101-2107.

TRAÜNMULLER, H. 1990. Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America 88 (1),* pp. 97-100.

VAISSIÈRE, J. 2005. Perception of Intonation. In PISONI, D. B. AND REMEZ, R. E. (eds.), *The Handbook of Speech Perception*. Oxford: Blackwell Publishing, pp. 236-263.

VAN HOOFF, J. A. R. A. M. 1975. A Comparative Approach to the Phylogeny of Laughter and Smiling. In HINDE, R. A. (ed.) *Non-Verbal Communication*. Cambridge: University Press, pp. 209-241.

VAN STRIEN, J. W. AND VAN BEEK, S. 2000. Ratings of Emotion in Laterally Presented Faces: Sex and Handedness Effects. *Brain and Cognition 44*, pp. 645–652.

WELLS J. C. 2006. *English intonation: an introduction*. Chapter 1, Cambridge: CUP, pp. 1-14.

Ilaria Torre
Department of Language and Linguistic Science
University of York
Heslington
York
Email: ilaria.torre11@gmail.com

BRADFORD PANJABI-ENGLISH: THE REALISATION OF FACE AND GOAT

JESSICA WORMALD

University of York

Abstract

FACE and GOAT realisations from six female Bradford Panjabi-English speakers were observed to follow the characteristic pattern of other Asian- (and multicultural-) Englishes with significantly lower F1s than three Bradford Anglo-English females. A qualitative distinction with the closer KIT and FOOT vowels is maintained, with younger PE speakers showing a greater degree of separation (more open FACE and GOAT). Both groups demonstrate variability in the degree of separation between the two vowel pairs (FACE and KIT; GOAT and FOOT). The current paper considers whether closer FACE and GOAT realisations provide evidence for transfer or innovation within Panjabi-English. Closer realisations are considered to index a non-Anglo ethnic identity, as suggested in other studies into Multicultural-Englishes, with the relationship to KIT and FOOT working in a complex way to index a further local identity.

1. Introduction

Midpoint F1 values for six female Panjabi-English (PE) speakers will be compared with three female Bradford 'Anglo' speakers to determine if PE realisations are closer. Female PE reading passage data from fieldwork undertaken in Bradford will be considered and compared with female 'Anglo' word list data from Watt and Tillotson (2001). Further, speakers' FACE and GOAT realisations will be compared to KIT and FOOT. These are traditionally closer than monophthongal FACE and GOAT so it might be expected that a qualitative difference between the two pairs may not be present in PE if FACE and GOAT are considerably closer.

PE will be defined with a brief discussion of the PE speaking community in Bradford. Consideration of the possible realisations of these variants in Panjabi and Bradford 'Anglo-English' (AE) will be followed by a review of previous work into PE relating to FACE and GOAT. Processes involved in dialect contact will be considered as a way to explain the patterns observed. Whether the features are evidence of transfer from Panjabi or innovation within PE will also be explored.

2. Background

2.1. Panjabi-English

Panjabi-English (PE) is the term used to refer to the native-English variety spoken by second- and future-generation speakers with Panjabi language heritage. 'Panjabi language heritage' refers to speakers who have at least one Panjabi speaking parent who is a first-generation immigrant from the Panjab region or another Panjabi speaking location. PE is spoken throughout the UK and has been explored in a number of locations.

2.2. Bradford

Bradford is a city in West Yorkshire with a history of in migration (Fieldhouse 1981). The number of individuals recording Panjabi as their main language in the 2011 census is 4%,

 ISSN 1758-0315

greater than the national average of 0.5% (ONS 2013:11; QS204EW). Panjabi is the third most widely spoken main language in England and Wales, following English (92.3%) and Polish (1%). Census data does not represent those who speak Panjabi but may not consider it their 'main' language[1]. The online ethnologue estimates the number of Panjabi speakers in the UK at 594,000 (Lewis et al. 2013) and Reynolds and Verma (2007: 307) comment that speakers of the main Indic languages in Britain (Panjabi, Hindi, Urdu, Gujarati and Bangla) outnumber speakers of indigenous minority languages.

According to the 2011 census, Bradford's Indian and Pakistani population is above average for England and Wales at 23%, the national average at 4.5% (ONS 2012a, 2012b). Bradford was therefore considered to have a potentially sizeable population of PE speakers, and also a relatively dichotomous population, with no other language variety or ethnic minority forming a large proportion of the population.

3. The input

3.1. Panjabi

The Panjab region is a geographical location in North West India and Northern Pakistan. Panjabi is the language of the Panjab region (Bhatia 1993: xxv) and is one of several national languages of India and Pakistan. There is a great deal of linguistic diversity in Panjabi, with some dialects being considerably different to the commonly referred to Modern Standard Panjabi (MSP) as spoken in India (Shackle 2003: 585).

Consequently, knowing the variety of Panjabi spoken by PE speakers' parents, and from which geographical area they come is important. PE speakers completed a questionnaire detailing native and other languages spoken. All were second-generation native English speakers and all listed Panjabi and Urdu as either 'native' or 'other languages'. Two of the six speakers (Zayna and Sadiyah) also listed Hindko and Pushto as native or other languages spoken. All of the PE speakers have family originating in Pakistan. Although the questionnaire did not ask participants to further specify the location, some commented that their families were from the Mirpur region, like most of Bradford's Pakistani community.

Lothers and Lothers (2012) comment that the majority of Pakistani immigrants in the UK speak a variety of Panjabi they term 'Mirpuri Pahari'. This is used as a cover term to include the varieties Pothwari, Pihari and Mirpuri, which are names given to varieties of Panjabi spoken in the Mirpur. Heselwood and McChrystal (1999) comment that their speakers from Manningham in Bradford originate from the Mirpur. Consequently, along with Hindko and Pushto, if there is literature available to suggest that these varieties differ from MSP it will be discussed.

All diphthongs in MSP are rising with a central and peripheral vowel (Shackle 2003: 588). The closest to the RP /eɪ/ diphthong is a rising /əi/. Panjabi also has a front mid /e/ and a central high-mid /ɪ/ (Shackle 2003; Bhatia 1993; Tolstaya 1981). This is mirrored at the back of the vowel space, with the diphthong /əu/ and the monophthongs /o/ and /ʊ/.

Western Pahari and Lahnda varieties have short counterparts for the long and peripheral /e/ and /o/ (Zograph 1982). Shackle (1983) comments in a review of Zograph's work that, 'caution is therefore required in using the book as a guide of linguistic facts' (1983: 372). It is

[1] The 2011 census asked 'What is your main language' with answering options 'English' or 'Other'. If 'Other', individuals were required to write in their 'main' language. 'Main language' was not further specified so it is not known how it was intended or interpreted.

difficult to know how much value to give to the statement made by Zograph. Bhardwaj (2012) comments in a book for learners of Panjabi, that the long vowels *e* and *o* should be pronounced as 'the Scottish pronunciation of g*a*te [and] g*o*' (2012: 7-8). Scottish monophthongal FACE and GOAT are traditionally close (Stuart-Smith 1999), this might suggest Panjabi /e/ and /o/ are also relatively close.

3.2. Bradford Anglo-English

This refers to the variety of English within Bradford as spoken by white British monolingual speakers. Petyt (1985) identified two lexical sets /ɛɪ/ and /e:/, which include words from the modern FACE set, with the incidence of /ɛɪ/ decreasing. The same is noted for the back of the vowel space, with /ɔʊ/ and /o:/ making up two different sets for modern day GOAT with decreasing /ɔʊ/. Petyt (1985) suggests that both of these pairs may be undergoing mergers and this appears to have taken place, with modern day Bradford AE having FACE and GOAT sets with similar incidence, but different phonetic realisations to RP. Petyt reports the short /ɪ/ and /ʊ/ of the modern day KIT and FOOT sets to be present in the Bradford vowel system (1985: 97).

Bradford monophthongal FACE and GOAT are characterised by Hughes et al. (2012), as [e:] and [o:], respectively. Diphthongal realisations may still occur. With FACE, this could be with words containing *'eigh'* in the spelling. For GOAT, this could be with words including 'ow' or *'ou'*. KIT and FOOT are not reported to vary from the close front /ɪ/ and the close centralised /ʊ/ of RP.

Watt and Tillotson (2001) considered fronting of monophthongal GOAT in Bradford. They analysed wordlist tokens from seven speakers (5 females, 2 males) and found GOAT fronting from more open [ɔ:] to centralised [ɵ:] to be more advanced in the speech of younger, particularly female, speakers.

3.3. Summary

The speakers are exposed to numerous inputs. The L2 English as spoken by first-generation speakers would constitute a third input. No data has been collected to be representative of this group. Consequently, it is not known how much of an influence this may have upon PE.

4. Multicultural-Englishes

An increasing amount of research is now being undertaken into the variation of multicultural-Englishes. Multicultural London English (MLE) (Cheshire et al. 2011; Torgersen & Szakay 2011; Cheshire et al. 2008; Kerswill et al. 2008) refers to the new and developing variety spoken in London which contains influences from many of the languages spoken in the capital today. Multicultural Manchester English (MME) has also been considered, with similar features being found to be characteristic of this variety, despite geographical distance from London (Drummond 2013a).

Increasingly, research is considering Asian-Englishes, with work in Leicester considering the influence of Gujarati (Rathore 2011a, 2011b), and in London exploring the prosody of Gujarati-English speakers (Zipp 2013). Harris (2006) introduced the notion of 'Brasian' with his work in a West London school referring to the simultaneous presence of British and Asian identities.

4.1 PE

PE is being considered in a number of UK locations including Sheffield (Kirkham 2011), Glasgow (Lambert et al. 2007), London (Sharma & Sankaran 2011), Bradford (Day 2009), and Edinburgh (Verma & Firth 1995). Numerous features have been observed to vary in different ways, with common variability occurring with /l/ realisation (Lambert et al. 2007), retroflexion of /t/, /d/ and /n/ (Heselwood & McChrystal 2000), and rhoticity (Hirson & Sohail 2007). Focus will be placed on those studies which have considered variation in FACE and GOAT with further research on the variety not being considered.

Stuart-Smith et al. (2011) observed closer and fronter FACE and GOAT with Glasgow Asian speakers compared to monolingual Glasgow speakers. Using word list and reading passage data, speech from five bilingual (English dominant & Panjabi) and two monolingual Glaswegian males was compared. Clearer separation between the groups was observed with the GOAT vowel. One Asian speaker patterned with the monolingual speakers, he was also found to have the lowest proportion of Asian friends (Stuart-Smith et al. 2011: 9).

Further analysis in Stuart-Smith et al. (2011) used a dataset of ethnographic interviews from six trilingual (English dominant, Panjabi & Urdu) eighteen year-old girls. They were considered representative of three Communities of Practice (CofP). The 'Moderns' were characterised by their identification with Western cultural practices and norms, including wearing make-up, dating, and educational aspirations. The 'Conservative' group were sub-divided into 'Conservative-Religious' and 'Conservative-Cultural' groups. The 'Conservative-Religious' speaker represented a group who identified closely with traditional Muslim values such as marrying young and educational equality. The 'Conservative-Cultural' speaker represented a group for whom traditional Pakistani values, including dressing modestly with a headscarf, and favouring marriage over relationships, were important. The final group, the 'Inbetweens', fell somewhere between the Moderns and Conservatives.

For FACE, individual speaker was found to be a stronger determining factor than CofP. The Moderns showed fronter realisations than the Conservatives, with the Inbetweens somewhere in between. The Conservative-Religious speaker had the closest realisation, differing from the Conservative-Cultural female who had a more open realisation. For GOAT, CofP was found to be a stronger determining factor. The height of this vowel was similar for all speakers, with the exception of the Conservative-Religious speaker who had a much closer realisation. Her realisation was also the most fronted, but was similar to one of the Moderns. As with the males, Stuart-Smith et al. found greater separation with GOAT. They suggest this may be indicative of greater weight carried by GOAT in its relationship to ethnicity and identity construction, although they highlight that this is based on a small amount of data. They also introduce the notion of 'Glaswasian' relating to Harris' (2006) 'Brasian', highlighting the complex identities of speakers indexing both local and ethnic identity.

In their work with bilingual children, Verma and Firth (1995) comment that speakers in West Yorkshire have adopted local monophthongal FACE and GOAT but do not further discuss the quality. Although monophthongal realisation here is attributed to an adoption of a local feature, qualitative variation may also be present as with Glaswasian speakers who demonstrate closer realisations (Stuart-Smith et al. 2011).

Sharma (2011) reports fieldwork from Southall, London. Four second-generation PE speakers took part in ethnographic interviews and self-recorded in different contexts with different interlocutors. Two age-groups are represented, with both males and females (one speaker per cell). The presence of monophthongal FACE or GOAT was considered to be an Indian feature,

differing from the diphthongal 'Anglo' realisation. Sharma (2011) does not provide further information on monophthongal quality so it is difficult to determine whether these are as close as monophthongal variants in Glaswasian. The PE speakers all used monophthongal FACE and GOAT at least some of the time with some interlocutors. The use of monophthongal Indian variants occurred mainly with Asian interlocutors. However, this varied both by speaker and context. There were also interlocutors with whom the speakers used no monophthongal FACE and GOAT. The monophthongs occur variably with 'Anglo' diphthongs, the standard /eɪ/ and /əʊ/ or 'Cockney diphthongal variants' (Sharma 2011: 471).

5. Dialect Contact

This refers to the process by which different varieties come into contact with new dialect formation being a potential result. The interaction of accommodatory effects and individual and group identities defines variants, contributing to the construction of a new variety. The linguistic situation in Bradford is one of contact, with the interaction of two English varieties (Bradford AE and the L2 English of first-generation immigrants) and the influence of Panjabi, resulting in PE.

First considered in depth by Trudgill (1986), he discussed the linguistic patterns observable when dialects come into contact and interact. Mixed intermediate forms arise as a result of many individual accommodation events. Trudgill introduced three processes which occur allowing for restructuring to take place before the new variety is formed. Firstly, levelling, which leads to a decline in the number of marked variants. Reallocation occurs when more than one variant remains, the levelling process being 'incomplete'. Variation between them instead occurring socially, stylistically or, if phonetic features, allophonically. Finally, simplification is discussed which refers to an increase in monomorphemic regularity in new dialect formation.

Watt (2002) observed levelling in Tyneside English. Monophthongal FACE [e:] and GOAT [o:] realisations had been levelled to a regional standard, over more local [ɪə] and [ʊə]. Reallocation was found by Dyer (2002) who illustrated that Scottish monophthongal GOAT had been reallocated for second- and third-generations to index a local Corby identity, whereas for first-generation speakers it formed part of their Scottish identity. Allomorphic simplification was observed by Britain (2002) in Fenland English, with a reduction in the number of past tense *be* morphemes from three to two. The combination of these processes can result in a newly formed dialect or 'koine'.

More recently, Trudgill (2008a) questioned the role of identity in new dialect formation, claiming sociolinguists too often rely on identity as an explanation for variation. He argued that the automaticity of accommodation meant that any influence of collective identity came later, once the new dialect had been established. The development of post-colonial varieties of English was presented as evidence for this. Trudgill claimed that a shared national identity would not have been present amongst new settlers therefore could not have contributed. Instead mechanical explanations of automatic accommodation would have led to the processes described in new dialect formation.

The publication received a number of responses generally agreeing with Trudgill's claim of the automaticity of accommodation (Coupland 2008; Mufwene 2008; Tuten 2008). However, many questioned Trudgill's dismissal of identity in its entirety and his assumption that it is only relevant on a national scale, highlighting the complexity and multiplicity of identity (Bauer 2008; Holmes & Kerswill 2008; Schneider 2008; Tuten 2008). Some discussed the relationship between accommodation and identity and their dependence on one another,

particularly amongst adults who cannot abandon identity (Coupland 2008; Holmes & Kerswill 2008; Schneider 2008; Tuten 2008). Further, Holmes and Kerswill (2008) comment that identity could be used to explain the choice of one variant over another when numerically equivalent options exist.

Trudgill's rejoinder (2008b) aimed to answer the queries raised in responses and further justify his position. He discussed the role children of the second- and future-generations play in new dialect formation, claiming children's full accommodation to the local speech community is a 'universal human behaviour' (2008b: 277). A lack of socialisation allowing them to work forming the dialect independent of identity. This is also considered in Trudgill (2004) where he claims third-generation[2] children are the ones who 'do the generating' (2004: 27).

Previous work into PE and other multicultural-Englishes has consistently observed the effects identity can have upon the developing variety. The following comment highlights this,

> It seems that certain features originally derived from language interference are now being *actively* deployed as English accent features by second- and later-generation speakers, though with rather different realisations and distributions from those expected in the original language.
>
> (Lambert et al. 2007: 1512, emphasis added).

As mentioned in responses to Trudgill, the absence of a shared national identity does not preclude the absence of a collective identity, nor does it take into account individual identities and how the interaction of large numbers of individuals may affect the variety. The above quote highlights how features are used creatively in new dialect formation as a reflection of speakers' identities.

6. Innovation or transfer?

Determining whether variability from AE provides evidence for transfer from Panjabi, or innovation (where realisations cannot have come from Panjabi or English) is another aim of the current paper. PE is a developing native English variety and innovations are expected. The adoption of features from the heritage variety through being part of speakers' extended linguistic repertoire, or 'feature pool' (Cheshire et al. 2011) is also expected.

Providing conclusive evidence for transfer based on qualitative vowel realisations could be difficult given the linguistic variability within Panjabi. Shackle comments that in Panjabi, 'Vowel length is taken to be phonetically less significant than vowel quality' (2003: 587). Transfer then, may be reflected by the retention of a qualitative distinction between vowels. If FACE and GOAT are found to be closer than those of the AE speakers their patterning with KIT and FOOT could be of interest. If the difference between FACE & KIT and GOAT & FOOT is one of quality, this could be evidence for transfer, with the Panjabi rule of qualitative over quantitative difference being employed. However, if the two vowel pairs are found not to be distinguished qualitatively, this could be evidence for innovation, with a new pattern contrary to AE and Panjabi being developed.

[2] Trudgill (2004) refers to these speakers as 'second-generation'; the second-generation of speakers born within the new community. For purposes of continuity they are referred to here as third-generation. Throughout this work 'first-generation' refers to the original members who migrated into the community.

7. Research Questions

The following questions will be addressed,

- Do PE speakers have closer monophthongal FACE and GOAT than AE speakers?
- How do FACE and GOAT pattern with KIT and FOOT?
- Are there differences between two second-generation PE age-groups?
- Can variation from AE be considered evidence of transfer or innovation?

8. Methods

Results from six PE and three AE females are reported. Five PE speakers live in the north west of Bradford. One speaker (Shelly) lives two miles north in Shipley but works and interacts with the other participants. PE speakers represent two age-groups and are all second-generation. The AE females included are from Watt and Tillotson (2001). Table 1 contains speakers' names and ages. PE speakers are grouped into categories defined for a larger project. AE speakers do not fit into these so ages are listed. Lisa and Christine were considered with the younger (18-25) speakers and Peggy with the older (35-45) speakers.

Speaker (PE)	**Age-Group**
Shelly	18-25
Zayna	18-25
Maysan	18-25
Afsana	35-45
Sumra	35-45
Sadiyah	35-45
Speaker (AE)	**Age**
Lisa	17
Christine	27
Peggy	55

Table 1. Speakers included in the study

PE speakers were recorded reading 'Fern's star turn', this took three to four minutes to complete and provided around 15-20 tokens per vowel per speaker. This data considered forms part of a larger data set including paired conversations and spot-the-difference tasks. The passage provides direct comparability across speakers with no style or topic variation.

PE Speakers were recorded with a Zoom H4n recorder and Beyerdynamic TG H54c neck worn microphones. The zoom recorder recorded at a 16 bit 44.1kHz sampling rate. Use of the neck worn microphone ensured speakers were recorded at a high quality with a consistent distance between the microphone and speaker's lips. Many of the women wore headscarves as a reflection of their Muslim faith. In these cases, participants were informed they did not have to wear the headset. These women held the microphone rather than wore it, or went to creative lengths to ensure they could be heard without having to hold the microphone during the interview.

AE speakers read wordlists and short phrases (Watt & Tillotson 2001: 302), providing around 10 tokens for FACE, KIT and FOOT and 15-20 for GOAT. AE speakers were recorded on audio cassette with a Sony WM-D6C Professional Walkman using a Sony directional stereo microphone (Watt & Tillotson 2001: 277).

9. Analysis

Midpoint F1 measurements were taken for FACE, KIT, GOAT and FOOT. Measurements were also taken for GOOSE, LOT and THOUGHT. These are being considered as potentially characteristic in PE. Although not considered further here, they provide points of reference in the vowel spaces. Vowels were marked out in Praat (version 5.3.45) measuring from the onset to offset of periodicity in the waveform and the offset of F2 energy in the spectrogram. Vowels were marked out in a text grid and the formant tracker was checked to ensure the red dots aligned with the formant bands in the spectrogram. If formants were not accurately identified by the formant tracker these were entered manually or omitted.

A formant dynamic script was used to extract midpoint F1 measurements. The script measures F1, F2 and F3 at nine equal intervals throughout the duration of the vowel. Only midpoint F1 will be discussed.

10. Results

10.1. PE speakers

Shapiro-Wilks tests were run for individual speakers' FACE and KIT. If distributions were normal (non-significant result), independent t-tests were carried out to observe within-speaker variation of midpoint F1. If the results were significant (non-normal distribution), a non-parametric unpaired Wilcoxon test was undertaken.

All but one speaker had significantly different midpoint F1 values for FACE and KIT, p-values (all independent t-tests) are listed in Table 2.

Younger		**Older**	
Shelly	<.0001***	Afsana	<.0001***
Zayna	<.0001***	Sumra	=.0006**
Maysan	<.0001***	Sadiyah	=.277

Table 2. Significance values for within-speaker difference between FACE and KIT

Although both older and younger speakers pattern similarly with significantly different midpoint F1s for FACE and KIT, younger speakers appear to have a greater degree of separation. Mixed effect linear regression in R confirmed this. Using the lmer() function in the lme4 package, mixed effects models were created with fixed effects of age group and lexical set included. Speaker and word were included as random effects. Markov-Chain Monte-Carlo (MCMC) simulations were then run using pvals.fnc() to determine a p-value (Baayen 2008: 248).

A t-value of over 2 in the lmer() output determines significance at the 5% level (Baayen 2008: 248). Across speakers, FACE is significantly different to KIT, with KIT having a significantly lower midpoint F1 (t=-2.28; p=.004). A significant interaction between age

group and lexical set is also observed (t=-4.84; p=.0002). Illustrated in Figure 1 this shows the older speakers on the left ("O") and the younger speakers on the right ("Y"). The degree of separation between vowels is greater for the younger speakers than the older speakers.

Figure 1. Degree of separation between FACE and KIT. Red circles = KIT; Black circles = FACE

Shapiro-Wilks tests were also run for individual speakers' GOAT and FOOT. A combination of unpaired Wilcoxon and independent t-tests were then carried out. All but one of the speakers showed significantly different midpoint F1 in GOAT and FOOT, with younger speakers appearing to have increased separation. Table 3 illustrates the pattern.

Younger		**Older**	
Shelly	<.0001*** independent t-test	Afsana	=.0014** independent t-test
Zayna	=.015* independent t-test	Sumra	=.001** independent t-test
Maysan	<.0001*** unpaired Wilcoxon	Sadiyah	=.92 unpaired Wilcoxon

Table 3. Significance values for within-speaker differences between GOAT and FOOT

Mixed effects regression was then undertaken in R, followed by MCMC simulations. Mirroring the front of the vowel space, a significant difference between GOAT and FOOT is present with the MCMC simulations (t=1.88; p=.019). Further, there is a significant interaction between age group and lexical set (t=2.9; p=.005). This is illustrated in Figure 2 which highlights the increased separation between GOAT and FOOT for younger speakers.

Figure 2. Degree of separation between GOAT and FOOT. Red circles = FOOT; Black circles = GOAT

Plots in Figures 3 through 8 illustrate the within-speaker variation reflected by the statistical results for the PE vowels.

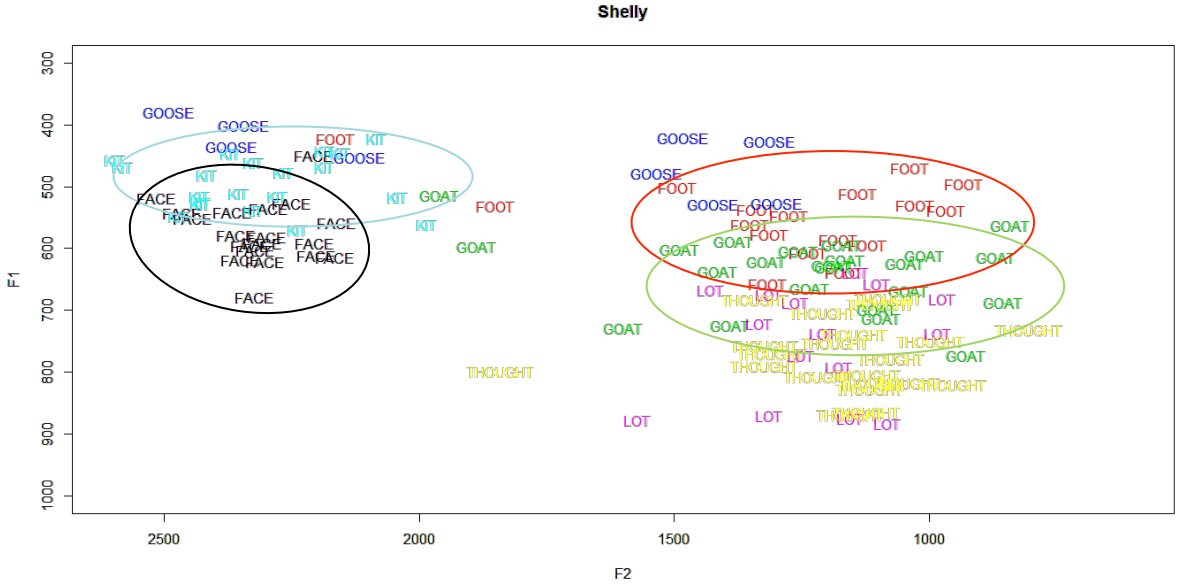


Figure 3. Vowel plot for Shelly.

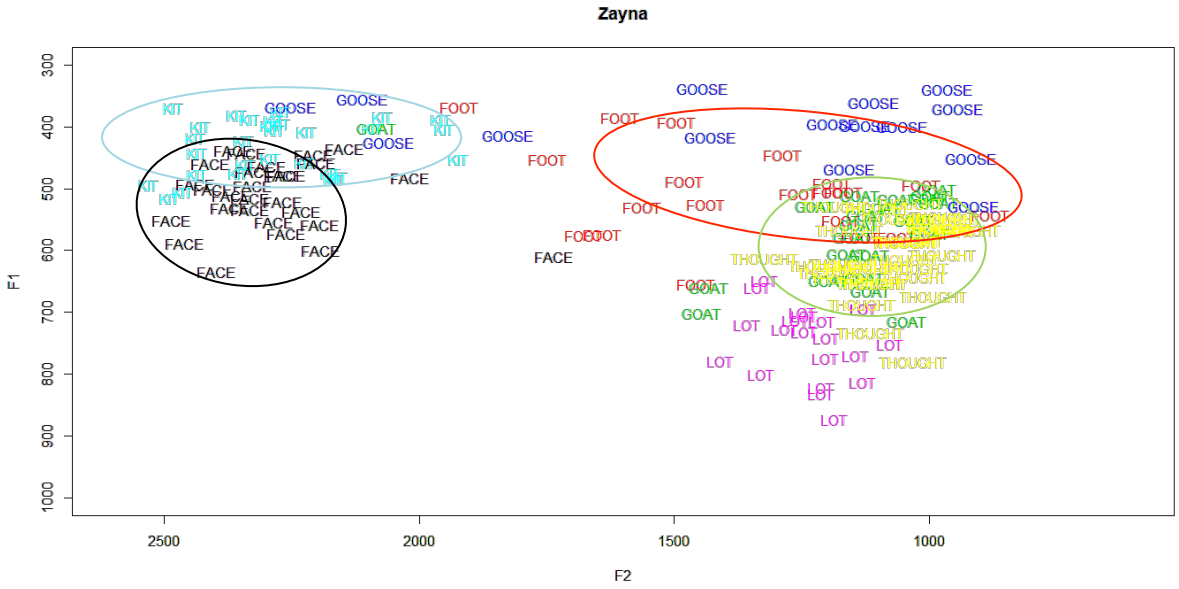


Figure 4. Vowel plot for Zayna.

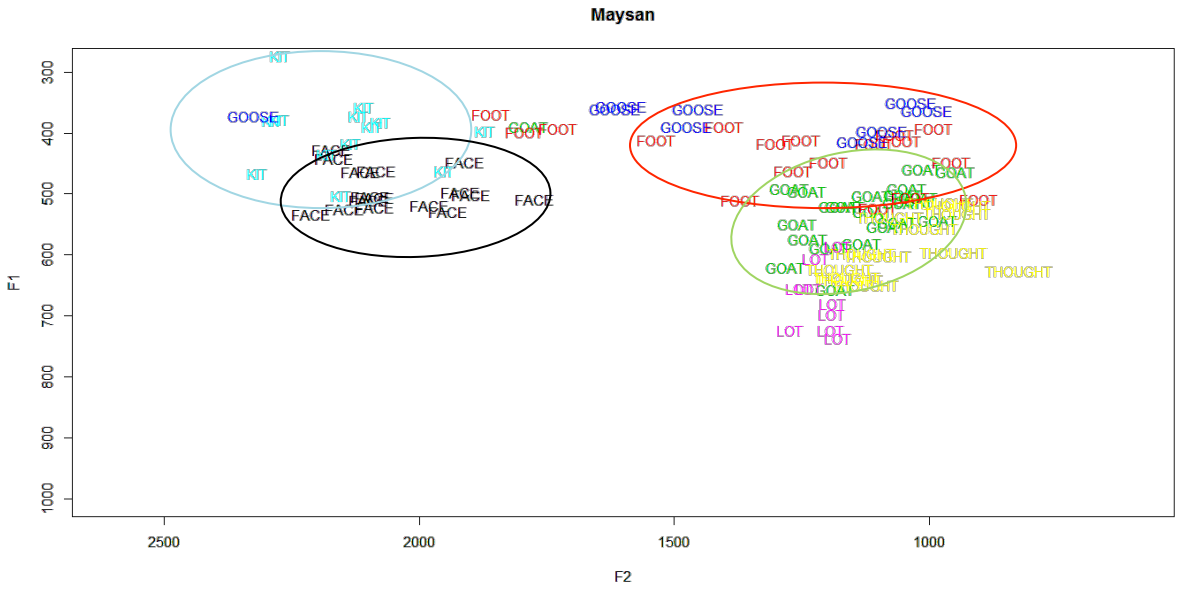


Figure 5. Vowel plot for Maysan.

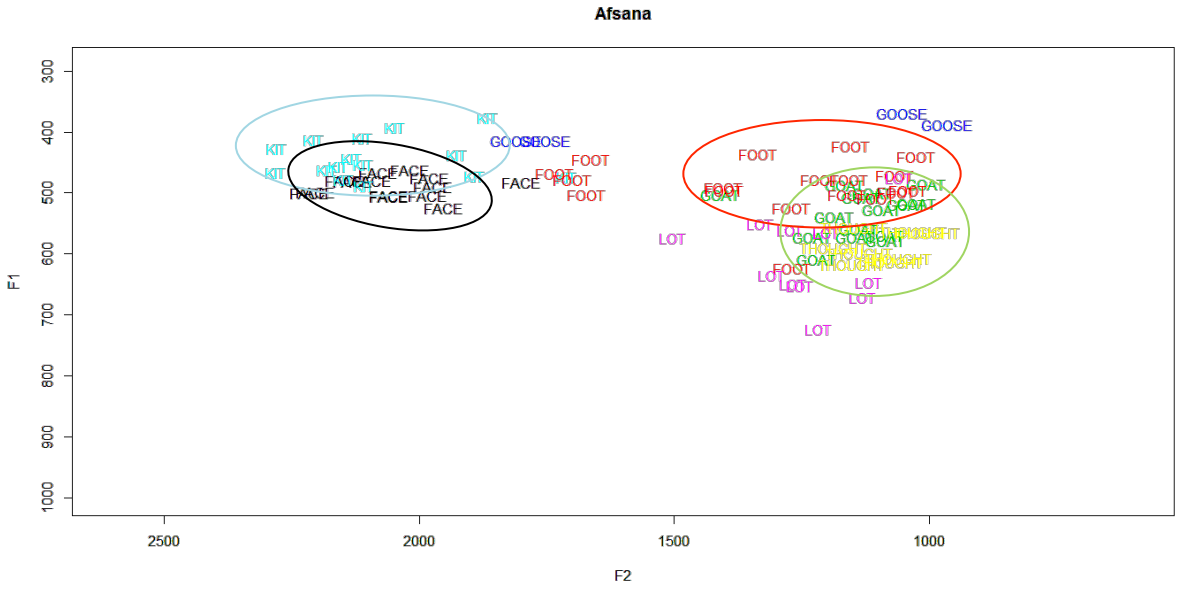


Figure 6. Vowel plot for Afsana.

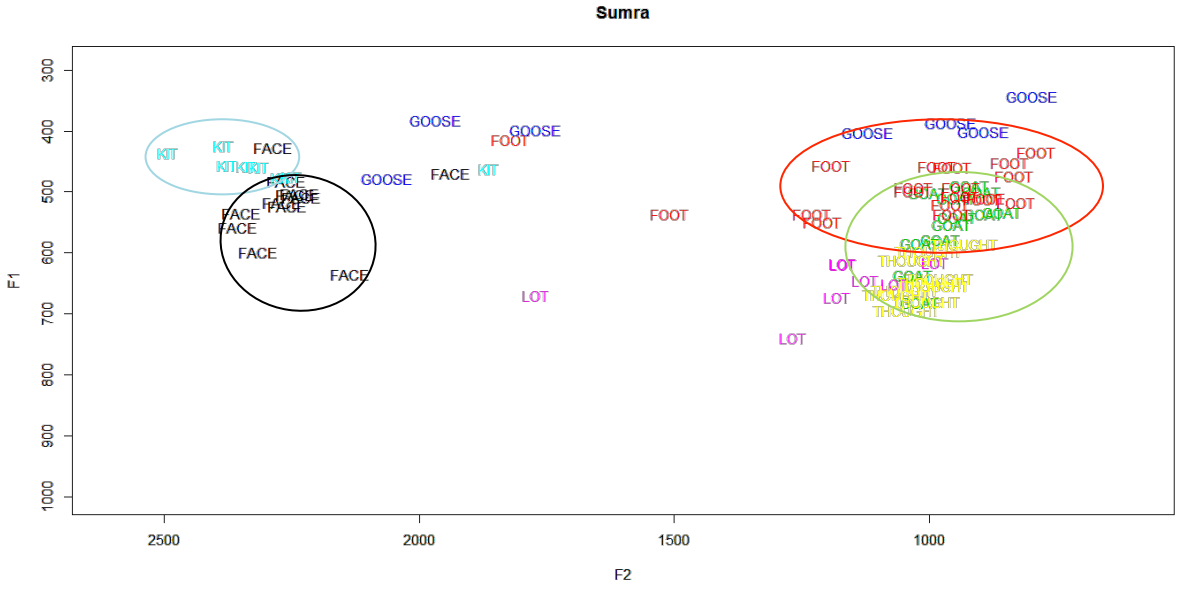


Figure 7. Vowels plot for Sumra.

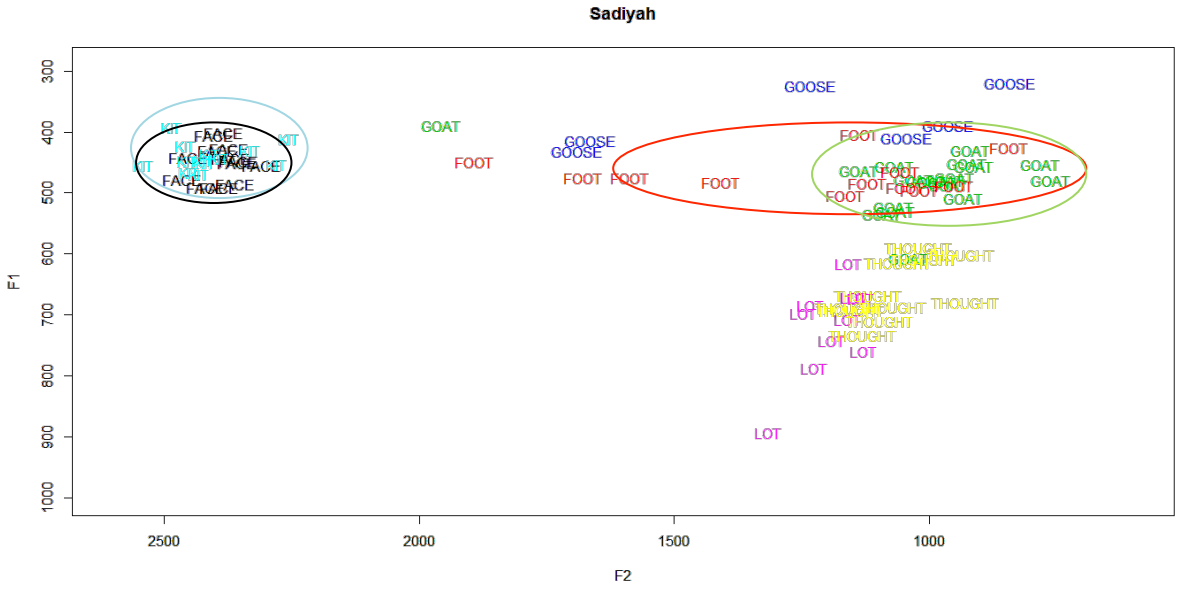


Figure 8. Vowel plot for Sadiyah.

Further statistical tests suggest the difference is in the position of FACE and GOAT. No significant difference was observed between the age-groups for KIT and FOOT (independent t-tests p=.75, p=.52, respectively). Significant differences were observed between the age-groups with FACE and GOAT midpoint F1s (unpaired Wilcoxon p=.0003 and independent t-test p<.0001, respectively).

The results presented here suggest that, monophthongal FACE and GOAT in these PE speakers maintain a qualitative distinction with KIT and FOOT. This qualitative distinction appears to be stronger for younger speakers.

10.2. AE speakers

A series of Shapiro-Wilks tests were undertaken for individual speakers' individual vowels. Table 4 illustrates that for all but one speaker FACE and KIT are significantly different in midpoint F1 (independent t-tests).

Speaker	
Lisa	p=.003 **
Christine	p=.04 *
Peggy	p=.099

Table 4. Significance values for within-speaker differences with FACE and KIT

Table 5 illustrates the same pattern with GOAT and FOOT, with all but one speaker having significantly different midpoint F1 values.

Speaker	
Lisa	p<.0001 *** unpaired Wilcoxon
Christine	p=.12 independent t-test
Peggy	p=.05 * unpaired Wilcoxon

Table 5. Significance values for within-speaker differences with GOAT and FOOT

These relationships are highlighted in Figures 9 through 11 showing formant plots for individual speakers.

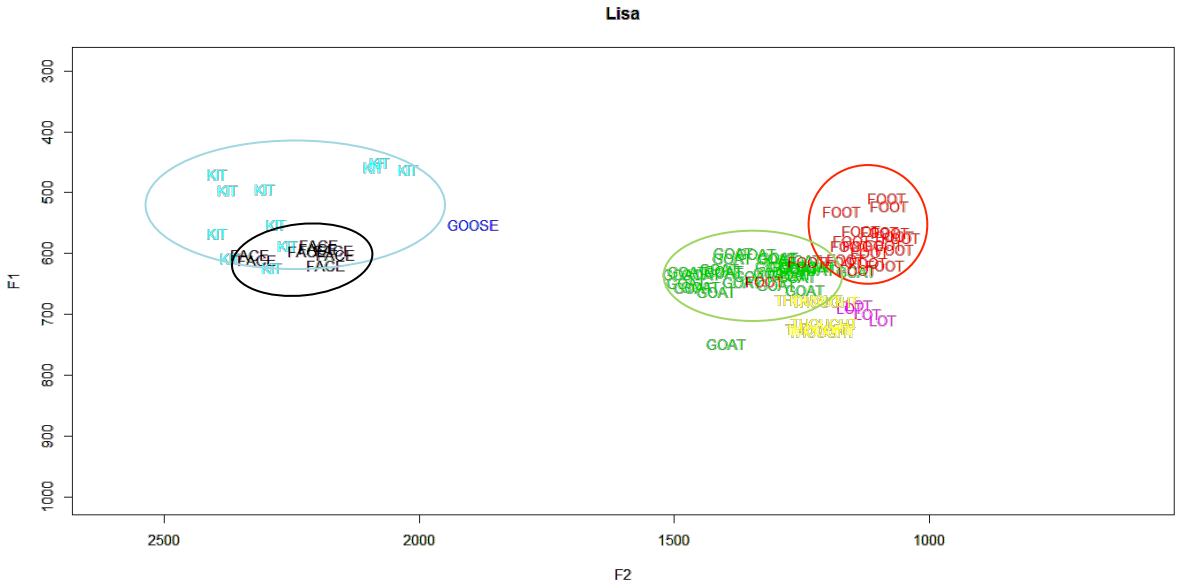


Figure 9. Vowel plot for Lisa.

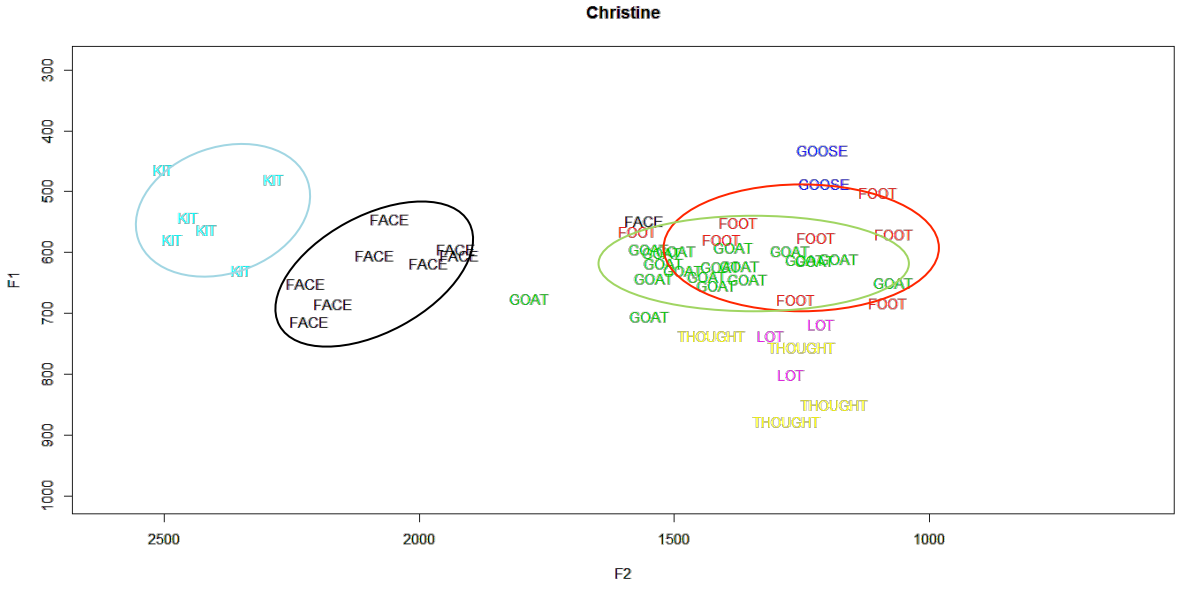


Figure 10. Vowel plot for Christine.

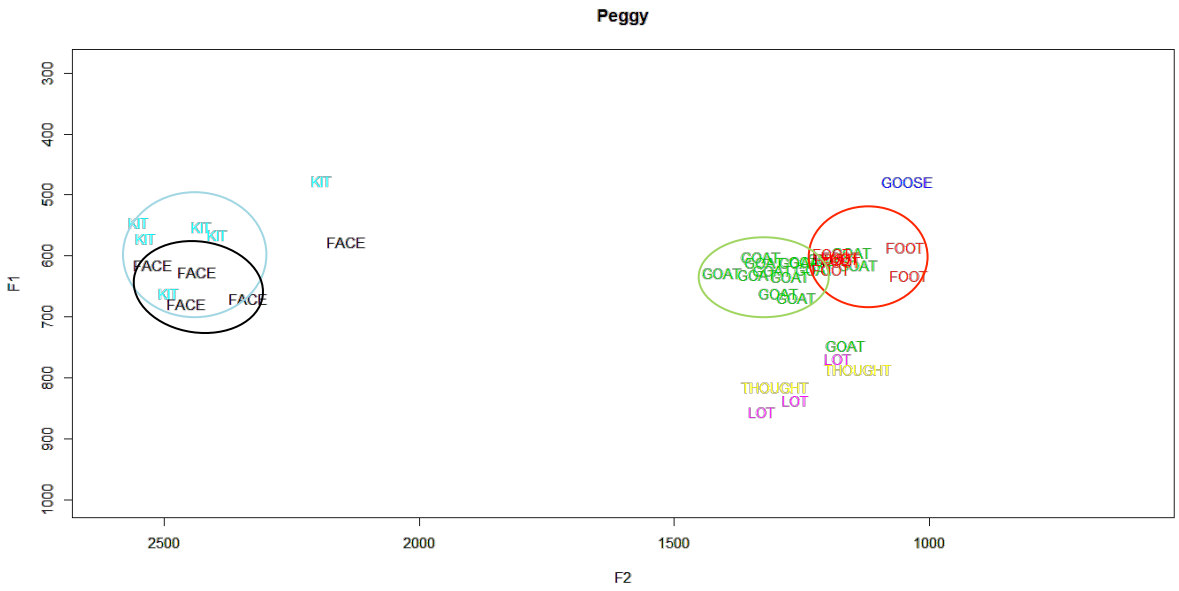


Figure 11. Vowel plot for Peggy.

Although mixed effects models were unsuccessful due to the small amount of data it seems AE speakers also show variability in the degree of separation between the two vowel pairs. Without further data and analysis it is not clear whether this is related to age in a similar way to that found with PE speakers. Here, the oldest speaker shows no difference between FACE and KIT, but the middle speaker shows no difference between GOAT and FOOT.

10.3. AE and PE speakers

Mixed effects linear regression models were undertaken with all speakers. Age group, lexical set and ethnicity were fixed effects with speaker and word included as random effects. Separate models were run for each vowel pair (FACE & KIT, GOAT & FOOT) followed by MCMC simulations.

A significant effect of ethnicity was observed in the FACE and KIT model (t=-3.14; p=.0002), indicating that PE speakers have a significantly lower F1 value for both FACE and KIT when compared with AE speakers. A significant difference was also observed between FACE and KIT across all speakers (t=-2.77; p=.007), suggesting KIT is significantly lower than FACE for all speakers. Figure 12 highlights the difference between PE and AE speakers, and also illustrates a lower KIT for both groups.

Figure 12. Significantly lower F1 for PE speakers. Significantly lower F1 for KIT.

A significant effect of ethnicity was observed with the GOAT and FOOT model (t=-2.5; p=.006). PE speakers have significantly lower midpoint F1 values than the AE speakers. Contrary to the front of the vowel space, no significant difference was observed across speakers between GOAT and FOOT, although FOOT is lower than GOAT (t=1.25; p=.23). Figure 13 illustrates the closer realisations for PE speakers overall.

Figure 13. Significantly lower F1 for PE speakers

11. Discussion

Variability in the degree of separation of FACE & KIT, and GOAT & FOOT is perhaps not restricted to PE speakers but could be evidence of more general variability in Bradford. The PE speaker who showed no significant difference between midpoint F1 values of both vowel pairs (Sadiyah) is a self-employed business woman with a great deal of contact in the wider Bradford community. If this feature was characteristic of PE, its presence in only Sadiyah's speech may be surprising. If considered as more general variation in Bradford and not constrained by ethnicity, its presence in Sadiyah's speech could illustrate her increased mobility, and potentially, increased integration.

It is unlikely that for speakers with no significant difference in midpoint F1 this is evidence of a merger. Further exploration of the ways in which the vowel pairs might be distinguished (duration, formant dynamics, F2 values, etc.) needs to be undertaken. Further, this paper considers only word list and reading passage data from nine female speakers.

Perhaps the most interesting observation is the significantly lower midpoint F1 values resulting in closer FACE, KIT, GOAT and FOOT realisations for all PE females when compared with AE females. The variation observed could be a consequence of the varying styles in which the two groups were recorded. AE speakers read a word list, whereas PE speakers read a passage. It is well known that speaking style can affect speakers' realisations (Labov 1978). However, considering patterns observed in other PE and multicultural-English varieties, the stylistic variation here is not considered to be the reason for the variation observed.

The closer realisations observed pattern with findings of Stuart-Smith et al. (2011). FACE and GOAT were observed to be significantly closer for Glaswasians than monolingual Glasgwegian speakers. The current work provides further evidence that this feature may be a more widespread characteristic of PE. Sharma (2011) also observed monophthongal FACE and GOAT with her second-generation speakers. Further, changes in the realisation of FACE and GOAT vowels in MLE could be patterning in a similar way. Cheshire et al. (2008) observed raised onsets and shorter trajectories in FACE and GOAT realisations for younger speakers in Hackney. They suggest that it is male non-Anglos who are leading the diphthong changes reported (Cheshire et al. 2008: 11). Drummond (2013a) observed monophthongal FACE and GOAT realisations in the speech of adolescent males in Manchester. He too comments on their close nature, although notes that the findings are currently based on a small amount of data (Drummond 2013b: pc).

It would appear that closer FACE and GOAT is a developing supra-local feature indexing non-Anglo ethnic identity. Sadiyah is the only PE speaker with no significant difference between midpoint F1 values of the vowel pairs. She still, however, shows close realisations of these vowels suggesting a complex interaction of local, and 'ethnic' identity. In the conversational data not considered here, she talks passionately about her faith and culture and how these are big parts of her life. Afsana is another of the older PE speakers. She shows highly significantly different F1 midpoints for FACE & KIT, and GOAT &FOOT but has some of the closest vowel realisations. Unlike Sadiyah, she has few friends and little contact outside of the local, predominantly Asian, community. It follows, therefore, that her speech patterns would be characteristic of PE, but not of more general Bradford variation. These results follow ideas discussed by Stuart-Smith et al. (2011) and Harris (2006) of a co-existent 'Brasian' identity, with speakers developing complex interactions between 'local' and 'ethnic' features. Figures 14 and 15 below contain boxplots illustrating the changing F1 at the front and back of the vowel space from AE to PE.

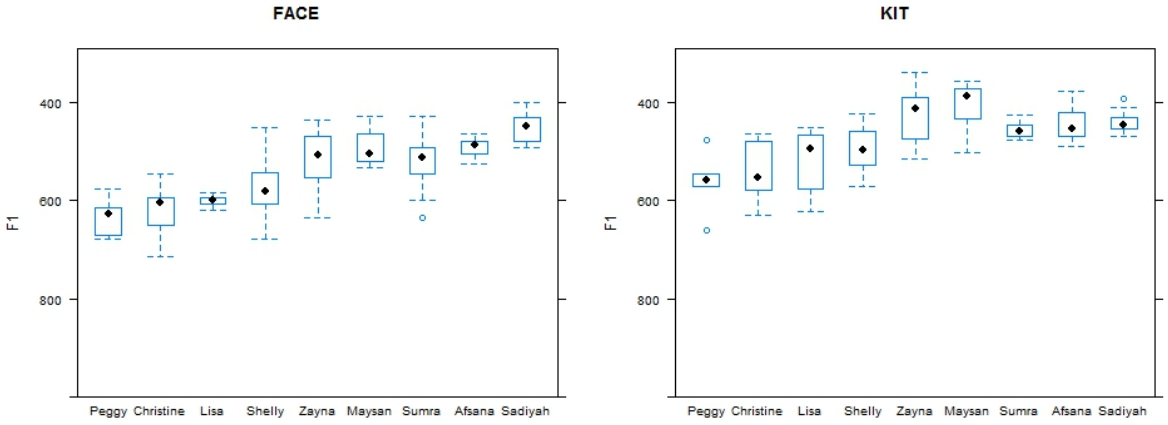


Figure 14. Boxplots for all speakers: F1 for FACE and KIT.

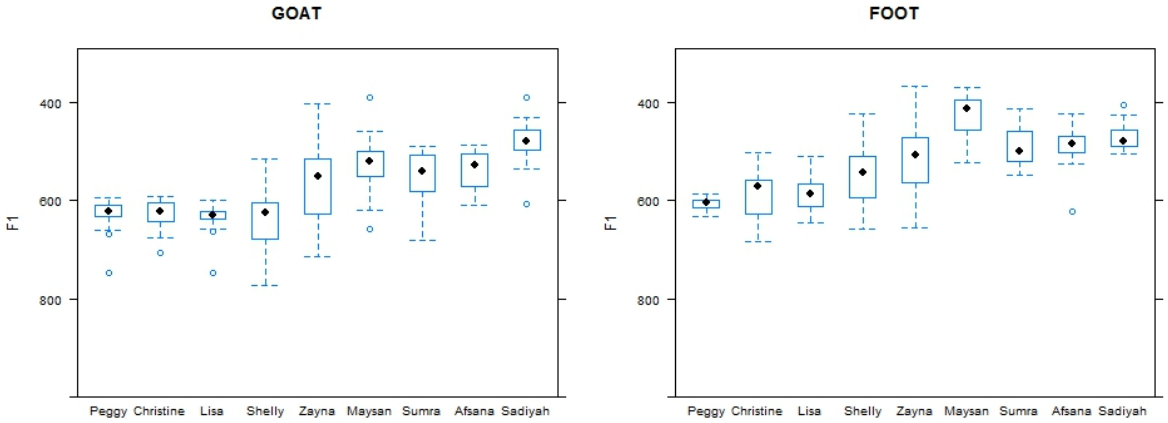


Figure 15. Boxplot for all speakers: F1 for GOAT and FOOT.

Determining whether the variability observed here is evidence of innovation or transfer from Panjabi is still unclear. The discussion of Panjabi above indicates that similar vowels with comparable qualities to those found in PE may be present in Panjabi. However, given the presence of this feature in other multicultural varieties of English a direct transfer from Panjabi would be surprising. Wells (1982: 626) notes that in Indian English, monophthongal FACE and GOAT are common, with realisations around /e/ and /o/. This suggests that monophthongal FACE and GOAT are not uncommon in English with Indic language influences. Wells (1982) attributes this to the influence of English, stating that long mid diphthongs probably arose around 1800 with an English presence in India existing before this time. This explanation does not account for the variation observed here in PE and other multicultural Englishes.

Without further knowledge of the first-generation realisations no categorical conclusions can be made. Reallocation of the potentially transferred feature may have taken place, aligning speakers with other varieties. Monophthongal FACE and GOAT could be indexing a non-Anglo ethnic identity. It would be useful to have a more comprehensive picture of the contact patterns of PE speakers, both within Bradford and throughout the UK. This could help address the question of how features have become characteristic in different localities and also determine whether the feature is a Panjabi transfer or a multicultural-English or PE innovation, or even if it is that clear cut.

References

BAAYEN, R. H. 2008. *Analysing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

BAUER, L. 2008. A question of identity: A response to Trudgill. *Language in Society,* 37(2), 270-273.

BHARDWAJ, M. R. 2012. *Colloquial Panjabi: The Complete Course for Beginners*. The Colloquial Series. Oxon: Routledge.

BHATIA, T. K. 1993. *Punjabi: A cognitive-descriptive grammar*. London: Routledge.

BRITAIN, D. 2002. Diffusion, levelling, simplification and reallocation in past tense BE in the English Fens. *Journal of Sociolinguistics,* 6(1), p. 16-44.

CHESHIRE, J., FOX, S., KERSWILL, P. AND TORGERSEN, E. 2008. Ethnicity, friendship network and social practices as the motor of dialect change: Linguistic innovation in London. *Sociolinguistica,* 22, p. 1-23.

CHESHIRE, J., KERSWILL, P., FOX, S. AND TORGERSEN, E. 2011. Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics,* 15(2), p. 151-196.

COUPLAND, N. 2008. The delicate constitution of identity in face-to-face accommodation: A response to Trudgill. *Language in Society,* 37(2), p. 267-270.

DAY, K. 2009. *The Intonation of Punjabi English*. Unpublished MA Dissertation; University of York.

DRUMMOND, R. 2013a. Adolescent identity construction through Multicultural (Manchester) English. *UKLVC 9*. 2nd - 4th September. University of Sheffield.

DRUMMOND, R. 2013b. Email received by Wormald, J., Multicultural Manchester English: UKLVC9 presentation. [15th September 2013].

DYER, J. 2002. 'We all speak the same round here': Dialect levelling in a Scottish-English community. *Journal of Sociolinguistics,* 6(1), p. 99-117.

HARRIS, R. 2006. *New Ethnicities and Language Use*. Language and Globalisation. Hampshire: Palgrave & Macmillan.

HESELWOOD, B. AND MCCHRYSTAL, L. 1999. The effect of age-group and place of L1 acquisition on the realisation of Punjabi stop consonants in Bradford: an acoustic sociophonetic study. *Leeds Working Papers in Linguistics and Phonetics,* 7, p. 49-68.

HESELWOOD, B. AND MCCHRYSTAL, L. 2000. Gender, accent features and voicing in Panjabi-English bilingual children. *Leeds Working Papers in Linguistics and Phonetics,* 8, p. 45-70.

HIRSON, A. AND SOHAIL, N. 2007. Variability of rhotics in Punjabi-English bilinguals. *16th International Congress of Phonetic Sciences*, Saarbrucken, p. 1501-1504.

HOLMES, J. AND KERSWILL, P. 2008. Contact is not enough: A response to Trudgill. *Language in Society,* 37(02), p. 273-277.

HUGHES, A., TRUDGILL, P. AND WATT, D. 2012. *English Accents and Dialects: Fifth Edition*. London: Hodder Education.

KERSWILL, P., TORGERSEN, E. N. AND FOX, S. 2008. Reversing "drift": Innovation and diffusion in the London diphthong system. *Language Variation and Change,* 20(3), p. 451-491.

KIRKHAM, S. 2011. The Acoustics of Coronal Stops in British Asian English. *International Congress of Phonetic Sciences*, Hong Kong, 17-21st August, p. 1102-1105.

LABOV, W. 1978. The Study of Language in its Social Context. In: *Sociolinguistic Patterns*. Oxford: Blackwell, p. 183-259.

LAMBERT, K., ALAM, F. AND STUART-SMITH, J. 2007. Investigating British Asian accents: studies from Glasgow. *16th International Congress of Phonetic Sciences*, Saarbrucken, p. 1509-1512.

LEWIS, M.P., SIMONS, G.F. AND FENNIG, C.D. eds. 2013. *Ethnologue: Languages of the World, Seventeenth Edition*. Dallas, Texas: SIL International. Online version: www.ethnologue.com (Accessed 11th October 2013)

LOTHERS, M. AND LOTHERS, L. 2012. *Mirpuri immigrants in England: A sociolinguistic survey*. SIL International: SIL Electronic Survey Report

MUFWENE, S. S. 2008. Colonization, population contacts, and the emergence of new language varieties: A response to Peter Trudgill. *Language in Society,* 37(2), p. 254-259.

ONS 2012a. Ethnicity and National Identity in England and Wales 2011. *Office for National Statistics*.

ONS 2012b. *KS201EW Ethnic group, local authorities in England and Wales*. Published Online 11/12/12 www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-286262.

ONS 2013a. 2011 Census: Quick Statistics for England and Wales, March 2011. *Office for National Statistics*.

ONS 2013b. *QS204EW Main language (detailed), local authorities in England and Wales*. Published online 30/01/13 www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-286348.

PETYT, T. K. 1985. *Dialect and accent in industrial West Yorkshire*. Amsterdam: J. Benjamins Pub. Co.

RATHORE, C. 2011a. East African Indians in Leicester, UK; phonological variation across generations. *ISLE2*. Boston.

RATHORE, C. 2011b. "I am more of an African than an Indian", East African Indian English in Leicester, UK? . *ICLaVE 6*. Freiburg.

REYNOLDS, M. AND VERMA, M. 2007. Indic Languages. In: Britain, D. ed. *Language in the British Isles*. Cambridge: Cambridge University Press, p. 293-307.

SCHNEIDER, E. W. 2008. Accommodation versus identity? A response to Trudgill. *Language in Society,* 37(2), p. 262-267.

SHACKLE, C. 1983. Review of Zograph (1982). *Bulletin of the School of Oriental and African Studies,* 46, p. 372-372.

SHACKLE, C. 2003. Panjabi. In: Cardona, G. and Jain, D. eds. *The Indo-Aryan Languages*. London: Routledge, p. 581-621.

SHARMA, D. 2011. Style repertoire and social change in British Asian English. *Journal of Sociolinguistics,* 15(4), p. 464-493.

SHARMA, D. AND SANKARAN, L. 2011. Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change,* 23(3), p. 399-429.

STUART-SMITH, J. 1999. Glasgow: Accent and Voice Quality. In: Foulkes, P. and Docherty, G. eds. *Urban Voices*. London: Arnold, p. 201-222.

STUART-SMITH, J., TIMMINS, C. AND ALAM, F. 2011. Hybridity and ethnic accents: A Sociophonetic analysis of 'Glaswasian'. Gregerson, F., Parrott, J. and Quist, P. eds. *Language Variation - European Perspectives 3: Selected papers from ICLaVE 5*, Copenhagen, p. 43-57. John Benjamins.

TOLSTAYA, N. I. 1981. *The Panjabi language: A descriptive grammar*. Trans. Campbell, G. L. Languages of Asia and Africa. Vol. 2. London: Routledge & Kegan Paul.

TORGERSEN, E. AND SZAKAY, A. 2011. A Study of Rhythm in London: Is Syllable-timing a Feature of Multicultural London English? *University of Pennsylvania Working Papers in Linguistics: Selected Papers from NWAV 39,* 17(2), p. 164-174.

TRUDGILL, P. 1986. *Dialects in contact*. Language in Society. Oxford, UK ; New York, NY, USA: B. Blackwell.
TRUDGILL, P. 2004. *New Dialect Formation: The Inevitability of Colonial Englishes*. Edinburgh: Edinburgh University Press.
TRUDGILL, P. 2008a. Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society,* 37(2), p. 241-254.
TRUDGILL, P. 2008b. On the role of children, and the mechanical view: A rejoinder. *Language in Society,* 37(2), p. 277-280.
TUTEN, D. N. 2008. Identity formation and accommodation: Sequential and simultaneous relations. *Language in Society,* 37(02), p. 259-262.
VERMA, M. AND FIRTH, S. 1995. Old Sounds and New Sounds: Bilinguals Learning ESL. In: Verma, M., Corrigan, K. and Firth, S. eds. *Working with Bilingual Children: Good Practice in the Primary Classroom*. Avon: Multilingual Matters, p. 128-142.
WATT, D. 2002. 'I don't speak with a Geordie accent, I speak, like, the Northern accent': Contact-induced levelling in the Tyneside vowel system. *Journal of Sociolinguistics,* 6(1), p. 44-64.
WATT, D. AND TILLOTSON, J. 2001. A spectrographic analysis of vowel fronting in Bradford English. *English World Wide,* 22(2), p. 269-304.
WELLS, J. C. 1982. *Accents of English 3: Beyond the British Isles*. Accents of English. Vol. 3. Cambridge: Cambridge University Press.
ZIPP, L. 2013. Stylistic variation in British Asian English prosody. *UKLVC 9*. University of Sheffield.
ZOGRAPH, G. A. 1982. *Languages of South Asia: A Guide*. Languages of Asia and Africa. Vol. 3. Britain: Routledge & Kegan Paul.

Jessica Wormald
Department of Language and Linguistic Science
University of York
Heslington, York
YO10 5DD
United Kingdom
Email: jessica.wormald@york.ac.uk