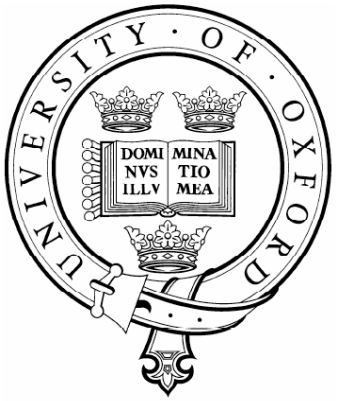
Oxford University Working Papers in Linguistics, Philology & Phonetics

Papers in Phonetics and Computational Linguistics



Edited by
ʻŌiwi Parker Jones and Elinor Payne

Volume 12
2009

Contents

Editorial Note

Oxford University Working Papers in Linguistics, Philology & Phonetics presents research being undertaken in these fields by staff, graduate students and other researchers in the Faculty of Linguistics, Philology & Phonetics at the University of Oxford. Each volume is devoted to a particular area of linguistic research in Oxford. The current volume presents a offering of papers form the Phonetics Laboratory, which includes work in phonetics and computational linguistics.

Comments on the papers included here are welcome. The first authors of each paper can be reached at the Lab:

Phonetics Laboratory
41 Wellington Square
Oxford
OX1 2JF
United Kingdom

Email: *firstname.lastname@phon.ox.ac.uk*

The editors can also be contacted by e-mail regarding the journal itself. To obtain further information regarding Linguistics, Philology & Phonetics at Oxford, please contact:

The Centre for Linguistics and Philology
Walton Street
Oxford
OX1 2HG
United Kingdom

This journal is currently distributed as part of an exchange arrangement involving similar journals from a number of university departments worldwide. We warmly welcome offers to institute further such agreements and invite university departments who express an interest to contact the editors.

As with other recent offerings, this volume of Oxford University Working Papers in Linguistics, Philology & Phonetics will be made available on the web (http://www.ling-phil.ox.ac.uk/pages/publications.html).

ʻŌiwi Parker Jones and Elinor Payne

Preface

Now into its 12th volume, the Oxford Working Papers in Linguistics, Philology & Phonetics (OWP) has established itself as a regular showcase for research being undertaken by current and former staff and postgraduate students from the University of Oxford. The current volume appears at a time of exciting change here in Oxford. Thanks to the untiring efforts and dedication of Steve Pulman, Anna Morpurgo Davies and John Coleman, we have become a *Faculty* of Linguistics, Philology & Phonetics. The Faculty is now headed by Professor Aditi Lahiri, whose arrival we are still celebrating. This volume of working papers has the honour of being the first published under the newly minted faculty.

According to OWP tradition, there is a loose thematic rotation between the four areas of Linguistics, Philology, Phonetics, and Romance Languages. Volume 1 (Dankovičová & Stuart-Smith, 1996), volume 5 (Coleman, 2000), and volume 8 (Grabe & Wright, 2003) showcased work from the Phonetics Laboratory, which is a tradition that we continue here. The present volume is subtitled *Papers in Phonetics and Computational Linguistics*, reflecting the thematic nature of its contents and the continued importance in the Phonetics Lab of work also done in computational linguistics. To see other research from the Phonetics Laboratory which has not been represented here, we invite you to visit our webpage (http://www.phon.ox.ac.uk).

Phonetics can trace a long and distinguished history at the University of Oxford. The 'study of sounds' (in one form or another) formed one of the first branches of traditional grammar. In this guise, phonetics would have been studied at Oxford from medieval times, first in relation to the 'holy' languages, and later more generally. Early work in what we might now call 'phonetics and phonology' includes that by John Wilkins and John Wallis in the late seventeenth century (see, e.g., McIntosh, 1956; Kemp, 1972; Subbiondo, 1987). A string of illustrious names in Phonetics are associated with the University in more modern times. At the dawn of the twentieth century, Henry Sweet 'squeezed into something called a Readership of phonetics' at Oxford (described diplomatically by George Bernard, in his preface to *Pygmalion*). Sweet published a number of highly influential works, including *A Handbook of Phonetics* (1877), *A Primer of Spoken English* (1890), and *The Sounds of English* (1908), earning himself the reputation of 'the man who taught Europe phonetics' (Howatt et al, 2004, pp. 198–207). After Sweet's death, Daniel Jones lectured in Oxford for a short period (1913–1914). Later, from 1930–1940, J. R. Firth lectured in phonetics (especially the phonetics of Indian

languages) for students at the Indian Institute who were training to go into the Indian Civil Service.

In his preface to *Pygmalion*, Shaw (1916) also described Sweet as 'a man of genius with a *seriously underrated subject*' [editors' italics]. Far from remaining an 'underrated subject', recognition and expansion of phonetics grew exponentially during the second half of the twentieth century. This resulted in part from the growth of linguistics as a whole and in part from an explosion of technological advances which profoundly changed both the methods and theoretical scope of the discipline. At Oxford, this was mirrored by the founding of the Phonetics Laboratory, in 1980, with Tony Bladon as its first director. John Coleman has been leading the Phonetics Laboratory forward as its director since 1993, following brief tenures by Ian Watson and Bruce Connell. Since its inception the Laboratory has gone from strength to strength, firmly establishing itself as a focal point for research as well as for undergraduate and postgraduate teaching. John Coleman became Oxford's very first Professor of Phonetics in 2008. It is therefore particularly fitting that the current volume of working papers should be a product of the Lab and the community associated with it.

We end with a few introductory words on each of our contributors:

Lluisa Astruc is a lecturer in Spanish at the Open University and an Affiliated Lecturer in the Department of Spanish and Portuguese, University of Cambridge.

As already noted, **John Coleman** is Professor of Phonetics and the Director of the Phonetics Laboratory. He is also a Fellow of Wolfson College, Oxford.

Greg Kochanski is a Research Fellow in the Phonetics Laboratory.

Anastassia Loukina is a Research Associate in the Phonetics Laboratory and a Junior Research Fellow at St Cross College, Oxford.

Nivi Mani is a former doctoral student in the Phonetics Laboratory and a current Postdoctoral Research Fellow at UCL.

Julien Mayor is a Postdoctoral Researcher in Oxford University's Psychology Department.

Paul O'Neill is currently finishing his doctorate in Oxford while having recently become University Teacher in Hispanic Linguistics at the University of Liverpool.

ʻŌiwi Parker Jones is a doctoral student in the Phonetics Laboratory and in the University of Oxford's Computational Linguistics Group. He has recently been awarded a non-stipendiary Research Fellowship at Wolfson College, Oxford.

Elinor Payne is the University Lecturer in Phonetics and Phonology, and a Fellow at St Hilda's College, Oxford.

Brechtje Post is a Senior Research Associate at Cambridge University's Research Centre for English and Applied Linguistics. She is also a Fellow at Jesus College, Cambridge.

Pilar Prieto is a research professor in the Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Barcelona.

Ros Temple is University Lecturer in French Linguistics and a Fellow of New College, Oxford.

Maria del Mar Vanrell is a doctoral student in the Department de Filologia Catalana, at the Universitat Autònoma de Barcelona.

Ian Watson is University Lecturer in French Language and Linguistics, and a Fellow of Christ Church, Oxford.

ʻŌiwi Parker Jones and Elinor Payne[1]

References

Coleman, J. (Ed.) (2000). *Oxford Working Papers in Linguistics, Philology & Phonetics*, *5*.

Dankovičová, J. & Stuart-Smith, J. (Eds.) (1996). *Oxford Working Papers in Linguistics, Philology & Phonetics*, *1*.

[1] We would like to thank John Coleman, David Cram, and Celia Glyn for helpful discussions on the history of the Lab. We also thank John Coleman and Katie Drager for feedback on an earlier draft of the preface, and Kate Dobson and Ranjan Sen for editorial advice.

Grabe, E., & Wright, D. G. S. (Eds.) (2003). *Oxford Working Papers in Linguistics, Philology & Phonetics*, *8*.

Howatt, A. P. R., & Widdowson, H. G. (2004). *A History of English Language Teaching*. Oxford: Oxford University Press

Kemp, J. A. (1972). *John Wallis's Grammar of the English Language, with an Introductory grammatico-physical Treatise on Speech*: *A new edition with translation and commentary*. London: Longman.

McIntosh, M. (1956). *The Phonetic and Linguistic Theory of the Royal Society School, from Wallis to Cooper*. B.Litt. Thesis, University of Oxford.

Shaw, G. B. (1916). *Pygmalion*. New York: Brentano.

Subbiondo, J. L. (1987). John Wilkins' theory of articulatory phonetics. In H. Aarsleff, L. G., Kelly, H. J., & Niederehe (Eds.), *Papers in the History of Linguistics* (pp. 263–270). Philadelphia: Benjamins.

Sweet, H. (1877). *A Handbook of Phonetics*. Oxford: Clarendon Press.

Sweet, H. (1890). *A Primer in Spoken English*. Reprinted in 1900 by Clarendon Press.

Sweet, H. (1908). *The Sounds of English: An Introduction to Phonetics*. Reprinted in 1929 by Clarendon Press.

Using Mimicry to Learn About Phonology

Greg Kochanski

Phonetics Laboratory, University of Oxford

Abstract

Phonology typically describes speech in terms of discrete signs like features. The field of intonational phonology uses discrete accents to describe intonation and prosody. But, are such representations useful? The results of mimicry experiments indicate that discrete signs are not a useful representation of the shape of intonation contours. Human behaviour seems to be better represented by a attractors where memory retains substantial fine detail about an utterance. There is no evidence that discrete abstract representations that might be formed that have an effect on the speech that is subsequently produced. This paper also discusses conditions under which a discrete phonology can arise from an attractor model and why – for intonation – attractors can be inferred without the implying a discrete phonology.

Keywords

Attractor, Intonation, Phonology, Phonetics, Experiment, Gradient, Binary

1.0 Introduction

We often think of the units of intonational phonology as discrete entities: accents which fall into just a few categories (Gussenhoven, 1999; Ladd, 1996; Beckman & Ayers Elam, 1997). In this view, accents in intonation are equivalent to phonemes in segmental phonology (except that they cover a larger interval). They have a rough correspondence to the acoustic properties of the relevant region and accents form a small set of atomic objects that do not have meaning individually but that can be combined to form larger objects that carry meaning. For segmental phonology, the larger objects are words; for intonation, the larger objects are tunes over a phrase.

However, the analogy is not strong, and there are many differences. For instance, there is no known useful mapping from intonational phonology to meaning. (Pierrehumbert & Hirschberg, 1990, point out some of the difficulties.) For words, the equivalent is accomplished by

dictionaries and internet search engines. These technologies have no intonational equivalents. To date, attempts to connect between intonation or fundamental frequency contours have not escaped from academia: the results are either probabilistic (Grabe, Kochanski, & Coleman, 2005), have been theoretical and primarily based on intuition, or have been conducted in tightly controlled laboratory conditions (Ladd & Morton, 1997; Gussenhoven & Rietveld, 1997).

Likewise, there is no known, reliable mapping between sound and intonational phonology. Probability distributions overlap (Grabe, Kochanski, & Coleman, 2007) and automated systems for recognizing intonation have not become commercially useful. In contrast, the connection between acoustics and segmental phonology is made by speech synthesis and recognition systems. The mapping between sound and segmental phonology is complicated, but it is reasonably well understood and reliable enough to be commercially useful. As a further contrast, transcription of intonation seems qualitatively different from transcription of segmental information. Intonational transcription (e.g. Grice et al., 1996; Jun et al., 2000; Yoon et al., 2004) is far more error-prone and slower than transcription of words, even after extensive training. Yoon et al. (2004) found an agreement of circa 85% between transcribers (depending on exactly what was being compared), but it is notable that at each point in the transcription, the transcribers had a choice between (typically) just two symbols. In a typical phonemic or orthographic transcription, the transcriber would attain comparable or higher precision while choosing between (about) 40 phones, or amongst thousands of possible words for each symbol.

So, in light of these differences, it is reasonable to ask whether intonation can be usefully described by a conventional discrete phonology or not. If it can be, what are the properties of the objects upon which the phonological rules operate? This paper lays out empirically-based answers to those questions and describes an experimental technique that can provide a reasonably direct exploration of the properties of phonological objects.

2.0 Modelling Mimicry

Figure 1 shows a simple model of speech mimicry. It is treated as a completely normal speech process: a person hears speech, perceives it, and generates a memory representation for it. Later, the person produces speech from the memory.

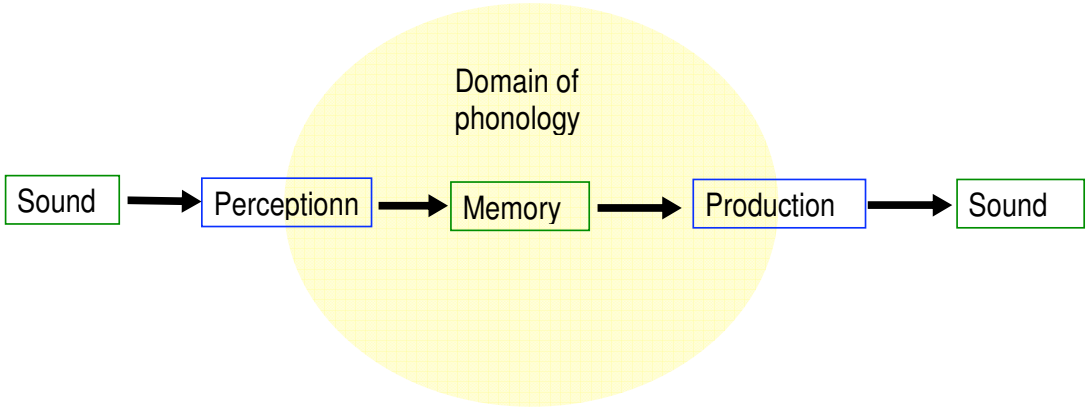


Figure 1: Mimicry via phonology. Sound is perceived, stored in memory in some form of phonological representation, then when the subject begins to speak, he or she articulates based on the stored memory representation.

The most contentious point might be the identification of the memory representation with a phonological representation. But, if we cannot usefully predict the acoustic properties of speech from phonology, how can phonology claim to be part of the study of language? Likewise, if phonological entities are not the end-product of the perceptual process, where do they come from?[1] This is a relatively strong interpretation: it asserts that there is some isomorphism between phonology, the mind, and the activity of the brain. In other words, that phonology can describe (at least in an approximate, abstract way) what is happening in the mind and the brain.

Some linguists would deny this biological connection, claiming that phonology is strictly a human invention that allows us to conveniently represent speech patterns in a way that humans can easily interpret and study. But, the denial does not follow from the invention: the self-evident fact that phonology is a human invention does not prohibit it from being isomorphic to processes in the brain. For example, secondary-school models of atoms are human constructs and some ideas of basic chemistry, such as "valence" are as abstract as phonology, but they reflect – in an approximate way – the underlying atoms and quantum mechanics. Thus, linguists who deny the biological relevance of phonology are not doing it out of necessity, but rather, they are making it an axiom of the field, based on tradition, history, and convenience.

[1] Of course, the perceptual process can be described at various levels of detail, and phonology is only one level of description. However, for phonology to be meaningful, there must eventually be a consistent description of the perceptual process (ideally, a quantitative, numerical model) that takes acoustics on one side and yields phonological entities on the other side.

Such a denial is a free choice, simply reflecting the researcher's view of where to set to be the academic boundary. Should phonology be determined by the behaviour it explains or by the representations that it uses? Here, my intent is to study linguistic behaviour of objects simpler than words, using whatever representation is most appropriate. The primary question is finding the representation that best describes human behaviour from among those representations that fit into the rest of linguistics.[2] One might reasonably hope that those representations which give the best description would have some analogy to the structure of the mind and/or brain.

Models of mimicry other than Figure 1 are possible, but they lead to a more complex description of human behaviour. For instance, if mimicry were a hard-wired part of early language learning, one might imagine that there were two separate parallel channels, one for non-phonological mimicry and one for speech that is treated phonologically. However, such a model would be more complex and evidence for a separate channel is weak.[3]

If we assume the model in Figure 1, the critical question is then the nature of the memory trace. Is it continuous in the sense that a small change in fundamental frequency always corresponds to a small change in the memory representation? This would imply that memory stores something analogous to pitch, suggesting some variety of Exemplar model (e.g. Goldinger, 1992; Johnson & Mullenix, 1997; Pierrehumbert, 2001). Or, alternatively, is the memory representation discrete, leading to a model close to Generative Phonology (e.g. Liberman, 1970). These two hypotheses will be considered below.

Below, I follow common practice (see discussion in Kochanski, 2006, §2) and approximate intonation by measurements of speech fundamental frequency. This approximation is undoubtedly imperfect: for instance

---

[2] The opposite viewpoint would be to separate the object and description, then simply accept that the description is based upon discrete categories while the object of study might be continuous. Certainly, one can operate this way in a self-consistent manner, but there is a cost: it becomes harder to distinguish good theories from bad by the process of prediction and experimental test. I would argue that for Phonology to define itself by the representations it uses (e.g. to freeze the field onto current phonological representations) would be analogous to Astronomy defining itself to exclude spectroscopy or Electrical Engineering defining itself via the gold-leaf electroscope. Should a field define itself by its tools, it will wither when important phenomena are found that cannot be studied with those tools.

[3] Most arguments for a separate mimicry channel assume that the phonological units are strictly discrete. Under that assumption, any early learning of speech before phonology is well-established would demand a specialised mimicry channel. However, in this paper, we are asking whether intonational phonology is discrete, so this assumption begs the question.

loudness and duration are important in defining accent locations (Kochanski & Orphanidou, 2008; Kochanski, 2006, and references therein). While I discuss continuous vs. discrete phonologies in terms of fundamental frequency, similar arguments could be made with respect to other acoustic properties. The two alternatives for phonology are cast as hypotheses to be tested and (potentially) rejected.

2.1 Hypothesis 1: The memory store is a continuous representation of fundamental frequency

In this hypothesis (schematized in Figure 2) nearby values of speech fundamental frequency in the input utterance are represented by nearby memory representations. Further, nearby memory representations yield nearby fundamental frequencies in the speech that is eventually produced. In other words, there is a continuous mapping between input fundamental frequency and the memory representation, a continuous memory representation, and a continuous mapping on the output.

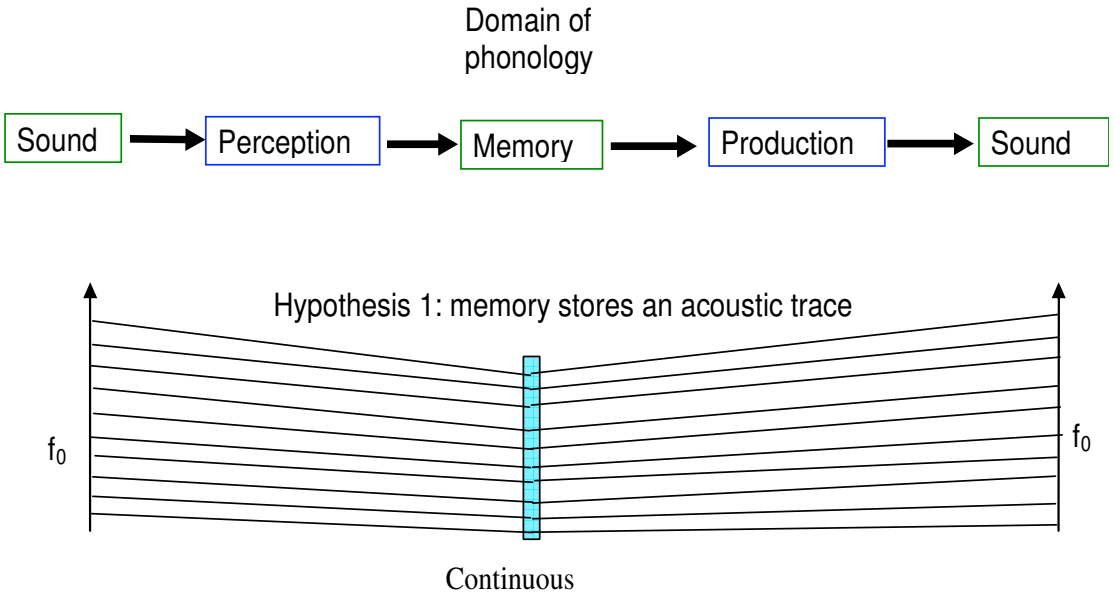


Figure 2: Hypothetical model of mimicry where the memory store is continuous. The lower half of the drawing represents speech fundamental frequency (increasing upwards) at some point in a phrase. The lines connect input fundamental frequency (left axis) to the corresponding memory representation (centre) to the fundamental frequency that is eventually produced (right axis).

Absent variability, the output would perfectly preserve any distinctions that were made in the input. This is not to say that the output

would necessarily equal the input, though. For instance, the human who is doing the mimicry might transpose all frequencies down to a more comfortable level, as in Figure 3.

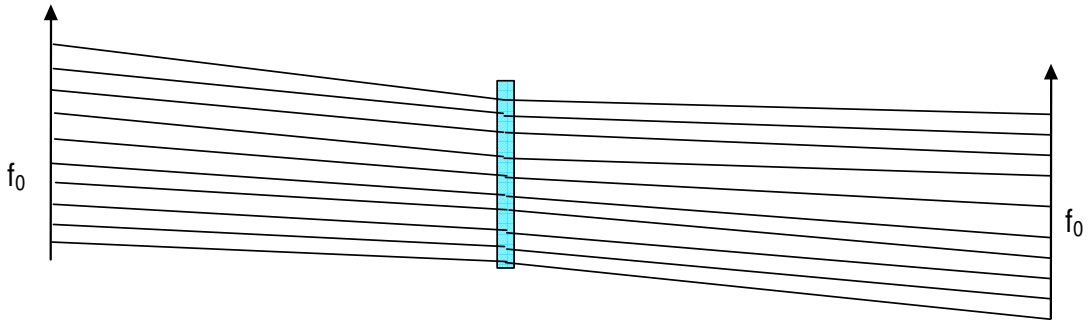


Figure 3: Continuous mappings and memory representation for a person who is transposing down to a lower pitch. Compare with Figure 2.

Utterance-to-utterance variation will limit the number and subtlety of distinctions that can be preserved by the mimicry process. Figure 4 shows this effect. In this example, any distinction between the bottom two frequencies is lost. This is an effect in language that will tend to prevent subtle phonetic distinctions from being used to represent any important phonological differences. Distinctions that are smaller than utterance-to-utterance variation will frequently be lost, leading to miscommunication and confusion. Presumably the language would evolve to avoid such unreliable distinctions.

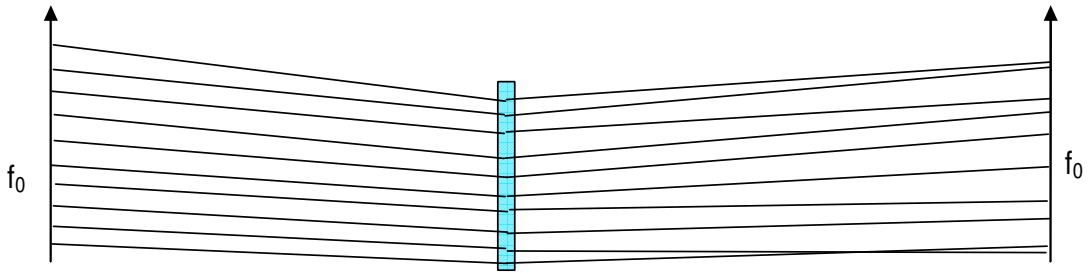


Figure 4: Mimicry with variation in production.

However, while language users are limited by variation, laboratory experiments need not be. Experiments can average over many utterances (a luxury that language users do not have in the midst of a conversation), thus reducing the variation as much as needed. If we average, we can construct an ideal variation-free model such as Figure 5. In that model, all input distinctions are preserved through the memory representation to the output. The pair of coloured lines show a distinction between two slightly different utterances in the average which might not have been

distinct in every observed utterance. At some point, these utterances have different fundamental frequency (left), which is perceived as two different memory representations (centre). These different memory representations lead to a measurable difference in the fundamental frequency that the subject produces (right).

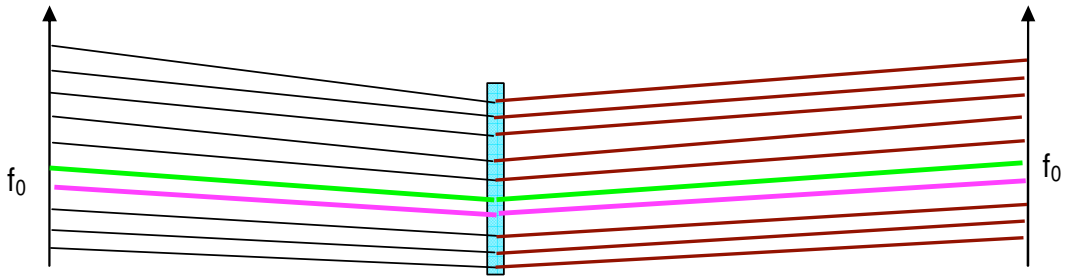


Figure 5: ideal model obtained by averaging results from many utterances (e.g. Figure 4 is one utterance) to reduce variation.

2.2 Hypothesis 0: The memory store is discrete

Intonational Phonology, like most of linguistics, assumes that its object of study can be represented well by discrete symbols. For the sake of argument, we assume that we can find a minimal pair of intonation contours that differ only by a single symbol, **H** vs. **L**.[4] Figure 6 shows this hypothesis schematically. Under the null hypothesis, the intonation is perceived (either categorically or not), then stored in memory as one or the other of two discrete representations. Finally, when the subject mimics the intonation contour, his/her speech is produced from the memory representation.

[4] However, the argument presented here does not depend upon having a minimal pair or upon having a simple difference. We will merely assume that there are a finite number of discrete memory representations. We also assume that these memory representations are not so numerous that perception is ambiguous.

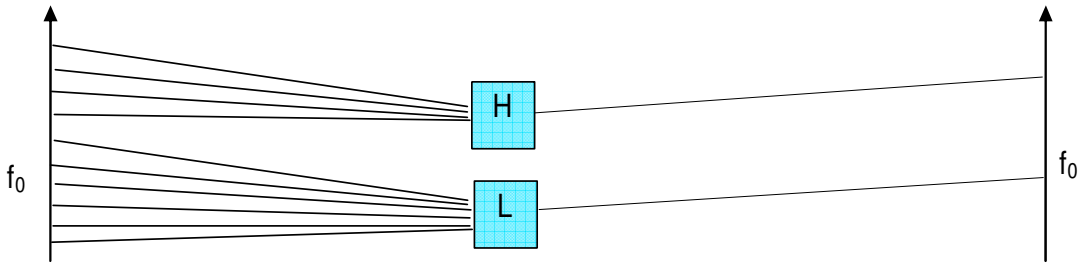


Figure 6: Hypothetical model of mimicry where the memory store is discrete. The drawing represents speech fundamental frequency (increasing upwards) at some point in a phrase. The lines connect input fundamental frequency (left axis) to the corresponding memory representation (centre) to the fundamental frequency that is eventually produced (right axis).

Now, on the basis of an individual utterance, production variation will yield a broad range of outputs for each memory representation. Figure 7 shows several potential outputs from the same phonology. Potentially, the resulting probability distributions produced from **H** and **L** could even overlap (though any substantial overlap would mean that the **H** vs. **L** distinction was not sufficiently clear to form a reliable minimal pair).

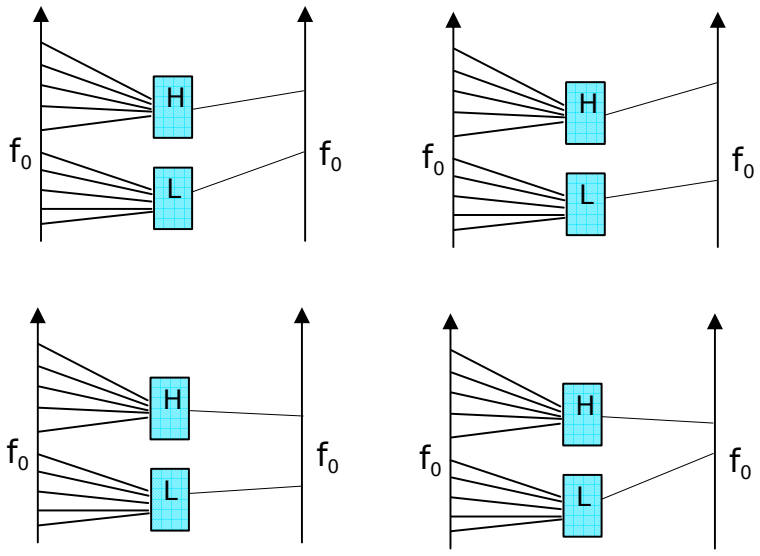


Figure 7: Hypothetical model of mimicry where the memory store is discrete. The drawing represents speech fundamental frequency (increasing upwards) at some point in a phrase. The lines connect input fundamental frequency (left axis) to the corresponding memory representation (centre) to the fundamental frequency that is eventually produced (right axis).

However, just as with Hypothesis 1, we can average over all productions from the same phonology and remove the effect of the variation. In this case, we see that the averaged productions form two well-separated values, different for **H** and **L**. However, the crucial difference between Hypotheses 0 and 1 lies in which distinctions are preserved. Hypothesis 1 preserves all input distinctions through to the output, but that is not the case for Hypothesis 0.

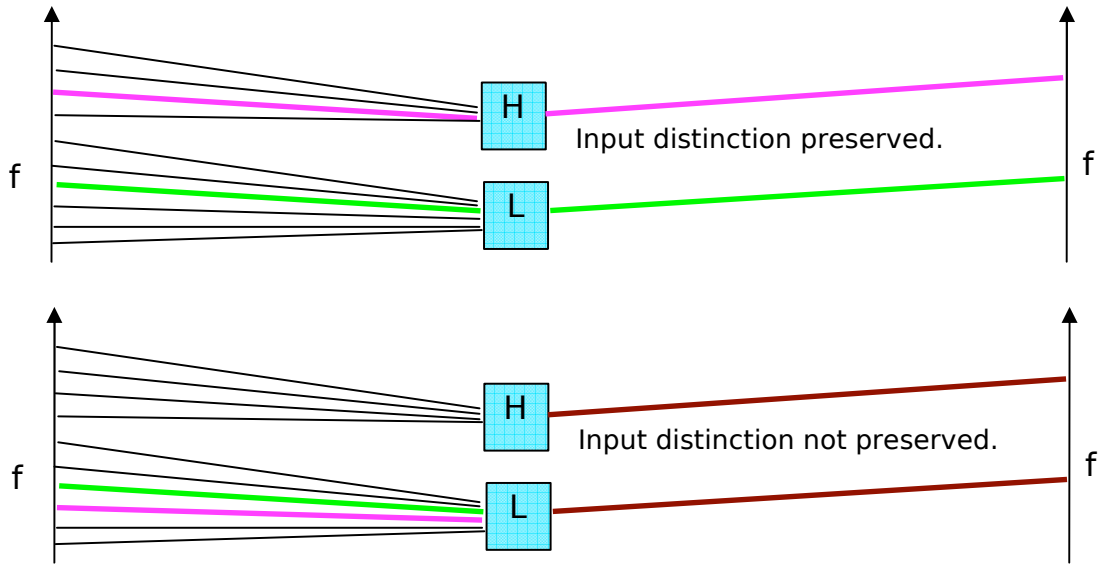


*Figure 8: If the memory representation is discrete, then only some input distinctions are preserved into the subject's mimicry output. The distinctions that are preserved are those that change the memory representation from one phonological entity to another. In the figure, the coloured lines show a pair of input stimuli (left). In the upper subfigure, the input distinction is preserved to the output because one activates **H** and the other activates **L**. In the lower sub-figure, both possible inputs (coloured/grey) lead to the same memory state, so the outputs will be identical, both produced from **L**.*

Figure 8 shows that distinctions between phonological entities are preserved but not input distinctions that produce the same phonological entity. In other words, any inputs that yield the same memory representation will produce the same output; distinctions within those sets are lost.

This behaviour is a general property of many systems and can be derived from Information Theory (e.g. Gray & Neuhoff, 2000, and references therein) as discussed in (Kochanski, 2006). It can be summarized as follows: *the memory representation must be complex*

enough to store all distinctions that are preserved to the output. Information Theory is well established and is the ultimate basis for all modern communication technology; so, this result can be derived with mathematical rigour, though one needs to be careful about the definitions involved.[5]

2.3 Summary of Hypotheses

The two hypotheses yield different predictions about which input distinctions people will be able to mimic reliably. This is exactly what is wanted because it will allow us to disprove one or the other hypothesis.

Equally important, we have a general principle that the memory representation must be able to store all the distinctions that people can mimic. This gives us a way to set a lower limit to the complexity of the memory representation of intonation based on observations of human behaviour. This allows us to experimentally measure at least one property of phonological entities.

3.0 Experiments on the Intonation of Speech

The main experiment discussed in this work have been reported in (Braun et al., 2006). The goal of this paper is not to present that work again, but rather to interpret it in the light of Hypotheses 0 and 1 to see what can be learned about human memory for intonation.

[5] Information theory is normally applied to long messages where each can be interpreted in isolation. Applying it to human speech implies that one must consider a "message" to be a sequence of speech that is long enough so that any context outside the sequence is relatively unimportant. In practise, this means that messages should be at least a sentence long (and possibly much longer under some circumstances). Specifically, it should be noted that the figures are schematic, and should not be interpreted to suggest that individual fundamental frequency values form a valid message from the viewpoint of Information Theory.

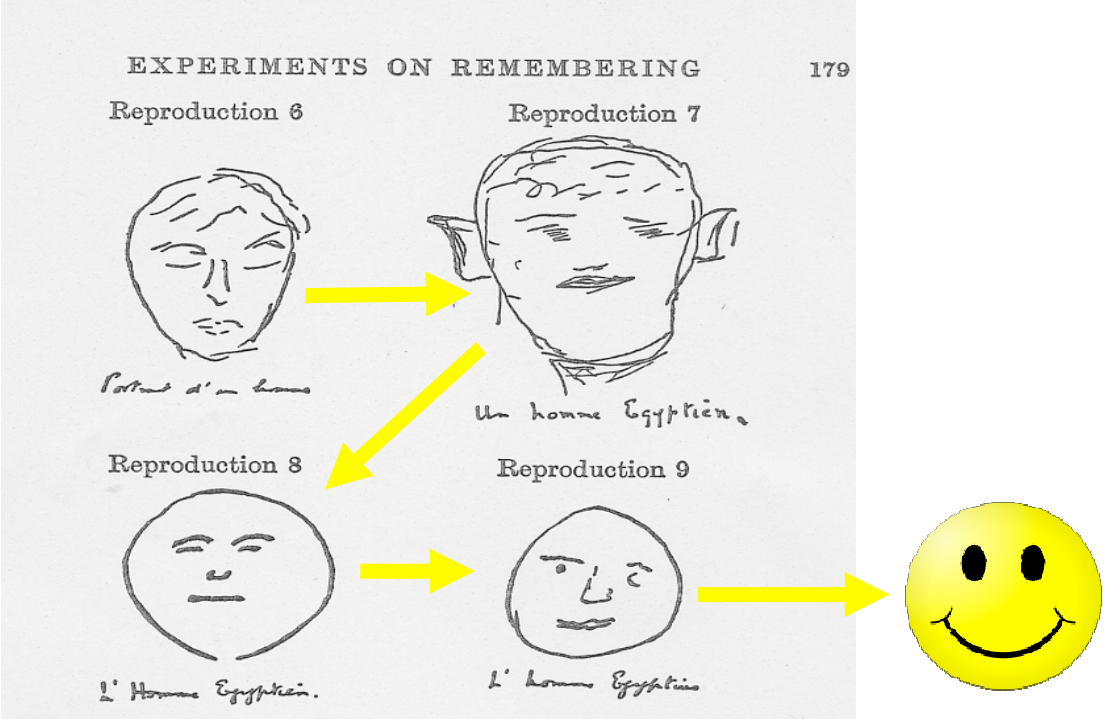


Figure 9: Bartlett's experiments on memory and mimicry of drawings. One of the more common changes was simplification. Continued simplification of a face could potentially lead to something like the modern "Smiley."

The Braun et al experiment was inspired by Bartlett 1932, Pierrehumbert & Steele 1989, and Repp & Williams 1987. Bartlett conducted a mimicry experiment on images, with a group of subjects. The first subject would be (briefly) shown a drawing, and then would be asked to sketch it. In turn, that drawing would be briefly shown to the next subject, et cetera. Bartlett found a variety of changes in the drawings, but one of the more common changes was simplification (Figure 9). If one extrapolates the simplifications forward, one might well obtain something like the modern smiley, a maximally abstract representation of the human face.

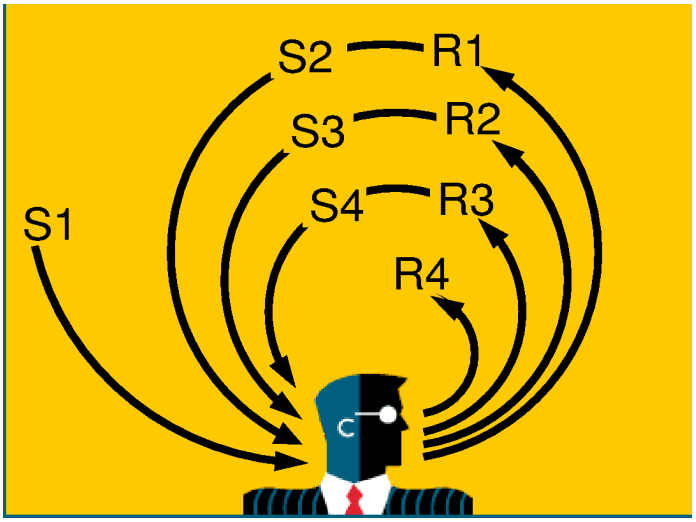


Figure 10: The general plan of the Braun et al mimicry experiment. Subjects were asked to imitate the speech and melody of each sentence, but to use their own voice. The first stimulus, S1, was synthesized to match the subject's normal pitch range. Further stimuli (S2, …) were the subject's own responses (after mild processing).

The Braun et al. experiment studied intonation contours rather than drawings, and it simplified the experiment by using only a single subject. (The experiment ran in blocks of 100 utterances, presented in random order, so that the subject would not be able to follow any particular utterance from iteration to iteration.) Figure 10 shows a schematic of the stimulus flow.

Following an utterance from one iteration of the Braun et al experiment to the next, one sees a combination of utterance-to-utterance variation and systematic change from one iteration to the next. A sample is shown in Figure 11. The question arises then, is this a secular decrease or does it have a target? A secular decrease might imply nothing more interesting than imperfect mimicry in that the subject has a tendency to produce speech with a frequency slightly lower than whatever he or she hears.

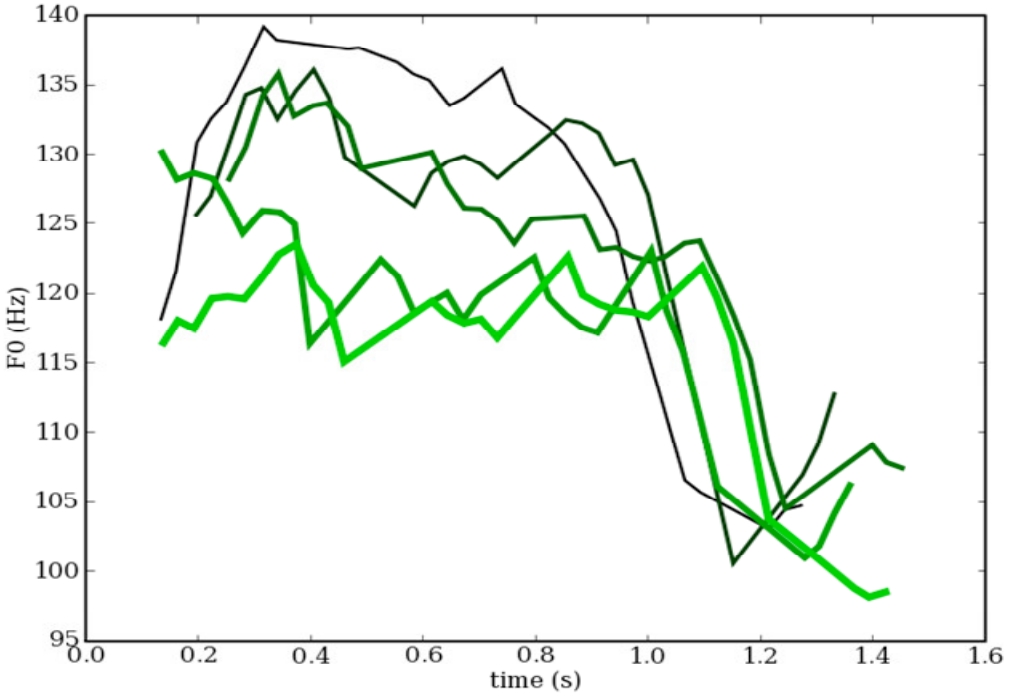


Figure 11: Stimulus 1, then Responses 1 ... 4 of the Braun et al mimicry experiment (dark and narrow → grey and broad lines, respectively). The horizontal axis is time in each utterance and the vertical axis is the fundamental frequency of the speech. At t=0.8 seconds, the utterances are in order from S1 at top down to R4 at bottom. In the central, relatively flat region, there is a systematic decrease in fundamental frequency.

The question can be answered by plotting the combined distribution of frequency measurements from all utterances and watching the distribution change from iteration to iteration. A downward shift would simply cause the histogram to move downward from one iteration to another. Instead, the histogram gradually became narrower and multimodal. Figures 12-15 show the intonation of a block of 100 utterances changing over four iterations. Figure 12 shows the stimuli (S1) which are linear combinations of three normal intonation contours. The feature to notice in the plot is that near the middle of the utterance (for τ between 0.3 and 0.6) the distribution of frequency measurements is broad and flat: in the stimuli, all frequencies are roughly equally probable.

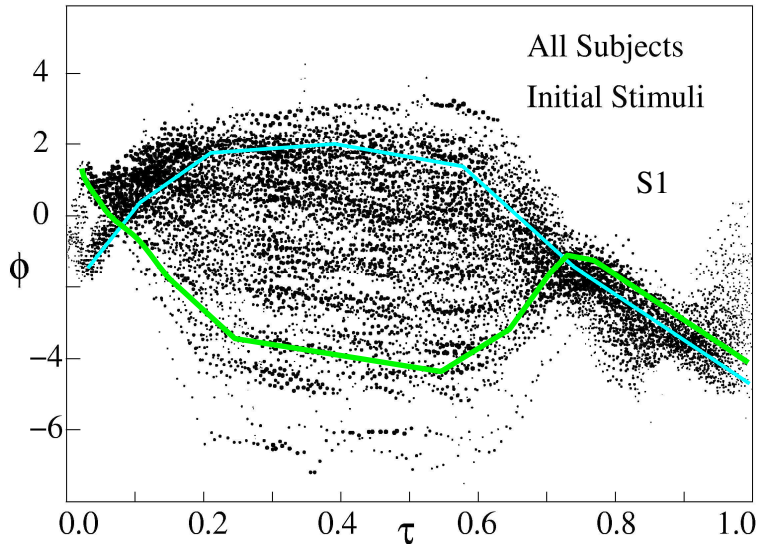


Figure 12: The distribution of all initial stimuli. Data from one hundred utterances are superimposed to make the plot. Each dot corresponds to one fundamental frequency measurement from one utterance. The coloured lines trace out two of the 100 utterances. The horizontal axis (τ) is normalized time and the vertical axis (φ) is frequency in semitones relative to the subject's average frequency.

However, after just one mimicry (iteration), the situation has changed. Figure 13 shows R1/S2. The variability of the fall where τ is near 0.8 has decreased, and the upper edge in the middle of the utterance has become denser.

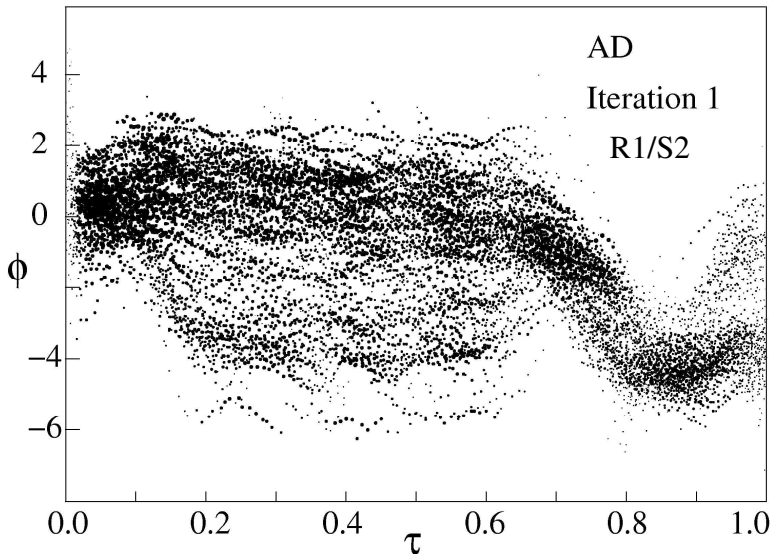


Figure 13: scatter plot of frequency measurements for subject AD after utterances have been mimicked once. Plotted as per Figure 12.

After a second mimicry (Figure 14), the upper edge, near the middle of the utterance is becoming a density peak about 1 semitone above the

speaker's average frequency, and another clump is forming, about three semitones below the speaker's average frequency. Another effect is that relatively few samples are found in between the clumps: the region where τ is near 0.25, one to two semitones below the speaker's average, is becoming sparse.

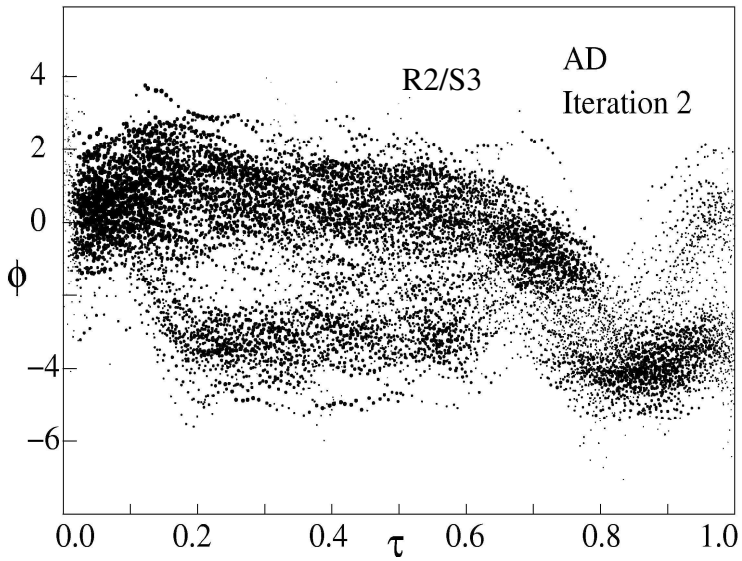


Figure 14: scatter plot of fundamental frequency measurements after two mimicries. Plotted as per Figure 12.

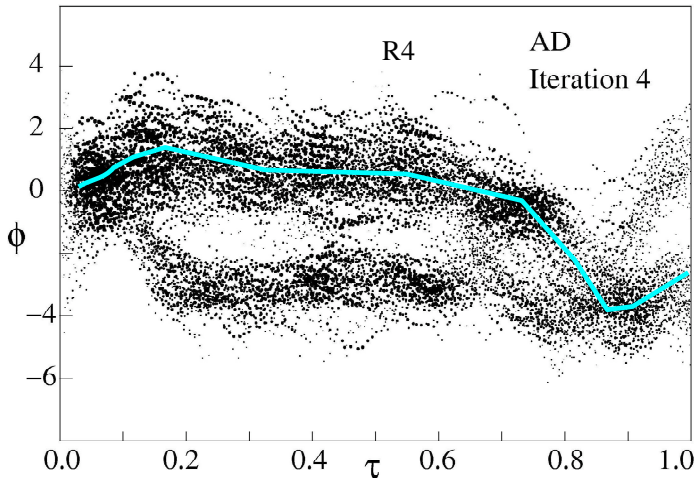


Figure 15: The scatterplot at the end of the experiment, after four mimicries. Plotted as per Figure 12. The blue line marks one utterance's intonation contour.

Finally, after four mimicries, Figure 15 shows that two separate groups of intonation contours have formed in the central part of the utterance. Utterances with intermediate frequencies have almost disappeared.

What is happening is that every time an utterance is mimicked, the produced intonation contour is biased towards one or the other of these two groups of contours. Figure 16 shows this by comparing an early and a late production. Aside from a certain amount of random variation, the contours approach either a high target or a low target, whichever they are closest to. In mathematical terms, from one iteration to the next, the contours are mapped towards one of these two attractors.

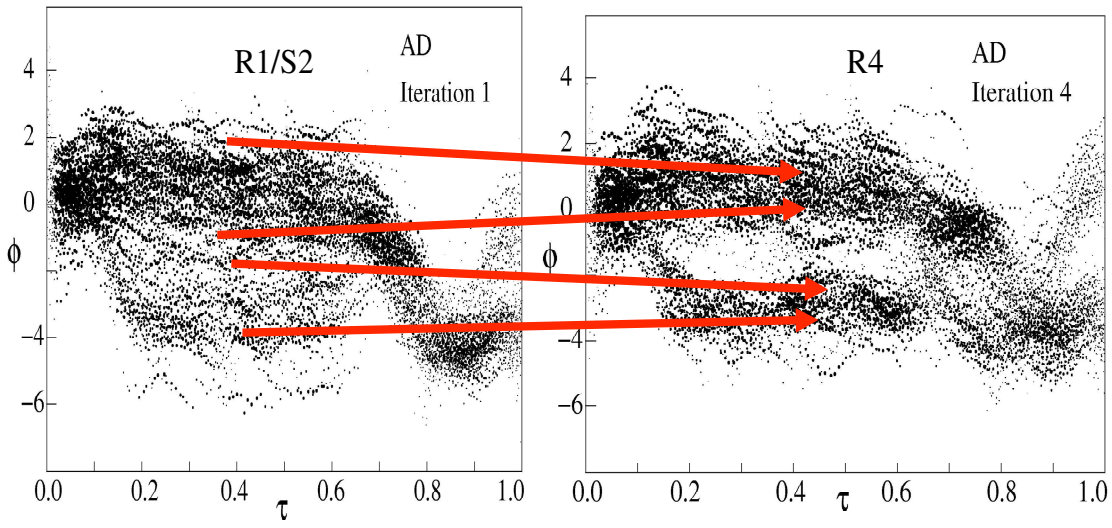


Figure 16: changes in the scatter plot between early and later productions in the mimicry experiment. From iteration to iteration, contours follow the red arrows: the highest stimuli are mimicked lower, the lowest are mimicked higher, and contours in the middle move up or down, depending on whether they are closer to the high group or the low group.

3.1 An Engineering Analogy

There is a close engineering analogy to this behaviour. It is called the "Static Discipline" and is taught in undergraduate electronic design classes. It is an essential design rule that makes digital electronics possible. One is tempted to suppose that an equivalent design rule evolved within the brain.

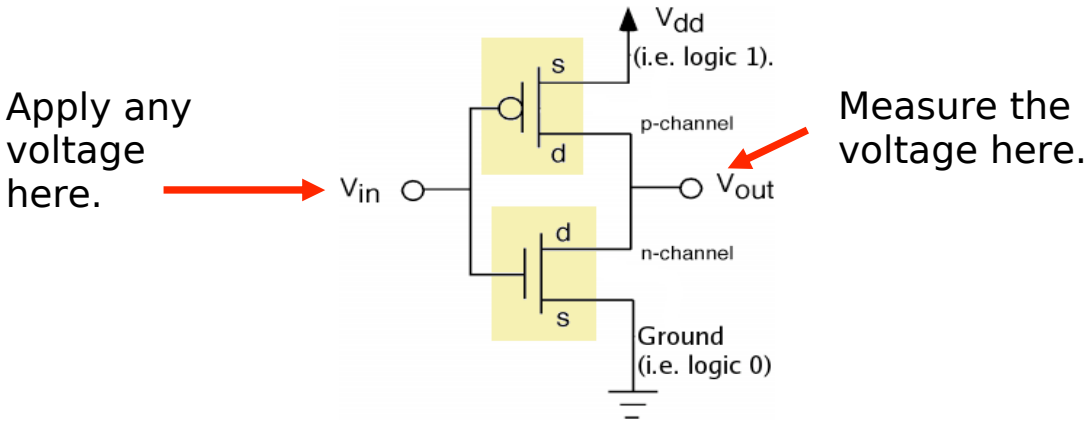


Figure 17: C-MOS inverter circuit.

Consider the simplest logic gate, an inverter (Figure 17). It is typically constructed out of two CMOS transistors, one N-channel and one P-channel. The two transistors have complementary properties so that when the input voltage is high, the lower transistor conducts and the upper transistor is off. As a result, the output voltage is pulled low. When the input voltage is low, the top transistor is turned on, the bottom one is turned off and the output voltage becomes high.

This device relates each input voltage to a corresponding output voltage. Mathematically, there is a mapping between the input and the output (Figure 18). There is also a small amount of noise, which plays the same role as utterance-to-utterance variation in language. Both sub-figures display the same input-to-output mapping; they just show it in different ways.

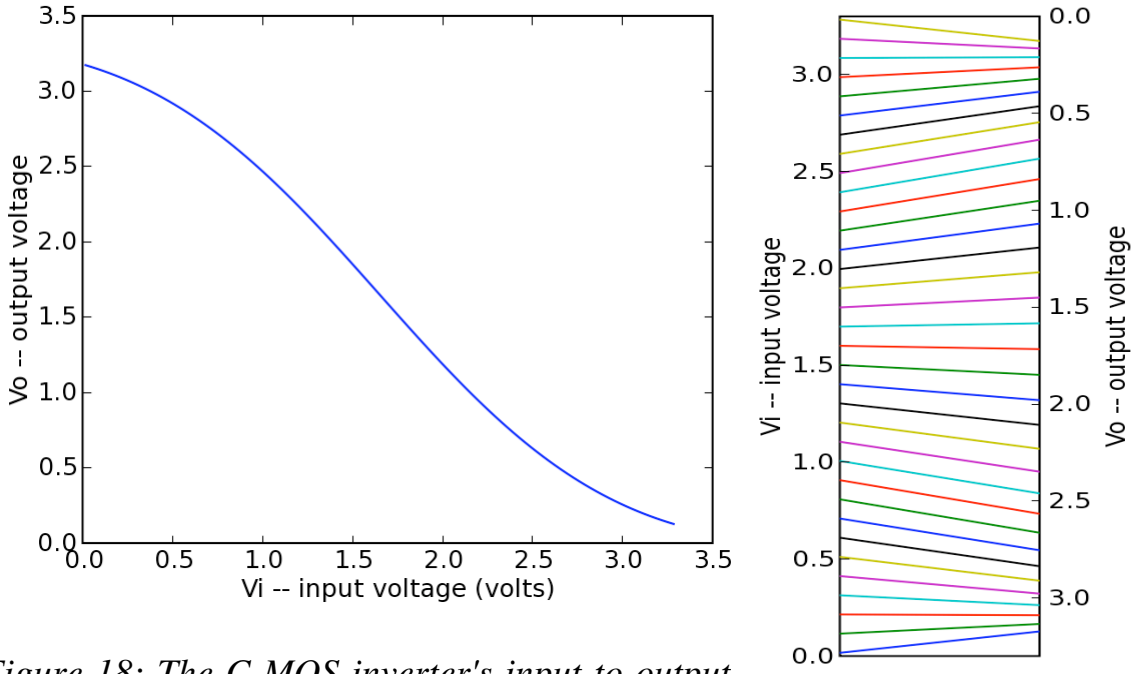


Figure 18: The C-MOS inverter's input-to-output mapping. The input voltage is placed on the left axis, and the output voltage is on the right axis. Lines connect corresponding input/output pairs. The mapping is compressive near the top and bottom where a given range of input voltages yields a smaller range of output voltages.

The static discipline requires that any digital logic element should have two regions where the mapping is compressive: one near zero volts input, and one at relatively high voltage. These compressive regions are important not so much in the context of a single logic gate, but rather for their effect on a large system composed of many logic gates connected in series. Computers, of course, are large and complex systems where any signal that is fed into one of the pins of a processor may propagate through at least dozens of logic gates before it comes out on some other pin. So, we can idealize a computer as a string of C-MOS inverters (Figure 19).

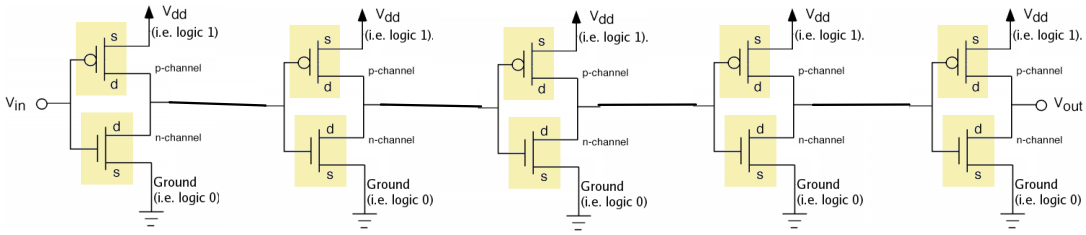


Figure 19: A string of C-MOS inverters. We will imagine putting a voltage on the first input, then measuring the voltage at all intermediate points in addition to the final output.

Each C-MOS inverter has a mapping from its input voltage to its output voltage. Likewise, every iteration of the Braun et al mimicry experiment reveals a mapping from the fundamental frequency of the stimulus to the fundamental frequency of the mimicked response. We can make an analogy between the two.

At this point, we have the tools needed to simulate a string of C-MOS inverters or (equivalently) a sequence of iterated intonational mimicries. The crucial ingredient is Figure 18, the mapping from input to output of each stage. One simply considers the stages (or iterations) one at a time, applying the Figure 18 mapping at each step. Since the output of one stage is the input to the next, we just take the output of the first mapping and use it as input for the second, then take the output of the second and use it as input for the third, ad infinitum. The result of this repeated mapping is shown in Figure 20. Each vertical line corresponds to the output of one inverter and the input of the next (or, by analogy) the response to one iteration of the mimicry experiment and the stimulus for the next.

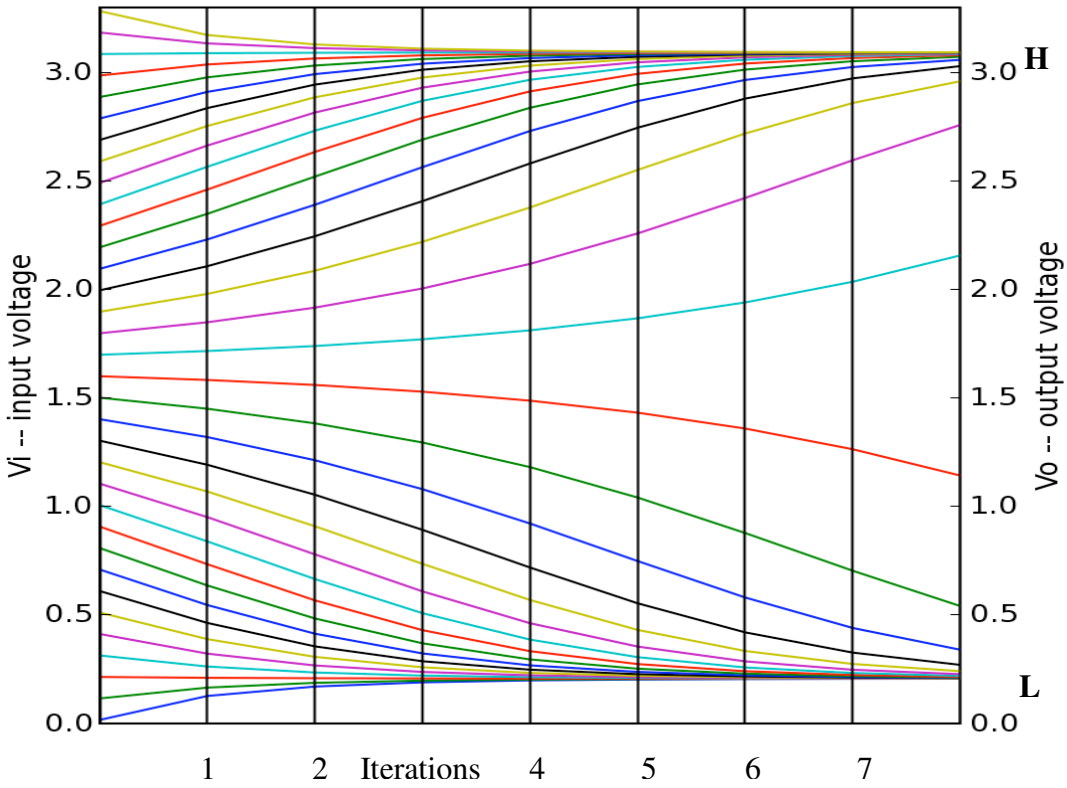


Figure 20: voltages within a string of C-MOS inverters. The output of each inverter drives the input of the next.

As can be seen toward the right side of Figure 20, this iterated system has an interesting behaviour: after enough stages, almost any input voltage gets mapped to either 0.2 V or 3.1 V. The system gradually becomes digital as it is made longer. This is the result of a series of compressive mappings. Each stage compresses voltages near 3.1 V together and it also compresses voltages near 0.2 V closer together. Conversely, the mapping of Figure 18 magnifies voltage differences near 1.7 V: different voltages near the mid-range get pushed further and further apart. In the limit of an infinite string of inverters, any input would yield an output voltage that could be precisely represented as a digital **H** or **L** state. This is an example where a discrete, digital system appears as an emergent property from analogue/continuous components.

Voltages between **H** and **L** do not stay there, they move away from the centre towards either the high attractor or the low attractor, whichever is closer. This result is analogous to what is seen experimentally in Figures 12-15, and it seems fair to interpret those figures as the result of an iterated mapping with two compressive regions. Each compressive region, after a few iterations, yields a dense group of intonation contours.

The static discipline is a design rule, and as such it has a purpose. The purpose is to force a complex system built out of these inverters to have two attractors. This allows the system to be thought of as digital, with discrete states. In a system built under the static discipline, there is no way to incrementally convert a low voltage into a high voltage by small changes because each C-MOS inverter will pull the voltage back towards the nearest attractor. This return toward the attractors is crucial in real systems because it means that small amounts of noise will not cause errors. Even if each stage of the system adds noise that pushes the voltage away from the attractors, the next stage will un-do the noise, pulling the voltage back towards the attractors. It is tempting to say that this is the mechanism by which discrete phonologies emerge from a continuous/analogue brain. It is tempting to see this as a victory for Hypothesis 0. While that might be the correct conclusion for segmental phonology or words, we will see that it is not true for intonation.

4.0 Discussion

4.1 Intonational Attractors are Slow

We saw already that it took several iterations of the mimicry experiment for the intonation contours to approach the high and low attractors. This can be quantified by measuring how strongly bimodal each scatter-plot of fundamental frequency is (e.g. Figure 15). Without going into the details (which can be found in Braun et al. 2006), the results can be seen in Figure 21. That figure is the answer to the question "How strongly bi-modal is the frequency distribution?" The vertical axis (valley depth) measures how empty is the middle of the scatterplot (e.g. Figure 15), relative to the density of fundamental frequency measurements near the high and low attractors. A value of zero implies that there is only a single maximum (not bimodal at all); values greater than one indicate two well-separated peaks with larger values indicating increasing separation.

The gradual increase in valley depth from iteration to iteration implies a slow and gradual separation of the scatter-plots into two peaks, over the course of several iterations. Recall that each iteration is a complete pass through the human subject involving on the order of 100 stages where one neuron triggers another[6], so if we equate a logic gate with a few

[6] A typical interval between neuron firings is about 10 milliseconds, and these intonation contours were remembered by the subjects for about 1 second. Thus, a memory of an intonation contour in the experiment is preserved across about 100 generations of neuron firings.

neurons, the rate of convergence per group of neurons (i.e. per logic gate) must be small indeed.

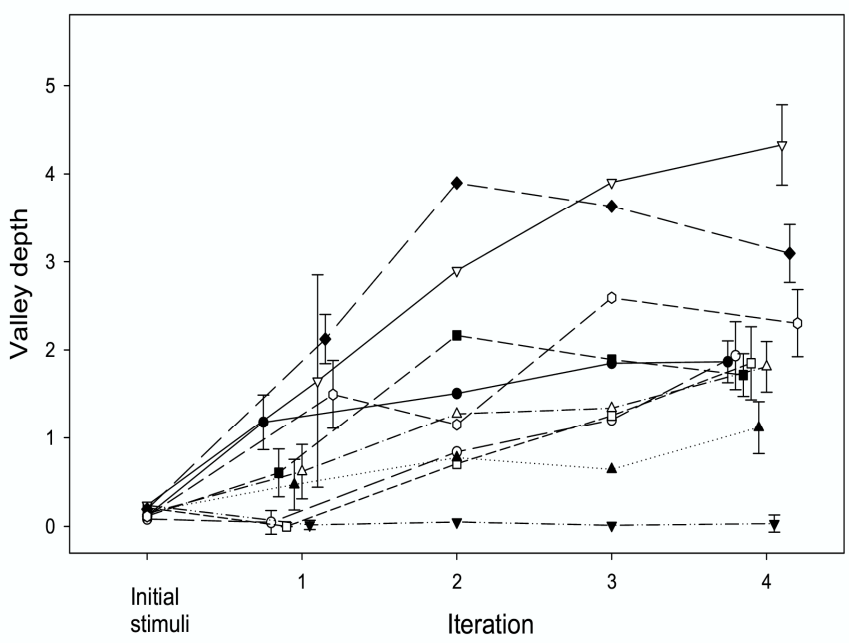


Figure 21: A measurement of the bimodality of f0 near the centre of the utterances. The horizontal axis shows the number of experimental iterations. Each curve corresponds to a different experimental subject. The gradual increase in valley depth values implies a slow and gradual separation of the scatter-plots into two peaks.

More practically, if it takes roughly four iterations for the fundamental frequency to converge toward a pair of almost-discrete states, then one certainly should not expect digital behaviour to emerge on a single trip between the ears and memory. The convergence that we see is approximately ten times too slow for intonational phonology to be accurately represented by a discrete memory representation.

4.2 What is stored in the memory representation?

One should also consider which distinctions the subjects can mimic. Recall that the memory representation must be at least rich enough to store all the distinctions that can be mimicked. A comparison of Figures 12 and 13 shows that subjects are able to mimic fine phonetic detail fairly accurately. Not only can subjects reproduce the contours that happen to be near the attractors, but they can reproduce the extreme contours and the contours between the attractors, too. So, all this detail is stored in memory and is plausibly part of the phonological entities.

Hypothesis 1 is actually the better approximation to our data, at least over a single iteration. All input distinctions are carried through to the output, although some distinctions may be emphasized and others reduced. Figure 22 shows one reasonable interpretation for mimicry behaviour. This model takes the view that the memory representation is essentially an acoustic memory, but biased slightly toward one or another special intonation contours. If interpreted literally, this model suggests that intonation contours might be stored in something like the phonological loop (Baddeley, 1997) and the gentle bias toward the attractors is due to interactions with something stable outside the phonological loop.

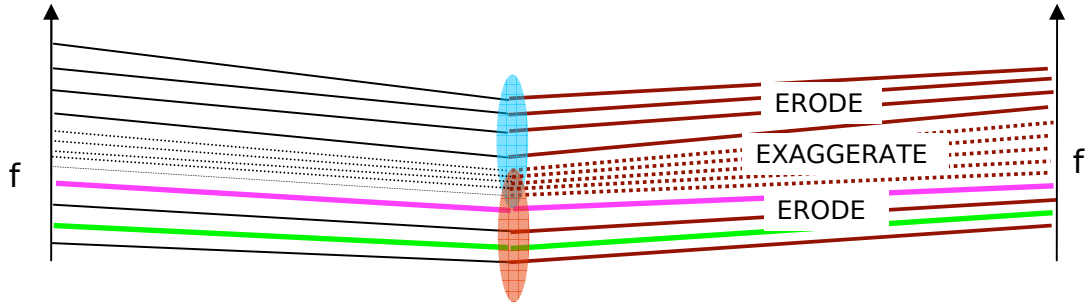


Figure 22: A plausible interpretation of the mimicry results, corresponding to an intermediate case between Hypothesis 0 and Hypothesis 1. All distinctions are preserved, but some are partially eroded and others are emphasised.

Another reasonable interpretation that is closer to the traditional phonological approach is to consider the memory to be a discrete phonological symbol along with substantial amounts of fine phonetic detail. This is a sort of "decorated object", shown in Figure 23. However, this interpretation does not carry a license to do traditional discrete phonology. The fine phonetic detail exists, stored in the memory representation, so one cannot arbitrarily ignore it. A proper phonological theory would include it, would involve it in the computations, and presumably, the fine phonetic detail would affect the answer generated in some phonological computations.

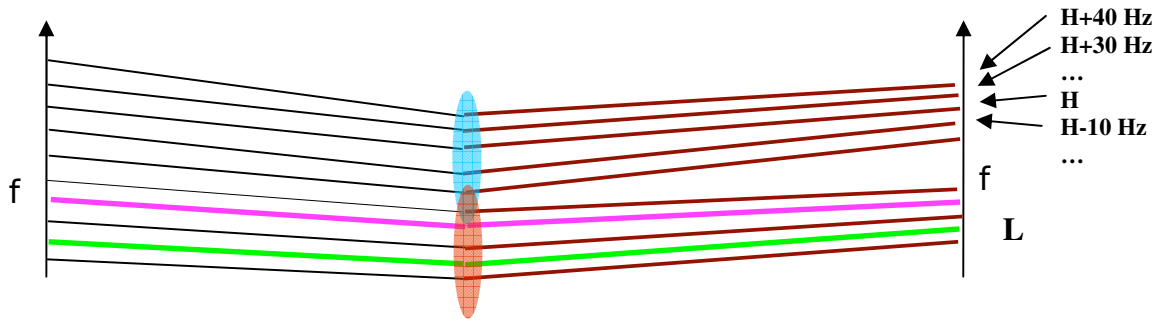


Figure 23: A plausible interpretation of mimicry results in terms of decorated categories or decorated symbols.

Given that some fine phonetic detail is stored, the onus is on the phonologists to show that their computations are useful descriptions to human language behaviour and that ignoring the phonetic detail is a valid approximation. Any phonological theory that uses discrete objects carries a implicit assumption that such discrete representations actually exist in the mind or at least are a good description of how the mind works. This is a strong assumption and needs to be justified, otherwise the resulting theory is built on sand.

We know the fine phonetic detail is used because we can hear the detail when a subject mimics an intonation contour. Since the detail is in the memory representation and accessible to conscious introspection, it seems likely that the phonological processes of speech production do not limit themselves to using only the discrete part of a decorated object. Speech production uses both the discrete part and the fine phonetic decoration, and presumably other phonological processes do too. The challenge is on the theorists to re-cast phonology in terms of either of these interpretations.

5.0 Conclusion

A straightforward interpretation of results from mimicry experiments shows interesting, complicated behaviour. The existence of attractors in intonation and their similarity to common intonation contours suggests that something like intonational phonology exists. However, the approach toward the attractors is far too slow for discrete phonological categories to be a good approximation to the way humans actually remember and reproduce intonation. To the extent that discrete phonological entities exist for intonation, they have only a weak influence on actual behaviour.

Humans do not behave as if their memory representation of intonation were a few discrete states. Memory certainly captures a much richer set

of distinctions than two phonological categories, and a reasonable interpretation is that a substantial amount of detailed information about the intonation contour is stored in memory, available for processing. Further, this detailed information is actually used in the mental processes of speech production.

Acknowledgments

The beginnings of this work were initially funded by the Oxford University Research Development fund and in later stages by the UK's Economic and Social Research Council under grant RES-000-23-1094. Both were greatly appreciated. I also thank the editors for asking awkward questions that improved this work.

References

Baddeley, A. (1997). *Human Memory: Theory and Practice*. (Revised ed.). Hove, East Sussex: Psychology Press.

Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press.

Beckman, M. & Ayers Elam, G. (1997). *Guidelines for ToBI labeling*. Linguistics Department, Ohio State University. Available online: http://ling.ohio-state-edu/~tobi/ame_tobi/labelling_guide_v3.pdf.

Braun, B., Kochanski, G., Grabe, E. & Rosner, B. (2006). Evidence for attractors in English intonation. *Journal of the Acoustical Society of America*, *119 (6)*, 4006–4015.

Goldinger, S.D. (1992). Words and Voices: Implicit and Explicit Memory for Spoken Words. Ph.D. thesis, Indiana University. Also available as *Research on Speech Perception Technical Report 7*, from Indiana University Press.

Grabe, E., Kochanski, G., & Coleman, J. (2005). The intonation of native accent varieties in the British Isles – Potential for miscommunication? In K. Dziubalska-Kolaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene*. Linguistic Insights Series 21. Oxford, Berlin, New York: Peter Lang.

Grabe, E., Kochanski, G., & Coleman, J. (2007). Connecting Intonation Labels to Mathematical Descriptions of Fundamental Frequency. *Language and Speech*, *50 (3)*, 281-310.

Gray, R. M., & Neuhoff, D. L. (2000). Quantization. In S. Verdú (Ed.), *Information Theory: 50 years of discovery*. Piscataway, NJ: IEEE Press. Reprinted from *IEEE Transactions of Information Theory*, *4* (1998).

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J., & Batliner, R. (1996). Consistency of transcription and labelling of German intonation with GToBI. *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP)*, 1716–1719.

Gussenhoven, C. (1999). Discreteness and Gradience in Intonational Contrasts. *Language and Speech*, *42(2-3)*, 283–305.

Gussenhoven, C., & Rietveld, A. M. C. (1997). Empirical evidence for the contrast between Dutch rising intonation contours. In A. Botinis, G. Kouroupetroglou, & G. Carayiannis (Eds.), *Intonation: Theory, Models and Applications, Proceedings of an ESCA Tutorial and Research Workshop*, Athens, Greece, September 18-20.

Johnson, K., & Mullennix, J. W. (1997). *Talker Variability in Speech Processing*. San Diego: Academic Press.

Jun, S.-A., Sook-Hyang, L., Keeho, K., & Yong-Ju, L. (2000). Labeler agreement in transcribing Korean intonation with K-ToBI. *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP)*, Beijing, China.

Kochanski, G. (2006). Prosody beyond fundamental frequency. In S. Sudhoff, P. Augurzky, I. Mleinek, & N. Richter (Eds.), *Methods in Empirical Prosody Research*. Language, Context and Cognition Series. Berlin, New York: De Gruyter.

Kochanski, G., & Orphanidou, C. (2008). *Journal of the Acoustical Society of America, 123(5)*, 2780–2791.

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.

Ladd, D. R., & Moreton, R. (1997). The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*, *25*, 313–342.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology (1)*, 301–323.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonation contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds), *Proceedings of the Third International Conference on Spoken Language Processing (ICSLP)*, *2*, 123–126, Yokohama.

Pierrehumbert, J. & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, *46*, 181–196.

Repp, B. H., & Williams, D. R. (1987). Categorical tendencies in self-imitating self-produced vowels. *Speech Communication*, *6*, 1–14.

Yoon, T., Chavarria, S., Cole, J., & Hasegawa-Johnson, M. (2004). Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. *Proceedings of the ICSA International Conference on Spoken: Language Processing (Interspeech 2004)*, 2729–2732, Jeju, Korea.

Phonetic Variation in Spontaneous Speech: Vowel and Consonant Reduction in Modern Greek Dialects

Anastassia Loukina

Phonetics Laboratory, University of Oxford

Abstract

The paper looks at phonetic variation in spontaneous speech in Athenian, Cypriot and Thessalian Greek. It is shown that while casual fast speech in all three varieties showed reduction of unstressed vowels and consonant lenition, the extent of these processes varied between the varieties. Therefore it is argued that although variation in time and effort is generally language-independent, it may be realized differently even in several varieties of the same language. The similarities between Greek dialects and the neighbouring languages suggest that language contact along with other factors may have contributed to the expansion of one of the variants which was also common to other languages involved in the contact.

Keywords

Greek Dialects, Phonetics, Segmental Reduction, Variation

1.0 Introduction

Human speech is inherently variable. Perkell (1990) discusses two major reasons for within-speaker variation in phonetics: the variability of the motor control system and speakers' adjustment depending on listeners' need for clarity of articulation. Due to the nature of the speech organs no sound can ever be pronounced in exactly the same way. Speakers also have a certain degree of control over the time and effort they 'invest' in various articulatory gestures depending on the situation in which the communication takes place. Lindblom (1990) describes such adjustment in terms of output-oriented control, or 'hyperspeech', and system-oriented control, or 'hypospeech'. On the one hand, system constraints require limiting what Lindblom (1983) calls 'energy expenditure per unit time', that is speakers tend to minimize articulatory effort to the extent that is possible. On the other hand, output constraints

ensure preservation of sufficient contrast necessary for lexical access and successful communication. The interaction between these constraints creates a continuum from clear speech, which requires greater articulatory effort (cf. Perkell, Zandipour, Matthies, & Lane, 2002 for experimental results), to quick casual speech, which shows greater tendency towards hypoarticulation and segmental reduction.

This study focuses on several specific cases of hypoarticulation common to faster speech. In vowels, shorter duration may lead to formant undershoot or greater assimilation of vowel to the adjacent segments. The resulting changes in quality are traditionally described as "vowel reduction" (cf. Lindblom, 1963; Moon & Lindblom, 1994). For consonants, less effortful articulation is often preferred in casual quicker speech leading to lenition of consonants. For example, stop consonants may be pronounced with incomplete closure or without closure (cf. Kirchner, 2001); intervocalic phonetically voiceless consonants are often phonetically voiced, in order to avoid the effort of "turning off" voicing and then "turning it on" again (cf. Ohala, 1983).

Barry and Andreeva (2001) suggest that such tendency for articulatory reduction is universal and language-independent. They analyzed spontaneous speech processes in six European languages including Greek and argued that the similarities between them were greater than the possible differences. Thus they found that all languages showed reduction of intervocalic clusters, lenition of stops, centralization of unstressed vowels and syllable loss. They conclude that comparable reduction phenomena are universal for all languages, at least in the context of the European languages covered in their study.

In this paper I will look at spontaneous speech processes in three regional varieties of Modern Greek in order to establish whether such processes are subject to regional variation or they operate universally as Barry and Andreeva suggested. The three varieties chosen for this study are Cypriot, Athenian and Thessalian Greek. Cypriot and Thessalian Greek represent respectively South-Eastern and Northern Greek dialects and show different treatment of most regional features (cf. for example Kontosopoulos, 2001; Newton, 1972b; Trudgill, 2003). Athenian Greek was chosen in order to provide some benchmark data which would be as close as possible to a natural colloquial form of Standard Modern Greek.

1.1 Vowel reduction in Modern Greek dialects

According to published descriptions, in Athenian and Cypriot Greek, the distribution of vowels is not dependent on stress (cf. Mackridge, 1985; Newton, 1972a) and all vowels can occur both in stressed and unstressed position without much variation in quality (Arvaniti, 1999a).

In contrast, in Thessalian Greek, like in most Northern Greek dialects, [o] and [e] are rare in unstressed position and usually alternate with [i] and [u]; etymological high vowels /i/ and /u/ are often dropped in unstressed position (cf. Papadopoulos, 1926; Tzartzanos, 1909). Thus χωράφι 'field' pronounced [xorˈafi] in Athenian Greek in Thessalian Greek appears as [xurˈafj], παιδί 'child' Athenian [peðˈi] corresponds to [piðˈi].

Nevertheless, some cases of vowel reduction have also been attested in areas outside the traditional Northern dialects area. Theophanopoulou-Kontou (1973), in her study of fast speech rules in Standard Modern Greek, refers to 'laxing' of unstressed high vowels as one of the general rules of the Modern Greek koiné, which occurs in almost all speech styles. According to Theophanopoulou-Kontou, unstressed /i/ and /u/ in all environments become shorter and 'lose a part of their sonority'. Devoicing or loss of unstressed /i/ and /u/ in Standard Modern Greek were also reported in experimental studies by Dauer (1980a) and Arvaniti (1999b). Chatzidakis (1892) noted that unstressed /i/ between consonants is sometimes lost also in Southern Greek dialects, for example in Crete; however, it cannot be compared to loss of /i/ in Northern Greek where it is much more regular. Recently Eftychiou (2007) reported that lenition of close vowels is very common in Cypriot Greek, at least in utterance final position.

Recent acoustic studies of vowel quality in Standard Modern Greek have shown that the difference in quality between stressed and unstressed vowels in this variety may also be greater than it is usually believed to be. Baltazani (2005) and Nicolaidis (2003) found that Standard Modern Greek shows a tendency for centralization of unstressed vowels as well as devoicing or loss of high vowels. Fourakis et al. (1999) and Nicolaidis (2003) also found upward shift of the vowel space for unstressed/shorter vowels in Standard Modern Greek.

1.2 Consonant lenition in Modern Greek dialects

Lenition of stop consonants has been reported both in Cypriot and in Athenian Greek. Newton (1972a) describes Cypriot voiceless stops as 'voiceless, unaspirated and quite lenis'; he notes that lenition is especially common near vowels, sonorants and /z/. In an experimental study of Cypriot geminates, Tserdanelis and Arvaniti (1999) noticed that single stops and affricates were lenited in intervocalic position, while geminates were not. However further investigation (Arvaniti & Tserdanelis, 2000) did not support this finding and there was no consistent difference either in the root mean squared amplitude (RMS amplitude) or in the difference in amplitude between first and second harmonic at the onset of the

following vowel (adopted as an indicator of the lenis-fortis distinction). Contrary to this finding, in a recent experimental study on Cypriot Greek by Eftychiou (2007), /t/ was most often pronounced as partially voiced stop. Other realizations included fully voiced stop, approximant and voiceless stops.

In one of the first experimental phonetic studies of Standard Greek, Dauer (1980b) noted that intervocalic consonants (especially /s/ and /t/) in casual speech and at rapid tempo may be voiced or partially voiced. Dauer found that the duration of consonants in Standard Modern Greek could be affected by stress, but there was substantial variation between speakers. Although medial voiceless stops may have longer durations in stressed syllables than in the unstressed syllables, this was not necessarily the case. Intervocalic stops in causal speech were voiced more often in unstressed syllables and in consonants with shorter durations. Dauer (1980b) also notes that stops between open vowels may be lenited, but only in very casual speech. In an experimental study of Standard Modern Greek stop consonants, Botinis et al. (2000) found that voiceless stops showed variability from partly voiced to completely voiceless. Greek consonants in spontaneous speech were also analyzed in a detailed articulatory study by Nicolaidis (2001), who found variation in the degree of constriction and the overall degree of contact in the pronunciation of plosive [t], depending on its duration. There were also tokens of [k] with incomplete velar constriction. Both [t] and [k] were often partially or fully voiced in intervocalic position or between vowel and voiced consonant.

2.0 Data and methodology

The present study is based on a data sample extracted from spontaneous monologues of 21 speakers: 7 speakers for each of the three varieties. The recordings were made respectively in Cyprus, Athens and Thessaly (Karditsa). All speakers in Cyprus and Thessaly were natives of the area; Athenian speakers have lived in Athens at least since the 1950s and did not show any noticeable regional features in their speech. The speakers in all three regions were selected following the same criteria: at the time of the recording all of them were over 70 years old; most speakers only had primary education, none of them had complete secondary education. The speakers were interviewed by the author in informal settings and were not instructed about the choice of language.

The data sample consisted of the same disyllabic words which occurred most frequently in all three varieties. The most frequent words were identified on the basis of a word index compiled for all recordings on the basis of orthographic transcription. The index consisted only of

nouns, adjectives, verbs and numerals. The data sample includes all occurrences of the chosen words in the recordings when they were part of a continuous monologue. Cases when a token was pronounced in isolation were excluded since they often filled the hesitation pause and thus showed specific rhythmic patterns. Tokens where the quality of the recording did not allow further acoustic analysis were also discarded.

The durations were measured manually on the spectrogram and double-checked on the waveform following the conventions suggested by Peterson and Lehiste (1960). Vowel amplitude was measured at 1 ms intervals and the highest value (peak amplitude) was used for the amplitude analysis. Peak amplitudes were normalized by dividing the absolute peak amplitude of each vowel by the peak amplitude of the word in which the vowel occurred. Formant frequencies were measured using Wavesurfer[1] speech processing software and manually checked against the spectrogram for accuracy. The formant frequency closest to the middle of the segment was used for further analysis. To compare the combine effect of both formants on differences between stressed and unstressed vowels, the Euclidean distance between stressed and unstressed vowels (ED_{stress}) was calculated using the following formula

$$(1) \qquad ED_{stress}=[(F1_{stressed} - F1_{unstressed})^2+(F2_{stressed} - F2_{unstressed})^2]^{1/2}$$

where $F1_{stressed}$ and $F2_{stressed}$ are the formant frequencies of the stressed vowel and the $F1_{unstressed}$ and $F2_{unstressed}$ are the formant frequencies of the unstressed vowel.

3.0 Results

3.1 Vowel reduction

Stressed and unstressed vowels were compared in words with vowels of the same phonemic quality in both syllables (285 tokens). The comparison of vowels within the same word reduced the impact of such factors as speech tempo or sentence stress. Unfortunately, the number of the most frequent words containing unstressed /u/ was insufficient for any statistical analysis.[2] Where possible the results obtained for two vowels within the same word were compared with those for vowels which occurred in a similar phonetic context in different words, since vowel

[1] http://www.speech.kth.se/wavesurfer/.

[2] The very low incidence of /u/ was also noticed by Nicolaidis (Nicolaidis, 2003), who excluded /u/ from some of her analysis of Greek vowels in spontaneous speech (cf. also Dauer, 1980a).

quality may depend on adjacent segments (cf. Perkell, 1990; Stevens & House, 1963).

In some tokens there was no proper vowel in the acoustic signal, preventing the measurement of formant frequencies. In all dialects vowel loss affected unstressed /i/ in /spˈiti/ *'house'*, which was omitted in 74% of cases in Thessalian Greek, 14% of cases in Athenian Greek and 10% of cases in Cypriot Greek. This shows that the so-called 'vowel loss' in Thessalian Greek does not affect all high unstressed vowels even within a given word. Furthermore, it also agrees with observations by Chatzidakis and Arvaniti and experimental results by Eftychiou (2007) that vowel loss is also present in Southern dialects, though it is not as common as in the Northern dialect. Words with vowel elision were excluded from further analysis and not included in the number of tokens given above.

A difference between stressed and unstressed /i/ (see Figure 1) was found only in Thessalian Greek, where the unstressed vowel appeared to be more central (F1 345 Hz vs. 362 Hz, F2 1517 Hz vs. 1812 Hz, Wilcoxon signed ranks test, $p<0.05$)[3]. In Cypriot and Athenian Greek there was no significant difference between the stressed and unstressed vowel. Furthermore, the Euclidean distance between stressed and unstressed /i/ in Cypriot Greek was significantly smaller than in the other two varieties (151 vs. 497 in Athenian and 505 in Cypriot, Mann-Whitney U test, $p<0.001$).

In Cypriot Greek unstressed /a/ (see Figure 1) was found to be closer to the centre of the vowel space than its stressed counterpart in the same word (F1 647Hz vs. 668 Hz, F2 1607 Hz vs. 1591 Hz, Wilcoxon signed ranks test, $p<0.05$). In Thessalian Greek unstressed /a/ had a lower F1 (713 Hz vs. 738 Hz, Wilcoxon signed ranks test, $p<0.05$) and lower F2 (1555Hz vs. 1659Hz, Wilcoxon signed ranks test, $p<0.05$). This supports the observation by Chatzidakis (1892) that the so-called vowel reduction in Northern dialects affects not only /o/ and /e/, but also /a/. There was no difference in formant frequencies of stressed and unstressed /a/ in Athenian Greek.

Comparison between stressed and unstressed /a/ in different words but in a similar phonetic context (after alveolars) revealed further differences. In Cypriot Greek comparison between stressed /a/ and unstressed /a/ after alveolar consonants showed a significant difference in F2 with the unstressed vowel being less retracted (1591 Hz vs. 1409 Hz, Mann-Whitney U test, $p<0.05$). Similarly in Athenian Greek unstressed /a/ had a lower F1; that is, it was slightly more centralized than stressed /a/ in the

[3] Although the mean F2 frequencies of stressed and unstressed /i/ differ by almost 300 Hz, this difference was not statistically significant (Wilcoxon signed-ranks test, $p=0.345$, cf. also quite large standard deviations for stressed and unstressed /i/).

same context (796 Hz vs. 660 Hz, Mann-Whitney U test, $p<0.05$). In Thessalian Greek unstressed /a/ had significantly higher F1 after velars than after alveolar: the difference between stressed and unstressed vowel was even greater if both vowels occurred after alveolars (738 Hz vs. 587 Hz for unstressed vowel, Mann-Whitney U test $p<0.001$). The mean formant frequencies of unstressed /a/ after alveolar consonants in Thessalian Greek (F1 587.26 Hz, F2 1433 Hz) were very close to the centre of the vowel space.

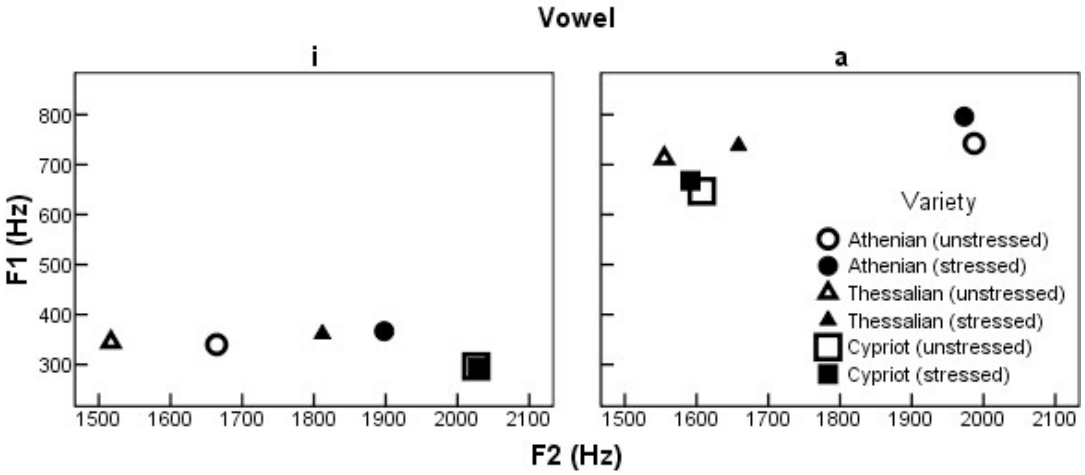


Figure 1: Mean formant frequencies (Hz) for stressed and unstressed /i/ (left) and stressed and unstressed /a/ (right) in all three varieties[4].

Unstressed /e/ (see Figure 2) differed from stressed /e/ in the same word in Cypriot and Thessalian Greek, although the difference was greater in Thessalian Greek (F1 ratio 1.3 vs. 1.09 Mann-Whitney U test, $p<0.001$). In Cypriot Greek there was a very small, but statistically significant difference in F2 between unstressed and stressed /e/ (1843 Hz vs. 1820 Hz, Wilcoxon signed ranks test, $p<0.01$), which was greater for unstressed vowels. As a result, unstressed /e/ was also more peripheral than stressed /e/. In Thessalian Greek as expected, there was a significant difference in F1 between stressed and unstressed /e/ (498Hz vs. 390Hz, Wilcoxon signed ranks test, $p<0.001$). Unstressed /e/ had a lower F1 and there was no statistical difference between the F1 of unstressed /e/ and stressed /i/. In Athenian Greek there were no significant differences between the quality of stressed and unstressed /e/ in a different phonetic context in the same word.

A comparison of stressed and unstressed /e/ when both were preceded by /p/ showed no difference in F1 or F2 between the stressed and

[4] In Cypriot Greek there was almost no difference between stressed and unstressed /i/ and therefore the markers overlap.

unstressed vowel in Cypriot Greek[5]. Unlike Cypriot Greek, the distinction between stressed and unstressed /e/ in Thessalian Greek was also preserved for vowels in different words, but identical phonetic context (Mann-Whitney U test, $p<0.001$). However, unstressed /e/ in Thessalian Greek differed in F2 from unstressed /i/ in the same phonetic context, with unstressed /i/ being more central (1517 Hz vs. 1728 Hz, Mann-Whitney U test, $p<0.05$). (It should be noted that the number of unstressed /i/ tokens was very small, since most unstressed /i/ in Thessalian Greek were elided). Notably, in Athenian Greek the F1 of unstressed /e/ was lower than in stressed /e/ in the same context (538 Hz vs. 479 Hz, Mann-Whitney U Test, $p<0.05$).

Cypriot Greek was the only variety where unstressed /o/ (see Figure 2) had a higher F1 than stressed /o/ (505Hz vs. 433 Hz, Wilcoxon signed ranks test, $p<0.05$); that is, unstressed /o/ was lower than stressed /o/. In Athenian Greek, (see Figure 2) there were no significant differences between the quality of stressed and unstressed /o/ in a different phonetic context in the same word. When the stressed and unstressed /o/ were compared in the same context, stressed /o/ had higher F1 than the unstressed /o/ (588 Hz vs. 508 Hz, Mann-Whitney U Test, $p<0.001$). As expected, in Thessalian Greek there was a significant difference between stressed and unstressed /o/ in both formant values (F1 495Hz vs. 425Hz, F2 1269 vs. 1095Hz, Wilcoxon signed ranks test, $p<0.001$). Unstressed /o/ had lower F1 and F2 and was more distant from the centre of the vowel space. This confirms that /o/ in Thessalian Greek is higher when in unstressed position. Nevertheless, there was a significant difference (Mann-Whitney U test, $p<0.001$) between the F1 of unstressed /o/ and stressed /u/. However, this difference is related to the difference in phonetic context between unstressed /o/ after a preceding velar consonant and stressed /u/ after a preceding labial consonant). Comparison of unstressed /o/ with the data for stressed /u/ in the same phonetic context showed no significant difference in F1 between the two vowels in this variety of Greek (Mann-Whitney U test, $p=0.879$). However, the F2 of stressed /u/ was lower than for unstressed /o/ (1041 Hz vs. 1438, Mann Whitney U test $p<0.001$); that is, unstressed /o/ was not quite as grave as stressed /u/. Due to the lack of tokens with unstressed /u/, it is not possible to say at this stage whether this difference would be preserved for unstressed /o/ and /u/.

[5] It should be noted that there was no significant difference in F2 of unstressed /e/ in different contexts.

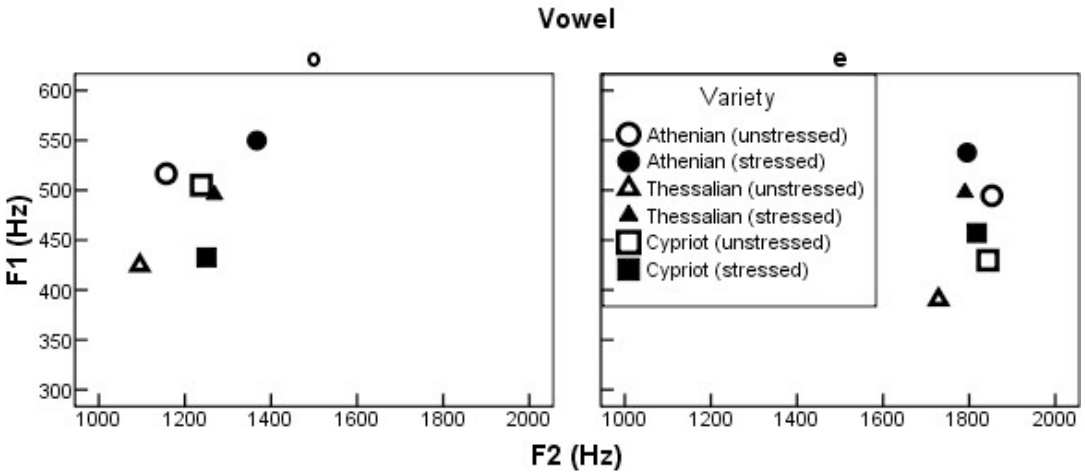


Figure 2: Mean formant frequencies (Hz) for stressed and unstressed /e/ (right) and stressed and unstressed /o/ (left) in all three varieties.

Acoustic analysis of formant frequencies of stressed and unstressed vowels in Thessalian, Cypriot and Athenian Greek has confirmed some earlier impressionistic descriptions. It has provided acoustic evidence for the so-called 'vowel raising', which is considered to be the most typical feature of the Northern Greek dialects. In Thessalian Greek unstressed /e/ differed significantly in F1 from the stressed /e/ and was similar to both unstressed and stressed /i/. Unstressed /o/ also had a lower F1 than the stressed /o/ and in a similar phonetic context was not distinguished from the stressed /u/. Both vowels in unstressed positions were higher than their stressed counterparts. However, it was found that in some cases mid vowels were preserved in unstressed position, that is vowel raising applies frequently, but not universally. Acoustic analysis showed that the two other vowels, /i/ and /a/, also differed in stressed and unstressed position, but in this case the unstressed vowel was more central than the stressed one (markedly for /i/, only slightly for /a/). In accordance with the previous descriptions, unstressed /i/ was elided in 74% of cases.

It was found that in Athenian Greek unstressed /e/ and /o/ had a lower F1 than the corresponding stressed vowel, but this difference was not consistent and depended on the phonetic environment. Unstressed /a/ in Athenian Greek also had a higher F1 than the stressed /a/.

In Cypriot Greek, unstressed vowels also differed from the stressed vowels, with the exception of /i/. The distance between stressed and unstressed vowels in Cypriot Greek was much smaller than in Thessalian Greek and unstressed vowels remained significantly different from each other. Notably, unstressed /o/ had a higher F1 and was more central than stressed /o/ only in Cypriot Greek. In the other two varieties unstressed /o/ had lower F1 than stressed /o/. Comparison of Euclidean distances between stressed and unstressed vowels for all varieties showed that the

stress effect was significantly smaller in Cypriot Greek than in the other two varieties (Cypriot Greek 165 Hz vs. Athenian Greek 321 Hz, Thessalian Greek 265 Hz, Mann-Whitney U test $p<0.001$).

Analysis of the correlations between vowel quality, vowel duration and vowel amplitude confirmed previously noted differences between stressed and unstressed vowels. It was found that in Athenian and Thessalian Greek, the higher F1 frequency of stressed /e/, /o/ and /a/ was correlated with longer duration and greater amplitude in comparison to unstressed vowels. Furthermore, in Athenian Greek variation in F1 of unstressed /e/ and /o/ was negatively correlated with duration; that is, the F1 of longer vowels was closer to the F1 of the corresponding stressed vowels. In Thessalian Greek this was the case for unstressed /a/, while the quality of unstressed /e/ and /o/ appeared to be independent of duration. Unlike Athenian and Thessalian Greek, Cypriot Greek exhibited little correlation between F1 and duration, but shorter vowels appeared to be more central than longer vowels.

3.2 Consonant lenition

Lenition of stop consonants was attested not only in Cypriot, but also in Athenian and Thessalian Greek. The data sample contained 628 tokens where stop consonants occurred in word-initial, although not necessarily phrase-initial, position. There were 355 occurrences of /p/, 215 occurrences of /t/ and 57 occurrences of /k/.

The percentage of cases when these consonants were pronounced with or without closure is shown in Table 1. In all varieties voiceless plosives were sometimes pronounced without closure. In most cases fricativization was accompanied by voicing, while stop consonants were usually voiceless. Though word-final /t/ was sometimes pronounced as a voiceless fricative, no such pronunciation was attested for word-initial /t/ in this data sample. Pronunciation with closure and voicing (voiced stop) only occurred in about 5% of cases in each variety, mainly after nasals and sometimes after vowels. Therefore in what follows I will focus on the distinction between voiceless stops and voiced fricatives or approximants.

Table 1: *Percentage of tokens pronounced with or without closure in three varieties of Greek.*

	/p/		/t/		/k/	
	Stop	No closure	Stop	No closure	Stop	No closure
Athenian	80.2%	19.8%	80.5%	19.5%	85.7%	14.3%
Thessalian	87.8%	12.2%	89.2%	10.8%	64.7%	35.3%
Cypriot	47.7%	52.3%	61.0%	39.0%	37.5%	62.5%

Table 1 shows that while in Athenian and Thessalian Greek all consonants were usually pronounced with closure, in Cypriot Greek the pronunciation without closure occurred as often or more frequently as the stop pronunciation (except for /t/, for which see below). The observed difference between Athenian and Thessalian Greek on the one hand and Cypriot Greek on the other hand was significant for all three consonants (chi-square test, $p<0.001$ for /p/ and /t/ and $p<0.05$ for /k/).

Interestingly, the pronunciation of these consonants did not depend on whether the token occurred after a pause or phrase-medially. Only in the case of /t/ in Cypriot Greek was the consonant more often lenited in phrase-medial position than phrase-initially (54% vs. 75%, chi-square test, $p<0.05$). In phrase-medial position, Cypriot Greek /p/ and /t/ were more often lenited in intervocalic position than when preceded by the consonant (61% vs. 33%, $p<0.001$ for /p/, 90% vs. 0%, $p<0.001$ for /k/). However, it should be noted that there were only 14 tokens of /k/ which occurred in phrase-medial position.

In Cypriot (and possibly Thessalian) Greek, consonants were more often lenited before unstressed vowels than before stressed vowels (58% vs. 41%, Chi-square test, $p<0.001$ for Cypriot Greek and 20% vs. 10%, $p<0.05$ for Thessalian Greek). The different stress pattern of tokens with /k/ and /t/ explains why in both these varieties /t/ was more often pronounced with closure than /k/.

In all three varieties consonants before stressed vowels were longer than consonants before unstressed vowels (Mann-Whitney U tests, $p<0.001$ in all cases), which agrees with the results reported for word-initial stop consonants by Botinis et al. (2001).

In Cypriot Greek, consonants pronounced with closure in some cases were longer than consonants pronounced without closure: /p/ before unstressed vowel 105 ms vs. 72 ms. (Mann-Whitney U test, $p<0.01$), /t/ before stressed vowel 95 ms vs. 56 ms (Mann-Whitney U test, $p<0.01$). However, one cannot talk about a consistent overall difference in duration between the two pronunciations. The difference in duration between different pronunciations was not consistent in all positions, except that /p/

in all varieties was longer before stressed vowels than before unstressed vowels.

It was also found that in agreement with results reported by Eftychiou (2007) for controlled speech, duration of consonants in Cypriot Greek depended on whether the following vowel was elided or not. Analysis of a further 84 tokens of /t/ in the word /spˈiti/ *'house'* showed that in Cypriot Greek there was a significant difference in the duration of the consonant depending on whether the final vowel was omitted or not (65 ms if the vowel was absent vs. 17 ms if the vowel was present, Mann-Whitney U test, $p<0.001$). In other varieties there was no significant difference, but it should be noted that in Athenian Greek the vowel was never omitted if the consonant duration was longer than 87 ms, which suggests the opposite relationships from the Cypriot Greek. In Thessalian Greek, where the vowel was omitted more often than in the other two varieties, the consonant was in most cases pronounced as a voiced fricative although its pronunciations as a voiced stop or a voiced fricative were also attested.

4.0 Discussion

Casual fast speech in Thessalian, Athenian and Cypriot Greek showed reduction of unstressed vowels and consonant lenition. At the same time, the extent of these processes varied between the varieties. In Thessalian and Athenian Greek, stop consonants lenition was less common than in Cypriot Greek, though unstressed vowels were usually reduced. In Cypriot Greek vowel quality showed almost no dependency on stress, while most stop consonants were lenited. Thus the observed differences between the varieties show that while variation in time and effort is generally language-independent, it may be realized differently even in several varieties of the same language.

Vowel reduction in Athenian Greek and consonant lenition in Cypriot Greek have not been attested in studies based on laboratory speech in these varieties, although the lenis quality of Cypriot consonants is mentioned in impressionistic descriptions. Therefore there is little doubt that in this case the pronunciation observed in spontaneous speech is the result of gestural undershoot of the articulatory target. However, in some cases hypoarticulation may eventually lead to a change of articulatory target. For example, in Thessalian Greek high vowels are pronounced instead of etymological mid vowels even in slow clear speech. Therefore it can be argued that in Thessalian Greek the reduction of unstressed vowels is no longer the result of more casual pronunciation, but has acquired conventional status (for futher discussion see Loukina, in press).

The idea that linguistic change originates in variation is not new (cf. for example Janse, 2000). Ohala suggested that in order to factor out variation, listeners apply what he called 'corrective rules' (cf. Ohala, 1981; Ohala, 1989, 1993). Inappropriate application of such rules may eventually lead to under-differentiation or over-differentiation of contrasts (cf. also Weinreich, 1968). The specific details of this model may differ. Thus Faber (1992) talks about the different sensitivity of listeners to small variation in speech rather than misperception as in Ohala's model. Blevins (2004) emphasizes that sound change occurs only when the results of gestural reduction or strengthening are re-interpreted and reproduced as new sounds. However, all these models share two basic assumptions: the inherent variability of speech production and categorical nature of speech perception, which may lead to structural changes in the sound system of the given language.

The results of this study showed that phonetic variation in the three varieties of Greek has taken different directions. Labov (1986) suggested that different direction of variation may be due to differences in properties of the specific linguistic systems and their development. The observed differences in spontaneous speech variation between the three varieties of Greek may be juxtaposed with other traits of the sound systems of these varieties. In Cypriot Greek consonants in general seem to be more prone to various reduction and strengthening processes than vowels or than the consonants of Athenian or Thessalian Greek[6]. Apart from the lenition of stop consonants, Cypriot also has geminates, which from a phonetic point of view can be described as fortis and which generally resist lenition in spontaneous speech. There is no such contrast in Thessalian, nor in Athenian Greek. It may be noted that other dialectal phenomena of Cypriot Greek which have not been discussed here have also sometimes been described in terms of weakening and strengthening. Thus according to some descriptions Cypriot shows (or used to show) loss of intervocalic fricatives, which can be seen as another case of lenition. On the other hand, consonant+/j/ clusters in this variety show occlusivization (that is strengthening) of /j/ to palatal stop [c] (cf. Drachman & Malikouti-Drachman, 1996; Malikouti-Drachman, 1999). On the contrary, in Cypriot Greek there is little difference in quality between stressed and unstressed vowels, but stressed vowels often lack acoustic prominence (Loukina, 2008). In this respect Cypriot Greek differs from Thessalian and Athenian Greek, where vowels in unstressed

[6] Chatzidakis (1892) suggested that the regional sound changes result from co-articulation in the Southern Greek and from greater prominence of stressed vowels in the Northern Greek. Notably, my results corroborate Chatzidakis' theory, though I argue that originally changes in both vowels and consonants are conditioned by hypoarticulation.

syllables are subject to reduction, while stressed vowels are consistently associated with acoustic prominence. In stressed syllables, Thessalian Greek vowels consistently had greater intensity than the unstressed vowels. Unlike Cypriot Greek, stressed vowels in both Athenian and Thessalian Greek stressed vowels mostly retain their acoustic prominence in spontaneous speech (Loukina, 2008).

The observed co-occurrence of consistent acoustic prominence of stressed vowels and vowel reduction on the one hand and gemination, lenition and relatively weak prominence of stressed vowels on the other agrees with previous typological observations about lenition and vowel reduction as historical processes. Martinet (1952), in his classic article on lenition, links Celtic lenition to phonetically weak accent and presence of gemination. On the other hand, vowel reduction is often associated with strong accent (cf. Van Coetsem, 1996 for discussion and multiple references). It is possible that the frequent co-occurrence of these features may be a result of causal relations between them, but an exact model of such interaction has yet to be developed.

Another factor that might have contributed to spread of one of the phonetic variants is language contact. Vowel reduction similar to the one described for Thessalian Greek has also been reported for eastern Bulgarian (cf. Pettersson & Wood, 1985; Tilkov & Boiadzhiev, 1981; Zhobov, 2004), eastern Macedonian (Sussex & Cubberley, 2006), Aromanian (Beis, 2001; Caragiu-Marioțeanu, 1968; Katsanes & Dinas, 1990; Kramer, 1989; Lazarou, 1986) and possibly Judeo-Spanish (Gabinsky, 1992). There also exists some evidence that Balkan varieties of Turkish display vowel reduction similar to the one found in Northern Greek dialects and some other Balkan languages, but it can not be considered conclusive at this stage (Brendemoen, 2002; Sawicka, 1997). In contrast, studies of Cypriot Arabic, Cypriot Turkish and Armenian do not mention vowel reduction of the type found in Balkan languages (Borg, 1985; Georgiou-Scharlipp & Scharlipp, 1997; Gürkan, 1997; Saracoğlu, 1989; Vaux, 1998).

Similarly, the systems of stop consonants of the contact languages for the three varieties of Greek show a clear division between what can be called 'Balkan' languages (Bulgarian, Macedonian, Albanian, Arvanitika, Aromanian and Judeo-Spanish) and 'South-Eastern' languages (Turkish, Arabic and Armenian). A voicing contrast is found in all "Balkan" languages, which distinguish between series of voiceless unaspirated plosives [p], [t], [k] and voiced plosives [b], [d], [g], at least in word-medial position. The only exception is Arvanitika, which has lost the opposition of voiced stops and nasal+voiced stop clusters, most probably under the Greek influence (cf. Hamp, 1961; Sasse, 1991). There is some evidence for lenition of stop consonants. According to Zhobov (2004) in

some Bulgarian dialects voiced stops may undergo optional spirantization in intervocalic position. In Judeo-Spanish like in Spanish /d/ in intervocalic position is pronounced as [ð] (Gabinsky, 1992). Noteworthy since the consonants are already voiced, in both cases lenition affects only manner of articulation.

All the contact languages of Cypriot Greek which are still spoken on the island show a contrast which is usually described as geminates vs. singletons, and probably a lack of a distinct voicing opposition for stop consonants. The system of stops of Cypriot Arabic according to Borg (1985) consists of 'voice-indifferent' unaspirated segments. Standard Turkish distinguishes between plain voiced and aspirated voiceless consonants (cf. Kornfilt, 1997). In an experimental study Kallestinova (2004) reported variation in voicing for orthographic voiced and voiceless stops in Istanbul Turkish. For Cypriot Turkish Gerogiou-Scharlipp and Scharlipp (1997) described variation in the voicing of stop consonants, which may also affect loanwords. Notably, in Turkish loanwords into Cypriot Greek Turkish voiceless stops correspond to Cypriot Greek geminates while Turkish voiced stops are represented as Cypriot Greek singletons (Newton, 1968): Cyp. [p$^{h|}$ullin] *'stamp'* < Turk. pul, Cyp. [put$^{h|}$in] '*thigh'* < Turk. but (cf. Standard Modern Greek μπούτι [b$^{|}$uti]). Western Armenian distinguishes between voiced plosives and voiceless aspirates (Vaux, 1998), that is lack of voicing is always concomitant with aspiration. In this respect Western Armenian differs from Eastern Armenian which also has voiceless unaspirated plosives.

The similarities between Greek dialects and the neighbouring languages suggest that language contact along with other factors may have contributed to the expansion of one of the variants which was also common to other languages involved in the contact. A similar mechanism was suggested by Friedman (1994) to explain some syntactic similarities between Balkan languages. Therefore the role of languages contacts in the development of Modern Greek dialects and possibly their contact languages was catalytic rather than causal: they enhanced the preference for one of the variants which already existed in the language.

This does not mean that the variation observed in Greek dialects will eventually result in structural changes as it may have happened for example in the case of Celtic lenition. As Kallen (2005) says, "general principles of phonology can at best only define points in the system which are open to change and establish probabilities that change will operate in a particular direction". At the same time it raises the question of the separation of 'phonetic variation' and 'structural' change and phonetics and phonology in general. The variation observed in the three varieties of Greek suggests the gradient nature of linguistic processes and

the importance of finer phonetic details for understanding the sound structure of a language or variety.

Acknowledgements

Financial support for this study was provided by the Oxford University Clarendon Fund, St Cross College and Oxford University Phonetics Lab and is gratefully acknowledged.

References

Arvaniti, A. (1999a). Cypriot Greek. *Journal of the International Phonetic Association, 29*(2), 173-178.

Arvaniti, A. (1999b). Standard Modern Greek. *Journal of the International Phonetic Association, 29*(2), 167-172.

Arvaniti, A., & Tserdanelis, G. (2000). *On the phonetics of geminates: evidence from Cypriot Greek.* Paper presented at the Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP 2000), Beijing.

Baltazani, M. (2005). Prosodic rhythm and the status of vowel reduction in Greek, *the 26th Meeting of the Linguistics Section, Aristotle University*. Thessaloniki, Greece.

Barry, W., & Andreeva, B. (2001). Cross-language similiarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association, 31*(1), 51-66.

Beis, S. (2001). *Το φαινόμενο της κώφωσης στα βόρεια ελληνικά ιδιώματα και στη βλάχικη γλώσσα.* Paper presented at the Recherches en linguistique grecque: Actes du 5e Colloque international de linguistique grecque, Sorbonne, 13-15 Septembre, 2001 = Γλωσσολογικές έρευνες για την Ελληνική, Paris.

Blevins, J. (2004). *Evolutionary phonology*. Cambridge: Cambridge University Press.

Borg, A. (1985). *Cypriot Arabic: a historical and comparative investigation into the phonology and morphology of the Arabic vernacular spoken by the Maronites of Kormakitii village in the Kyrenia district of north-western Cyprus*. Marburg: Deutsche Morgländische Gesellschaft

Botinis, A., Fourakis, M., & Bannert, R. (2001). Prosodic interactions on segmental durations in Greek. *Working papers. Department of Linguistics, Lund University., 49*, 10-13.

Botinis, A., Fourakis, M., & Prinou, I. (2000). Acoustic structure of the Greek stop consonants. *Glossologia, 11-12*, 167-199.

Brendemoen, B. (2002). *The Turkish dialects of Trabzon: their phonology and historical development*. Wiesbaden: Harrassowitz.

Caragiu-Marioţeanu, M. (1968). *Fonomorfologie Aromânâ: studiu de dialectologie structuralâ*. Bucureşti: Editura Academiei Republicii Socialiste România.

Chatzidakis, G. N. (1892). *Einleitung in die neugriechische Grammatik*. Leipzig: Breitkopf und Härtel.

Dauer, R. (1980a). The Reduction of Unstressed High Vowels in Modern Greek. *Journal of the International Phonetic Association, 10*(1-2), 17-27.

Dauer, R. (1980b). *Stress and Rhythm in Modern Greek.* Unpublished Dissertation submitted to the University of Edinburgh for the degree of Doctor of Philosophy.

Drachman, G., & Malikouti-Drachman, A. (1996). Dissimilation in Cypriot Greek: competing analysis. *Μελέτες για την Ελληνική γλώσσα. Πρακτικά της 17ης Ετήσιας συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστήμιου Θεσσαλονίκης, Θεσσαλονίκη, 22-24 Απριλίου 1996*, 57-71.

Eftychiou, E. (2007). *Stop vowel coarticulation in Cypriot Greek.* Paper presented at the Proceedings of the ICPhS XVI, 10-18 August 2007, Saarbrücken.

Faber, A. (1992). Articulatory variability, categorical perception and the inevitability of sound change. In G. W. Davis & G. K. Iverson (Eds.), *Explanation in historical linguistics* (pp. 59-76). Amsterdam: John Benjamins.

Fourakis, M., Botinis, A., & Katsaiti, M. (1999). Acoustic characteristics of Greek vowels. *Phonetica, 56*(1-2), 28-43.

Friedman, V. (1994). Variation and Grammaticalization in the Development of Balkanisms. In K. Beals (Ed.), *CLS 30: papers from the 30th Regional Meeting of the Chicago Linguistic Society, 1994* (Vol. 2). Chicago: Chicago Linguistic Society.

Gabinsky, M. A. (1992). *Сефардский (еврейско-испанский) язык: балканское наречие*. Кишинев: Штиинца.

Georgiou-Scharlipp, K., & Scharlipp, W.-E. (1997). Some remarks on the Tukish dialect of the village of Potamya. In G. Hazai, B. Kellner-Heinkele & P. Zieme (Eds.), *Studia Ottomanica: Festgabe für György Hazai zum 65. Geburtstag* (pp. 141-146). Wiesbaden: Harrassowitz.

Gürkan, A. (1997). *Kıbrıs ağizında edatlar, bağlaçlar, ve ünlemlerin kullanım özellikleri*. Cyprus: K.K.T.C. Millî Eğitim ve Kültür Bakanlığı

Hamp, E. (1961). On the Arvanitika dialects of Attica and the Megarid. *Балканско езикознание/Linguistique Balkanique, 3*(2), 101-106.

Janse, M. (2000, 2001). *Cappadocian variables.* Paper presented at the Proceedings of the First international conference of Modern Greek dialects and linguistic theory (Patras, Greece, Oct. 12-14, 2000), Patras.

Kallen, J. L. (2005). Internal and external factors in phonological convergence: the case of English /t/ lenition. In P. Auer, F. Hinskens & P. Kerswill (Eds.), *Dialect change: convergence and divergence in European languages* (pp. 51-80). Cambridge: Cambridge University Press.

Kallestinova, E. (2004). Voice and aspiration of stops in Turkish. *Folia linguistica: acta Societatis Linguisticae Europaeae, 38*(1-2), 117-144.

Katsanes, N., & Dinas, K. (1990). *Γραμματική της κοινής Κουτσοβλάχικης*. Θεσσαλονίκη.

Kirchner, R. (2001). *An effort based approach to consonant lenition*. New York: Routledge.

Kontosopoulos, N. G. (2001). *Διάλεκτοι και ιδιώματα της Νέας Ελληνικής* (3rd. ed.). Αθήνα: Εκδόσεις Γρηγόρη.

Kornfilt, J. (1997). *Turkish*. London: Routledge.

Kramer, J. (1989). Rumänisch: Areallinguistik, II: Aromunisch/Les Aires linguistiques, II: Aroumain. In G. Holtus, M. Metzeltin & C. Schmitt (Eds.), *Lexikon der Romanistischen Linguistik (LRL), III: Die einzelnen romanischen Sprachen und Sprachgebiete von der Renaissance bis zur Gegenwart: Rumänisch, Dalmatisch/Istroromanisch, Friaulisch, Ladinisch, Bündnerromanisch/Les Différentes Langues romanes et leurs régions d'implantation de la Renaissance a nos jours: Le Roumain, dalmatico/istroromanzo, friulano, ladino, le romanche* (pp. 423-435). Tübingen: Niemeyer.

Labov, W. (1986). Sources of inherent variation in the speech production. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 402-423). London: Lawrence Erlbaum Associates.

Lazarou, A. G. (1986). *Βλάχοι: ιστορική φιλολογική μελέτη*. Αθήνα.

Lindblom, B. (1963). Spectrographic Study of Vowel Reduction. *Journal of the Acoustical Society of America, 35*, 1773-1781.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 216-245). New York: Springer-Verlag.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht: Kluwer Academic Publishers.

Loukina, A. (2008). *Regional phonetic variation in Modern Greek.* Unpublished Thesis submitted for the degree of Doctor of Philosophy, Oxford University, Oxford.

Loukina, A. (in press). *Phonetic variation in regional varieties of Modern Greek: vowel raising*. Paper presented at the Proceedings of the 3rd International Conference on Modern Greek Dialects and Linguistic Theory, Nicosia, 14-16 June 2007.

Mackridge, P. A. (1985). *The modern Greek language: a descriptive analysis of standard modern Greek*. Oxford: Oxford University Press.

Malikouti-Drachman, A. (1999). Παρατηρήσεις σε διαλεκτικές υποχωρήσεις τησ Κυπριακής. *Μελέτες για την Ελληνική γλώσσα. Πρακτικά της 20ης Ετήσιας συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, Θεσσαλονίκη, 23-25 Απριλίου 1999*, 292-302.

Martinet, A. (1952). Celtic Lenition and Western Romance Consonants. *Language, 28*(2), 192-217.

Moon, S.-J., & Lindblom, B. (1994). Interaction between Duration, Context, and Speaking Style in English Stressed Vowels. *Journal of the Acoustical Society of America, 96*(1), 40-55.

Newton, B. (1968). Spontaneous gemination in Cypriot Greek. *Lingua, 20*, 15-57.

Newton, B. (1972a). *Cypriot Greek. Its phonology and inflections*. The Hague: Mouton.

Newton, B. (1972b). *The generative interpretation of dialect: a study of modern Greek phonology*. Cambridge: University press.

Nicolaidis, K. (2001). An Electropalatographic Study of Greek Spontaneous Speech. *Journal of the International Phonetic Association, 31*(1), 67-85.

Nicolaidis, K. (2003). Acoustic variability of vowels in Greek spontaneous speech, *15th ICPhS*. Barcelona.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. Hendrick & M. F. Miller (Eds.), *Papers from the parasession on language and behaviour* (pp. 178-203). Chicago: Chicago linguistics society.

Ohala, J. J. (1983). The origin of sound patters in vocal tract constraints. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 189-216). New York: Springer-Verlag.

Ohala, J. J. (1989). Sound Change is Drawn from a Pool of Synchronic Variation. In L. Breivik & E. Jahr (Eds.), *Language change: contributions to the study of its causes* (pp. 173-189). Berlin: Mouton de Gruyter.

Ohala, J. J. (1993). The Phonetics of Sound Change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives,* (pp. 237-278). London: Longman.

Papadopoulos, A. A. (1926). *Γραμματική των βόρειων ιδιωμάτων της νέας Ελληνικής γλώσσης*. Αθήναι: P.D. Sakellarios.

Perkell, J. S. (1990). Testing theories of speech production: implications of some detailed analyses of variable articulatory data. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 263-288). Dordrecht: Kluwer Academic Publishers.

Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America, 112*(4), 1627-1641.

Peterson, G. E., & Lehiste, I. (1960). Duration of Syllable Nuclei in English. *The Journal of the Acoustical Society of America, 32*(6), 693-703.

Pettersson, T., & Wood, S. (1985). *A spectrographic study of vowel reduction in Bulgarian.* Paper presented at the X Nordiska Slavistmötet, 13-17 August 1984, Åbo.

Saracoğlu, E. (1989). *Kıbrıs Türk halk edebiyatı ve folkloru: bildiriler* (1. bask¸. ed.). Cyprus: K.K.T.C. Millî Eğitim ve Kültür Bakanlığı.

Sasse, H.-J. (1991). *Arvanitika: die albanischen Sprachreste in Griechenland*. Wiesbaden: Harrassowitz.

Sawicka, I. (1997). *The Balkan Sprachbund in the light of phonetic features*. Warszawa: Wydawnictwo Energeia.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: an acoustical study. *Journal of speech and hearing research, 6*(2), 111-128.

Sussex, R., & Cubberley, P. V. (2006). *The Slavic languages*. Cambridge, UK ; New York: Cambridge University Press.

Theophanopoulou-Kontou, D. (1973). Fast speech rules and some phonological processes of Modern Greek: a preliminary investigation. *Επιστημονική επετηρίς Φιλοσοφικής Σχολής του Πανεπιστημίου Αθηνών, 23 (Second period)*, 372-390.

Tilkov, D., & Boiadzhiev, T. (1981). *Българска фонетика*. София: Наука и изкуство.

Trudgill, P. (2003). Modern Greek dialects: a preliminary classification. *Journal of Greek linguistics, 4*, 45-64.

Tserdanelis, G., & Arvaniti, A. (1999). *The acoustic characteristics of geminate consonants in Cypriot Greek.* Paper presented at the Greek linguistics '99. Proceedings of the 4th international conference on Greek linguistcs, Nicosia, September 17-19, 1999, Thessaloniki.

Tzartzanos, A. A. (1909). *Περί της συγχρόνου Θεσσαλικής διαλέκτου*. Αθήναι: Τυπ. Π.Α. Πετράκου.

Van Coetsem, F. (1996). *Towards a typology of lexical accent: 'stress accent' and 'pitch accent' in a renewed perspective*. Heidelberg: Universitätsverlag C. Winter.

Vaux, B. (1998). *The phonology of Armenian*. Oxford: Clarendon Press.

Weinreich, U. (1968). *Languages in contact: findings and problems*. The Hague: Mouton & Co.

Zhobov, V. (2004). *Звуковете в българския език*. София.

The Role of Prosodic Prominence in Disambiguating Word Pairs

Nivedita Mani

Phonetics Laboratory, University of Oxford

Abstract

This research tests the early incorporation of prosodic information during disambiguation of word pairs such as *Packing cases*. The word pair is ambiguous between a noun or verb phrase interpretation. However, the two interpretations are prosodically distinct. The current study presents the results of an on-line, cross-modal, response-time task suggesting that subjects can disambiguate these word pairs using differences in contrastive focus. A second experiment swapped the timing, f0, and amplitude of the noun phrase versions with the verb phrase versions. If prosodic information were guiding parsing, swapping the prosody of the alternatives should change subjects' parses of the word-pairs. Indeed, the results of Experiment 2 confirmed that swapping the prosody of the word pairs reversed subjects' parses of the word pairs. The findings presented here, therefore, provide strong evidence in favour of the on-line use of prosodic prominence in guiding interpretation.

Keywords

Prosody, Syntactic Ambiguity, Parsing, Contrastive Focus

1.0 Introduction

Most studies on the use of prosody in speech processing accept that prosody can be used to interpret sentence structure and unpack syntactic content (Lehiste, Olive & Streeter 1976; Klatt 1976; Cooper & Paccia-Cooper 1980; Nespor & Vogel 1983; Price et al. 1991; Marslen-Wilson et al. 1992; Nagel et al. 1995; Shapiro & Nagel 1995; Speer et al. 1996; Watt & Murray 1996; Schepman 1997; Schafer 1997; Kjelgaard & Speer 1999; Schafer et al. 2000; Carlson et al. 2001). A majority of these studies have focussed on the role of prosody in guiding the parsing of sentences with ambiguous attachment such as *Mary murdered the man with the gun*. Sentences such as these provide a test of prosody alone guiding processing, in the absence of parsing cues from syntactic, and sometimes, even semantic sources. The findings of these studies suggest a

role for prosody in sentence processing at least as far as ambiguous speech stimuli such as these are concerned.

However, a number of these studies have focussed on testing the role of prosody in guiding parsing off-line (Speer, Shih & Slowiaczek 1989; Beach 1991; Price et al. 1991; Speer, Crowder & Thomas 1993; Ferreira et al. 1996; Kjelgaard & Speer 1999; Schafer et al. 2000; Cohen et al. 2001, among others). Typically, these studies presented subjects with ambiguous sentences, and compared subjects' parses of the sentences when spoken with or without disambiguating prosody, or with conflicting or concurring prosody. These off-line tests do not, therefore, provide much information about the real-time incorporation of prosody in speech processing. A more appropriate examination of the early incorporation of prosodic cues can be obtained from on-line tests of the influence of prosody in speech processing (e.g., Marslen-Wilson et al. 1992; Nagel et al. 1995; Shapiro & Nagel 1995; Speer, Kjelgaard & Dobroth, 1996; Kjelgaard & Speer 1999; Watt & Murray 1996; Schepman 1997). On-line tasks judge subjects' reactions to incomplete sentences or clauses to ascertain which interpretation of an ambiguous sentence subjects prefer during initial parsing.

For instance, Marslen-Wilson et al. (1992) examined whether listeners can use prosodic information during on-line processing of sentences such as (1) and (2) below.

(1) The workers considered the last offer from the management was a real insult
(2) The workers considered the last offer from the management of the factory

In (1), the phrase *last offer from the management* is the complement clause of the verb considered. In (2), the phrase is the direct object of the verb. The differences between the two sentences might be signalled by prosodic differences, marked in (3) and (4) below.

(3) (The workers considered H* L%) IPh (the last offer from the management H* L%) IPh (was a real insult H* L%) IPh
(4) (The workers considered H* the last offer H* L%) IPh (from the management of the factory H* L%) IPh

Subjects were played only the ambiguous segments of sentences (1) and (2) (underlined). At the end of each of the auditory fragments, a visual probe (*was*) appeared on the computer screen in front of them. The probe was an appropriate continuation of the ambiguous fragment in (1) but not of (2) – given the syntactic differences cued by the prosody of the two

sentences. In (2), the prosodic content would cue the direct object interpretation of *the last offer from the management*. However, the direct object structure would conflict with the complement clause continuation suggested by the probe *was*. Given this conflict, subjects should take less time to respond to the probe following (1), where the prosodic input would have already cued the complement clause interpretation consistent with the probe was. Indeed, their results confirmed that subjects took longer to name the same probe following (2) than (1). Their findings indicate that prosodic cues may be incorporated during initial parsing.

However, the delay in testing ambiguity till the word *management* does not fully assess how parsing is progressing following the verb *considered*, at which point the two interpretations of the structure are apparent, thereby not testing subjects' interpretation at the point at which the structure becomes ambiguous (after *considered*). As the authors themselves acknowledge, what is required is a test "earlier in the string" (Marslen-Wilson et al. 1992, 84) when adequate prosodic information first becomes available to guide parsing.

2.0 Experiment One

In an attempt to address these issues, the current study tests whether listeners can use prosodic cues alone to differentiate between the ambiguous interpretations of the sentence-initial word pairs in (5) and (6).

(5) Sailing ships are always newsworthy
(6) Sailing ships is always newsworthy

The stimuli tested in the current study are all ambiguous between the noun and verb phrase interpretation of the word pairs. The difference between the word pairs is prosodically characterised with the first word being stressed in the noun phrase reading and the second word being stressed in the verb phrase reading. The ambiguity in these word pairs is apparent upon presentation of the second word of the word pairs. Importantly, in contrast to the Marslen-Wilson et al. (1992) study, disambiguating prosodic information is presented by the completion of the second word at the point where subjects' disambiguation of the words pairs is tested. This provides a more suitable test of the immediacy of use of prosodic cues in disambiguating meaning. In addition, this constitutes a test of the use of prosody in resolving phrase-internal ambiguity, providing a more appropriate test of the local use of prosodic prominence as a cue to meaning. Indeed, relatively little is known about the role of prominence, aside from work by Warren et al. (1995), whose findings

indicate an influential role for prosodic boundaries, but not for word stress in interpretation.

2.1 Method

2.1.1. Participants

Fifteen adults took part in the experiment (8 female, 7 male). All subjects were native speakers of British English from Southern England and were between 21 to 25 years of age. All subjects were right-handed. Subjects were paid for their participation.

2.1.2. Stimuli

The contrasting word pairs were taken from Tyler & Marslen-Wilson (1977). Twelve pairs of syntactically ambiguous word pairs were recorded in sentences where the word pair was always followed by the phrase *are always newsworthy* for the noun phrase version and *is always newsworthy* for the verb phrase version. The noun and verb phrase versions of six word pairs were prosodically contrastive, in the manner outlined above. The noun and verb phrase alternations of the other six pairs were prosodically non-contrastive, in that the change from a noun to a verb phrase reading made little difference prosodically (e.g. *Flying planes*). The control word pairs are not contrastive in focus, because the noun and verb phrase interpretations of these word pairs do not cue completely independent meanings: *Cutting boards* are different kinds of boards compared to other boards, such as wooden boards. Consequently, the stress on *cutting* emphasises the type of board being referred to. Conversely, *Breaking glasses* are merely glasses that are breaking, and not necessarily specific to a particular type of glass, where there is little need for emphasis on breaking in the noun phrase version to indicate the kind of glass being referred to.

The stimuli were produced by an untrained male native speaker of British English, in response to sentences implicitly cuing either the noun or verb phrase interpretation. For e.g. the sentence *Modern ships are never interesting* cues the noun phrase interpretation of the word pair. The word pairs were then spliced from the sentences and loaded into an in-house response time program, similar to the one used by Marslen-Wilson et al. (1992).

2.1.3. Lexical bias measures

While the word pairs do not present subjects with syntactic information to guide disambiguation, it is possible that the frequency of occurrence of the noun and verb phrase interpretations of ambiguous stimuli in speech influences the ease with which either alternative is accessed (Binder & Morris 1995; Binder & Rayner 1998; Carpenter & Daneman 1981; Rayner, Pacht & Duffy 1994; Tabossi et al. 1987). Consequently, it is important to monitor the frequency of occurrence of the alternative interpretations of the word pairs. The relative frequency of the alternative interpretations of the word pairs was computed from the British National Corpus (www.natcorp.ox.ac.uk). Relative frequency was computed as the difference between the log of the frequency of occurrence of the noun phrase form and the log of the frequency of occurrence of the verb phrase form. The noun phrase form was the dominant interpretation of all of the prosodically contrastive word pairs, apart from one (*Playing cards*). Conversely, the verb phrase form was the dominant interpretation of all of the prosodically non-contrastive word pairs; apart from one (*Melting glaciers*).

2.1.4. Procedure

Subjects were seated in a sound proof room and provided with headphones over which they heard the fragments. As soon as each fragment finished, a word appeared on a screen in front of them. This word was always either *is* or *are*. In half of the presentations, the word that appeared on the screen was an appropriate continuation of what they had just heard – noun phrases followed by *are* and verb phrases followed by is. In the other half, the word that appeared on the screen was an inappropriate continuation of what they had heard. Subjects were provided with a button set with an *is* and an *are* button, and instructed to press the button that corresponded to the word that appeared on the screen.

Each subject heard sixteen repeats of each of the chosen fragments – a total of 384 sentence fragments per subject. The 384 fragments were presented to subjects in two different random orders. Subjects were randomly assigned to one of the two orders. The buttons corresponding to the is and are probes were counterbalanced across subjects, in order to avoid any left or right hand bias. Subjects were put through a practice run of 16 different fragments to familiarise themselves with the task.

If prosodic prominence can cue the intended interpretation of the word pairs, then, response times to appropriate probes (noun phrases followed by *are*, verb phrases followed by *is*) should be faster than response times

to inappropriate probes (noun phrases followed by *is*; verb phrases followed by *are*) following the prosodically contrastive word pairs. Conversely, the prosodically non-contrastive word pairs do not provide adequate prosodic information to disambiguate the word pairs. Consequently, subjects' responses to appropriate and inappropriate probes following the prosodically non-contrastive word pairs should be comparable.

2.2. Results

4.9% of all responses were excluded in order to ensure that only responses which could reasonably constitute immediate reactions to the visual probes (i.e. < 1000 ms) were analysed. The response time data were then aggregated by condition for each subject. The difference in response times to appropriate and inappropriate probes following the prosodically contrastive and non-contrastive stimuli is given in Table 1. Subjects took longer to respond to inappropriate probes than to appropriate probes following the prosodically contrastive stimuli. Conversely, subjects took the same amount of time to respond to appropriate and inappropriate probes following the prosodically non-contrastive stimuli.

We ran a 2X2 repeated measures ANOVA with condition (prosodically contrastive, non-contrastive) and match (match, mismatch) as within-subjects factors. There was a significant main effect of match, $F(1, 14) = 10.34$; $p = .006$, and a significant interaction between match and condition, $F(1, 15) = 7.78$; $p = .014$, but no significant main effect of condition, $F(1, 14) = .13$; $p = .71$. Planned comparisons revealed that there was a significant difference between response times to appropriate and inappropriate probes in the prosodically contrastive condition, $t (14) = 3.31$; $p = .005$, but not in the non-contrastive condition, $t (14) = .63$; $p = .53$.

Table 1 Mean Response Times to probes in Experiment 1

Stimuli	RT to appropriate probes (ms)	SE[a]	RT to inappropriate probes (ms)	SE[a]
Prosodically Contrastive	533	28.6	555	29.2
Prosodically Non-contrastive	544	26.6	546	25.5

[a] Standard Error

2.3. Discussion

Subjects' responses to appropriate probes following the prosodically contrastive word pairs were significantly faster than their responses to inappropriate probes. This suggests that subjects were able to differentiate the ambiguous word pairs and predict the number of the verb to follow. Since the main difference between the noun and verb phrase versions of the stimuli were prosodic, it is reasonable to assume that subjects' ability to discriminate the stimuli was guided by prosodic prominence.

The results of analysis of the responses to the prosodically non-contrastive stimuli support the conclusion that prosodic cues can guide disambiguation. There was a significant difference between subjects' responses to the prosodically contrastive compared to the non-contrastive word pairs. Subjects were unable to accurately disambiguate the prosodically non-contrastive word pairs. Since the main difference between the prosodically contrastive and non-contrastive word pairs was that the former could be disambiguated using prosodic cues, subjects' inability to accurately disambiguate the prosodically non-contrastive word pairs suggests that prosodic information was, indeed, guiding parsing of the prosodically contrastive word pairs.

However, there is an important caveat to this conclusion. The recordings of the word pairs were taken from sentences where the word pairs were followed by either *is* or *are* continuations (noun phrase versions followed by *are always newsworthy* and verb phrase versions followed by *is always newsworthy*). It is possible that the last segment of the word pairs [s/z] co-articulated with the initial vowel of the verb to follow (*is*/*are*), giving subjects an additional segmental cue to the appropriate continuation of the fragments (Shadle & Scully 1995). Therefore, subjects' accurate disambiguation of the word pairs in Experiment 1 could have been caused by co-articulation and not prosodic information. In addition, since all the word pairs were biased towards one interpretation of the word pair, there was the possibility that the lexical bias of the word pairs might have influenced subjects' responses in some of the trials. Experiment 2, therefore, controls for the influence of coarticulatory cues and lexical bias in guiding subjects' responses.

3.0 Experiment Two

Experiment 2 presents subjects with cross-synthesised versions of the original word pairs, in which the timing, amplitude, and f0 of each of the noun and verb phrase versions of word pairs were swapped with the timing, amplitude, and f0 contour of the alternative versions. The newly cross-synthesised stimuli retain the segmental and spectral content of the

original word pairs. However, the main prosodic characteristics of the stimuli – timing, amplitude, and f0 – have been swapped over. This should lead subjects to alternative interpretations of the word pairs, i.e. noun phrase interpretations of the cross-synthesised verb phrases, and verb phrase interpretations of the cross-synthesised noun phrases. Experiment 2 tests whether subjects' prefer parses that are consistent with the cross-synthesised prosodic characteristics over-parses consistent with the spectral characteristics of the stimuli. Therefore, response times to appropriate probes (probes inconsistent with the segmental content of cross-synthesised word pairs) would be longer than response times to inappropriate probes (probes consistent with the segmental content of cross-synthesised word pairs).

3.1. Method

3.1.1. Participants

All subjects were native speakers of English from Southern England. Twelve subjects took part in Experiment 2 (6 male; 6 female). All subjects were aged between 22 to 29 years. All subjects were right-handed.

3.1.2. Stimuli

The word pairs presented to subjects in Experiment 2 were cross-synthesised versions of the word pairs presented to subjects in Experiment 1. First, the word pairs were time-warped, so that the durations of all parts of the noun phrase forms of the word pairs were either lengthened or shortened to the durations of the corresponding verb phrase versions. Similarly, the durations of the verb phrase versions were time-warped to fit the durations of the original noun phrase versions of the word pairs. Time-warping repeats or deletes the required number of frames from one of the versions of the word pair to match the number of frames in the other version of the word pair. The time-warped stimuli were then re-synthesised with the f0, and amplitude contour of the alternative reading of the word pair. The re-synthesised stimuli thus had the timing, f0, and amplitude of the alternative interpretation of the word pair.

3.1.3. Procedure

The procedure was the same as in Experiment 1. Subjects were paid to participate in the experiment.

3.2. Results

6.6% of all responses were omitted because of incorrect responses (pressing the *is* button for an *are* probe or vice versa) or because of excessive delay in pressing the response buttons: any responses over 1000 msec were omitted. The mean response times of the remaining measures are given in Table 2.

Table 2 indicates that subjects had faster mean response times to appropriate probes than to inappropriate probes following the prosodically contrastive stimuli. Conversely, subjects had marginally slower mean response times to appropriate than to inappropriate probes following the prosodically non-contrastive stimuli. We ran a 2X2 repeated measures ANOVA with condition and match as within-subjects factors. There was a significant interaction between match and condition, $F(1, 11) = 6.05$; $p = .032$, but no significant main effect of condition, $F(1, 11) = .01$; $p = .91$, or match, $F(1, 11) = 3.16$; $p = .11$. Planned comparisons revealed that there was a significant difference between response times to appropriate and inappropriate probes in the prosodically contrastive condition, $t (11) = 2.39$; $p = .035$, but not in the non-contrastive condition, $t (11) = .28$; $p = .78$.

Table 2 Mean Response Times to probes in Experiment 2

Stimuli	RT to appropriate probes (ms)	SE[a]	RT to inappropriate probes (ms)	SE[a]
Prosodically Contrastive	579	13.8	604	12.7
Prosodically Non-contrastive	590	14.6	590	13.4

[a] Standard Error

3.3. Discussion

Experiment 2 presented subjects with cross-synthesised auditory stimuli – while the segmental content and, consequently, co-articulatory cues and lexical biases of the word pairs remained the same, the prosodic content of the alternative versions of the stimuli was swapped around. Subjects were then tested on their responses to appropriate and inappropriate probes, where appropriateness of the probe was decided by the prosody of the stimuli, and not the segmental content. Consequently, appropriate probes would be acceptable continuations of the stimuli if

subjects were using prosody to disambiguate the stimuli and inappropriate continuations of the stimuli if subjects were using co-articulatory cues to disambiguate the stimuli. Contrary to the segmental hypothesis, subjects displayed faster responses towards appropriate probes (inappropriate segmentally) than towards inappropriate probes (appropriate segmentally) following the prosodically contrastive stimuli. Conversely, there was no difference in subjects' responses to appropriate and inappropriate probes following the prosodically non-contrastive stimuli.

In addition, there was a difference in subjects' responding to the prosodically contrastive word pairs compared to the non-contrastive word pairs. If subjects were being guided by co-articulatory cues, then they should have been able to disambiguate the prosodically non-contrastive word pairs in both experiments in the current study. The difference in subjects' ability to discriminate the prosodically contrastive and non-contrastive word pairs suggests that subjects were, indeed, guided by the prosodic differences between the two interpretations of the contrastive word pairs.

Similarly, if subjects' responses were driven by the lexical biases of the word pairs, then subjects' responses should have been the same as in Experiment 1. The reversal of subjects' responses in Experiment 2 provides strong evidence in favour of a role for prosody in guiding on-line processing.

4.0 Conclusions

Typically, research investigating the influence of different sources of linguistic information in guiding parsing has focussed on the processing of ambiguous stimuli, since these stimuli provide the experimenter with a naturalistic test case of the importance of one kind of linguistic source over another. In addition, most of this research has tested the interaction between syntactic and semantic (or pragmatic) information in influencing the interpretation of ambiguous stimuli. Spanning the last quarter of a century, this research has found that there is an influence of both syntactic and semantic (or pragmatic) information in guiding parsing on-line; i.e., concurrently with the identification of linguistic units in the speech stream. Conversely, until the early 90s, the role of prosody in guiding parsing was relatively under-explored.

Since previous off-line studies tested the influence of prosody in guiding subjects' paraphrases of already completed sentences, no conclusions could be drawn regarding the on-line deployment of prosody in disambiguation. Previous on-line studies, on the other hand, typically, test disambiguation at the introduction of syntactic ambiguity, rather than

at the point of recognition of syntactic ambiguity. Note that the term recognition here does not refer to the listener's recognition of syntactic ambiguity but rather entails a reference to the point at which syntactic ambiguity is first noticeable.

Conversely, the current study tested subjects' disambiguation of sentence-initial word pairs, where the introduction of ambiguity is concurrent with the presentation of adequate prosodic information to guide parsing and testing the influence of prosodic information in guiding parsing. This provides a much required test of the deployment of prosody in guiding parsing immediately upon perception of the linguistic units in the speech stream.

The results of Experiment 1 indicated that subjects were able to disambiguate the noun and verb phrase versions of the prosodically contrastive word pairs alone, suggesting the importance of the prosodic contrast between the two versions of the word pairs. Experiment 2 ruled out the possibility that subjects' ability to disambiguate the word pairs was driven by the segmental cues provided by co-articulation between the final segment of the word pairs and the initial vowel of the main verbs following the word pairs. Subjects showed a preference for parses consistent with the prosodic content of the stimuli over parses consistent with the segmental content of the stimuli. Admittedly, this lack of preference for parses consistent with the segmental content of the stimuli could have been driven by the presence of strong prosodic cues to the contrary.

There was also the possibility that subjects' responses were influenced by the likelihood of the occurrence of one of the interpretations of the ambiguous word pairs: the dominant interpretation being more easily accessed than the less frequent interpretation. However, the lexical biases of the stimuli, obtained from the British National Corpus, would necessarily conflict with the intended interpretation in half of the presentations – when a stimulus biased towards the noun phrase interpretation was presented with verb phrase prosody and a stimulus biased towards the verb phrase version was presented with noun phrase prosody. Consequently, the lexical bias of the word pairs could not have guided subjects towards accurate disambiguation of the stimuli. The finding of directly contrasting parses of the same word pairs in Experiments 1 and 2 provide stronger evidence against a lexically based motivation for subjects' responses, since any lexical biases should have remained constant across the two experiments.

Some previous research has argued against an influence of prosody on interpretation because of the lack of a simple correlation between prosody and grammatical structure: speakers do not produce consistent differences between two versions of an ambiguous syntactic structure (Pierrehumbert

& Hirschberg 1990; Albritton et al. 1991; Beckman 1996; Schepman 1997). This variability in the production of consistent prosodic cues does not appear to interfere with listeners' ability to use prosodic information on-line. However, the stimuli presented to subjects in this experiment were carefully controlled so as to provide subjects with adequate prosodic information to guide parsing. Consequently, the current study can only conclude that when adequate prosodic information is present, it can be used to disambiguate structures on-line. In the absence of such disambiguating prosodic information, pragmatic, segmental, or lexical biases may control structure assignment. Neither does the current study advocate a privileged role for prosodic information in guiding parsing. The stimuli tested in the experiments presented here provided subjects with few syntactic or pragmatic cues towards disambiguation. For instance, the possibility of pragmatic biases guiding parsing away from the prosodically favoured interpretation remains unexplored. At the same time, the prosodically favoured interpretation might equally override listeners' pragmatic biases. The current study does not provide a test of the relative importance of prosodic, lexical, syntactic, and pragmatic biases in guiding parsing on-line.

Importantly, the word pairs presented to subjects in this study differed merely in prominence. As noted in the introduction, there are few studies on the influence of prosodic prominence on structure assignment. While the results of the current study suggest local and immediate use of prosodic information, it is acknowledged that the word pairs provide subjects with adequate information to compare the level of prominence of both words. Consequently, something akin to more global processing may be in use, albeit at a smaller scale. The latter mechanism of processing suggests that listeners may be paying attention to the relation between the levels of prominence on the words presented, as concluded by research on prosodic boundaries (Carlson et al. 2001; Clifton et al. 2002; Frazier et al. 2004), rather than prominence on one word cuing disambiguation.

Finally, there must be some discussion of the interaction between prosodic, syntactic, and semantic sources in on-line parsing. The current paper does not have evidence that there are intrinsic correlations between the prosody and the syntactic structure of the word pairs which listeners are using to guide disambiguation. A more plausible explanation may be found by examining more closely the relation between the prosodically contrastive and non-contrastive word pairs. As highlighted above, the latter are not expected to have contrastive focus because they do not cue completely unrelated interpretations. *Flying kites* do not differ from non-flying kites in any way intrinsic to the nature of kites. In contrast, *cutting boards* are different from other boards that may also be cut, such as

wooden boards, or black boards. The processes guiding the use of prosodic prominence in disambiguating these word pairs may be more linked to the stress accent on the noun phrase version of the word pairs highlighting the adjective and therefore, cueing this noun phrase interpretation of the word pair.

In conclusion, Experiments 1 and 2 provide strong evidence in favour of the on-line use of prosodic cues in guiding sentence processing. Importantly, both experiments test subjects' parsing at the point of ambiguity when adequate prosodic information is available to guide parsing, thereby providing a more robust test of the immediate use of prosody on-line. The incorporation of prosodic prominence in guiding parsing has been relatively under explored. Differences in prosodic prominence tend to be more subtle than differences caused by the movement of prosodic boundaries from one syntactic boundary location to another. The latter usually involve highly perceptible changes in the duration of pauses between words, in addition to changes in prosodic prominence. Experiments 1 and 2 strongly suggest that differences in prosodic prominence can cue disambiguation of word pairs in an on-line response time task. This suggests an extremely important role for prosody even in immediate interpretation of ambiguous phrase-internal stimuli.

Acknowledgements

The author thanks Prof. John Coleman for his supervision of the current study, which was undertaken as a part of her PhD thesis. In addition, the author thanks Andrew Slater for this technical assistance in designing the experiments and preparing the materials for presentation.

References

Albritton, D., McKoon, G., & Ratcliff, R. (1996) Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22.3*: 714-735.

Anderson, S. W., & Cooper, W. E. (1986) Fundamental frequency patterns during spontaneous picture description. *Journal of the Acoustical Society of America*: 1172-1174.

Beach, C.M. (1991) The interpretation of prosodic patterns at points of syntactic structure ambiguity: evidence for cue trading relations. *Journal of Memory and Language, 30*: 644-663.

Beckman, M.E. (1996) The Parsing of Prosody. *Language and Cognitive Processes, 11(1/2):* 17-67.

Binder, K.S., & Morris, R.K. (1995) Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 21: 1186-1196.

Binder, K.S., & Rayner, K. (1998) Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin & Review, 5:* 271-276.

Blasko, D. G., & Hall, M. D. (1998) Influence of prosodic boundaries on comprehension of spoken English sentences. *Perceptual and Motor Skills*, *87:* 3-18.

Boersma, P., & Weenink, D. (2000) Praat computer program. On-line, Institute of Phonetic Sciences, University of Amsterdam, http://www.praat.org/.

Carlson, K. (2001) The effects of parallelism and prosody on the processing of gapping structures. *Language and Speech, 44:* 1-26.

Carlson, K., Clifton, C., Jr., & Frazier, L. (2001) Prosodic boundaries in adjunct attachment. *Journal of Memory and Language,* 45: 58-81.

Carpeneter, P.A., & Daneman, M. (1981) Lexical retrieval and error recovery in reading: A model based on eye fixations. *Journal of Verbal Learning and Verbal Behaviour, 20:* 137-160.

Clifton, C., Jr., Carlson, K., & Frazier, L. (2002) Informative prosodic boundaries. *Language and Speech, 45:* 87-114.

Cooper, W.E., & Paccia-Cooper, J. (1980) *Syntax and Speech*. Cambridge Massachusetts: Harvard University Press.

Ferreira, F., Anes, M.S., & Horine, M.D. (1996) Exploring the use of prosody during language comprehension using the auditory moving window technique. *Journal of Psycholinguistic Research, 25.2*: 273-289.

Fodor, J.A. (1983) *The Modularity of the Mind: An Essay on Faculty Psychology*. Cambridge Massachusetts: MIT Press.

Frazier, L., & Rayner, K. (1982) Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14:* 178-210.

Frazier, L. (1979) *On Comprehending sentences: Syntactic Parsing Strategies*. Bloomington: Indiana University Linguistics Club.

Frazier, L. (1989) Sentence processing: a tutorial review. In Coltheart, M. (ed) *Attention and Performance XII: The Psychology of Reading*. Erlbaum.

Frazier, L. (1990) Exploring the architecture of the language-processing system. In Altmann, G.T.M. (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge Massachusetts: MIT Press.

Frazier,L. Clifton, C., Jr., & Carlson, K. (2004) Dont break, or do – Prosodic boundary preferences. *Lingua*, *114.1:* 3-27.

Hirschberg, J., & Ward,G. (1992) The Influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics*, *20:* 241-251.

Kjelgaard, M.M., & Speer, S.R. (1999) Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40: 153-194.

Klatt, C. (1976) Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, *59:* 1208-1221.

Lehiste, I., Olive, J.P., & Streeter,L.A. (1976) The role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, *60:* 1199-1202.

Marslen-Wilson, W.D., Tyler, L.K., Warren, P., Grenier, P., & Lee, C.S. (1992) Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology*, *45A*: 73-87.

Nagel, H., Shapiro, L., Tuller, B., & Nawy, R. (1995) Prosodic influences on the processing of attachment ambiguities. *Journal of Psycholinguistic Research, 25.2:* 319-343.

Nespor, M.A. & Vogel, I. (1983) Prosodic structure above the word. In Cutler, A. & Ladd, D.R. (ed) *Studies in Sentence Perception*. New York: Wiley.

Pierrehumbert, J.B., & Hirschberg, J. (1990) The meaning of intonational contours in the interpretation of discourse. In Cohen, J.M.P & Pollack, M. (Eds.), *Intentions in Communication*. Cambridge Massachusetts: MIT Press.

Price, P.J., Ostendorf, M., Shattuck-Hafnagel, S., & Fong, C. (1991) The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90:* 2956- 2970.

Rayner, K., Pacht, J.M., & Duffy, S.A. (1994) Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language, 33:* 527-544.

Rayner, K., Carlson,M., & Frazier, L. (1983) The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behaviour*, *22:* 358-274.

Schafer, A., Carlson, K., Clifton, C.Jr., & Frazier, L. (2000) Focus and the interpretation of pitch accent: disambiguating embedded questions. *Language and Speech*, *43.1:* 75- 105.

Schafer, A.J. (1997) *Prosodic Parsing: The Role of Prosody in Sentence Comprehension*. University of Massachusetts Doctoral Dissertation. Amherst.

Schepman, A.H.B.C. (1997) *Prosody and On-Line Parsing*. D.Phil Thesis, University of Sussex. Downloading available from: http://www.ling.ed.ac.uk/~astrid/thesisdownload.html.

Shadle, C.H., & Scully, C. (1995) An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics*, *23:* 53-66.

Shapiro,L. & Nagel, H. (1995) Lexical properties, prosody, and syntax: implications for normal and disordered language. *Brain and Language*, *50:* 240-257.

Speer, S.R., Crowder, R.G., & Thomas, L.M. (1993) Prosodic structure and sentence recognition. *Journal of Memory and Language*, *32:* 336-358.

Speer, S.R., Kjelgaard, M.M., & Dobroth, K.M. (1996) The influence of prosodic structure on the resolution of syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25:* 249-272.

Speer, S.R., Shih, C.L., & Slowiaczek, M. (1989) Prosodic structure in language understanding: evidence from Tone Sandhi in Mandarin. *Language and Speech, 32.4*: 337-354.

Tabossi P., Colombo L., & Job R. (1987) Accessing lexical ambiguity: Effects of context and dominance. *Journal of Psycholiguistic Research*, *49:* 161-167.

Tyler, L.K., & Marslen-Wilson, W.D. (1977) The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, *16:* 683-692.

Warren, P., Nolan, F., Grabe, E., & Holst, T. (1995) Post-lexical and prosodic phonological processing. *Language and Cognitive Processes*, *10:* 411-417.

Watt, S.M., & Murray, W.S. (1996) Prosodic form and parsing commitments. *Journal of Psycholinguistic Research*, *25.2:* 291- 317.

S-Aspiration and Occlusives in Andalusian Spanish: Phonetics or Phonology?

Paul O'Neill

Oxford Research Centre for Romance Linguistics, University of Oxford
School of Cultures, Languages, and Area Studies, University of Liverpool

Abstract

This paper examines the pronunciation of the underlying sibilant in the word internal sequences Vowel + Sibilant + Voiceless Occlusive + Vowel (e.g pasta 'pasta') in the Spanish spoken in Seville. In this position the sibilant can not only be pronounced as some type of aspiration e.g [pa$^{\eth}_{\Bumpeq}$tha] but also there is an alternative pronunciation in which no aspiration occurs and the following occlusive consonant is heavily post-aspirated, e.g [pa$_{\Bumpeq}^{\text{t}^{\text{h}}}$a]. I hypothesize that this latter pronunciation cannot be derived from the former via an interplay of timing of gestures of a purely phonetic nature but rather corresponds to an underlying aspirated occlusive phoneme. On the basis of this hypothesis I form a number of predictions regarding the phonetic nature of these sequences and via statistical methods come to the conclusion that the Spanish of Seville possesses a set of aspirated occlusive phonemes.

Keywords

Andalusian Spanish, Acoustic Phonetics, Phoneme Inventory

1.0 Introduction

Traditionally, it is considered that in Andalusian Spanish /s/ in coda position undergoes a phonological process whereby it loses its oral features and is pronounced as a voiceless glottal fricative [h] (Goldsmith 1981). This is illustrated in Table 1 with respect to Castilian, the prestige form of the language in Spain.

Table 1 The traditional view of /s/ aspiration in Andalusian Spanish.

Gloss	Orthographic form	Castilian	Andalusian
pasta	pasta	[pas̪ta]	[pah̪ta]
swan	cisne	[θisne]	[sihne] , [θihne]
island	isla	[isla]	[ihla]

Such aspiration is well documented in the literature (Rodríguez Castellano & Palacio, 1948; Alarcos Llorach, 1958; Zamora Vicente, 1969; Goldsmith 1981; Hualde, 1987; Mondéjar, 1991). Some authors have, in addition, drawn attention to the fact that the aspiration can be voiced [ɦ] and very often the phonetic outcome is not a glottal fricative but rather can appear as breathy voice on the vowel (Gerfen, 2002), cause compensatory lengthening of the following consonant (Bishop, 2007; Gerfen, 2002), or even be deleted.

O'Neill (2008) found no evidence of a full segment [h] in a spectrographic analysis of a corpus of 11 minimal pairs of the sequences Vowel + Sibilant + Voiceless Occlusive + Vowel (henceforth VSOV) vs. Vowel + Voiceless Occlusive + Vowel (henceforth VOV) produced by 2 subjects in 4 different major cities in Andalusia (a total of 352 words analysed). He did find evidence of breathy phonation in the vowel or pre-aspiration of the stop, or both. The cases of breathy voice and pre-aspiration were originally classed separately since it was hypothesised that although the former may be due to an anticipatory opening of the vocal folds during the final part of the vocalic gesture, phonologically it belonged to the timing of the vowel whilst the pre-aspiration was part of the consonant. However, it was shown that the portion of breathy voice added significantly to the duration of the vowel1 and therefore the pre-aspiration and breathy voice were two different phonetic realisations of an opening of the glottis after the vocalic gesture; so, for future studies their durations were calculated together and termed GLOT. Another frequent production for VSOV sequences was one in which the vocal fold abduction was timed to coincide with the stop closure and thus, in these cases the sequence VSOV was either pronounced with a long occlusive or an aspirated occlusive.

In summary, the various possible pronunciations for the sequence SO of VSOV can be categorised into three groups as illustrated in Table 2:

[1] More specifically, following the observation that cross-linguistically vowels in closed syllables have a shorter duration than vowels in open syllables (Maddieson, 1985) it was shown that in the sequences VSOV the duration of V, was only significantly different than that of the first vocalic element in VOV sequences when the breathy voice was taken **not** to be part of the vowel.

Table 2 Three variant pronunciations of VSOV.

(a) GLOT + O + VOT (all of varying lengths) [pa$^{\tilde{h}}$t̪^{h}a]
(b) long O + VOT [pat̪:^{h}a]
(c) occlusive + long VOT [pat̪$^{h\cdot}$a]

These are to be the focus of the present study.

1.1. A phonetic based explanation of the variation

The different pronunciations of the VSOV sequence could be given a purely phonetic explanation based on the interplay of the timing of articulatory gestures and trade-off relationships therein. To understand the reasoning behind this assumption I shall describe, using the terminology and representations of Articulatory Phonology (Browman & Goldstein, 1986; 1989; 1992), what could be the physiological mechanisms of the vocal tract when producing these various phonetic outcomes.

In Browman and Goldstein's theory, the assumption is that the behaviour of the vocal tract can be decomposed into the constrictions made by the different articulatory organs. These movements are called 'gestures' and are simply bags of information about how, when, and to what extent a constriction is to be produced. Gestures have an internal duration and can be combined and super-imposed upon each other. Graphically gestures are represented as blocks and each articulator is labelled and is represented on a different level; the length of each block denotes the time in which this articulator is actively controlled.

With regard to the Andalusian sequences, VSOV, breathy voice could be the result of a glottal gesture, which is relatively slow, in that the vocal folds are not abducted brusquely but rather gradually, allowing for an increased flow of air through the glottis, which at the same time is still, to some extent, vibrating. Pre-aspiration, the other phonetic realisation of GLOT, is, on this interpretation, the result of the overlap between the glottal gesture and the onset of the occlusion: while the tongue approaches the passive articulator to form a stop, the glottis widens and therefore the air passing through the glottis becomes turbulent at this constriction. GLOT is absent when the onset of the consonantal occlusive gesture occurs exactly as the vocalic gesture ends. The presence of a heavily aspirated occlusive depends crucially on the timing of the transition between this consonantal gesture and the following vocalic

gesture. If there is a large lapse of time during which the glottal gesture is still active, the result is a post-aspirated occlusive. The presence of a long consonant depends upon the combination of two factors: the onset and the release of the occlusion. This is represented in Figure 1.

As to the variation in the duration of each segment, in accordance with the above it follows naturally that the sum of the duration of GLOT and VOT will be inversely proportional to the duration of the occlusive. This follows from the fact that this consonant will be longer the earlier the onset of the occlusion and the later its release. As to why there can be such variation in the synchronisation of these gestures, a possible explanation appeals to the notion of the phase windows (Byrd, 1996) and the model of the syllable as expressed by Browman & Goldstein (1989). The basic idea is that gestures which take place in the onset of the syllable tend to couple into an in-phase relationship and the relative timing of gestures is fixed, whilst gestures in the coda position exhibit more freedom, and although segments belong to two different phase-windows there can be variation as to the organisation and duration of the gesture constellations, with the proviso that the total length of the windows ought to remain constant (Byrd, 1996; Docherty, 1992). In this way, if two gestures are mutually exclusive, as is the case of the glottal gesture and the occlusive gesture, then if one is prolonged it automatically follows that the other is shortened.[2]

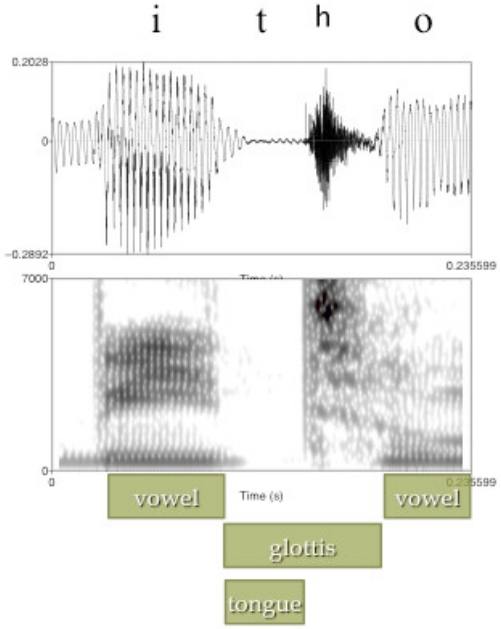
The alternative proposed here is that in the pronunciations (2b) and (2c), where there is no GLOT after the vowel in the VSOV sequences (which henceforth shall be termed $VSOV_{NOGLOT}$), there occurs a different phoneme. In (2b) this would be a phonologically aspirated consonant (VO^hV), whilst in (2c) a phonologically geminate consonant (VO:V). More specifically, in such pronunciations the situation is not one in which a VSOV sequence undergoes a phonological operation whereby /s/ loses its oral features in the coda of the syllable and the resultant pronunciation is due to the interplay between the timing of the glottal gesture (of the underlying /s/) and the occlusive gesture. Rather, these pronunciations correspond to the phonetic realisation of a different sequence of phonemes. Such phonologization of sounds which are initially phonetically conditioned is well attested in diachrony[3].

[2] In the Andalusian data, since there are three variables, it could be the case that only one is shortened (e.g. the hypothetical situation that length of the consonant is always damaging to the GLOT and not the VOT). This shall not be considered in the present study.

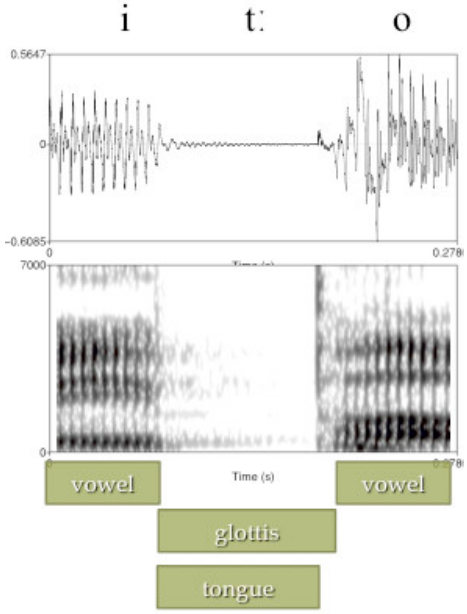
[3] A relevant example is that of Gujarati, a language in which there is a synchronic phonemic difference between breathy and modal voiced vowels which diachronically is due to the loss of a portion of aspiration between two syllables (Ladefoged. 2001:126).

Figure 3 Schematic depictions of the organisation of gestures for the sequences VSOV for the 3 variant pronunciations in Table 2.

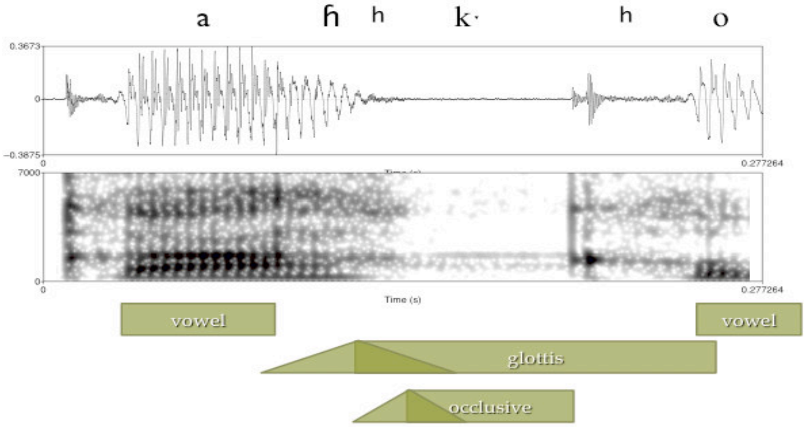
(a) pisto (a typical dish) with an aspirated occlusive



(b) pisto produced with a long occlusive



(c) casco, 'helmet' produced with breathy voice, pre-aspiration, a fairly long and aspirated consonant.



The hypothesis advanced here is that in Andalusian Spanish, in the sequences $VSOV_{NOGLOT}$, the phonetic realisation of the stop is not predictable from trade-off relationships between the timing of articulatory gestures and as such cannot be considered the outcome of a phonological operation but rather the aspirated stop and the geminate stop in the

sequences (2b) and (2c) (VO^hV and VO:V) are in direct phonemic opposition to the non aspirated stop in the sequences VOV, and thus could be considered to be phonemes. Due to limits of space, I shall restrict my study to the Spanish spoken in the city of Seville and only to the pronunciation of type (2b) in which there occurs an aspirated consonant, since in this particular city this is the pronunciation which, from a merely impressionistic point of view, seems to be the most dominant.

2.0 Summary of Hypotheses

The overall hypothesis is that in the variety of Andalusian Spanish spoken in Seville, the sequence of phonemes which in Standard Castilian is VSOV e.g. pasta 'pasta' [pas̪ta], has two main pronunciations which correspond to two different underlying phonemic sequences: The phonological sequence VSOV, e.g [pa$^{\tilde{n}}$t̪^{h}a] which corresponds to (2a) and shall be termed $VSOV_{GLOT}$.

The phonological sequence VO^hV (termed $VSOV_{NOGLOT}$) in which the occlusive is realised as an aspirated consonant, e.g. [patha] 'pasta', and is in phonemic opposition to the sequence VOV, e.g. [pata] 'paw'. Based on this hypothesis, the following predictions are made:

2.1 Prediction 1

In the sequences $VSOV_{GLOT}$ if the variation of the duration of the pronunciations of GLOT and VOT is dependent on the timing of the occlusive gesture then the duration of this consonant ought to be inversely proportionate to the sum of the durations of GLOT and VOT. That is, the duration of the occlusive has a negative correlation with the sum of the durations of GLOT and VOT.

$GLOT_{GLOT} + VOT_{GLOT} = k/O_{GLOT}$.

2.2 Prediction 2

If the sequences $VSOV_{NOGLOT}$, and $VSOV_{GLOT}$ correspond to phonologically different underlying sequences, VSOV and VO^hV respectively, then this should be manifested phonetically in that the population mean of the duration of the sequence VSO_{NOGLOT} ought to be significantly less than that of VSO_{GLOT}

$VSO_{GLOT} - VSO_{NOGLOT} > 0$

2.3 Prediction 3

Following on from this, if the sequence $VSOV_{NOGLOT}$ consists of three phonemes VO^hV whilst the sequence $VSOV_{GLOT}$ has four phonemes, in accordance with Spanish syllabification the first vowel in $VSOV_{NOGLOT}$ (underlyingly VO^hV) is in an open syllable and therefore following the principle that vowels in open syllables have a greater duration than vowels in closed syllables (cf. Maddieson, 1985) there ought to be no significant difference between the population mean of the first vowel in $VSOV_{NOGLOT}$ when compared to its near minimal pair sequences VOV. This however, should not be the case in the $VSOV_{GLOT}$ sequences (underlyingly VSOV) since the first vowel is in a closed syllable and therefore the population mean of the duration of the first vowel in its minimal pair sequences VOV ought to be significantly greater than that of $VSOV_{GLOT}$.

(a) $V_{VOV} - V_{NOGLOT} = 0$
(b) $V_{VOV} - V_{GLOT} > 0$

3.4 Prediction 4

Finally, if there is a phonemic opposition between aspirated and non aspirated occlusives[4] pasta ‘pasta’ [paṯ^{h}ˑa] vs. pata ‘paw’ [paṯa], the population mean of the duration of the VOT in the sequence $VSOV_{NOGLOT}$ ought to be significantly greater than that of the sequences VOV, whilst the population mean of the duration of the occlusive in both sequences ought not to be significantly different.

(a) $VOT_{NOGLOT} - VOT_{VOV} > 0$
(b) $O_{NOGLOT} - O_{VOV} = 0$

3.0 Experimental Design

For the experiment, I designed a corpus based on 11 words which contained the sequence VSOV and their near minimal pairs which lacked the sibilant. The words used in the experiment are given in Table 3.

[4] The sequences $VSOV_{NOGLOT}$ vs. VOV.

Table 3 List of words used in the corpus.

consonant	VSOV	gloss	VOV	Gloss
/p/	avispa obispo caspa	wasp bishop dandruff	pipa tipo capa	Pipe Type cape
/t/	pisto motorista pasta susto	a typical dish motorist pasta a scare	pito margarita pata luto	whistle daisy paw mourning
/k/	busca bizco casco mosca	he / she seeks cross-eyed helmet fly	nuca pico caco boca	back of neck beak thief mouth

I recorded 4 subjects (2 male and 2 female), all 18 years old and native inhabitants of the city of Seville. Subjects were presented with photographs depicting the words in the corpus, along with a number of control words that contained neither a sibilant nor an occlusive. The subjects were requested to name what they saw[5] in the following carrier phrase *dame un* ____ *pa mí* 'give me a ____ for me'. In this way, the subjects produced the 11 pairs of words in the corpus giving 44 near minimal pairs and a total of 176 words.

The data were analysed acoustically using Praat (Boersma & Weenink, 2005). The following measurements were taken for each word: i) duration of the first vowel (V); ii) duration of the glottal gesture before the consonant (GLOT); iii) duration of the consonant (O); iv) duration of the VOT. A statistical analysis was conducted using R (R Development Core Team, 2007). For Prediction 1, a linear model was created between the two variables in order to test for correlations. For hypotheses 2, a one tailed t-test was applied to establish the probability that the true mean of the first element was greater than the second. The same process was carried out for hypothesis 3b and 4a, but since these involved minimal pairs, a paired t-test was applied. For hypothesis 3a and 4b, two tailed paired t-test was performed.

[5] For those words in the corpus which it was impossible to solicit via the naming of images, subjects read the words embedded in a carrier phrase, from a card.

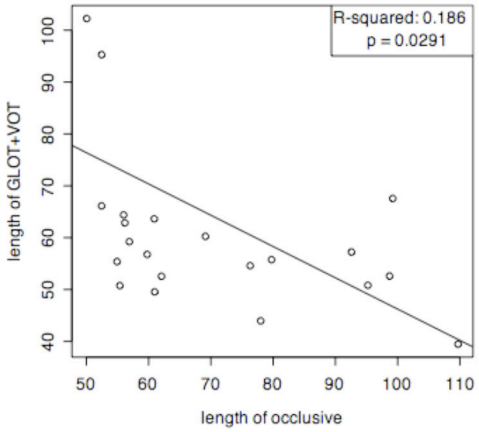
4.0 Results

In this section each prediction shall be repeated formally, briefly explained and the relevant graphs and results shall be presented.

Prediction 1: $GLOT_{GLOT} + VOT_{GLOT} = k/O_{GLOT}$

The claim here was that in $VSOV_{GLOT}$ sequences the occlusive duration is inversely proportionate to the sum of the durations of GLOT and VOT. The results are presented in Figure 2 in the form of a regression.

Figure 2 Regression of occlusive duration and the combined duration of (GLOT+VOT) for the $VSOV_{GLOT}$ *sequence.*

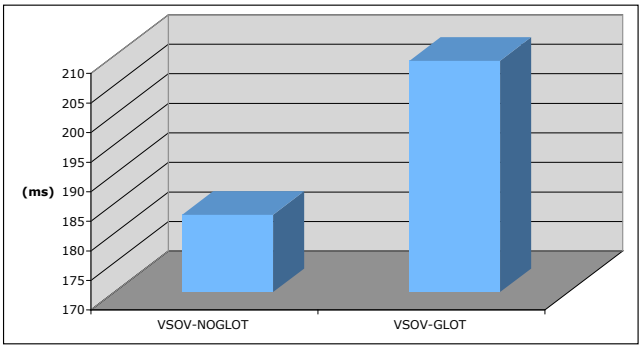


The significant negative regression (p=0.0291) in (5) does not falsify the prediction and therefore provides evidence to support the claim above and the idea that the different durations of GLOT, O and VOT are due to an interplay of the timing of articulatory gestures.

Prediction 2: $VSO_{GLOT} - VSO_{NOGLOT} > 0$

The claim here was that the duration of VSO_{GLOT} was significantly greater than that of VSO_{NOGLOT} (including the VOT in both cases). In Figure 2 is a graph of the sample mean durations of these values.

Figure 2 Mean durations of the sequence VSO for $VSOV_{GLOT}$ and $VSOV_{NOGLOT}$.

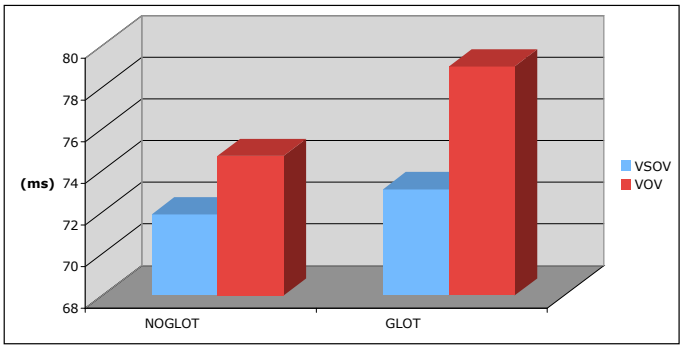


This duration of $VSO(V)_{GLOT}$ is significantly greater (p=0.01) than $VSO(V)_{NOGLOT}$ and therefore prediction 2 is not falsified. This lends support to the idea that the differences between the sequences $VSOV_{GLOT}$ and $VSOV_{NOGLOT}$ are not differences in degree but rather differences in kind. More specifically, the results suggest that the duration of the occlusive + VOT is not due to a premature occlusive gesture which precludes any realisation of GLOT of the underlying /s/ but maintains the timing for this segment but rather could be due to an underlying aspirated occlusive phoneme

Prediction 3a, 3b: $V_{VOV} - V_{NOGLOT} = 0$; $V_{VOV} - V_{GLOT} > 0$

Here it was hypothesised that the first vowel in the sequences VOV is significantly greater in duration than that of its minimal pair sequences $VSOV_{GLOT}$ but not of those of the sequences $VSOV_{NOGLOT}$. The sample mean durations of these values are given in Figure 3.

Figure 3 Mean durations of the first vowel for $VSOV_{GLOT}$ and $VSOV_{NOGLOT}$ together with their near minimal pairs of the sequences VOV.



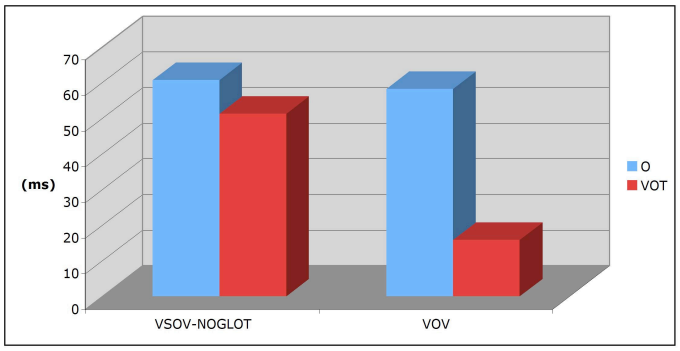
Although, in this graph, the duration of the first vowel in VOV sequences is greater than the corresponding vowel both in the sequences

$VSOV_{NOGLOT}$ and $VSOV_{GLOT}$, the difference between the that of VOV and the sequence $VSOV_{NOGLOT}$ is not statistically significant (p=0.74) whilst it is in the sequence $VSOV_{GLOT}$ (p=0.01). As such, both predictions 3a and 3b are not falsified, which lends support to the proposition that the initial vowels in the $VSOV_{NOGLOT}$ sequences are phonologically considered to be part of an open syllable and thus strengthens the argument that the $VSOV_{NOGLOT}$ sequences consist of three phonemes and not four.

Prediction 4a, 4b: $VOT_{NOGLOT} - VOT_{VOV} > 0$; $O_{NOGLOT} - O_{VOV} = 0$

Simply stated, these predictions claimed that the stop consonants in $VSOV_{GLOT}$ sequences would be more aspirated than those of the VOV sequences but the duration of the occlusion would be the same. In Figure 4 is a graph with the relevant values.

Figure 4 Mean durations of of VOT and O for both VSOVNOGLOT and VOV sequences.



The statistics corroborate the impression of the data in the graph in that the duration of the VOT in the $VSOV_{NOGLOT}$ sequences is significantly greater (p=0) than that of the VOV sequences whilst there is no significant difference (p=0.41) between the duration of the period of occlusion.

5.0 Conclusion

The conclusion to be drawn from these results is that, in the Seville variety of Andalusian Spanish, in words of the type pasta 'pasta' there are two variant phonological strings for the same word, the first consisting of four phonemes VSOV (even though the sibilant conventionally is not produced) and the second consisting of three phonemes, VO^hV, which is in phonemic opposition to VOV sequences of the type pata 'paw'.

Regarding the first variant VSOV, which in this study has been termed $VSOV_{GLOT}$, given the absence of a canonical sibilant in the production of these words, the distinction between the sequences VSOV and VOV is phonetically cued in the following way: stops in the VSOV sequences are voiceless, longer, more aspirated and can be preceded by a portion of breathy voice or pre-aspiration (GLOT). Although there is a great deal of phonetic variation in the duration of each of these pronunciations, there is a trading relationship between the duration of the occlusive and the durations of GLOT and VOT. The timing for the four phonemes is always maintained.

In the $VSOV_{NOGLOT}$ sequences, the evidence suggests that these sequences cannot consist of four phonemes, since these words, with regard to duration, are significantly different from those in which GLOT appears and it has been argued that their phonetic outcome cannot in any way be derived from an interplay of timing of gestures. Furthermore the duration of the preceding vowel suggests that phonologically this vowel belongs to an open syllable and not a closed one. As such, my conclusion is that these $VSOV_{NOGLOT}$ sequences consist of three phonemes: a vowel, an aspirated consonant and a vowel (VO^hV) and thus there is a distinct set of voiceless aspirated occlusive phonemes /p^h, t^h, k^h/ in the Spanish spoken in Seville.

Acknowledgements

The author would like to thank Greg Kochanski for his help with the statistical analysis and Elinor Payne, ʻŌiwi Parker Jones and Iggy Roca for their useful comments.

References

Alarcos Llorach, E. (1958). Fonología y fonética: a propósito de las vocales andaluzas. *Archivum*, *8*, 191-203.

Browman, C. & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219-252.

Browman, C. & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 201-251.

Browman, C. & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, *49*, 55-180.

Byrd, D. (1996). A phase window framework for articulatory timing. *Phonology*, *13*(2), 139-169.

Bishop, J. (2007). Incomplete neutralization in Eastern Andalusian Spanish: perceptual consequences of durational differences involved

in s-aspiration. *Proceedings of the International Congress of Phonetic Sciences*, *16*, 1765-1768.

Gerfen, C. (2001). A critical view of licensing by cue: codas and obstruents in Eastern Andalusian Spanish. In L. Lombardi (Ed.), *Segmental Phonology in Optimality Theory*. New York: Cambridge UP.

Gerfen, C. (2002). Andalusian Codas. *Probus*, *14*, 247-277.

Goldsmith, J. (1981). Subsegmentals in Spanish phonology: an autosegmental approach. In W. Cressey & D. J. Napoli (Eds.), *Linguistic Symposium on Romance Languages* 9 (pp. 1-16). Washington DC: Georgetown University Press.

Hualde, J. I. (1987). Delinking processes in Romance. In C. Kirschener & J. DeCesaris (Eds.), *Studies in Romance Linguistics* (pp. 177-193). Philadelphia: John Benjamins.

Ladefoged, P. (2001). *A Course in Phonetics*. Fort Worth: Harcourt College Publishers.

Maddieson, I. (1985). Phonetic Cues to Syllabification. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays In Honor of Peter Ladefoged* (pp. 203-219). Orlando: Academic Press Inc.

Mondéjar, J. (1991). *Dialectología andaluza*. Granada: Editorial Don Quijote.

Martínez Melgar, A. (1986). Estudio experimental sobre un muestreo de vocalismo andaluz. *Estudios de Fonética Experimental*, *2*, 195-248

Navarro Tomás, T. A. (1939). Desdoblameiento de fonemas vocálicos. *Revista de Filologia Hispánica*, *1*, 165-167.

O'Neill, P. (2008). La fonación murmurada y la preaspiración en andaluz. *Language Design*, *9*.

R Development Core Team (2007). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Rodríguez Castellano, L. & Ana Palacio. (1948). El habla de Cabra. *Revista de dialectología y tradiciones populares*, *4*, 378-418.

Torreira, F. (2006). Coarticulation between aspirated-s and voiceless stops in Spanish: an interdialectal comparison. In Nuria Sagarra & Almeida Jacqueline Toribio (Eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium* (pp. 113-120). Somerville MA: Cascadilla Press.

Is Speech Rate Lexical?

ʻŌiwi Parker Jones and John Coleman

Phonetics Laboratory, University of Oxford

Abstract

This paper reports on a word-recognition experiment which was designed to test the question of how lexical items are stored in the mind. Focusing on speech rate, we investigated the possibility of a priming effect for it—using semantic priming in order to ensure lexical access in subjects. What we found was an unexpected and highly significant correlation between target rate and response time. While this is an interesting result with important implications, it also means that the experiment was ultimately inconclusive with regard to the question of whether or not speech rate is lexical. We are continuing to explore the question in ongoing work.

Keywords

Speech Rate, Mental Lexicon, Priming Experiment

1.0 Introduction

In this paper, we address an ongoing theoretical debate about the nature of the mental lexicon. One camp (Generative Grammar) holds that lexical entries are redundancy free and separate from performance factors, like speech rate (Chomsky, 1965; Halle, 1997). The other camp (Exemplar Theory) boldly claims that such so-called 'performance' factors, such as speech rate, ought to be encoded in the lexicon (Johnson, 1997, 2007; Pierrehumbert, 2001).[1] So, how much detail do phonological representations in the mental lexicon contain? Are lexical items abstract, symbolic representations, or do they encode fine grained, possibly speaker-specific, phonetic details?

[1] A word of warning: the term *Exemplar Theory* is many-ways ambiguous (cf., e.g., Lahiri and Reetz (2002, p. 670) who interpret the 'exmplars' posited by Medin and Schaffer (1978) and Nosofsky (1986) to represent 'category prototypes', from the variety of 'full listing' approaches such as those offered by Data-Oriented Parsing (Bod, 1998) and Memory-Based Learning (Daelemans & van den Bosch, 2005)). Here, we refer to what Hay and Bresnan (2006) call 'Phonetic Exemplar Theory' (e.g., Johnson, 1997, 2007; Pierrehumbert, 2001).

To date, there is mounting evidence to suggest that listeners store more information about word pronunciation than Generative Grammar would lead one to believe. For example, Goldinger (1996, 1997) investigated speaker voice and memory in a series of landmark experiments. He asked his subjects to listen to recorded words and to indicate whether they had heard the words before in the experiment. They first heard 150 words; then, in later sessions, they classified 300 words as either 'old' or 'new'. Some 'old' words were spoken by the same voice as the subject had previously encountered, while others were spoken by different voices. Subjects heard 2, 6, or 10 stimulus voices, and between sessions, they waited 5 minutes, 1 day, or 1 week. In all conditions, half of the voices for the 'old' words were changed between practice and test. Goldinger found that same-voice repetitions were identified more accurately than different-voice repetitions for all delays, implying that, in general, information about *what was said* is stored together with information about *which voice said it*.

Goldinger's results crucially run counter to Generative Grammar. On the generative account, listeners analyse incoming words into at least two separate parts:

(1) the abstract phonological code used to store words (*competence*)
(2) a memory of the speakers' voices (*performance*)

Since the mental lexicon is assumed to be a component of competence, Generative Grammar locates speaker voice outside of the lexicon.

What Goldinger's results suggest is that (1) and (2) may be more closely related than the generative account would imply. Goldinger advocates an *episodic* model of speech perception, in which each presentation of a word—even a familiar word—lays down a new 'trace' in auditory memory, complete with fine phonetic details which relate to the individual speaker's voice (cf., e.g., McClelland & Rumelhart, 1985).

Still, many linguists remain unswayed. For example, Myers throws down the following gauntlet:

> 'As an English speaker, I know the word *dog*. One thing I know about it is how to pronounce it in my dialect. My intuition is that there is only one word *dog*, although this word can be used to refer to quite different sorts of animals, and although it can be produced in quite different ways under different speaking conditions. There's only one entry in my long-term memory for this item.

> ʻThis seems like a plausible assumption. […] I know of no clear counter-evidence to this [*uniqueness*] hypothesis. It might be possible to use standard experimental paradigms from the lexical access literature to probe whether each lexical entry has just a single phonological entry. Priming, for example, is the facilitation of word identification (as measured in response time) triggered by the previous presentation of a related form […]. If there are different entries of fast-speech versions and slow-speech versions of a lexical entry, for example, one might expect that a version at the same rate would be more efficient at priming identification than a version at a different rate.ʼ
>
> (Myers, 2000, pp. 260–262)

In summary, an episodic account predicts that lexical entries will be stored complete with phonetic detail, such as the specific speech rate at which an utterance was produced. This as yet untested prediction was the focus for the experiment that we present in this paper.

We explored the theoretical debate over the mental lexicon by running a word-recognition experiment—with a focus on speech rate. Basically, we wanted to see if we could find a significant effect for speech rate facilitating response time (RT). We also wanted to control for lexical access while doing so, to be able to say something about the nature of the mental lexicon. Therefore, we replicated an auditory semantic priming experiment following Andruski et al. (1994), who also used semantic priming to measure the effect of phonetic differences on lexical access (in their case, Voice Onset Time).

Semantic priming requires lexical access to determine the relatedness of prime and target. When a prime and target are semantically related (e.g., doctor–nurse), subject RTs generally improve on word recognition tasks (cf. doctor–snorkel)—this is called semantic facilitation (Meyer & Schvaneveldt, 1971; see, e.g., Neely, 1991, for a review).

We wanted to see if we could observe an additional facilitation effect, like those of phonological priming paradigms, even when we manipulate a non-phonological variable such as speech rate. The null hypothesis was that the speech rate of the prime and the target in a prime–target pair would be inconsequential in all cases, as predicted by Generative Grammar. The alternative hypothesis was that, when the prime semantically facilitates the target, thereby ensuring lexical access, we should observe an additional facilitation if the prime and the target are played at the same speech rate, as predicted by Exemplar Theory. This would suggest that the lexicon contains a phonetic dimension for speech

rate in which the distance between word tokens (as memory traces) can be measured.

In fact, neither of these hypotheses were borne out by the results. Instead, we found an unexpected and highly significant correlation between target rate and RT. While this is an interesting result with important implications, it also means that the experiment was ultimately inconclusive with regard to the question, as it introduces a new confound which we did not control for. We are, therefore, continuing to explore the question of whether or not speech rate is lexical in ongoing work.

2.0 Experiment

We tested whether reaction times decrease when listeners hear same rate prime–target pairs versus different rate prime–target pairs, over and above the generally-expected reduced reaction times for semantically related prime–target pairs.

2.1. Methods

2.1.1. Materials

We recorded 576 three- and four-syllable words, as produced by a young female speaker of Southern British English. These reduced to 288 prime–target pairs, where half of the targets were real words and half were nonsense words. Some of our nonsense words were:

**thulzhaquians*	/θʌlˈʒakwɪənz/
**nanbappize*	/ˈnanbəpɑɪz/
**dinmorrel*	/dɪnˈmɒɹəl/

These were automatically-generated from a long list of novel combinations of attested English onsets and rimes which were chosen by the consensus of a group of native English speakers in an experimental phonetics graduate seminar at the University of Oxford.

Of the 144 pairs with both real word primes and real word targets, 72 were semantically related while the other 72 were semantically unrelated. We determined semantic relatedness by reference to the Edinburgh Associative Thesaurus of word association norms (Kiss et al., 1973). We matched the real word primes and targets for frequency with reference to

the British National Corpus. The average target frequency was 44 parts per million and the average prime frequency was 55 parts per million.[2]

Half of the pairs had primes and targets at the same speech rate (i.e., fast–fast, slow–slow, or medium–medium). The other half had primes and targets that mismatched for speech rate (i.e., fast–medium, fast–slow, etc.) in equal numbers. We synthetically manipulated the timing for all of the recordings, without altering the pitch, by using an implementation of overlap-add resynthesis (Moulines & Charpentier, 1990) in Praat (Boersma, 2001). One third of the words were thereby sped up by a factor of 1.5 to sound like fast-speech, another third were slowed down by a factor of 0.5 to sound like slow or laboured speech, and the final third were not temporally altered (though we passed them through the same software, with a factor of 1, so that they would be subjected to the same analysis-resynthesis process).

In summary, the stimuli included the following:

(a) 72 targets with 72 semantically related primes, at various rates (36 pairs at the same rate, 36 at different rates):

6 slow–medium pairs
12 slow–slow pairs
6 slow–fast pairs

6 medium–slow pairs
12 medium–medium pairs
6 medium–fast pairs

6 fast–slow pairs
6 fast–medium pairs
12 fast–fast pairs

(b) 72 targets with 72 unrelated primes, with same pattern of rate variation as given above for (a)

(c) 144 non-word targets with 144 unrelated, real-word primes

The experiment was conducted in an acoustically insulated booth in the Oxford University Phonetics Laboratory. We used in-house software to present the experiment over a PC to the subjects, who listened to the

2 A complete unrandomized list of the paired primes and targets, together with their BNC frequencies, is available online (http://www.phon.ox.ac.uk/~jcoleman/rate-priming-pairs.html).

stimuli through high quality headphones and responded by using a button box.

2.1.2. Procedures

We played word pairs to the participants who were directed to attend only to the targets, which they were instructed to categorize as a real words or as a nonsense words by pressing the buttons on the button box.

Participants experienced the experiment as a game, called RT-Bonus, where they made more 'money' by responding quickly and correctly. Incorrect responses were penalized by a deduction from their 'earnings'. Slow responses paid out increasingly less 'money' until the winnings reached zero. We used 'money' as an incentive for the participants to attend to the task by telling them that they would be rewarded proportionally to their final 'earnings'. In fact, we paid all participants a flat rate which was greater than the 'earnings' that any of them accumulated by playing the game.

Before playing, participants were given written instructions and two stimulus pairs for practice. One of the practice pairs required a 'yes' response (i.e., 'this is a real word'), while the other required a 'no' response ('this is not a real word'). The presentation programme played a cash-register sound (*cha-ching!*) after every correct answer, or a game show buzzer when participants answered incorrectly. The participants also received visual aid on the screen to monitor their current 'earnings'.

On the button box, one button was colored green for the 'yes' responses, while the other was coloured red for 'no' responses. To avoid a possible handedness bias, the 'yes' button appeared on the left for half of the subjects, and on the right for the other half. We asked the subjects to pick just one finger to press the buttons, and we encouraged them to leave it somewhere between the buttons between responses. The motivation was to prevent a possible bias towards faster responses with one finger (e.g., if they used an index finger for one button and a thumb for another).

The experiment lasted about 15 minutes for each subject. Afterwards, we paid them £5 and were happy to answer any questions they might have about the experiment.

2.1.3. Participants

20 undergraduates from the University of Oxford participated in this study.3 All were speakers of Southern British English. We excluded Psychology and Linguistics students as insufficiently naïve. No subject reported any speech or hearing disorders.

2.2. Results

All of the participants completed the task with over 90% accuracy, only mis-identifying 10% of the real word targets as nonsense words, and vice versa. As shown in Figure 1, by far the most common error was to misidentify the Old English name *Ethelred* as a nonsense word. This mistake was made by 13 subjects (i.e., 65%). **plutenit* was most frequently mis-identified as a real word, with 8 subjects (40%) getting it wrong. Most errors, however, were one-offs.

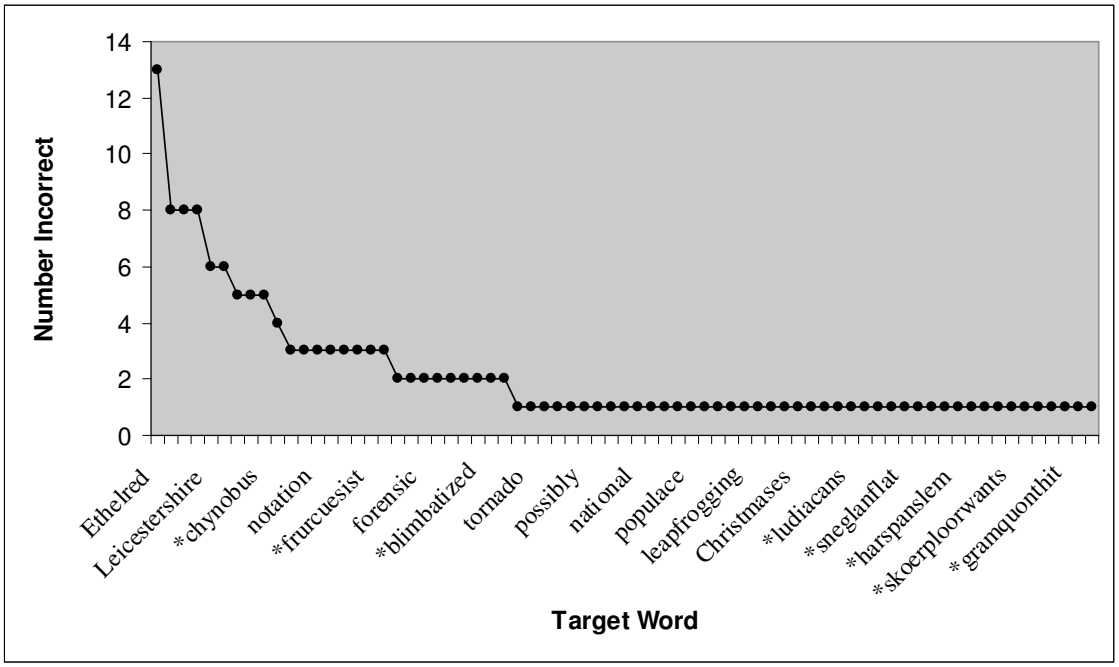


Figure 1: A graph of the 71 target words that our subjects misidentified, from the most misidentified to the least.

We only modeled RT for correct results. These were fit to a linear mixed-effect model (see, e.g., Baayen 2008) using R (R Development Core Team, 2009). We treated subject and stimulus number as random effects so as to guard against any subject carrying a trend as well as to

3 This figure was culled from the methodology of Andruski et al. (1994) who used 21 subjects.

guard against artifacts from the order in which the stimuli were presented. The fixed effects that we considered were:

> the speech rate of the prime (i.e., slow, medium, or fast)
> the speech rate of the target (i.e., slow, medium, or fast)
> whether the speech rates of the prime and targets were the same (i.e. same vs. different)
> whether the prime and target were semantically related (i.e. yes or no)

These were evaluated in terms of how well they predicted the subject RT.

The only factors to reach significance were the rate of the target and the semantic relatedness of the prime and the target. RT decreases when the target rate decreases ($p<0.001$) and RT decreases when the target and prime are semantically related, as opposed to when prime and target are semantically unrelated ($p<0.001$). These factors are shown in Tables 1 and 2.

Table 1: ANOVA of Model.

	Df	Sum Sq	Mean Sq	F	Df2	p
Target rate	1	13692307.32	13692307.32	712.93	5603	<0.001
Semantic relatedness	1	1493563.41	1493563.41	77.77	5603	<0.001

Table 2: Coefficients of Fixed Effects.

	Estimate
(Intercept)	1604.98
Target rate	–298.05
related=1 (true)	–185.56

Figures 2 and 3 present the results graphically. Both the speech rate of the prime as well as whether the prime and target were at the same rate failed to reach significance.

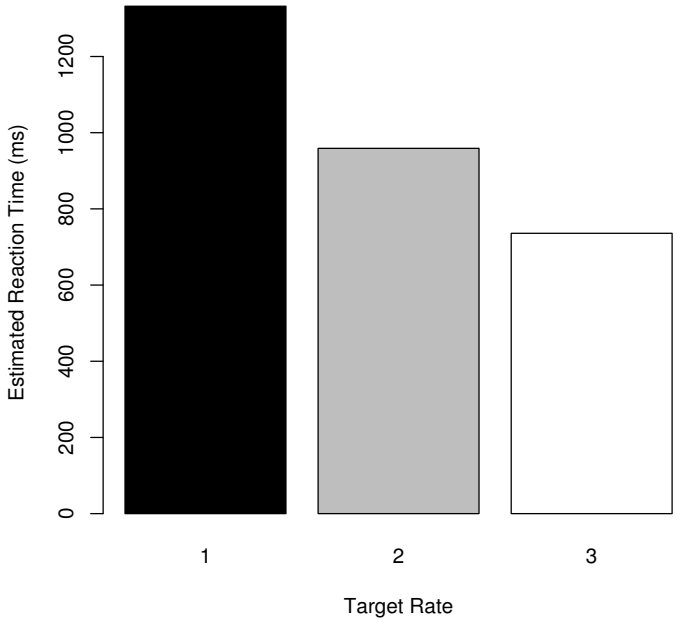


Figure 2: Faster target rates make RT significantly faster, where 1 is ʻslow', 2 is ʻmedium', and 3 is ʻfast'.

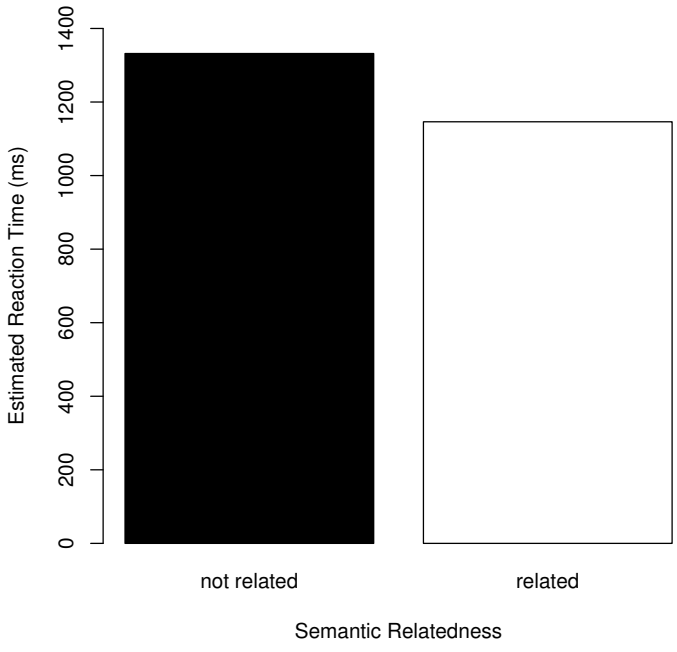


Figure 3: Semantic priming makes RT significantly faster.

3.0 Discussion

Myers's challenge was to demonstrate whether or not speech rate facilitates RT. In order to make sure that any effect we found was lexical, we designed our experiment so that any such speech rate priming effect would get piggybacked on top of a semantic priming effect. But we did not anticipate a highly significant correlation between target rate and RT, so we did not control for it. For example, we would have expected participants to respond faster to a slow–slow word pair (because prime and target are at the same speech rate) than they would to a slow–fast word pair (because prime and target are at different speech rates). However, because of the correlation between target rate and RT, participants would respond faster the slow–fast word pair because the target word is fast. As a consequence, our result was ultimately inconclusive with regard to Myers's challenge. It neither confirms nor undermines the possibility of a priming effect for speech rate because the effect of target rate on RT could be overshadowing an effect of same rate on RT.

While there is a literature on how RT differs in phoneme monitoring experiments for different phoneme types (van Ooijen, Cutler, & Norris, 1992; Cutler, van Ooijen, Norris, & Sanchez-Casas, 1996) and how RT increases with phoneme duration (West, Slater, Coleman, & Cortina Borja, 2000), to our knowledge, no previous research has demonstrated the effect of target speech rate on RT. Although unexpected, it is an interesting result that carries some important implications. For example, had we or Myers known about it ahead of time, then we might have explored the lexical status of speech rate by using some other measure, like accuracy (which is what Goldinger used in his voice experiments); we are currently exploring such alternatives in ongoing work.

Further work on the unexpected correlation between target rate and RT could also be interesting. For example, what does it tell us about processing? We note that, although it did not reach significance, we also observed a trend for faster prime rate to correlate with faster RTs. Together with the target rate effect, this might suggest a linguistic processing effect that degrades rapidly over time. One line of research might investigate just how quickly the effect degrades. On the other hand, it is also possible that the effect of target rate may be merely *strategic*—an artifact of humans sitting experiments which tells us little or nothing about linguistic processing 'in the wild'.

Finally, we successfully replicated the well-known effect of semantic priming (e.g., Meyer & Schvaneveldt, 1971; Neely, 1991; Radeau, 1983; Andruski et al., 1994).

4.0 Conclusion

It is important to report negative results. Ours was an interesting one, because it brought to light an unexpected and highly significant effect for target rate and RT. While this is not the effect we initially set out to test, it provides insight into participants' behaviour on semantic priming experiments and has methodological implications for experiments that use similar paradigms.

The experiment was, however, ultimately inconclusive with regard to the original question of whether speech rate is encoded in the lexicon. Further work will be needed to settle the debate between Generative Grammar and Exemplar Theory.

Acknowledgements

Thanks to Andrew Slater for writing the RT-Bonus software; to Elinor Keane, the voice of the stimuli; to Katie Drager, for helpful comments and help with R; to Kim Plunkett and Elinor Payne for helpful comments; to all of the participants; and to all of the graduate students who helped with the experiment: Anna Goodhand, Anna Silk, James Webb, May Chan, Sabine Wahl, Soyoung Lee, Stephen Kaye, Thomas Culetto, and Zoe Sternberg.

References

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*, 163–187.

Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Bod, R. (1998). Beyond Grammar: An Experience-Based Theory of Language. Cambridge: Cambridge University Press.

Boersma, P. (2001). Praat: A system for doing phonetics by computer. *Glot International*, *5 (9/10)*, 341–345.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Cutler, A., van Ooijen, B., Norris, D., & Sanchez-Casas, R. (1996). Speeded detection of vowels: A cross-linguistic study. *Perception and Psychophysics*, *58*, 807–822.

Daelemans, W., & van den Bosch, A. (2005). *Memory-Based Language Processing*. Cambridge: Cambridge University Press.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166–1183.

Goldinger, S. D. (1997). Words and voices: perception and production in an episodic lexicon. In K. Johnson and J. W. Mullenix (Eds.), *Talker Variability in Speech Processing* (pp. 33–66). New York: Academic Press.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, Joanne K. (1992). Form-based priming in spoken word recognition: the roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 1211–1238.

Halle, Morris. (1997). Some consequences of the representation of words in memory. *Lingua*, *100*, 91–100.

Hay, J., & Bresnan, J. (2006). Spoken Syntax: The Phonetics of *giving a hand* in New Zealand English. *The Linguistic Review: Special Issue on Exemplar-Based Models in Linguistics*, *23*, 321–349.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson, & J. Mullenix (Eds.), *Talker variability in speech processing* (pp. 146–165). San Diego: Academic Press.

Johnson, K. (2007). Decisions and mechanisms in Exemplar-based phonology. In M.-J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental Approaches to Phonology* (pp. 25–40). Oxford: Oxford University Press.

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The Computer and Literary Studies.* Edinburgh University Press. Electronic version re-structured and documented by Michael Wilson and available from the Oxford Text Archive (http://www.ota.ox.ac.uk/).

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, N. Werner, & T. Rietveld (Eds.), *Labphon 7* (pp. 637–676). Berlin: Mouton.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed Memory and the Representation of General and Specific Information. *Journal of Experimental Psychology: General, 114 (2)*, 159–188.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*, 453–467.

Myers, S. (2000). Boundary disputes: the distinction between phonetic and phonological sound patterns. In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological knowledge: conceptual and empirical issues* (pp. 245–272). Oxford: Oxford University Press.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G. W. Humphreys (Eds.), *Basic Processes in Reading: Visual Word Recognition* (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: Benjamins.

Radeau, M. (1983). Semantic priming between spoken words in adults and children. *Canadian Journal of Psychology*, *37 (4)*, 547–556.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Software available online (http://www.R-project.org).

van Ooijen, B., Cutler, A., & Norris, D. (1992). Deletion of vowel and consonants with minimal acoustic variation. *Speech Communication*, *11*, 101–108.

West, P., Slater, A, Coleman, J., & Cortina Borja, M. (2000). Reaction time in phoneme monitoring varies with segment duration. *Oxford University Working Papers in Linguistics, Philology and Phonetics*, *5*, 87–95.

The Hawaiian Passive:
A Neural Network Simulation

‘Ōiwi Parker Jones[a,c] and Julien Mayor[b,c]

[a] Phonetics Laboratory, University of Oxford
[b] Department of Experimental Psychology, University of Oxford
[c] Centre for Connectionist Modelling, University of Oxford

Abstract

The passive provides arguably the most famous dataset in Polynesian linguistics. While most explorations of the passive over the past 40 years have been framed in terms of Hale’s (1968) phonological and morphological analyses, we instead frame the Hawaiian passive in terms of a word-based analysis. As Parker Jones (2008) has done for New Zealand Māori, we model passivization in Hawaiian as a mapping from active verbs to passive verbs in a feed-forward neural network. Unlike Māori, the Hawaiian passive exhibits productivity for multiple categories with thematic consonants. By scrutinizing the model, we conclude that passivization in Hawaiian is exemplar-driven.

1.0 Introduction

The Polynesian passive provides arguably the most famous dataset in Polynesian linguistics. While the literature has generally focused on New Zealand Māori (e.g. Hale, 1968, *et seq*.), similar patterns have been widely observed in other Polynesian languages, such as Hawaiian (Elbert & Pukui, 1979). The problem seems to have been inherited in all of these Polynesian languages from a shared ancestor language. For example, consider the Proto-Austronesian passive reconstructions in (1).

(1) Active and passive verbs in Proto-Austronesian

a.	**inum*	‘to drink’	(active verb)
b.	**inum-ia*	‘to have drank’	(passive verb)

The active verb in (1a) appears to be related to the passive verb in (1b) by suffixation of a passive *-ia* morpheme. However, Polynesian languages like Hawaiian crucially differ from (1) by not allowing word-final consonants (e.g., Blevins, 2004). Consequently, the *m* in **inum* or **inum-ia* has no correspondent in Hawaiian *inu*, although it does have a

correspondent in the Hawaiian *inu-m-ia*, since the *m* is not word-final in the passive.

(2) Active and passive verbs in Hawaiian

a.	*inu*	‘to drink’	(active verb)
b.	*inu-m-ia*	‘to have drank’	(passive verb)

Diachronically stranded consonants, like *-m-*, are called ‘thematic consonants’ (Milner, 1966). The crux of the problem is, then, to explain, in synchronic terms, how a speaker of a language like Hawaiian knows which subset of thematic consonants goes with which passive form, especially in productive generalizations, like passivized loanwords—such as *laka-ʻ-ia* ‘to have been locked’ (< English *lock*).

There are six thematic consonants in Hawaiian (i.e., *-h-*, *-k-*, *-l-*, *-m-*, *-n-*, *-ʻ-*). The lack of a thematic consonant can also be observed, as in *mala-ia* ‘to have soured’. Finally, some passives appear to take another set of suffixes, *-a* and *-na*, as in *nane-a* ‘to have riddled’ and *haʻi-na* ‘to have told’. These are all exemplified in (3) (Pukui & Elbert, 1986).

(3)

Active	Passive	Gloss
malu	*maluhia*	‘peace’
holo	*holokia*	‘run’
kaʻa	*kaʻalia*	‘roll’
inu	*inumia*	‘drink’
kuhi	*kuhinia*	‘point’
hume	*humeʻia*	‘put on a loincloth’
mala	*malaia*	‘sour’
nane	*nanea*	‘riddle’
haʻi	*haʻina*	‘tell’

In this paper, we will ignore passive verbs like *nane-a* and *haʻi-na* to focus on the thematic consonants.

The question is about the choice of thematic consonants in passive verbs. Why, for instance, is *malu-h-ia* the passive for *malu*, while **malu-k-ia*, **malu-l-ia*, and **malu-m-ia* are not? In particular, how does a speaker of Hawaiian know which thematic consonants are acceptable when passivizing a new verb? Our answer will be to propose a relatively detailed and probabilistic view of the Hawaiian lexicon, expanding on previous work on the Māori passive (Parker Jones, 2008). We model this proposal as a connectionist neural network, which maps active verbs to sets of passive verbs. For the cases in which we have enough data to warrant any conclusions, we rule out the null hypothesis that the Hawaiian passive is not probabilistic. In the discussion section, we

consider the importance of this result within the literature on the Polynesian passive.

2.0 Data

In this section, we consider our dataset. We mined our dataset from the standard Hawaiian dictionary (Pukui & Elbert, 1986), which represents the culmination of Pukui and Elbert's famous lexicographic collaboration (Pukui & Elbert, 1957, 1964, 1971, 1986). It also signifies the continuation of a deeper tradition, drawing from the work of Andrews (1865), Lyons (1878), Parker (1922), and Judd, Pukui, and Stokes (1945) (see Schütz, 1994, pp. 206–240).

Pukui and Elbert's (1986) dictionary contains about 29,000 entries. Passive forms are marked (with 'Pas/imp.' for passive/imperative) as in (5).

(5) **inu.mia**. Pas/imp. of *inu*. (PPN *inumia*.)

Unfortunately, we were only able to find 105 passive forms from Pukui and Elbert's dictionary. These can be summarized by Category (i.e., the set of passive suffixes that each word takes) and Count (the number of words belonging to each category) as in Table 1. For example, active verb *malu* belongs to the {*-hia*, *-ʻia*}-category as it is only associated with the passive forms *malu-hia* and *malu-ʻia*.

Table 1: Target Categories at a Glance (Pukui-Elbert corpus)

Category	Count
{*-hia*, *-ʻia*}	65
{*-lia*, *-ʻia*}	22
{*-hia*, *-lia*, *-ʻia*}	6
{*-nia*, *-ʻia*}	4
{*-kia*, *-ʻia*}	2
{*-hia*, *-mia*, *-ʻia*}	2
{*-mia*, *-ʻia*}	1
{*-kia*, *-lia*, *-ʻia*}	1
{*-hia*, *-lia*, *-nia*, *-ʻia*}	1
{*-hia*, *-kia*, *-ia*, *-ʻia*}	1

Because of the limited size of this dataset, we also mined a second Hawaiian dictionary (Kōmike Huaʻōlelo, 2003). Assembled in 1987, the Kōmike Huaʻōlelo ('Lexicon Committee') functions to record Hawaiian words that have entered the Hawaiian language since Pukui and Elbert completed their dictionary in 1986. The Kōmike Huaʻōlelo also records

old Hawaiian words (or extended definitions) that were somehow excluded from Pukui and Elbert 1986. We were particularly interested in this second dictionary as textual evidence for the productivity of the Hawaiian passive. As Pukui and Elbert do, the Kōmike Huaʻōlelo mark passives in their dictionary.

(6) **hiʻoꞏhia** *ʻa‘* Discovered, existence revealed. *Ua hiʻohia ʻo Hawaiʻi nei no kō nā ʻāina ʻē Kāpena Kuke*. Hawaiʻi was discovered for foreigners by Captain Cook. Comb. *hiʻo* (*Tah*., look) + *-hia*. Cf. *kaunānā*. *Mea hiʻohia*. Find, as an archeological find.

The example in (6) stands out because the active verb *hiʻo* 'look' (< Tahitian *hiʻo*) combines with the passive suffix *-hia*, rather than with *-ʻia*. Elbert and Pukui (1979) described three passive suffixes as productive: *-ʻia*, *-a*, *-na*. Of these, only *-ʻia* contains a thematic consonant. Therefore, the entry in (6) provides textual support for multiple productive thematic consonants in Hawaiian.

The yield from the Kōmike Huaʻōlelo dictionary was even smaller than from the Pukui-Elbert dictionary: we found 11 passives. Table 2 provides a descriptive overview of these.

Table 2: Target Categories at a Glance (Kōmike Huaʻōlelo corpus)

Category	Count
{*-hia*,*-ʻia*}	7
{*-lia*,*-ʻia*}	2
{*-mia*,*-ʻia*}	2

In sum, we only found 116 passive forms, which raises a concern about the sparseness of the dataset. What statistical generalizations can we extract from such a small sample? Indeed, there are methods for dealing with small datasets, like using leave-one-out validation, as we do in §3. Still, as we shall see in §4, the small dataset limits our ability to say much about most thematic consonants. The real test cases were the most frequent thematic consonants in our dataset: *-hia*, *-lia*. *-ʻia*. Of these, *-ʻia* was the most frequent in the corpus. As a rule, it occurs with every word (Elbert & Pukui, 1979). Because of its ubiquity, the dictionaries do not waste any ink listing it in the entries. So, we added it to every word in the dataset ourselves.

3.0 Method

In this section, we describe our methods: how we pre-processed the data (§3.1); how we designed the network's architecture (§3.2); how we measured error in the network (§3.3); and how we defined the baselines against which to evaluate the network's performance (§3.4).

3.1. Pre-processing

In pre-processing, we organized the input data into syllables using a three-cell span. The first position in the span was reserved for an onset consonant, the second for a primary vowel, and the third for a secondary vowel (i.e., the second vowel in a diphthong).

Parker Jones (2008) considered two syllable coding schemes for New Zealand Māori. However, unlike Māori there is no Hawaiian controversy about sequences of a long vowel followed by a short vowel (i.e., about whether the vowels are tautosyllabic or not). For instance, Māori has been described as stressing the first of three light syllables in a three syllable word, like *tá.ŋa.ta* 'man' (de Lacy, 2004, p. 498). So, there is a spurious ambiguity in Māori words like *haere* 'go': there the adjacent vowels could be analyzed as part of a diphthong *háe.re* or in hiatus *há.e.re*. In either case, the stress pattern is predicted to be the same. So, there was some question as to the correct syllabification of such words.

This is never the case in Hawaiian. A three light-syllable word in Hawaiian like *ka.ná.ka* 'man' is stressed on the penultimate syllable. On the other hand, the first vowel in a diphthong receives the stress, as in *'ái* 'food'. So, stress position can be used as a diagnostic to determine the syllable structure of words like *liona* 'lion' and *loina* 'custom' (for details see Schütz, 1981). The vowel cluster in the former exemplifies hiatus, as the stress appears on the second vowel: *li.ó.na*. The vowel cluster in the latter exemplifies a diphthong, as the stress appears on the first vowel: *lói.na*. Incidentally, it is not the case that *o* attracts stress, as it is not always stressed: *ko.é.na* 'remainder'. Nor do loanwords pattern differently: *li.ó.na* (< English *lion*) exemplifies hiatus, while *kái.la* (< English *style*) exemplifies the diphthong pattern. Parker Jones (2005) ran a large-scale computational study of the stress patterns of Hawaiian, and found these patterns to be stable. Indeed, there appears to be no ambiguity in Hawaiian syllabification analogous to the Māori case. So, as there was no controversy, we did not bother to try different syllabic coding schemes for Hawaiian.

The Kōmike Huaʻōlelo dataset is small enough to include in its entirety, to illustrate our approach to coding. Table 3 represents the input coding before we transformed each cell into an 11-bit vector.

Table 3: Input Coding for Illustrative Data-subset (before transformations)

SYLL			SYLL			SYLL			SYLL			Active Forms:
C	V	V	C	V	V	C	V	V	C	V	V	
–	–	–	–	–	–	h	i	–	ʻ	o	–	*hiʻo*
–	–	–	–	–	–	h	a	–	k	u	–	*haku*
h	o	–	ʻ	o	–	k	ā	–	w	a	i	*hoʻokāwai*
–	–	–	h	o	–	ʻ	o	–	m	a	u	*hoʻomau*
–	–	–	k	a	u	–	o	–	m	o	–	*kauomo*
–	–	–	k	u	–	ʻ	i	–	p	ē	–	*kuʻipē*
–	–	–	h	o	–	ʻ	o	–	h	a	u	*hoʻohau*
–	–	–	–	–	–	ʻ	o	–	p	i	–	*ʻopi*
–	–	–	–	–	–	–	–	–	p	ā	–	*pā*
–	–	–	p	a	–	ʻ	a	–	p	ū	–	*paʻapū*
–	–	–	–	–	–	p	ē	–	ʻ	ā	–	*pēʻā*

The inputs in Table 3 map to the target vectors in Table 4. The output can be read as -C*ia* where each segment fills C, except for the null set ∅.

Table 4: Target Codings

h	k	l	m	n	∅	ʻ	Passive Forms:
1	0	0	0	0	0	1	{*hiʻo-hia*, *hiʻo-ʻia*}
1	0	0	0	0	0	1	{*haku-hia*, *haku-ʻia*}
1	0	0	0	0	0	1	{*hoʻokāwa-hia*, *hoʻokāwai-ʻia*}
0	0	1	0	0	0	1	{*hoʻomau-lia*, *hoʻomau-ʻia*}
0	0	1	0	0	0	1	{*kauomo-lia*, *kauomo-ʻia*}
1	0	0	0	0	0	1	{*kuʻipē-hia*, *kuʻipē-ʻia*}
0	0	0	1	0	0	1	{*hoʻohau-mia*, *hoʻohau-ʻia*}
1	0	0	0	0	0	1	{*ʻopi-hia*, *ʻopi-ʻia*}
0	0	0	1	0	0	1	{*pā-mia*, *pā-ʻia*}
1	0	0	0	0	0	1	{*paʻapū-hia*, *paʻapū-ʻia*}
1	0	0	0	0	0	1	{*pēʻā-hia*, *pēʻā-ʻia*}

Each input cell was subsequently transformed into an 11-bit vector where the bits corresponded to feature specifications, assuming a fairly rich feature set (i.e. 11 binary features providing 2^{11} contrasts, where 2^5 contrasts would more than suffice). We assumed a rich set of features: the idea was to give the network more than it might need to work with, as we did not know ahead of time that it would extract any statistical patterns at

all. Many analyses, for example, would collapse the features *labial* and *round*, as both bear on lip articulation. In our model, both features were posited as initially unrelated (though the network was free to figure out that they go together, if indeed they do).

The feature analysis is summarized in Table 5, where segments are shown across the rows and features down the columns. Perhaps the first feature to note is the *vocalic* one, distinguishing between consonants and vowels. We chose *vocalic* rather than *consonantal*, because with *consonantal* no feature distinguished between the null-segment and *e* (which would be 0 for *consonantal* but 1 for *vocalic*). The null-segment (–) is an 11-bit vector of 0s. We can thus consider the remaining feature specifications for consonants and vowels.

Table 5: Segment to Feature Correspondences

	labial	*lingual*	*glottal*	*nasal*	*plosive*	*continuant*	*vocalic*	*long*	*round*	*high*	*low*
h	0	0	1	0	0	1	0	0	0	0	0
k	0	1	0	0	1	0	0	0	0	0	0
l	0	1	0	0	0	1	0	0	0	0	0
m	1	0	0	1	0	0	0	0	0	0	0
n	0	1	0	1	0	0	0	0	0	0	0
p	1	0	0	0	1	0	0	0	0	0	0
w	1	0	0	0	0	1	0	0	0	0	0
‘	0	0	1	0	1	0	0	0	0	0	0
a	0	0	0	0	0	0	1	0	0	0	1
ā	0	0	0	0	0	0	1	1	0	0	1
e	0	0	0	0	0	0	1	0	0	0	0
ē	0	0	0	0	0	0	1	1	0	0	0
i	0	0	0	0	0	0	1	0	0	1	0
ī	0	0	0	0	0	0	1	1	0	1	0
o	0	0	0	0	0	0	1	0	1	0	0
ō	0	0	0	0	0	0	1	1	1	0	0
u	0	0	0	0	0	0	1	0	1	1	0
ū	0	0	0	0	0	0	1	1	1	1	0
–	0	0	0	0	0	0	0	0	0	0	0

We assumed six consonant features: three were place features (*labial*, *lingual*, *glottal*) and three were manner features (*nasal*, *plosive*, *continuant*). The segments *m*, *p*, *w* are *labial*; *k*, *l*, *n* are *lingual*; *h* and ‘ are *glottal*. The segments *m*, *n* are *nasal*; *k*, *p*, ‘ are *plosive*; *h*, *l*, *w* are *continuant*.

Vowels were specified using four features (*long*, *round*, *high*, *low*). Although the feature *long* went out of fashion in phonology with the ascension of moraic analyses in the 1980s, we found it more expedient to encode phonemic length with *long* than by introducing a moraic tier (a column representation does not lend itself to modelling re-entrance, besides it is not clear what, if anything, a moraic tier would buy in this instance). The *long* vowels are *ā*, *ē*, *ī*, *ō*, *ū* (hence, the short vowels are *a*, *e*, *i*, *o*, *u*). The *round* vowels are *o*, *ō*, *u*, *ū*. The *high* vowels are *i*, *ī*, *u*, *ū* and the *low* vowels are *a*, *ā*—*e*, *ē*, *o*, *ō* are neither high nor low.

Finally, in addition to the feature-based coding scheme, we also tried a *random* coding of segments into 11-bit vectors. The motivation for this second coding scheme was the question of how much a feature analysis (even a very richly detailed one, like the one described above) adds to the model. The use of features thus becomes an empirical question.

3.2. Architecture

We used a standard multi-layer neural network with one hidden layer. The activity of the neurone was a sigmoid function F of its inputs, with the addition of a bias term θ:

$$(7) \qquad act_i(t+1) = F\left(\sum_j \omega_{ji}(t) act_j(t) + \theta_i(t)\right).$$

All neurones within a layer were connected to all neurones in the adjacent layer. In other words, every input node was connected to every hidden node and every hidden node was additionally connected to every output node. There was no lateral connectivity (i.e., connections between nodes in a layer). See Figure 1 for a schematic representation of the proportion of nodes in each layer (bias term not illustrated).

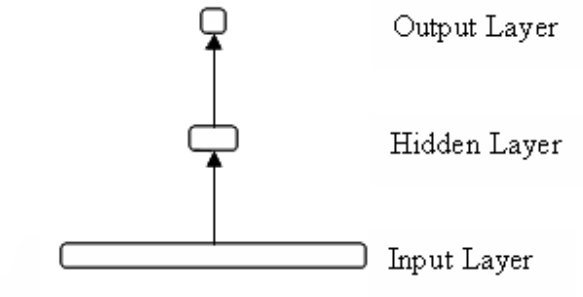


Figure 1: A Schematic of the Network's Architecture

In a preliminary study, the number of nodes in the input layer was explored in order to optimize the number of hidden units needed for the final simulation. Data were split into training, test, and validation sets, and only the training and test sets were used to optimize the number of hidden units, holding out the validation set for later. The results reported in this paper were obtained with 20 nodes in the hidden layer. The input layer contained 132 units and the output layer contained 7 units, as explained below.

All weights were trained with the classical algorithm of backpropagation though time (Werbos, 1974; Rumelhart, 1986), with the following parameters: learning rate=0.1 and momentum=0.05. The weights were initialised with a random distribution centred on zero, within the range of [–0.1, 0.1]. The network was trained for 500 epochs (52,500 sweeps).

3.3. Error measures

In simulation 1, we trained the network using the leave-one-out cross-validation technique. Given a limited dataset, this technique ensures that the model does not over-fit the data. In other words, the network extracts the *statistical regularities* of the data instead of simply *memorizing* the training set.

Using leave-one-out cross-validation, the network was trained on all but one pattern. The hold-out pattern was then presented to the network in the test phase and the associated error was computed by comparing the network's output against the target output. The procedure was repeated for every pattern separately (complete with re-initialization of the weights in the network), ensuring that the test patterns never smuggled any information back to the training sets. A virtue of the leave-one-out method is that it provides an error for each input-output pattern, thereby allowing analysis of each pattern, and class of patterns, separately. We used two measures of performance for the network: the Mean Squared Error; and the Classification Error.

3.3.1 Mean Squared Error

The Mean Squared Error is a continuous error measure: it measures, in the Euclidian space, the distance between the vector of values in the output layer and the vector of their corresponding target values. The Mean Squared Error for input pattern *i* is

$$(8) \qquad E_i = \frac{1}{7}\sum_{j=1}^{7}[\text{out}(i,j) - \text{target}(i,j)]^2 ,$$

where *i* is the active verb and *j* is an output unit (i.e., a passive class). We can define the Mean Squared Error for each of the classes by averaging the error of each input pattern that belongs to the class, as in

$$(9) \qquad E_C = \frac{1}{7}\sum_{j=1}^{7}\frac{1}{c}\sum_{i\in C}[\text{out}(i,j) - \text{target}(i,j)]^2 ,$$

where *c* is the number of patterns that belong to class *C*.

3.3.2 Classification Error

A second useful measure of performance is the Classification Error. Following a forced choice procedure on each of the output node – any value larger than 0.5 is set to 1 and any value lower than 0.5 is set to 0 – we can measure whether to output pattern of activation corresponds to the correct target or not. The overall classification success can be computed for each class as a percentage of correct classification,

$$(10) \qquad P^C = \frac{100}{c}\sum_{j\in C}\delta(\text{output},\text{target}) ,$$

where δ(output, target) equals 1 if the forced choice output matches the target and 0 otherwise. At this point, it is useful to note that the forced choice is enforced on each of the output node, separately. Therefore, the output pattern produced may well correspond to none of the possible targets. The procedure does not force the network to choose between a set of possible targets. Consequently, the measure used here is more conservative.

3.4. Baseline estimates

Having reviewed our training parameters and error measures, we turn now to the set of baselines that we used to compare against the performance of the network. We will define (and later evaluate) two strategies that the network could find in order to minimise errors without actually discovering the mappings between active verbs and classes.

3.4.1 Random guess: adaptation to the range *of target values*

The simplest method the network can develop is to adapt the range of its output values to the range of the target values, namely 0 and 1. A simple method for the network would then be to generate a random output with a flat distribution *p(x)* between 0 and 1. The 'random error' (E_r) – or the Mean Squared Error to a binary output with random values (using the symmetry between 0 and 1) – would then be as follows:

(11) $$E_r = \int_0^1 p(x)[1-x]^2 dx = \frac{1}{3}.$$

The equation assumes that all output nodes can take values of 0 or 1. However, since one of the target values is constant for all input patterns (i.e., the 'default' -*'ia*), this category can be learnt very easily simply by setting the bias unit's value to 1 and by letting all weights from inputs to this output unit be 0. In practice, backpropagation networks excel at learning such simple mappings (e.g., Bishop, 1995), so we can safely assume that the network will find this solution. A more generous baseline would therefore set the error at 0 for this output unit. This gives us the following average 'pseudo-random error' (E_{pr}), which takes into account the constant class:

(12) $$E_{pr} = \frac{6E_r + 0}{7} = 0.286.$$

The result is a model of 'random guess', embodying the literature's assumption that the thematic consonants are unpredictable.

3.4.2 Adaptation to the distribution *of target values*

One shortcoming of the 'random guess' approach is that it assumes that the target values in the *range* are equiprobable (i.e., as often 0 as 1). However, for each of the output nodes, the *distribution* of 0 and 1 turns out to be uneven.

Therefore, by sampling the statistics of the target values (or teacher signals), the network can tune the output bias to minimise the Mean Square Error on each of the nodes, without even looking at the input patterns. Given the proportion of target values p of 0 for a given node j, we can find the optimal activation of the node in order to minimise the Mean Squared Error:

(13) $$E = pa^2 + (1-p)(1-a)^2.$$

A minimisation with respect to the activation $\frac{\partial E}{\partial a}$ gives the optimal value a_{dist} which minimises E:

(14) $a_{dist} = 1 - p$.

The Mean Squared Error associated with this node over all patterns is thus:

(15) $E_{dist} = p(1-p)$.

Similarly, the network can extract the distribution of target values in order to optimise the Classification Error. In this case, the strategy is even simpler; for each output node, the network would simply produce the most frequent value. For example, if the target values for input node *j* take a value of 1 for more than 50% of the input patterns, then the network should always produce an output of 1 (after the forced choice procedure). In this way, the model would be free to ignore the input patterns and focus entirely on the output nodes, and the value of the both the neural network and the exemplar interpretation would be up to question. We will refer to this classification baseline as being P_{dist}.

4.0 Results

In this section, we consider the results for our simulations. We beginning with the overall results, aggregating over all patterns. We then consider a per-item analysis, exploring the network in some detail in order to understand what it has learned.

Table 6 shows the average Mean Squared Errors (and their standard deviations) following leave-one-out cross-validation for the two types of encoding (random bit vectors and phonological encodings); it also shows the baseline errors produced by the random guess (11), pseudo-random guess (12) and the error produced by a network that would adapt to the distribution of target values (15).

Table 6: Overall Results

	Phonological	Bit-vector	Random	Pseudo-random	Distribution
Error	0.094 (0.113)	0.098 (0.119)	0.286	0.333	0.076

We found no statistical difference between the phonological encoding scheme and the random (bit-vector) coding. Both encoding methods lead to similar results, being markedly lower than random and pseudo-random baselines.

The network's overall performance does not reach the level that could be obtained by looking at the distribution of the classes in the data-set. So, we cannot claim that the network has extracted any statistical regularities in the data-set beyond the relative proportion of the different classes without doing some additional work. Here is the problem in a nutshell: is the success of the 'distribution' baseline a general result across the classes, or is it a particular result dependent on a subset of classes which are carrying the trend? To answer this question, we will consider a detailed analysis of the network's performance per class. The goal will be to identify whether the network has managed to build mappings between input active verbs and corresponding classes for any of the classes—remembering from §2 that our dataset was impoverished for a number of suffixes (i.e., all but *-hia*, *-lia*, *-'ia*).

4.1 "Simple" classes

Gladly, we can make use of the power of leave-one out validation to analyse different classes separately. First, we can focus on the Mean Squared Errors associated with the 'simple' classes, by which we mean active verbs accepting one class only, apart from the *-'ia* class (class 7) which is common to all active verbs.

Table 7: Mean Squared Errors associated with "Simple" Classes

Class (learnt)	Form	Occurrence	Phonological (dispersion)	Bit-vector (dispersion)	Distribution
1 (yes)	*-hia*	65	0.053 (0.090)	0.061 (0.093)	0.026
2 (no)	*-kia*	2	0.258 (0.040)	0.286 (0.001)	0.211
3 (yes)	*-lia*	22	0.105 (0.104)	0.106 (0.120)	0.143
4 (no)	*-mia*	1	0.286 (N/A)	0.283 (N/A)	0.214
5 (no)	*-nia*	4	0.289 (0.010)	0.284 (0.003)	0.209
6 (no)	*-ia*	0	N/A	N/A	N/A
7 (yes)	*-'ia*	0	N/A	N/A	N/A

Table 7 shows that for classes 2, 4, and 5, the errors lie in the range of the pseudo-random values (E = 0.286). For these classes, the network's only achievement is to associate all active verbs to the 'default' class (class 7).

Only classes 1 and 3 possess sufficient occurrences for potential mappings from actives verbs to classes to be learnt. Results for class 1 are only outperformed by the distribution baseline (E_{dist}) given in (15), which takes into an account the distribution of classes. In other words, the level of performance obtained for class 1 can be reached by adjusting to the statistics of the target values, thereby ignoring the input patterns. Mean Squared Errors for class 3 are lower than any baseline method, suggesting that the network discovered some statistical structure in that class, which indicates productivity for class 3. However, further analysis is needed as the network's performance does not differ statistically from the distribution baseline.

For another perspective, we can also approach these results in terms of Classification Errors. Recall that, for every output unit, the network forced the choice of having either 0 (for any value below 0.5, before the forced choice) or 1 (for values above 0.5). The network's Classification Errors for each class is shown in Table 8.

Table 8: Classification Errors associated with "Simple" Classes

Class	Occurrence	Phonological	Dist. (noise=0)
1	65	70.77%	100%
2	2	0	0
3	22	40.91%	0
4	1	0	0
5	4	0	0
6	0	0	0
7	0	0	0

The Classification Errors in Table 8 confirm the intuition based on the Mean Squared Error analysis in Table 7. The network manages to classify stems for only two classes correctly: class 1 and class 3. Again, network performance has to be compared to the baselines, in order to ensure that the network has developed a mapping between input active verbs and output classes, rather than merely adjusting its strategy to minimise the error based on the distribution of the output classes alone. The baseline with which we will compare the network's performance is the optimal strategy given that there has been no extraction of information from the input patterns. This strategy would be to reproduce in the output nodes the most frequent class in the training data-set—that is, class 1. As a consequence, all input patterns that belong to class 1 would be classified correctly (100%), while all input patterns that belong to any other class would be misclassified (0%). We should like to rule out the possibility that the network has learned this brute-force strategy which would let it ignore the inputs.

In fact, the overall performance of the network, in terms of classification, is lower than the optimal baseline strategy. The network classifies 70.77% of the patterns in class 1 and 40.91% in class 3, altogether classifying 55 patterns correctly. The E_{dist} baseline correctly classified 65 patterns, all of them belonging to class 1, which reflects the fact that more than 50% of the dataset was associated with class 1.

The strategy of identifying the relative frequency of classes introduces a correlation between output patterns. For example, the addition of noise to output node activation reduces the number of patterns that are classified correctly for class 1 while allowing some of the class 3 patterns to be classified appropriately. In other words, it is possible *a priori* that some intermediate level of performance for the two dominant 'simple' classes (1 and 3) may be an artifact of the addition of noise to E_{dist}. Thus, in order to rule out this explanation for the network's performance, we investigated the role of different noise strengths in the output layer of the network on top of the optimal node activation described in (15). The output activation of nodes 1 and 3 thus becomes

$$a_{dist} = (1 - p) + \Phi_\sigma , \tag{16}$$

where Φ_σ is a normally distributed variable centred in zero and with variance σ^2.

We ran extensive simulations, investigating the role of noise strength on the performance of the baseline method. Figure 2 depicts the number of patterns classified correctly in class 1 and class 3 as the noise level increases (from 0 in the top left corner, to 1).

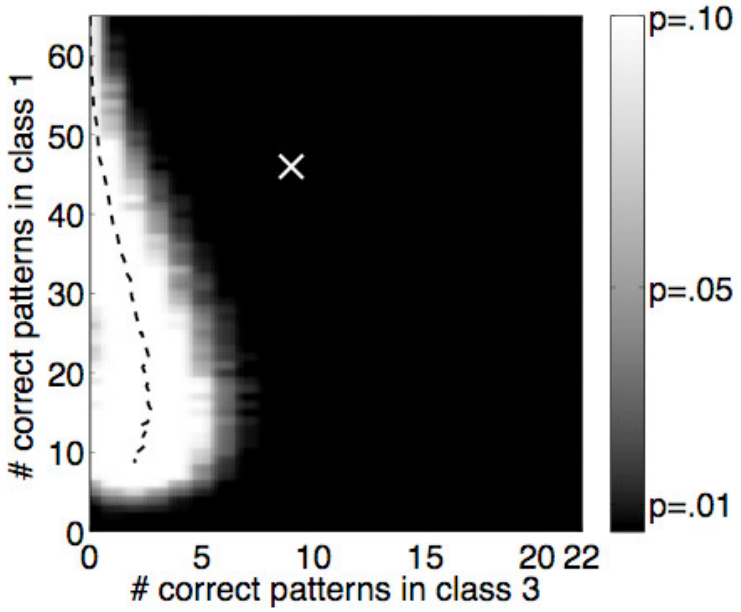


Figure 2: Co-variation of classification performance for classes 1 and 3 for different noise levels, where the white cross is the network's performance and shading indicates p-values (white is insignificant, mid-gray is borderline significant, and black is highly significant)

The dashed line corresponds to the average performance of the baseline method for a given amount of noise and the grayscale indicates the likeliness of reaching a given pair of values (i.e., mid-gray is the threshold $p = 0.05$). If the network’s performance were to have been within the white region of Figure 2, then we could not conclude any role for the input exemplars in the model; the output-based baseline in (16) probably would have accounted for the network’s performance.

However, given that the network’s performance does in fact lie within the black region of Figure 2 (pictured as a white cross), we must conclude with high significance ($p << 0.0001$) that the network made use of the input exemplars and learned mappings between them and the output categories. The likelihood that such an extreme pair of values as achieved by the network could be attributed to noise on top of the E_{dis} baseline is vanishingly small ($p << 0.0001$).

To further confirm that the network managed to discover the statistical regularities in the input patterns, we performed a *t*-test between the activation of the output nodes of class 1 and class 3. In theory, if the network only found a solution in terms of relative proportion of classes, ignoring the input patterns, then output patterns should not be influenced by the type of pattern presented to the network. In contrast, if output vectors are statistically different between class 1 and class 3, it should

mean that the network is taking into account information presented to it in its input layer. Table 9 displays the *p*-values for each of the output nodes.

Table 9: t-*test between class 1 and 3 for each node*

Node	1	2	3	4	5	6	7
p-value	<<0.0001	0.36	<<0.0001	0.86	0.28	0.12	0.29

As expected, the *t*-test based on the activation value of the output nodes between the class 1 input patterns and the class 3 input patterns shows that the output nodes 1 and 3 are the only nodes that differ statistically. All other nodes have activation values that do not differ if the input patterns are from classes 1 or 3. So, the network is effectively filtering the input so as to predict which class the stems belong to.

Given that class 3 is to some extent productive, we can claim that the phonotactics of the active verbs is a predictor of class membership. As such, the network relies on input exemplars as 'memory traces' (i.e., adjustments to the weights); it does not ignore the input in favour of a purely output-based approach.

On the other hand, class 1's performance can still be attributed to a purely output-based dominance effect, since more than 50% of the data belongs to class 1. A simple system ignoring all the input patterns can achieve good performance for class 1, simply by defaulting to it. One way to find out if class 1, like class 3, is also being learnt with reference to the inputs would be to boost the dataset so that less than 50% of the data were associated with class 1. Further work is also required to clarify the status of class 1 in order to discover if it is productive or not.

4.3. "Complex" classes

In a sense, the "simple" classes of the previous subsection are also "complex" insofar as they all include two classes, since the -*'ia* class (class 7) is constant. So, what we refer to here as "complex" classes are those in which active verbs map to three or more categories. With this definition in mind, Table 10 provides an overview of the results for the "complex" classes.

Table 10: "Complex" Classes

Classes (learnt)	Forms	Occurrence	Phonological	Bit-vector
1+3+7 (yes)	*-hia*, *-lia, -ʻia*	6	0.132 (0.088)	0.084 (0.092)
1+3+5+7 (no)	*-hia*, *-lia*, *-ia, -ʻia*	1	0.286 (N/A)	0.286 (N/A)
1+4+7 (no)	*-hia*, *-mia, -ʻia*	2	0.274 (0.185)	0.285 (0.201)
2+3+7 (no)	*-kia*, *-lia, -ʻia*	1	0.143 (N/A)	0.337 (N/A)
1+2+6+7 (no)	*-hia*, *-kia, -ʻia*	1	0.311 (N/A)	0.286 (N/A)

Perhaps the first observation to draw from Table 10 is that the occurrence of all the classes is extremely low. So, it is somewhat surprising that *any* of the "complex" classes were learnt at all. Nonetheless, one "complex" class was learned. The network was able to beat the pseudo-random error for those active verbs that map to both *-hia* and *-lia* (in addition to *-ʻia*). The network error for that "complex" class was 0.084 with a dispersion of 0.092, which is well below the E_{pr} = 0.286.

We would actually like to offer two possible interpretations for the learning of the "complex" class (which we do not follow up here). The first is that the network managed to learn the statistical regularities in these active verbs, even though the number of examples in the training sets was very low. The other interpretation is that the network made use of some knowledge about class 1 and class 3 *independently*, and then successfully applied this knowledge to the "complex" class. Further, since the occurrence of the learnt "complex" class was far below 50% of the dataset, the output-based default is a non-issue.

5.0 Discussion

For the suffixes with sufficient data (i.e., more than 4 tokens in the Pukui-Elbert corpus), the neural network significantly outperformed all baseline measures, including the random guess and the output-only baseline. In fact, the result held for both the "simple" and the "complex" classes, though we did not explore the relationship between the "simple" and "complex" classes. For example, it is possible that the individual successes of both *-hia* and *-lia* contributed to the success of the "complex" {*-hia*, *-lia*, *-ʻia*} class. Alternatively, the model might not have associated the success of the "simple" classes with that of the "complex" ones; in both cases, the successful classes were the ones which provided sufficient training data for the network.

Either way, our simulations support a role for probabilistic phonotactics in the phonology and morphology of Hawaiian. Classical Conditioning is a possible explanation for the neural network's success

(Pavlov 1927). Hawaiian speakers learn to associate active and passive verb pairs through experience. Learning is, furthermore, sensitive to statistical regularities in the phonotactics of the active verbs that take similar passive forms. Speakers then draw on these regularities during generalization. This is a radically different view of the Polynesian passive than has generally been considered (through see Parker Jones, 2008). For example, it requires no concession to 'derivations' or 'underlying forms', or even to 'morphemes'.

Though we have used the term 'suffix' in this paper to describe sets of passive verbs that share a thematic consonant, to do so was merely for convenience. Nothing hinges on the term 'suffix' or indeed 'morpheme' in our model. It suffices to take a purely word-based perspective to the problem, whereby active verbs map to their passive counterparts.

To appreciate the non-morphological—and radical—nature of our position, we shall review the received analyses of Polynesian passive. Most of the literature, which has followed Hale (1968), has assumed that the thematic consonants are unpredictable in the passive forms. Thus, Blevins (1994, p. 29), for example, writes that 'a consonant of *unpredictable* quality appears in the passive' (our italics).

The standard analysis of Polynesian passives has thus been to *lexicalize* them along with the base. The debate following Hale (1968), having assumed that the thematic consonants are unpredictable from the active verbs, has been preoccupied by the orthogonal question of whether it is better lump the thematic consonants along with the stem or with the suffix. These two analyses have been labelled the 'phonological' and the 'morphological' analyses. To situate our model in the wider literature, we therefore summarize that debate below.

In the 'phonological' analysis, active and passive verbs are assumed to derive from stems which may be consonant-final. Thus, to derive an active verb, something like a deletion rule is required, where a sample derivation would be *inum* $\Rightarrow$ *inu*. To derive a passive verb, however, requires only one suffix form: *-ia*. For example, *inum* $\Rightarrow$ *inum-ia*, *kuhin* $\Rightarrow$ *kuhin-ia*, etc. In effect, the 'phonological' analysis synchronically replicates the loss of Proto-Austronesian word-final consonants in Eastern-Polynesian languages (see §1).

In the 'morphological' analysis, passive verbs derive from active verbs (or, if you like, from stems that exactly resemble active verbs). An important motivation for this alternative is that a deletion rule, like the one in the 'phonological' analysis, which is perceived to complicate the account, thereby becomes unnecessary. Nevertheless, unlike the 'phonological' analysis which only required one suffix form, the 'morphological' analysis must assume multiple suffix forms. In order to

curtail overgeneration, the 'morphological' analysis also appends features to its verb stems.

For example, by associating *inu* with the feature [+*m*], the 'morphological' analysis rules out unattested forms like **inu-hia*. A passive derivation might therefore look something like *inu* [+*m*] ⇒ *inu-mia*, while an active derivation would look something like *inu* [+*m*] ⇒ *inu*, where the underlying feature simply does not surface. When no features are specified, a default feature may be assumed such as [+*t*] in some dialects of Māori (see Parker Jones 2008 for critique). Thus, a borrowed verb like *taraiwa* (< English *drive*) would get the feature [+*t*] by default, and correctly derive the passive form *taraiwa-tia*. Consequently, a native stem like *awhi* 'embrace', which takes the suffix -*tia*, also does not need to be explicitly associated with the feature [+*t*], because not associating it with any feature amounts to the same thing.

Crucially, both the 'phonological' and 'morphological' analysis *lexicalizes* the thematic consonants. First, in the 'phonological' analysis, they are lexicalized by being represented as part of the underlying representation of the stem (e.g., *n* in *inum*). Second, the 'morphological' analysis lexicalizes them by the use of features, which are also associated with the stem (e.g., [+*n*] in the lexical entry: *inum* [+*n*]). Even the default feature [+*t*] is effectively lexicalized by virtue of being the only underspecified feature.

By contrast, our analysis is rather different. First, it rejects the partitioning of the analysis into 'underlying' and 'surface' levels. The representations are all *words*. In a derivational model, an underlying form (e.g., *inum*) derives both the active verb (*inu*) and the passive verb (*inumia*). By contrast, in our model, the network simply relates an active verb (e.g., *inu*) to its corresponding set of passive verbs ({*inumia*, *inuʻia*}). Therefore, out network does away with underlying representations and simply models passivization as a relation between words.

A second important way in which our analysis is different from the standard 'phonological' and 'morphological' analyses is that it does not lexicalize the thematic consonants with the active verbs. Although we did not assume underlying representations in which to lexicalize thematic consonants, we still could have talked about lexicalizing them on the active verbs. But instead we asked ourselves how predictable the passive forms were from the phonotactic patterns of the active forms. So, contrary to the standard assumption that the thematic consonants are *unpredictable* from the active verbs, we in fact found that they were *predictable* for those cases in which we had enough data to test.

Unlike the 'phonological' or 'morphological' analyses, which may only postulate a single default case (i.e., *-tia* or *-ngia* in Māori, depending

on the dialect) which can apply to productive cases (e.g., passivization of loanwords), our analysis can handle *multiple* productive classes. In fact, we predict that all of the classes should be more or less productive. This is a critical strength of our model when you consider that Hawaiian presents at least two productive classes (-*ʻia* and -*hia*), as illustrated by the following examples:

(16) *kalaiwa-ʻia* 'driven' (< English *drive*)
(17) *hiʻo-hia* 'discovered' (< Tahitian *hiʻo*)

While we concede that the evidence for the second productive class in (17) is scant, text-based, and open to other criticisms, we still think that our model's ability to accommodate this variation is interesting and should be explored experimentally with native speaking Hawaiians. Rather than assume that all productive forms ought to belong to just type, our model predicts that there should be gradient acceptable productive forms.

Our model also outperforms the standard analyses in its ability to predict the class of known words, as we did using leave-one-out rotation estimation. This shows robustness within the system. It also provides an opportunity to investigate analogical change. For example, many passive forms have equivalents in Hawaiian and Māori, such as *inu-m-mia* (Hawaiian) and *inu-m-mia* (Māori), or even *ʻau-l-ia* (Hawaiian) and *kau-r-ia* (Māori) since *l* and *r* are ultimately reflexes of the same historical phoneme (as are *ʻ* and *k*). But there are cases in which the languages diverge. For instance, the passive form of *ʻimi* 'to seek' in Hawaiian is *ʻimi-h-ia*, whereas the passive form of the equivalent verb *kimi* 'to seek' in Māori is *kimi-t-ia*. Crucially, *h* and *t* are not reflexes of the same historical phoneme. Therefore, they suggest drift by one or even both languages away from the ancestral form. This provides yet another opportunity to test our neural network of Hawaiian, in addition to Parker Jones's (2008) neural network of Māori.

The 'phonological' and 'morphological' analyses would only predict drift from *h* to *t* or *ng* in Māori (where either *t* or *ng* is the Māori default), but precludes the possibilities that Hawaiian, rather than Māori, may have drifted, or that they have both drifted. The theory would be that different lexical frequencies in the languages could create analogical pull in different directions, when speakers try to generate a passive form for an active verb—without knowing what the historical form of the passive verb had been. Such an investigation could, data permitting, be extended to other related languages, as well. Or, it could even be extended to

dialectical differences, like the Māori preference in some parts of New Zealand for *t* and others for *ng*.

In general, this work adds to a growing connectionist, data-oriented, and exemplar-based literature (e.g., Rumelhart et al., 1986; McClelland et al., 1986; Daelemans & van den Bosch, 2005). For more on the relationship between exemplars and Connectionism, see McClelland and Rumelhart (1985).

6.0 Conclusion

In this paper, we challenged the status quo analysis of Polynesian passives by asking how predictable passive verbs from their corresponding active verbs. In so doing, we were extending previous work on Māori (Parker Jones, 2008) to Hawaiian, a language with interesting differences. For example, Hawaiian shows evidence for multiple passive forms with thematic consonants being productive.

We modelled a word-based and probabilistic analysis in an artificial neural network, which we tested against two baselines. One baseline (random guess) represented the null hypothesis that thematic consonants are unpredictable. This was shown to be false for all of the classes that we had enough data to test.

The other baseline (adaptation to the *distribution* of the target values) represented the possibility that the best strategy would be to ignore the active verbs and just count how many passive verbs belongs to each class—using the most likely class for all passive verbs. By running extensive simulations we were able to conclude that the network was indeed paying attention to the active verbs.

Finally, we considered the context of this work within the broader literature and pointed to a number of interesting empirical implications of our approach. For example, we predict that unobserved passive forms should demonstrate gradient relative acceptability. Also, we suggested that a cross-linguistic application of similar models could be used to explore historical drift of passive forms between Polynesian languages.

Acknowledgements

Cheers to John Coleman, Greg Kochanski, Elinor Payne, and the audience members at The 83rd Annual Meeting of the Linguistic Society of America.

References

Andrews, L. (1865) A *Dictionary of the Hawaiian Language, to which is appended an English-Hawaiian Vocabulary and a Chronological Table of Remarkable Events*. Honolulu: Whitney. Reprinted in 2003 with new introductions by N. K. Silva & A. J. Schütz. Honolulu: Island Heritage.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.

Blevins, J. 1994. A phonological and morphological reanalysis of the Maori passive. *Te Reo*, *37*, 29–53.

Blevins, J. (2004). The mystery of Austronesian final consonant loss. *Oceanic Linguistics*, *43(1)*, 208–213.

Chomsky, N. (1975). *The Logical Structure of Linguistic Theory*. Chicago: University of Chicago Press. Harvard manuscript, 1955.

Daelemans, W., & van den Bosch, A. (2005). *Memory-based language processing*. Cambridge: Cambridge University Press.

de Lacy, P. (2004). Maximal words and the Maori passive. In J. McCarthy (Ed.), *Optimality Theory in Phonology: A Reader* (pp. 495–512). Blackwell, Oxford.

Elbert, S. H., & Pukui, M. K. (1979). *Hawaiian Grammar*. Honolulu: University of Hawai‘i Press.

Hale, K. (1968). Review of Hohepa (1967). *Journal of the Polynesian Society*, *77*, 83–99.

Judd, H. P., Pukui, M. K., & Stokes, J. F. G. (1945) *Introduction to the Hawaiian Langauge: An English-Hawaiian Vocabulary Comprising Five Thousand of the Commonest and Most Useful English Words and their Equivalents, in the Modern Hawaiian Speech, Correctly Pronounced, with a Complementary Hawaiian-English Vocabulary*. Honolulu: Tongg.

Komike Hua‘ōlelo. (2003). *Māmaka Kaiao: A Native Hawaiian Vocabulary*. Honolulu: University of Hawai‘i Press.

Lyons, L. (1878). *The Friend*, 2 September, p. 73. Letter from Waimea, Hawai‘i, 9 August 1878 on the Hawaiian language and Andrews’s dictionary.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed Memory and the Representation of General and Specific Information. *Journal of Experimental Psychology*, *114 (2)*,159–188.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* Cambridge, MA: MIT Press.

Milner, G. (1966). *Samoan Dictionary: Samoan-English, English-Samoan*. London: Oxford University Press.

Parker, H. H. (1922). A Dictionary of the Hawaiian Language. Bernice P. Bishop Museum Special Publication 8. Honolulu: Board of Commissioners of Public Archives of the Territory of Hawaii. This is also known as the Andrews-Parker dictionary, because Parker 1922 is a revision of Andrews 1865.

Parker Jones, ʻŌ. (2005). *The Stress Patterns of Hawaiian*. M.Phil. thesis, University of Oxford.

Parker Jones, ʻŌ. (2005). Phonotactic Probability and the Maori Passive. In J. Eisner, & J. Heinz (Eds.), *Proceedings of the Tenth Meeting of the ACL Special Interest Group on Computational Morphology and Phonology* (pp. 39–48).

Pavlov, I. P. (1927). *Conditioned Reflexes*. London: Oxford University Press.

Pukui, M. K., & Elbert, S. H. (1957). *Hawaiian-English Dictionary*. Honolulu: University of Hawaiʻi Press.

Pukui, M. K., & Elbert, S. H. (1964). *English-Hawaiian Dictionary*. Honolulu: University of Hawaiʻi Press.

Pukui, M. K., & Elbert, S. H. (1971). *Hawaiian Dictionary: Hawaiian-English, English-Hawaiian*. Honolulu: University of Hawaiʻi Press.

Pukui, M. K., & Elbert, S. H. (1986). *Hawaiian Dictionary: Hawaiian-English, English-Hawaiian*. Revised and Enlarged Edition. Honolulu: University of Hawaiʻi Press.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processing. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 194–248). Erlbaum, Mahwah, New Jersey.

Ryan, P. M. (1989). *The Revised Dictionary of Modern Māori*. Heinemann Education, Auckland. Third edition.

Schütz, A. J. (1994). *Voices of Eden: A History of Hawaiian Language Studies*. Honolulu: University of Hawaiʻi Press.

Schütz, A. J. (1981). A reanalysis of the Hawaiian vowel system. *Oceanic Linguistics*, *20 (1)*, 1–43.

Werbos, P. J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Ph.D. thesis, Harvard University.

Rhythmic Modification in Child Directed Speech

Elinor Payne[a], Brechtje Post[b], Lluïsa Astruc[c], Pilar Prieto[d], Maria del Mar Vanrell[e]

[a] Phonetics Lab and St Hilda's College, University of Oxford
[b] RCEAL and Jesus College, University of Cambridge
[c] Open University and Dept of Spanish and Portuguese, University of Cambridge
[d] Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra
[e] Departament de Filologia Catalana, Universitat Autònoma de Barcelona

Abstract

Interval-based rhythm metrics were applied to the speech of English, Catalan and Spanish female adults addressing their children (aged 2, 4 and 6 years) and compared with their speech when addressing adults. Results reveal that when mothers speak to their children, their speech is more vocalic and more even-timed than when they address other adults. However, cross-linguistic differences are still maintained, suggesting that indices specific to both speech style and language co-exist.

Keywords

Rhythm, Child Directed Speech, Cross-linguistic, English, Spanish, Catalan

1.0 Introduction

This paper examines the extent to which the rhythmic properties of adult speech vary according to situational context, and specifically whether the speaker is addressing a child or another adult. To measure rhythmic properties we apply a set of well known 'rhythm metrics' which extract and quantify certain (predominantly durational) properties[1] from the acoustic signal (cf Ramus, Nespor & Mehler, 1999; Dellwo, 2004; Low et al, 2000; Grabe & Low, 2002; White & Mattys, 2007). While these metrics have been widely applied to capturing perceived cross-linguistic differences in rhythm in (largely non-spontaneous) reading styles of adult speech, little attention has been paid to determining

[1] Following White and Mattys (2007), other rhythm metrics, based on variability in the duration of phonological constituents (e.g. Deterding, 2001), were not applied.

whether such differences are evident in different speech contexts known to trigger different speaking styles, e.g. child-directed-speech (CDS). We examine the rhythmic properties of CDS for English, Spanish and Catalan as a function of child age, and compare them with adult-directed speech (ADS) in the same languages. We consider the possible interplay of structural and implementation factors in the generation of rhythmic percept, and also the degree to which temporal modifications in CDS may be interpreted as accommodations towards the speech of the child in question.

1.1 Background on speech rhythm measurement

Though the acoustic basis of perceived cross-linguistic differences is elusive, the percept of a distinction between two broad language categories (i.e. between the 'Morse Code' rhythm of e.g. Germanic languages and the 'machine gun' rhythm of e.g. Romance languages) is strong, and empirically supported. For example, Ramus, Dupoux and Mehler, 2003, show adult ability to distinguish *between* rhythmic categories, but not *within* them. The validity and efficacy of rhythm metrics in capturing perceived differences have been widely discussed and evaluated (see in particular an overview and comparison of all metrics by White and Mattys, 2007), and only a brief outline of their development will be given here.

Empirical studies (cf. Pointon, 1980; Lea, 1974; Dauer, 1983; Roach, 1982 found no evidence to support the early thesis (Pike, 1945; Abercrombie, 1967) that cross-linguistic differences result from top-down timing strategies selecting different units (foot versus syllable[2]) for isochrony. Following the empirical discreditation of the isochrony account, an alternative explanation (see Bertinetto, 1981; Dauer, 1983; 1987; Roach, 1982; Dasher & Bolinger, 1982) proposed that rhythmic distinction emerges from distinct sets of phonological and phonetic properties found across languages, most notably the complexity of syllable structure and the presence versus absence of vowel reduction. In brief, it was observed that in so-called 'stress-timed' languages, like English, there is a greater range of syllable structures (permitting complex codas and onsets), heavier syllables are more likely to attract stress, and unstressed vowels tend to be reduced (both durationally and quantitatively, see Delattre, 1966). In contrast, in so-called 'syllable-timed' languages like Spanish, open syllables are far more common, and vowel reduction is much less evident. This 'phonologically-derived' hypothesis of rhythmic difference in effect captures differences in the

[2] Whence the terms 'stress-timed' and 'syllable-timed'.

way languages execute prosodic prominence. On this interpretation, the execution of prosodic prominence is not dependent (or not merely dependent) on transparent timing strategies, hence the failure to find the basis of rhythm in isochrony. Instead it interacts in a complex manner with the segmental string. To the extent that the proposed rhythm metrics 'work', it is because they have access to the acoustic effects of this interaction, without having to 'know' it.

These claims find strong support in psycholinguistic studies (Nazzi, Bertoncini & Mehler, 1998; see also Ramus et al, 2003; Nazzi & Ramus, 2003; Nazzi et al, 2000) showing that infants attend to rhythmic differences from birth, indicating that they reflect something that can be perceived before linguistic analysis is available to the infant, and therefore can be captured objectively from the acoustic stream. From this, Ramus, et al (1999) argue that "[the] infant primarily perceives speech as a succession of vowels of variable durations and intensities, alternating with periods of unanalysed noise (i.e. consonants)." This insight led Ramus et al (1999) to develop three measures of utterance rhythm which can be extracted purely on acoustic grounds, on the basis of vocalic and consonantal intervals: i) the standard deviation of vocalic intervals (ΔV); ii) the standard deviation of intervocalic (i.e. consonantal) intervals (ΔC); and the proportion of utterance duration which comprises vocalic intervals (%V). Application of these metrics (Ramus et al, 1999; Ramus et al, 2003) to languages of perceived different rhythmic categories revealed a combination of ΔV and either ΔC or %V to be the most useful in distinguishing categories.

A different approach to measuring the basis of rhythmic difference, but one nevertheless derived from acoustic intervals, uses the pairwise variability index (PVI), proposed by Low et al (2000). Rather than taking global 'vocalic-ness' and variability in that 'vocalic-ness', the PVI attempts to capture *sequential* differences in vocalic interval duration, which arise for example between stressed and unstressed syllables. The motivation for looking at the sequential nature of the contrast is that prosodic prominence depends on syntagmatic contrast: what counts is a comparison with what has gone immediately before and with what lies immediately ahead. The PVI is calculated as the mean of the differences between successive intervals, and is normalised (nPVI) for variability of speech rate by dividing by the sum of intervals. Grabe and Low (2002) also propose an intervocalic PVI (PVI-C), but warn against normalising this. The rationale for this is that since the size and variability of intervocalic intervals largely reflect a language's phonotactics and since these are claimed to be an underlying source of that language's rhythmic properties, normalising eliminates rhythmic difference. Thus, normalising

speech variability over an utterance is deemed to be helpful for capturing vocalic interval variability, but unhelpful for capturing consonantal interval variability.

The decision as to whether to normalise for speech rate is non-trivial. Evidence that ΔC and ΔV are inversely proportional to speech rate (e.g. Barry, Andreeva, Russo, Dimitrova & Kostadinova, 2003; Dellwo & Wagner, 2003; Lee & Todd, 2004) led Dellwo (2004) to propose the normalised version of ΔC, VarcoC, and Ferragne and Pellegrino (2004) and White and Mattys (2007) to propose the normalised version of Δ ΔC, VarcoV. Indeed, these normalised metrics performed better at distinguishing rhythmic categories than their non-normalised counterparts. Conversely, Dellwo and Wagner (2003) found %V not to vary consistently with speech rate. Thus, there is general consensus that metrics for variability in vocalic intervals should be normalised, but greater uncertainty about the equivalent consonant metrics. However, a potential confound, as noted by White and Mattys (2007), is that variation in speech rate may itself contribute to the rhythmic percept. This may be particularly important when comparing different speech styles within a given language, where the underlying structural properties (presumably) remain constant[3], but parameters of phonetic implementation, such as speech rate, may be systematically different. In this case, one could argue that the difference is one of rate rather than rhythm. However, if a different rate systematically alters the rhythmic percept, and is systematically used for a given speech style, the boundary is a fuzzy one. Ideally, language-dependent rate differences would be distinguished from individual, style-dependent rate differences. At present, we know of no satisfactory method of distinguishing the two and so are obliged to conflate them in the present study.

With the proliferation of different metrics, the question arises as to how well they perform, comparatively. White and Mattys (2007) carried out a direct comparison of the usefulness of rhythm metrics in capturing perceived differences between 'stress-timed' English and Dutch on the one-hand and 'syllable-timed' French and Spanish on the other (for adult speech). They found that %V[4] and the rate-normalised vowel metrics (VarcoV and nPVI-V) to be the most effective (with a slight favouring of VarcoV over nPVI-V, also for within-category discrimination).

Measures of consonantal interval variation proved to be far less effective or consistent in discriminating between rhythmic types. However, as White and Mattys (2007: 18) point out, this may be due to the idiosyncrasies of particular materials, and they cite previous studies

[3] There is no reason to predict that syllables or segments are deleted in CDS.
[4] Or %C, since these amount to the same thing.

(e.g. Ramus et al, 1999) that did achieve greater discriminatory success with consonant interval metrics. Furthermore, as has been noted, the appropriateness of rate normalisation for consonant interval metrics is also questionable. The role of consonant variability in the creation of the percept of rhythmic difference remains, therefore, rather elusive. Further investigation will require careful consideration of the role of speech rate and of the segmental composition of the materials used.

1.2 Characteristics of Child Directed Speech

From a purely descriptive point of view, it is well documented that certain properties of Child-Directed-Speech (CDS) may differ markedly from those observed in Adult Directed Speech (ADS), notably in prosodic patterns and speaking rate (see e.g. Albin & Echols, 1996; Fernald 2000; Fernald & Mazzie, 1991; Fernald & Simon, 1984; Garnica, 1977; Grieser & Kuhl, 1988), but also in segmental properties (see e.g. Watt, Docherty & Foulkes, 2003). Common features of CDS that have been identified by these and other studies include higher and greater range of pitch, especially in stressed syllables; longer duration of individual words; prominent final lengthening; slower speech rate; higher amplitude; longer pauses, and more reliable positioning of pauses at phrase boundaries.

Some of these modifications, in particular final lengthening, could reasonably be expected to influence rhythm, and therefore to impact on the rhythm metrics. CDS may, however, present other, more generalised, modifications to vocalic and consonantal interval timing, which, one would expect, would also be reflected in the metrics. Any modifications may (and equally may not) be of a systematic nature. If they turned out to be systematic, it could reflect, for example, an accommodation towards temporal properties exhibited in the speech of the child interlocutor, or a particular strategy aimed at bonding or some kind of perceptual facilitation or instruction. One possibility is that adults exaggerate temporal properties characteristic of ADS. Indeed, Dominey and Dodane (2004: 128) claim that 'the essential acoustic property of CDS is the exaggeration or modulation of characteristics that are *already present* in ADS'[5], though they refer to general prosodic characteristics and not to rhythm specifically.

If this exaggeration were true for rhythm, we would predict Spanish CDS to be more even-timed and English CDS to be less even-timed than their ADS equivalents, particularly for speech towards younger infants, and, as a result, for the cross-linguistic differences to be greater for this

[5] Echoed elsewhere in the CDS literature, e.g. Albin and Echols (1996).

style of speech. However, cross-linguistic differences in ADS are claimed to emerge from structural differences (at least in part), and such differences are discrete and categorical (syllables are either predominantly open or they are not). It would be difficult to conceive of a way in which such structural properties could be exaggerated. More problematically still, there is simply no reason to expect any structural differences (a difference in syllable structure, for example) between CDS and ADS: unless an adult were mimicking child phonology, why would she delete or simplify consonant clusters? Any observed rhythmic differences, therefore, can be presumed to stem from variation in phonetic implementation[6].

From a functional point of view, it has been argued that the richness of prosodic cues in CDS may actively function to attract attention, highlight linguistic structure and aid word identification. In support of this thesis, Kemler Nelson et al (1989) showed that 7- and 10-month old infants were sensitive to prosodic cues to clause boundaries in CDS but not Adult-Directed Speech (ADS). Evidence also suggests (Fernald, 2000; Shockey & Bond, 1980) that properties of CDS evolve as a function of the child's development. Determining whether any CDS modifications are used or indeed are even of any potential linguistic use to children is no trivial task, and it is highly debatable whether this factor could be successfully isolated and therefore whether a truly useful test be conducted. Nevertheless, the application of rhythm metrics to CDS allows us to test for variation between adult speaking styles/contexts.

2.0 Objectives of present study

The present study aims to determine the extent to which cross-linguistic differences in certain temporal properties (claimed to give rise to specific rhythmic percepts) are detectable also in CDS. Our prediction was that such differences would be present to some degree, given the unlikelihood of structural divergence. It also aims to determine whether any cross-linguistic *similarities* are detectable in CDS, when compared with ADS. Here, our prediction was that differences between CDS and ADS would be observable, but that these would not necessarily be systematic, or similar across languages, given the performative nature of their source.

The languages chosen for this study differ in certain phonological and prosodic properties, and are said to differ according to rhythmic class (for ADS).

[6] If we keep to a purely segmental/durational account of rhythmic source. The possibility remains that intonational structure differs between ADS and CDS.

i) **English** displays a wide variety of syllable structure types, quantitative vowel reduction (Delattre, 1966), stress attraction to heavier syllables, and substantial final lengthening (Wightman et al, 1992). Rhythmically, it is classically defined as 'stress-timed'.

ii) **Spanish** is dominated by CV syllable structure, and displays weak vowel reduction (Delattre, 1966), a weak correlation between stress and syllable weight, and weaker final lengthening. Rhythmically, it is classically defined as 'syllable-timed.'

iii) **Catalan** displays mostly CV structure (though has a greater incidence of more complex syllables than Spanish), weak vowel reduction, a moderate correlation between stress and syllable weight, and more final lengthening than Spanish (Ortega-Llebaria & Prieto, 2007). The fact that some of these properties are more typical of 'syllable-timed' languages and some more typical of 'stress-timed' languages has led some researchers to conclude that Catalan is rhythmically 'intermediate' (Nespor, 1990; Ramus et al, 1999). However, there is no firm consensus on its rhythmic status. Prieto, Vanrell, Astruc, Payne & Post (2008) show that when syllable structure properties are controlled for, no durational planning differences arise between Catalan and Spanish. Further evidence for the 'syllable-timed' status of Catalan is given by Gavaldà-Ferré (2007), who shows that different degrees of vowel reduction found in different dialects of Catalan make no difference to rhythm.

2.1 Summary of hypotheses

We hypothesised that:

H1 CDS scores would show some degree of cross-linguistic divergence, as observed in ADS;

H2 CDS may still show rhythmic divergence from ADS, but in an unsystematic way cross-linguistically.

3.0 Method

3.1 Participants

For each language, we recorded twelve mother-child dialogues (to elicit CDS), and made separate recordings of the mothers interacting with an adult interviewer (to elicit ADS). The children fell into three age groups (i.e. there were four children in each age group for each language): 2-, 4- and 6-year-olds. Some mothers were recorded in two different dialogues, with children of different ages. In total, 36 children and 26 adults were recorded[7].

3.2 Materials and elicitation

The data consisted of short question-and-answer dialogues, elicited through the medium of a structured game, based on short, animated clips, shown on Powerpoint slides on a laptop screen. The animations showed simple, everyday scenes, which could easily be described in words that were highly familiar to the children. For example, one scene showed a little girl blowing soap bubbles, another showed a little boy playing with building blocks.

The mother was instructed to ask her child to describe what was happening in each clip, then praise the child for getting it right, and repeat what the child had said. A typical dialogue went thus:

Mother: "*What's happening here? What's the little girl doing?"*
Child: *"(She's) blowing bubbles!"*[8]
Mother: "*That's right! She's blowing bubbles!*"

The mothers were also recorded doing the same task, in the same role, interacting with an adult (the interviewer).

Table 1

	CDS 2-year olds	CDS 4-year olds	CDS 6-year olds	CDS total	ADS
English	69	77	68	214	226
Catalan	69	64	68	201	154
Spanish	65	33	-	98	138

[7] This study was part of a wider study which also investigated rhythm in child speech.
[8] Or an approximation.

Table 1 shows the number of utterances produced by mothers in dialogue with each age group, and in dialogue with other adults, for each language. Spanish CDS utterances directed at 6-year olds are missing: this is because, in this particular recording session, the mothers were mistakenly not instructed to repeat what their child had said.

Recordings were made respectively in Cambridge, Madrid and Barcelona, using a Marantz PMD660 recorder and Shure PG81 microphones, in the participants' homes.

3.3 Analysis

Vocalic intervals were segmented and labelled (start-points and end-points) from the waveform and spectrogram in Praat using standard criteria (e.g. Peterson & Lehiste, 1960). Vocalic and intervocalic (consonantal) intervals were then extracted using a Praat script, and the following metrics were calculated (after Grabe & Low, 2002 and White & Mattys, 2007): %V; ΔV; VarcoV; ΔC; VarcoC; rPVI-V; nPVI-V; nPVI-C and nPVI-C.

Using the SPSS software package, a series of univariate ANOVAS was run separately on the CDS and ADS scores, and a repeated measures analysis was run to compare CDS and ADS scores, with speaking style as the repeated measure. Bonferroni post hoc comparisons were performed on language and age. The p-level is only reported for significant (i.e. $p < .05$) or near-significant differences (i.e. $05 < p \leq .10$).

4.0 Results and Discussion

As a first step, we checked the robustness of the cross-linguistic differences for ADS, to verify whether the elicitation technique used could replicate previous findings for these languages for adult speech. For ADS, 'language' was found to be a main effect for all metrics (see Table 4, Appendix, for the results of statistical tests). However, the post hoc comparisons reveal that the only robust difference is between English on the one hand and Spanish and Catalan on the other, while the latter two are more similar. English is distinct from Catalan along all measures, and from Spanish along all measures except nPVI-C. As we would expect for a 'stress-timed' language, English shows greater variability in both consonant and vocalic intervals (see Figures 1a and 1b, and Tables 5 and 6), and a lower %V (see Figure 1b). The choice of parameters represented graphically in Figures 1a and 1b follows White and Mattys (2007), for their adult rhythm data, in plotting %V against VarcoV and rPVI-C against nPVI-V, and reproduce a very similar pattern to the results for

Spanish and English in that study[9] (compare with their Figures 1 and 2, p. 511-12). White and Mattys chose to represent these parameters in particular because they proved to be the most successful in capturing perceived differences. Essentially, nPVI-V plotted against rPVI-C shows sequential variability in vocalic and consonantal interval durations, while VarcoV plotted against %V shows global variability in vocalic interval intervals and amount of 'vocalicness' overall.

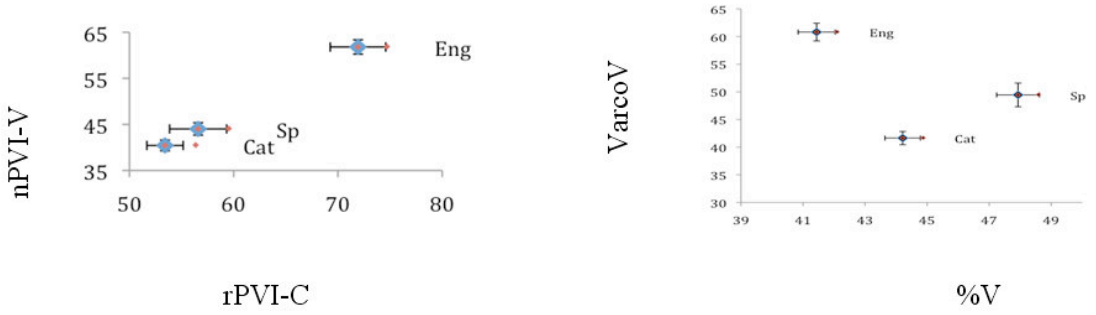


Figure 1a (left): Distribution of Catalan, Spanish and English ADS over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean. Figure 1b (right): Distribution of Catalan, Spanish and English ADS over the %V, VarcoV plane. Bars represent one standard error around the mean.

Table 2: Means (standard errors) of rhythm metrics for Spanish, Catalan and English ADS.

	Spanish	Catalan	English
Interval measures			
ΔV	48 (3.5)	34 (1.2)	58 (2.1)
ΔC	54 (1.6)	50 (1.7)	73 (2.9)
%V	48 (0.7)	44 (0.6)	41 (0.6)
VarcoV	50 (2.1)	42 (1.2)	61 (1.6)
VarcoC	57 (1.9)	49 (1.2)	63 (1.5)
Pairwise Variability Indices			
nPVI-V	44 (1.4)	40 (1.2)	62 (1.6)
rPVI-V	48 (3.5)	34 (1.3)	60 (2.1)
nPVI-C	61 (1.6)	57 (1.5)	63 (1.5)
rPVI-C	57 (2.7)	53 (1.7)	72 (2.7)

[9] Catalan was not investigated in that study.

Some distinction between Catalan and Spanish is detectable from the metrics, but this is significant in the main only for vocalic scores (excepting nPVI-V, see Table 4). For consonant scores, the two languages are only significantly different for VarcoC. %V is significantly lower in Catalan, as would be expected if Catalan does indeed fall in an intermediate rhythmic category. However, for *variability* scores, even where distinctions are detected, Catalan scores are actually *lower* than Spanish, and this is contrary to claims of Catalan being less syllable-timed than Spanish (Nespor, 1990).

4.1 Hypothesis I: Cross-linguistic differences in CDS

Our first hypothesis was that CDS would show the cross-linguistic divergence in rhythmic scores observed in ADS. The results support this hypothesis: as with ADS, language was a main effect for all metrics (see Table 7, Appendix). For all metrics, as with ADS, English shows greater variability in both consonant and vocalic intervals (see Figures 2a and 2b), and a lower %V (Figure 2b)[10]. Post hoc comparisons showed that, as for ADS, Catalan is distinct from English in all measures. However, the distinction between the other two languages is made along fewer parameters. English is distinct from Spanish only for rPVI-C, and nPVI-V, while Catalan is distinct from Spanish only in the normalised vocalic metrics (VarcoV and nPVI-V). In other words, the associated rhythmic categories for ADS in these languages are maintained to some degree in CDS, although more weakly for the distinction between English and Spanish. Child age is shown to be a main effect for all consonant metrics, but for none of the vocalic metrics, or for %V (see Table 7).

[10] As with the results for ADS, the choice of parameters represented graphically in Figures 1a and 1b follows White and Mattys (2007), for their adult rhythm data, in plotting %V against VarcoV and rPVI-C against nPVI-V.

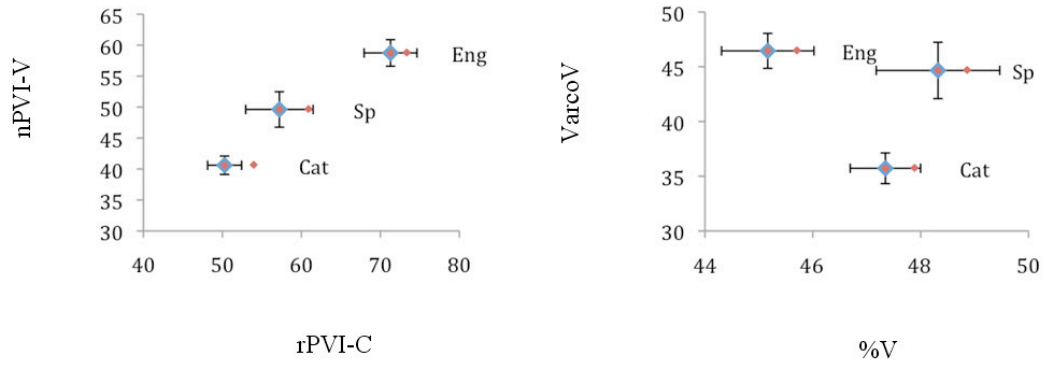


Figure 2a (left): Distribution of Catalan, Spanish and English CDS over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean. Figure 2b (right): Distribution of Catalan, Spanish and English CDS over the %V, VarcoV plane. Bars represent one standard error around the mean.

Table 3: Means (standard errors) of rhythm metrics for Spanish, Catalan and English CDS.

	Spanish	Catalan	English
Interval measures			
ΔV	44 (3.9)	36 (2.4)	46 (2.1)
ΔC	48 (3.9)	43 (1.5)	55 (2.3)
%V	48 (1.2)	47 (0.7)	45 (0.9)
VarcoV	45 (2.6)	36 (1.4)	47 (1.6)
VarcoC	50 (2.6)	45 (1.4)	54 (1.6)
Pairwise Variability Indices			
nPVI-V	50 (2.9)	41 (1.5)	59 (2.1)
rPVI-V	51 (4.7)	42 (2.4)	58 (2.1)
nPVI-C	62 (3.2)	53 (2.0)	68 (2.1)
rPVI-C	57 (4.3)	50 (2.2)	71 (3.4)

4.2 Hypothesis 2: rhythmic properties of CDS

Our second hypothesis, relating to rhythmic type, was that CDS scores would differ, cross-linguistically, from ADS scores, and specifically that they would exaggerate the rhythmic properties of ADS in the language concerned.

4.2.1 %V

%V in CDS proved to be (significantly) higher than in ADS.

4.2.2 Vocalic interval variability

Comparing ADS and CDS, for vocalic interval variability, we find that adult speech style had a significant main effect when variability was measured globally (ΔV and VarcoV), with variability significantly *lower* in CDS than in ADS in all languages (see Figures 3-5b for VarcoV, and Table 8 for statistical results). For sequential measures (PVI), however, there was no significant difference between CDS and ADS (see Figures 3a, 4a and 5a for nPVI-V).

4.2.3 Consonant interval variability

Comparing adult speech styles, consonant interval variability was significantly *lower* in CDS than in ADS (see Table 8, and Figures 3b-5b), when measured globally (i.e. ΔC and VarcoC). For sequential measures (PVI), however, there was no significant difference between CDS and ADS (see Figures 3-5a, for rPVI-C). These mirror the results for vocalic interval variability.

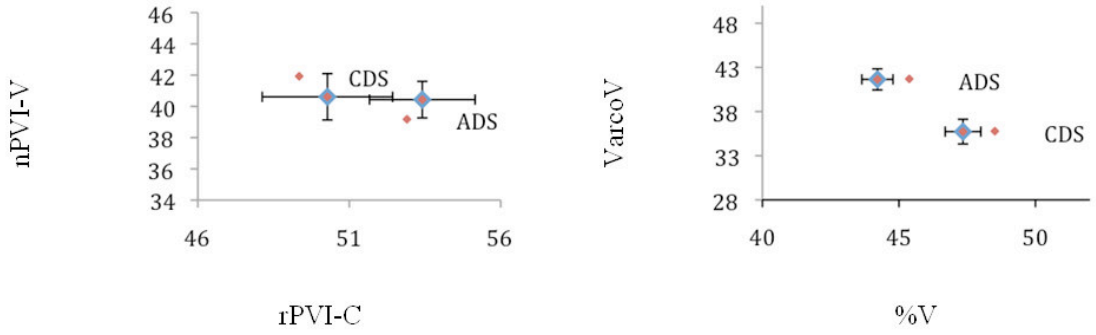


Figure 3a (left): Distribution of CDS and ADS in Catalan over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean. Figure 3b (right): Distribution of CDS and ADS in Catalan over the %V, VarcoV plane. Bars represent one standard error around the mean.

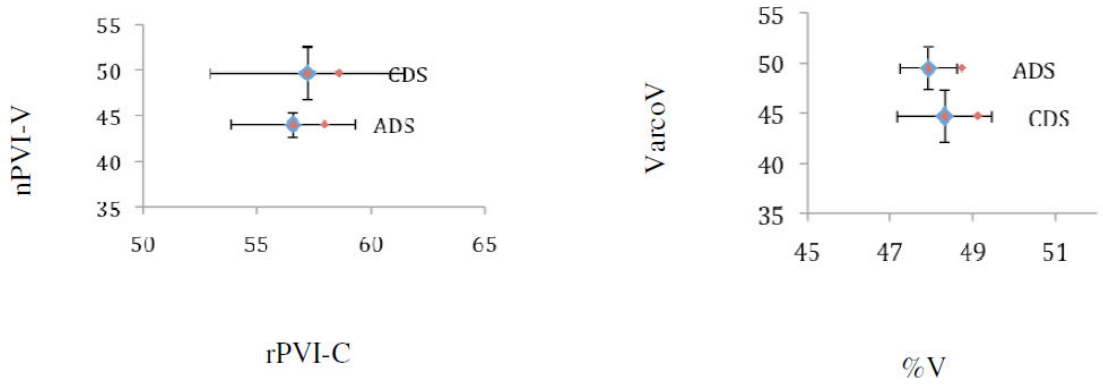


Figure 4a (left): Distribution of CDS and ADS in Spanish over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean. Figure 4b (right): Distribution of CDS and ADS in Spanish over the %V, VarcoV plane. Bars represent one standard error around the mean.

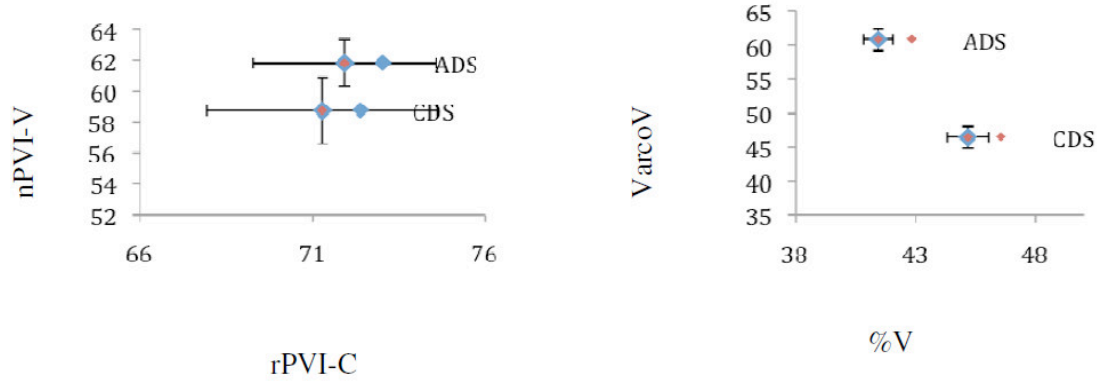


Figure 5a (left): Distribution of CDS and ADS in English over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean. Figure 5b (right): Distribution of CDS and ADS in English over the %V, VarcoV plane. Bars represent one standard error around the mean.

4.2.4 Summary of CDS rhythmic characteristics

For adult speech style, the metrics reveal that, cross-linguistically, mothers showed less *global* variability in both consonant and vowel interval duration, and had more vocalic speech when speaking to their children than when speaking to the adult interviewer. Since these modifications occurred regardless of the language spoken, they also can be presumed to have occurred regardless of claimed rhythmic class. Higher %V and lower variability are supposedly both characteristic of 'syllable-timing' (more even timing), and therefore it would appear to be the case that generalised rhythmic modifications take place in CDS.

Interestingly, when sequencing in taken into account in the variability measures (PVI), the variability differences between CDS and ADS disappear. This implies that overall, CDS is more variable than ADS, but when segmental intervals are compared pairwise the proportion between longer and shorter does not change.

5.0 Conclusion

5.1 Summary of main findings

A clear finding of this study is that rhythmic class distinctions detectable in ADS are also clearly observable in CDS, though in some cases not to the same degree. The differences between English and Catalan are particularly robust: CDS in English has a higher variability and lower %V (i.e. more 'stress-timed' rhythm), than in other languages. There is a weaker distinction between English and Spanish, and a very weak distinction between Catalan and Spanish. Contrary to some claims (e.g. Nespor, 1990), there is little evidence from these metrics, for either ADS or CDS, to support the view that Catalan should be classed as intermediate between 'stress-timed' and 'syllable-timed': although Catalan has a higher %V than Spanish, it has *lower* variability in both consonant and vowel interval duration, and, if anything, is 'further away' from English. Our findings support those of Prieto et al (2008) which show that when syllable structure properties are controlled for, no durational planning differences arise between Catalan and Spanish, and of Gavaldà-Ferré (2007), which shows that different degrees of vowel reduction found in different dialects of Catalan make no difference to rhythm.

Cross-linguistically, mothers increase the vocalic proportion and decrease the interval variability of their speech when they address infants. This modification results in a slightly weaker distinction between languages for this style of speech (as reported above). For some parameters at least, how mothers modify their speech depends in part on the age of the child they are addressing: this is particularly evident for consonant interval variability, which is already lower than in ADS, but also decreases as the age of the child increases. This is interesting because it has been claimed by some (cf Grabe & Low, 2002) that consonant interval variability is an artefact of a language's phonotactic properties. The fact that adults vary this for different speaking styles in the same language would suggest that this parameter of timing *can* be controlled for speech style. Furthermore, if adults are making fine distinctions in their speech directed at different ages of children, it would

appear that control of this parameter is actually quite subtle. Higher %V and lower variability are both purportedly characteristics of so-called 'syllable-timing' rhythm, suggesting that adults in all three languages modify their speech towards this end of a hypothesised rhythm continuum (though at different points along this continuum, since the language distinction is maintained). By doing so, the mothers investigated in the study were possibly accommodating towards aspects of the speech of their child (by increasing %V and lowering vocalic variability, see Payne et al, *submitted*), at the same time as exaggerating the lower consonant variability of normal adult speech.

5.2 Next steps

The interval-based metrics applied in this study yield a clear picture of cross-linguistic rhythmic distinctions (previously established for ADS) co-existing with cross-linguistic similarities in CDS, which is more categorically even-timed than speech directed to other adults. It should be borne in mind that the differences investigated here are in acoustic indices which, it is claimed, *form the basis* of a cross-linguistic perceptual distinction in ADS. Since it is not known the *degree* to which acoustic parameters must vary in order to trigger the percept of rhythmic difference, we cannot be sure that the observed variation between CDS and ADS for a given language, and cross-linguistically within CDS itself, translates into a perceived difference. Thus, the perceptual validity of these differences in CDS needs to be tested. Furthermore, other potential dimensions to the perception of rhythm, such as intensity, vowel quality and F0, lie beyond the scope of this paper, but remain to be explored more fully[11]. Differences in interval-based indices might not be perceived as rhythmic in CDS – they could, for example, prove perceptually weak or incoherent.

The fact that, from the production perspective, temporal characteristics deemed at least to *contribute* to the percept of rhythm, vary systematically between two styles of adult speech within a given language, shows that the percept of rhythm does not emerge from structural properties alone. It would seem, therefore, that a full model of speech rhythm should be able to account for and incorporate a performance dimension (i.e. phonetic implementation). It is reasonable to conjecture that such a dimension would accommodate also an individual, speaker-specific parameter, in addition to the style-specific (but speaker-

[11] In a rare study of this kind, Lee and Todd (2004) report evidence that 'stress-timed' English and Dutch show greater variability in vocalic intensity than 'syllable-timed' Italian and French, for ADS.

shared) parameters as investigated in the present study. Additionally, it would be interesting to examine the degree to which speaker-shared performative characteristics are, at the same time, language-specific; in other words, are there linguistic-phonetic factors which contribute to the percept of speech rhythm? That the speech context investigated here is shown to exhibit a degree of similarity cross-linguistically is of particular interest, since it implies that cross-linguistic implementation patterns can and do exist, and would suggest the possibility that manipulation of rhythm has a more general functional purpose (or at least aim).

Acknowledgements

This research has been funded by a Batista i Roca research project entitled "The acquisition of rhythm in Catalan, Spanish and English (Ref.: 2007 PBR 29), awarded by the Generalitat de Catalunya, with Pilar Prieto as PI. We are grateful to Núria Argemí, Anna Barberà, Ana Estrella, and Melanie Bell for recording the data in different languages, and especially to Naomi Hilton for performing the segmentation of the data. Many thanks to Greg Kochanski for helpful comments on an earlier draft of this paper, and to Bill Wells and other colleagues who provided helpful discussion of related presentations at BAAP 2008 and EASR 2008.

References

Abercrombie, D. (1967). Elements of general phonetics. Edinburgh: Edinburgh University Press.

Albin, D. D., and Echols, C. H. (1996). Characteristics of stressed and word-final syllables in infant-directed speech: Implications for word-level segmentation. Infant Behavior and Development, 19, 401-418.

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., and Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? In Proceedings of the 15th international congress of phonetics sciences (pp. 2693–2696). Barcelona.

Bertinetto, P.M. (1981). Strutture Prosodiche dell'Italiano. Firenze: Accademia della Crusca.

Dasher, R., and Bolinger, D. (1982). On pre-accentual lengthening. Journal of the International Phonetic Association, 12, 58–69.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. Journal of Phonetics, 11, 51-62.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. In Proceedings of 11th ICPhS, Tallinn, 447-450.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. International Review of Applied Linguistics in Language Teaching, 4, 183-198.

Dellwo, V. (2004). Rhythm and speech rate: A variation coefficient for deltaC. Proceedings of the 38th Linguistic Colloquium, Budapest.

Dellwo, V., and Wagner, P. (2003). Relations between language rhythm and speech rate. In Proceedings of the 15th international congress of phonetics sciences (pp. 471–474). Barcelona.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. Journal of Phonetics, 29, 217–230.

Dominey, P. F. and Dodane, C. (2004). Indeterminacy in language acquisition: the role of child directed speech and joint attention. Journal of Neurolinguistics, 17, 121-145.

Fernald, A. (2000). Speech to Infants as Hyperspeech: Knowledge-Driven Processes in Early Word Recognition. Phonetica 57: 242-254.

Fernald, A. and Mazzie, C. (1991). Prosody and focus in speech to infants and adults. Developmental Psychology, 27(2), 209-221.

Fernald, A. and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. Developmental Psychology, 20(1), 104-113.

Ferragne, E. and Pellegrino, F. (2004). A comparative account of the suprasegmental and rhythmic features of British English dialects. In Proceedings of ‘‘Modelisations pour l’Identification des Langues’’ Paris.

Garnica, O., (1977). Some prosodic and paralinguistic features of speech to young children. In Snow, C.E. and Ferguson, C.A., (Eds.), Talking to children: Language input and acquisition, (pp. 63-88), Cambridge, MA: Cambridge University Press.

Gavaldà-Ferré, N. (2007). Vowel reduction and Catalan speech rhythm. Unpublished MA Thesis, University College London.

Grabe, E. and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis” in N. Warner and C. Gussenhoven (Eds.), Papers in Laboratory Phonology 7, (pp. 515-546), Berlin: Mouton de Gruyter.

Grieser, D. L., and Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. Developmental Psychology, 24(1), 14-20.

Kemler Nelson, D.G., Hirsh-Pasek, K., Jusczyk, P.W., and Cassidy, K.W. (1989). How the prosodic cues in motherese might assist language learning. Journal of Child Language, 16, 55-68.

Lea, W. A. (1974). Univac Report PX10791. Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding. St. Paul, Minnesota: Sperry Univac.

Lee, C. S. and Todd, N. P. M. (2004). Towards an auditory account of speech rhythm: application of a model of the auditory 'primal sketch' to two multi-language corpora. Cognition, 93, 225-254.

Low, E.L., Grabe, E. and Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'syllable-timing' in Singapore English. Language and Speech, 43, 377-401.

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. Journal of Experimental Psychology: Human Perception and Performance, 24, 756-766.

Nazzi, T., Jusczyk, P.W., and Johnson, E.K. (2000). Language discrimination by English learning 5-month-olds: Effects of rhythm and familiarity. Journal of Memory and Language, 43, 1-19.

Nazzi, T., and Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. Speech Communication, 41, 233-243.

Nespor, M. (1990). On the rhythm parameter in phonology. In I. M. Roca (ed.) Logical issues in language acquisition, (pp. 157-175), Dordrecht: Foris.

Ortega, M. and Prieto, P. (2007). Disentangling stress from accent in Spanish: production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró, and M.-J. Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32, 693–703.

Solé (ed.) Segmental and Prosodic Issues in Romance Phonology, Amsterdam/Philadelphia, John Benjamins.

Pike, K. (1945). The intonation of American English. Ann Arbor: University of Michigan Press.

Pointon, G.E. (1980). Is Spanish really syllable-timed? Journal of Phonetics, 8, 293-304.

Prieto, P., Vanrell, M, Astruc, L., Payne, E. and Post, B. (2008). Is there any evidence for top-down planning of rhythm? Evidence from Catalan, English and Spanish, poster presented at UCL Workshop on Rhythm, March 2008.

Ramus, F., Dupoux, E., and Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, (pp. 337-342).

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. Cognition, 73, 265-292.

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (ed.) Linguistic controversies, London: Edward Arnold.

Shockey, L. and Bond, Z. S. (1980). Phonological processes in speech addressed to children. Phonetica, vol. 37, 267-274.

Watt, D., G. Docherty and P. Foulkes (2003). First accent acquisition: a study of phonetic variation in child-directed speech. In Proceedings of the 15th ICPhS, Barcelona, 1959-1962.

White L. and Mattys, S. (2007) Calibrating rhythm: First language and second language studies. Journal of Phonetics, 35, 501-522.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. Journal of the Acoustical Society of America, 91, 1707-1717.

Appendix

Table 4: Inter-language differences in scores for ADS.

metric	Language as main effect	En v Ca *p*=	En v Sp *p*=	Ca v Sp *p*=
%V	F(2) = 26.694 p = 0.000	0.004	0.000	0.000
ΔC -	F(2) = 24.850 p = 0.000	0.000	0.000	*1.000*
VarcoC	F(2) = 20.318 p = 0.000	0.000	0.037	0.002
rPVIc	F(2) = 16.913 p = 0.000	0.000	0.000	*1.000*
nPVIc	F(2)= 3.812 p = 0.023	0.018	1.000	*0.308*
ΔV	F(2) = 29.240 p = 0.000	0.000	0.008	0.000
VarcoV	F(2) = 36.079 p = 0.000	0.000	0.000	0.008
rPVIv	F(2) = 32.168 p = 0.000	0.000	0.002	0.000
nPVIv	F(2) = 68.308 p = 0.000	0.000	0.000	*0.336*

Table 5: Mean consonant interval metric scores for ADS and CDS scores (ADS scores cannot be categorised by age).

metric	Age	En ADS	En CDS	Cat ADS	Cat CDS	Sp ADS	Sp CDS
ΔC	2		62.25		45.07		46.32
	4	73.42	57.17	49.59	44.26	53.6	49.87
	6		44.37		40.14		
VarcoC	2		54.5		50.19		51.16
	4	62.5	56.86	49.11	46.33	57.04	47.71
	6		50.85		38.67		
rPVI-C	2		81.86		56.43		56.93
	4	71.93	74.43	53.41	48.03	56.59	57.78
	6		57.0		46.16		
nPVI-C	2		69.0		62.48		64.06
	4	62.88	70.83	57.1	51.45	60.95	58.27
	6		64.28		46.06		

Table 6: Mean vocalic interval metric scores for ADS and CDS (ADS scores are not categorised by age).

metric	Age	En ADS	En CDS	Cat ADS	Cat CDS	Sp ADS	Sp CDS
%V	2		42.95		48.39		48.7
	4	41.44	46.53	44.21	46.72	47.93	47.56
	6		45.61		47.02		
ΔV	2		44.5		44.91		43.59
	4	58.03	48.84	33.84	31.32	48.18	46.11
	6		43.61		32.19		
VarcoV	2		43.59		39.25		43.5
	4	60.82	47.86	41.64	36.82	49.45	47.00
	6		47.42		31.58		
rPVI-V	2		54.59		50.92		49.11
	4	59.47	62.35	33.49	40.03	47.81	55.85
	6		55.35		35.99		
nPVI-V	2		54.09		42.61		47.25
	4	61.81	61.35	40.43	44.32	44.02	54.36
	6		59.91		35.31		

Table 7: Results of ANOVA for CDS scores.

Metric	language	Post hoc lang	age	Post hoc age
%V	F(2)=3.413; p=0.034	-	F(2)=0.032; p=0.968	-
ΔC	F(2)=7.867; p=0.000	C v E: 0.000	F(2)=4.816; p=0.008	2 v 6: 0.024 4 v 6: 0.038
VarcoC	F(2)=9.012; p=0.000	C v E: 0.000	F(2)=5.043; 0.007	2 v 6: 0.011 4 v 6: 0.033
rPVIc	F(2)=14.898; p=0.000	C v E: 0.000 S v E: 0.015	F(2)=5.704; 0.004	2 v 6: 0.008
nPVIc	F(2)=12.487; p=0.000	C v E: 0.000	F(2)=4.597; 0.011	2 v 6: 0.009
ΔV	F(2)=4.171; p=0.016	C v E: 0.008	F(2)=0.964; p=0.382	-
VarcoV	F(2)=11.622; p=0.000	C v E: 0.000 C v S: 0.004	F(2)=0.813; p=0.444	-
rPVIv	F(2)=7.382; p=0.001	C v E: 0.000	F(2)=1.206; p=0.300	-
nPVIv	F(2)=22.025; p=0.000	C v E: 0.000; C v S: 0.020; E v S: 0.017	F(2)=2.340; p=0.097	-

Table 8: Results of Anovas for CDS versus ADS.

metric	ADS versus CDS
%V	F(1)=6.375; p=0.012
ΔC	F(1)=14.043; p=0.000
VarcoC	F(1)=13.265; p=0.000
rPVIc	F(1)=0.017; p=0.895
nPVIc	F(1)=0.157; p=0.692
ΔV	F(1)=4.626; p=0.032
VarcoV	F(1)=25.670; p=0.000
rPVIv	F(1)=1.251; p=0.264
nPVIv	F(1)=0.436; p=0.509

(t,d): the Variable Status of a Variable Rule

Rosalind A. M. Temple

New College, University of Oxford

Abstract

Tagliamonte and Temple (2005) describes an attempt to replicate on a variety of British English some of the many and varied North American studies of the sociolinguistic variable known *inter alia* as "-t,d deletion" or "coronal stop deletion", that is the variable deletion of word-final /t/ or /d/ in two-consonant clusters. The results of that study were not entirely compatible with previous accounts and could not be explained away by reference to extra-linguistic variables. The present paper represents the first stage of investigating alternative explanations of the apparent incompatibility of the British and North American findings by exploring some of the methodological and analytical questions raised during the production of Tagliamonte and Temple, but which the authors were not able to address or develop explicitly in that paper.

Key Words

Sociophonetics, Coronal Stop Deletion, Variable Rules

1.0 Introduction

(t,d) is a well known variable phonological rule which deletes the second consonant in word-final clusters ending with a coronal stop, as in *fact* > [fakt]/[fak] or *walked* > [wɔːkt]/[wɔːk], and is said to apply to all varieties of English. The great level of interest in (t,d)[1] since it was first explored in, for example, Labov *et al* (1968), Wolfram (1969) and Fasold (1972) stems from the fact that this phonetic/phonological variable occurs in morphologically complex contexts as well as morphologically simple ones and therefore provides a potentially interesting locus for exploration of the interaction between variationist and (morpho-)phonological theory. Tagliamonte and Temple (2005, henceforth T&T) examined the three

[1] It will become clear in the course of this paper why I consider terms such as "*-t,d* deletion" problematic. Although this variable notation also implies acceptance of the fact of consonant deletion it should not be taken as such: it is used purely for the sake of convenience, as is the word "deletion".

independent linguistic variables[2] found to be most robust in conditioning patterns of (t,d) variation in North American studies: the following phonological segment, the preceding phonological segment, and the morphological structure of the word[3]. Their data were taken from sociolinguistic interviews with 38 speakers of British English resident in the city of York recorded for the York English Corpus described in Tagliamonte (1998). After careful transcription by two independent researchers the data were coded and analysed in various configurations using Goldvarb 2.0 (Rand & Sankoff, 1990) to perform multivariate analysis. The overall results are reproduced here as Table 1.

Corrected mean:	.18		
Total N:	1118		
	FACTOR WEIGHT	%	N
FOLLOWING PHONOLOGICAL SEGMENT			
Obstruent	.84	55	325
Glide	.69	38	106
/r/	.60	28	29
/l/	.50	25	24
Vowel	.29	8	507
Pause	.20	6	127
Range	65		
PRECEDING PHONOLOGICAL SEGMENT			
/s/	.68	42	303
Other sibilant	.58	31	64
Nasal	.50	21	329
/l/	.40	21	126
Stop	.40	16	169
Other fricative	.27	12	127
Range	41		
MORPHOLOGICAL CLASS			
Monomorpheme, e.g. *mist*	[.53]	30	602
Irregular past, e.g. *kept*	[.50]	21	128
Regular past, e.g. *missed*	[.45]	19	388

Table 1. Results of Variable rule analysis of the contribution of factors selected as significant to the probability of –t,d deletion. After Tagliamonte and Temple (op. cit., p. 293, Table 4). Factor groups not selected as significant are shown in square brackets.

[2] T&T also tested extra-linguistic variables, but these are not central to the discussion.
[3] Detailed explanation of these variables can be found in T&T. Because that paper is recent and easily available, details which can be found there will be kept to a minimum in the present paper.

In Table 1, factor groups (in linguistic terms, the independent variables) are presented in descending order of their significance in accounting for the patterns of variability in the data. For each factor group, the factors (variants) are listed in descending order of their tendency to favour deletion of final /t,d/. The rightmost column gives the number of tokens with that particular factor, the middle column gives the percentage of tokens with that factor whose /t,d/ is deleted and the leftmost numerical column gives the probability of deletion occurring with that factor as assigned by Goldvarb on a scale of 0 to 1. Thus, the first line of figures shows that there were 325 tokens with following obstruents (e.g. *old carpets*); of these 55% had deleted /t,d/ and when the whole pattern of variation is taken into account, these tokens have a 0.84 chance of being pronounced without a final surface reflex of /t,d/. The range of probabilities, given at the end of each significant factor group, is the difference between the highest and lowest for that group and provides an indication of how important the group is in accounting for the patterns of variation: the greater the range, the more important the relative contribution of that factor group.

The results for phonological context were broadly consistent with other studies and provided further evidence pertinent to ongoing debates in the literature. Following segment has been found to have the strongest effect in most if not all studies of (t,d), as it is here. The hierarchy of factor weights was consistent with previous studies, except for the proximity of /r/ and /l/, which lent further support to Labov's (1997) argument that the patterning of following effects cannot be explained in terms of resyllabification, as proposed in Guy (1991). Preceding phonological segment has been considered a relatively weak constraint (e.g. by Labov, 1989, 1995) but one for which it is possible to draw broadly consistent language-wide generalisations. Thus Labov identifies /s/ > stops > nasals > other fricatives > liquids as a generally consistent cross-dialectal pattern (1989, p. 90). This is not the hierarchy produced in T&T's results, nor do their results sit comfortably with an account in terms of the Obligatory Contour Principle, as proposed in Guy & Boberg (1997). T&T considered that fact in itself not to be unduly problematic, since it is generally acknowledged that the strength of effect and hierarchy of variants have varied from study to study (*cf*., e.g., Guy, forthcoming). However, we shall return to this constraint below.

The results for morphological context in Table 1 are altogether more perplexing. Guy (1991) elaborated an explanation for the frequently observed effect of the morphological makeup of any given word containing a final $CC^{[+cor]}$ cluster within the framework of Lexical

Phonology. The analysis predicts that deletion will occur most frequently in monomorphemic forms such as *round* and least frequently in regular past tense forms ending in *–ed*, such as *trashed*. So-called semi-weak verbal forms, with a past-tense suffix but also a vowel alternation in the stem, for example *kept*, will pattern intermediately between the other two categories[4]. Many subsequent studies have provided support for this analysis, which has become generally accepted as correct (e.g. Santa-Ana, 1992; Bayley, 1995). However, as Table 1 shows, this was not the case for T&T: although the trend was in the expected direction, morphological class was not selected as significant in their analysis. Moreover, T&T found that other predictions of the Lexical Phonology-based account were not borne out in their data. Whereas the hierarchy of factor weights for following phonological segment was consistent across morphological classes, as predicted, the range of those factor weights was not (T&T: 294-5, Tables 5a, 5b), which runs counter to expectations. The morphological effect did not show the expected consistency across speakers even when the category with the smallest number of tokens (semi-weak forms) was disregarded.

T&T concluded that although their study clearly confirmed that the second consonant in word-final $CC^{[+cor]}$ clusters behaves variably, none of the major theoretical explanations of the variability (resyllabification, the OCP, Lexical Phonology) held for their data, despite the fact that they had made every effort to replicate the methodology of previous studies. Their suggestion was that the most fruitful way to move towards a more successful explanation would be to start from a "bottom-up" investigation of the combinatorial phonetic properties of these word-final clusters, given that there is plenty of evidence to show that speakers are capable of manipulating fine phonetic detail (e.g., Docherty, 1992; Docherty et al., 1997; Temple 2000). The purpose of the present paper is to explore further some of the issues which led to that conclusion as a preliminary to a further bottom-up study. These issues initially arose as methodological difficulties encountered by T&T, about which there appeared to be little or no discussion in the available literature, but as we shall see, they have both methodological and theoretical implications. They will be explored under three broad headings, distributional issues, issues concerning the nature of "deletion" and issues of how the variable rule fits into the phonology as a whole. However, as will become obvious, questions within and across these categories interact with each other creating a

[4] Although there are explanations for why they might pattern with one of the other classes (e.g. Guy & Boyd, 1990), they should not show more deletion than monomorphemes or less than regular past tense forms.

complex web which appears to indicate the need for some radical rethinking about variationist approaches to data such as these.

2.0 Distributional Issues

T&T used Goldvarb 2.0 (Rand & Sankoff, *op. cit.*), a multiple regression-based statistical package designed for linguistic analysis, and they followed a strict protocol in selecting tokens for analysis, taking for each speaker the first twenty tokens from each morphological category to maximise even distribution across categories, and only the first three tokens of any given lexical item to control the type-token ratio (following, e.g., Wolfram, 1993, p. 214). However, the morphological categories were still somewhat uneven, with particularly low numbers of tokens in the semi-weak category. Since Goldvarb is designed to cope with such uneven data sets this was not considered too problematic in itself. What does seem potentially problematic, however, is the distribution of preceding phonological context across the morphological categories. Table 2 shows this distribution for preceding (underlying) segments, ordered according to their factor-weight rankings in Table 1, with those most favouring deletion at the top. Sibilants other than /s/ are grouped together because they have the same (restricted) distribution across morpheme categories, whereas this is not the case with stops or weak fricatives, which are shown individually. Combined cells in the Factor Weight column indicate that the relevant tokens were tested as a single factor for Table 1. Cells with bold outlined borders are those representing around 20% or more of the tokens for that particular morphological group. The cells for /s/ and other sibilants are outlined together in the regular past tense column because although the factor weight assigned to the two groups was different when the whole data set was analysed (Table 1 above), when morpheme categories were tested separately (cf. T&T: 294, Table 5a), all the sibilants were assigned the same weight (0.69) for this group, which is the only one to have sibilants other than /s/[5].

[5] This is a consequence of the distribution of /s/ versus /z, ʃ, ʒ/ across the vocabulary of English rather than a function of T&T's particular data set. It means that the factor weights generated in for Table 1 (and in other studies) are in some sense rather misleading.

PRECEDING SEGMENT	N	PROPORTION OF TOTAL DATA SET	OVERALL RATE OF *–T,D* DELETION	FACTOR WEIGHT (TABLE 1)	PROPORTION OF MORPHEME CATEGORIES		
					MONO-MORPHEMES	SEMI-WEAK	REGULAR PAST
/s/	303	27%	42%	.68	41%	11%	10%
other sibilants	64	6%	31%	.58	-	-	17%
/n/	430	29%	21%	.50	46%	5%	12%
/l/	130	11%	21%	.40	7%	33%	11%
/k/	109	10%	17%	.40	3%	-	23%
/p/	53	5%	15%		0.5%	21%	6%
/b/	4	0.4%	0		-	-	1%
/g/	3	0.3%	0		-	-	<1%
/v/	74	7%	7%	.27	-	-	19%
/f/	53	5%	19%		2%	30%	0.8%

Table 2. Distribution of preceding phonological contexts across morpheme categories (percentages higher than 2 have been rounded up to the nearest whole number).

Comparison across categories shows that only the regular past tense forms have a fairly even distribution across favouring and disfavouring preceding phonological contexts, with 27% of tokens in contexts most favouring deletion, 20% in contexts most disfavouring it and the rest distributed across neutral and mildly disfavouring contexts. Almost half the monomorphemes (46%) are preceded by /n/, which has a neutral effect (factor weight 0.5); the vast majority of the remaining 54% of tokens (82%, i.e. 41% of the total) are preceded by /s/, which highly favours deletion, whereas very few tokens occur in moderately disfavouring contexts (10.5%) and only 2% have strongly disfavouring preceding /f/. By contrast, the majority of semi-weak tokens are preceded by moderately or highly disfavouring preceding contexts (51% and 31% respectively). Thus, in preceding contexts having a favouring or disfavouring effect on the variability, arguably 80% of monomorphemic tokens have preceding consonants which favour deletion, whereas 80% of semi-weak tokens have preceding consonants which disfavour it, as do well over 60% of regular past tokens. This would appear to explain why in Table 1 the hierarchy of frequencies of deletion is apparently consistent with the Lexical Phonology account of (t,d) but the factor group is not selected as significant in accounting for the variability, suggesting that the frequency differences between morphological categories are an artefact of the distribution of favouring and disfavouring

phonological contexts across those categories. The restricted set of preceding phonological contexts which can occur in semi-weak forms is acknowledged by some authors but the fact that monomorphemes too have a somewhat skewed set of preceding contexts does not seem to figure in discussions of this variable.

A further run, replicating Table 1 but without testing preceding phonological context, produced the same significant range and hierarchy of effect for following context, but a different result for morphological category: the factor group was selected as significant and the rank ordering of factors was monomorphemes (.58) > semi-weak forms (.42) > regular past-tense forms (.39). This is strongly suggestive of an interaction between the preceding segment and morphological category factor groups[6]. Disregarding the numerically small semi-weak category does not affect the flipping between significance and non-significance: when all three factor groups are included morphological category is not selected as significant (monomorphemes (.57) > regular past-tense forms (.40)) whereas when preceding context is not tested morphological category is selected as significant with exactly the same distribution of factor weights. As a control exercise, the same procedure was followed disregarding the following context. This made no difference to the non-selection of morphological category, with or without the semi-weak forms, indicating that any interactions there may be between morphological context and following context are well within the capacity of logistic regression to correct (*cf.*, e.g., Sigley, 2003, p. 229).

This brief sketch of the distributional problem raised by T&T's findings does not prove anything but it does demonstrate that morphological category, upon which the Lexical Phonology account of (t,d) crucially depends, is inherently subject to interaction effects with preceding phonological context, effects which seem to have received little attention in the literature on the variable. Rather than exploring these interactions in greater depth, we now turn to another methodological problem area at the opposite end of the spectrum, that of the classification of the data which are input to the variable rule analysis.

3.0 Problems with the interpretation of natural(istic) data

The statistical modelling of variation in speech crucially depends on accurate categorisation of the raw data. On the face of it, (t,d) is a relatively straightforward variable to model, involving as it does a

[6] In the sense of Sigley's (2003) second type of interaction effect, that is associations between factors in different factor groups which lead to unevenly occupied cross-tabulation cells.

categorical alternation between the absence and a phonetic surface realisation of a word-final coronal stop. It is generally acknowledged that an apical stop following a token constitutes a "neutralizing environment" (Guy, 1980, p. 4) and tokens in such contexts are excluded from analyses on the grounds that it is not possible to tell whether a stop produced in that context is just a reflex of the following stop or a reflex of both that and the word-final stop. However, the phonetic analysis and coding of the data for T&T showed that this kind of difficulty arose in far more cases than merely the tokens which are conventionally excluded on the grounds of neutralisation. This section will firstly review what constitutes neutralisation and then examine some other phenomena which can make it difficult to determine where a deletion may or may not have applied. Since the account critiqued here is the one grounded in Lexical Phonology, the working assumption is that if it is correct, (t,d) must be a phonological rule; thus, any phonetic reflex of underlying /t,d/ must mean that the rule has not applied and any ambiguities in the phonetics must raise a question mark over whether it has applied.

3.1 Neutralisation

As already mentioned, so-called "neutralising" environments are a context where problems in identifying variants have long been acknowledged: "... in word-final consonant clusters it is necessary to exclude clusters which are immediately followed by a homorganic stop (e.g. *test day*) from the tabulation since it is sometimes impossible to determine whether the final consonant of the cluster is present or absent." (Wolfram, 1969, p. 48). The exclusion of "neutralisation" contexts seems to have been normal practice since Wolfram's study, although half the studies referred to in T&T give no information about their treatment of clusters in these contexts. Only one of the remaining studies T&T consulted (Bailey, 1995) also excludes tokens with following interdental fricatives, on the grounds that they are frequently realised as stops by Bailey's Tejano subjects. As it is well known that these consonants are frequently realised as lax stops in British English, they were also excluded by T&T. However, there are other following consonants which could arguably also have this kind of neutralising effect on the variation, but which, to our knowledge, are never mentioned. The most notable is [n], which is also articulated with apical/laminal occlusion at the teeth/alveolar ridge. It might be argued that the presence of nasality would always differentiate the nasal from the preceding stop, and stops, particularly voiceless ones, are often clearly audible even if there is no release before the following nasal. However, nasality as a phonetic property is notoriously non-segmental, that is it is rarely strictly co-

temporal with all the other properties of the segment to which it "belongs". In (1), for example, the [s] is followed by a brief, nasalised puff of aspiration and a partially devoiced nasal consonant (the transcription is somewhat misleading because of the sequential limitations of the font).

(1) *they try their <u>best</u> not* [bɛ^{h}˜˳snɑʔ] *to stay on*[7]

As with [t#d] and other accepted "neutralisation" sequences, release of the word-final plosive would not be expected in normal casual, unscripted speech. The nasality is clearly audible from the end of the [s], but it is very difficult to say whether there is actually a reflex of an underlying /t/ with nasal assimilation or whether the /t/ has been deleted and the nasal, which does not sound unduly long, is merely devoiced. Such decisions cannot be made on an *ad hoc* basis: decisions of principle need to be taken as to what is to be deemed a sufficient cue to the surface presence of /t,d/. Discussions of these principles tend in the literature to be limited to consideration of segmental variants such as flaps or glottal stops, whereas (1) illustrates a context where the question is what subsegmental properties are sufficient to cue a /t/, in this case whether the voicelessness is ascribable to the juxtaposition of /n/ and /s/ alone.

With all following consonants sharing alveolar or dental articulation with /t,d/, one might consider a definition of neutralisation closer to the conventional structuralist one and ask whether in some sequences [t] or [d] on the one hand and zero on the other are both truly possible pronunciations. For example, in /sts/ sequences in certain syntactic / discourse contexts (e.g. "at the <u>last</u> second"), where one might ask whether [sts] is a possible pronunciation in natural, fast speech. These problems are, however, not limited to such "neutralisation" contexts and we now turn to examine some areas which, I would argue, also need principled decisions to be taken about how to interpret the data and which in some cases are impossible to interpret definitively with only auditory and acoustic information.

3.2 Masking Effects

[7] All numbered examples are taken from T&T's data. In each case the word with (t,d) is underlined in the orthographic transcription and the phonetic transcription is of that word and the following word only. It is not practicable to give spectrographic illustrations for all examples, so we rely on detailed transcription and description for most.

The problems T&T encountered with the raw data are grouped here somewhat arbitrarily: other groupings are possible and the problems illustrated for each group overlap sometimes to a considerable degree. They all concern phenomena which are instantly recognisable as normal to phoneticians familiar with continuous speech processes (CSPs) and which have been much studied since the early invention of such articulatory techniques as static palatography, since supplanted by electropalatography and more recent techniques such as Electromagnetic Articulography (EMA). General comments regarding CSPs here should be taken as referring to varieties of British English; No detailed knowledge of the phonetics of other varieties studied with reference to (t,d) is claimed. The term "masking" is used to denote the possibility of an articulatory gesture, possibly an incomplete one, which is physiologically and/or acoustically hidden by the articulation of surrounding consonants.

Where there is a following vowel, the duration of the stop closure, the audible release and the visible formant transitions into the vowel make the reflex of the (t,d) token easy to identify, as in (2) and (3):

(2) *er Simon and I <u>kept</u> in touch* [kʰɛpˀtʰɪntʊtʃ]

(3) *if if a <u>project</u> or* [pˌɹəʊdʒɛʔtʰɔˑ] *contract comes up*

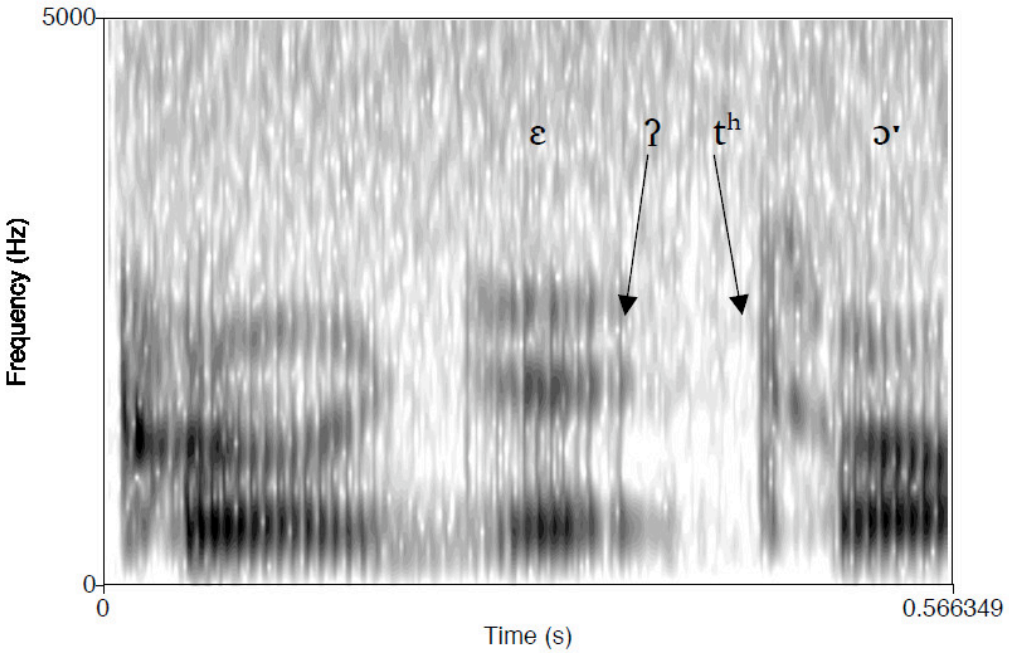


Figure 1. Spectrographic representation of "<u>project</u> or" (3); male speaker.

Figure 1 is a spectrogram of part of (3) showing the preceding /k/ realised as a glottal, a clear closure period and a release with formant transitions consistent with an alveolar plosive reflex of the word-final /t/ of *project*.

In the absence of a release, however, the unambiguous identification of the deletion of word-final /t,d/ is much more difficult, as is the case with the token in (4), which is illustrated in Figure 2:

(4) *having this lego <u>kept</u> me* [k^hɛ̰p˺mi^j] *occupied for years.*

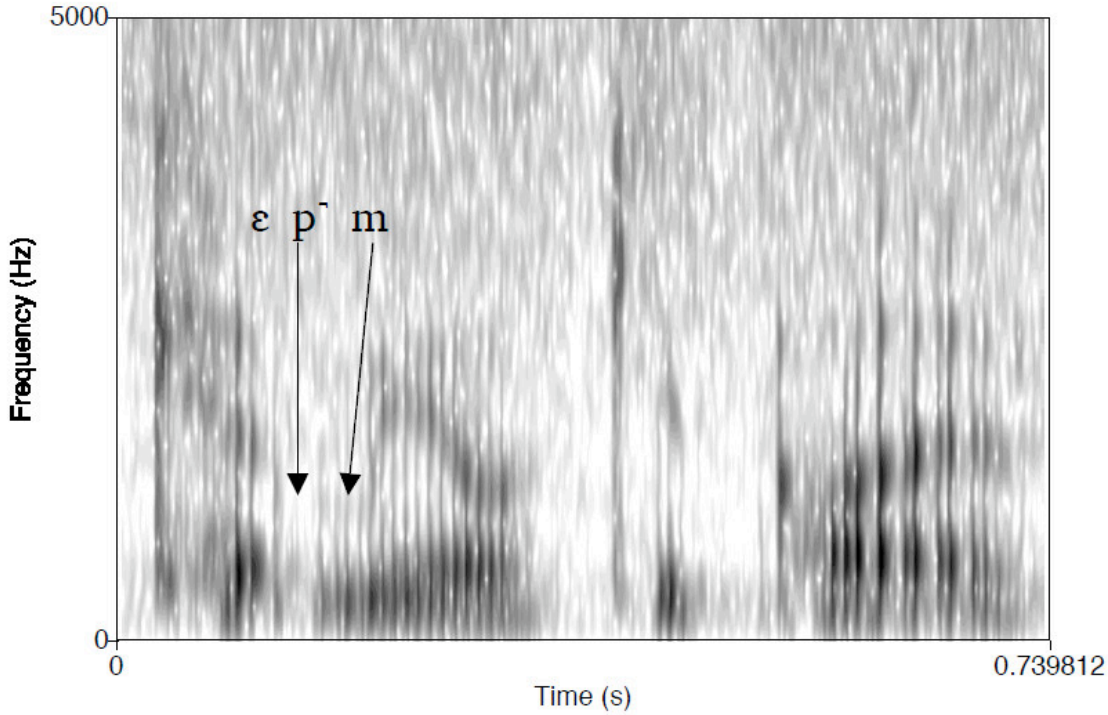


Figure 2. Spectrographic representation of "<u>kept</u> me occupied" (4); male speaker.

As Figure 2 shows, there is glottalisation of the vowel of *kept* and possibly glottal reinforcement of the [p], but auditory analysis reveals that there is also unambiguous bilabial closure. The following [m] is clearly visible. There is no evidence in the spectrogram or auditorily of a [t] between the [p] and the [m], but it is not possible to state categorically that there is or is not a stop gesture present. It is quite possible that an apical closure gesture could occur between the two, but unless the preceding bilabial closure was released before the /t/ gesture, and the following bilabial closure happened after it, it would not be perceived auditorily[8]. The unreleased /p/-to-homorganic-/m/ sequence is, of course, exactly what one would expect from a fluent native speaker of English

[8] The relatively short duration of the closure in *kept* compared to the /p/ of *occupied* is ascribable to a rapid deceleration of speech rate and cannot necessarily be taken as an indication of /t/ deletion.

and it is impossible to tell for certain whether the /t/ has truly been deleted or whether a residual gesture might remain. Even assuming the absence of a lingual gesture, the presence of glottalisation could be interpreted as a reflex of /t/ in a glottal stop, but this interpretation is no more straightforward: the presence of a masked glottal stop is no easier to identify, and the creak on the preceding vowel and in the diphthong of *occupied*, clearly apparent in Figure 2, means that this could just be a function of the speaker's register.

There were many tokens which showed this masking effect in T&T's data. In (4) the place of articulation of the preceding and following consonant is the same, but (5) and (6) demonstrate how this is not necessary for masking to occur:

(5) *well it was all <u>pressed</u> bits of* [pˌɹɛsbɪʔ͡tə] *meat you know*

(6) *but there was all <u>old</u> carpets* [ɔ̰lkʰa̲pʰɪʔs] *and pictures.*

In each case there is a preceding coronal gesture towards the alveolar ridge. Since word-final stops are not obligatorily accompanied by oral release (and, I would argue, not normally so in this type of context), the absence of an audible or visible release burst cannot be taken as the unambiguous absence of /t,d/: in (5) the blade and tip of the tongue could have raised from their fricative position to form a closure during the articulation of the "following" [b], just as the side(s) of the tongue could have raised to complete a post-lateral closure in (6). In both cases, the coronal release would have been masked by the closure of the following stop. It is, of course, equally possible that the tongue tip/blade was never raised further than for a fricative in (5) and was released as the dorsum (and sides) raised for [k] closure in (6). The problem is that it is impossible to tell either way without fine-grained articulatory data.

Masking is particularly problematic where there is glottalisation of the preceding consonant and with combinations of preceding nasals and following plosives or nasals. (7) is taken from the same subordinate clause as (6), focusing on the second (t,d) token; the relevant extract is shown in Figure 3:

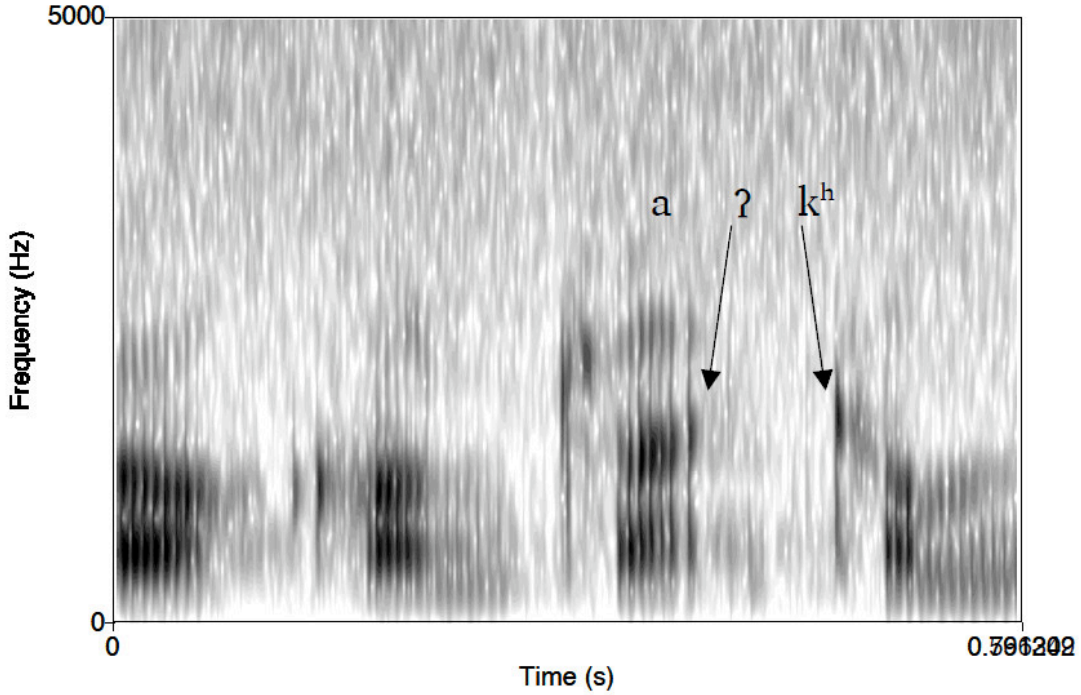


Figure 3. Spectrographic representation of "contract comes" (7); male speaker.

(7) *if if a project or contract comes* [kɒntɹ̥aʔˈkʊmz] *up*.

Again, the preceding and following segments are unproblematic: there is a clear closure into a glottal reflex of the preceding /k/ of *contract* and a clear velar release of the initial plosive of *comes*. Again it is not possible to state categorically that there is not a [t] gesture present, but if this were the case the glottal gesture would have to be released before the release of a [t] and crucially before the velar closure for the following /k/, for the presence of the /t/ to be perceived independently or show up on the spectrogram. Alternatively, given that a glottal stop is a common reflex of /t/, this could be construed as a further neutralising context since the presence of a preceding glottal stop makes it impossible to detect whether the glottal reflex is present or not (or, more accurately, it is impossible to tell whether the glottal is a reflex of /k/ or /t/ or both – see 3.3.4 below).

The parallel problem with nasals is illustrated in (8) to (10):

(8) *you know we were educated, trained people* [tɹ̥eˈnpiˈpl̩] / [tɹ̥eˈndˈpiˈpl̩]

(9) *they've found me asleep* [faʊnmiʲəslɪp] *in their bedroom*

(10) *they were over a thousand quid* [θaʊˈzn̩kwɪd] *each*

Occasionally, such cases could be disambiguated from spectrographic evidence, for example a sharp cessation and resumption of voicing with word-final /t/ followed by a voiced stop, but unsurprisingly, the majority are more like (8), represented spectrographically in Figure 4. The energy showing faintly between the [n] and the [p] release in Figure 4 is from the interviewer speaking over the informant; the informant's closure period between the bold vertical lines crossing the x-axis is unambiguously voiceless. Prior to that it is possible to see the nasal energy falling off in frequency, but

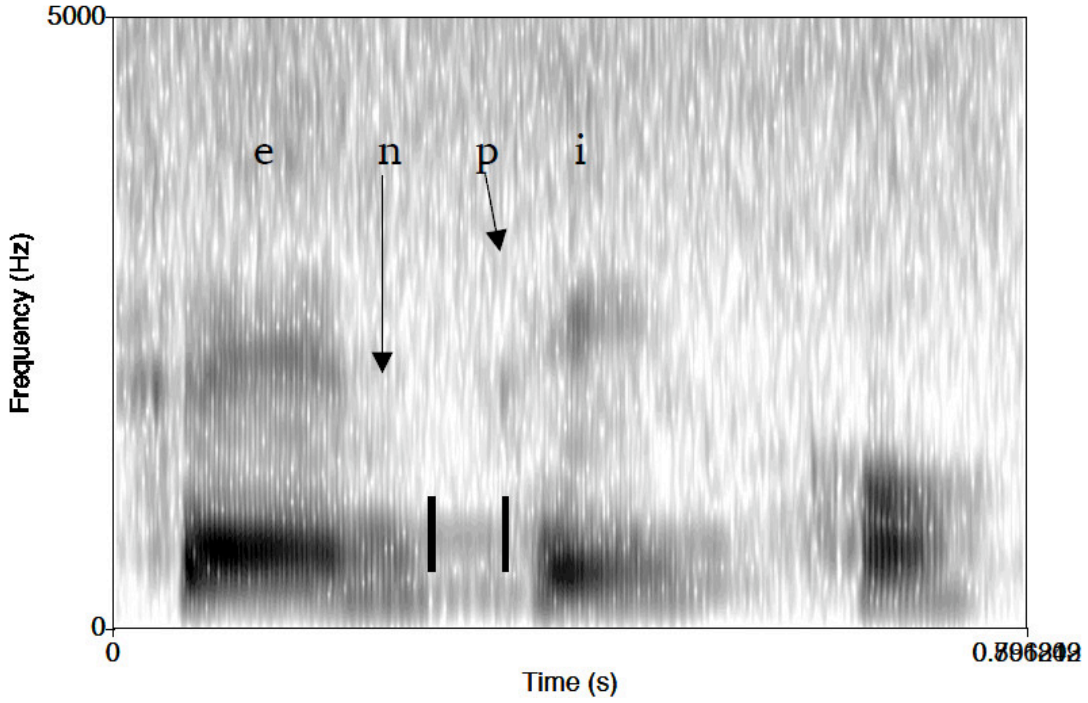


Figure 4. Spectrographic representation of "trained people" (8); female speaker.

there is no stretch of non-nasalised voicing consistent with a fully voiced [d]. The lack of voicing could be explained by the word-final assimilatory devoicing characteristic of many Yorkshire speakers, but in the absence of a release this potential explanation is of no help in determining whether or not the word-final stop is present.

Tokens in these contexts rarely have released [t,d], and those which do have audible release usually involve hesitation or a prosodic pattern signalling a pragmatic or discourse effect. This is the case in (11) and Figure 5, where the speaker is introducing the computer game Minesweeper as the source of his friend's problems with distraction at work and produces a micro pause after *found* followed by a lengthened diphthong in the first syllable of *Minesweeper*:

(11) *and he <u>found</u> Minesweeper* [faʊn̥d maɪnswiːpʰə]*, have you played Minesweeper?*

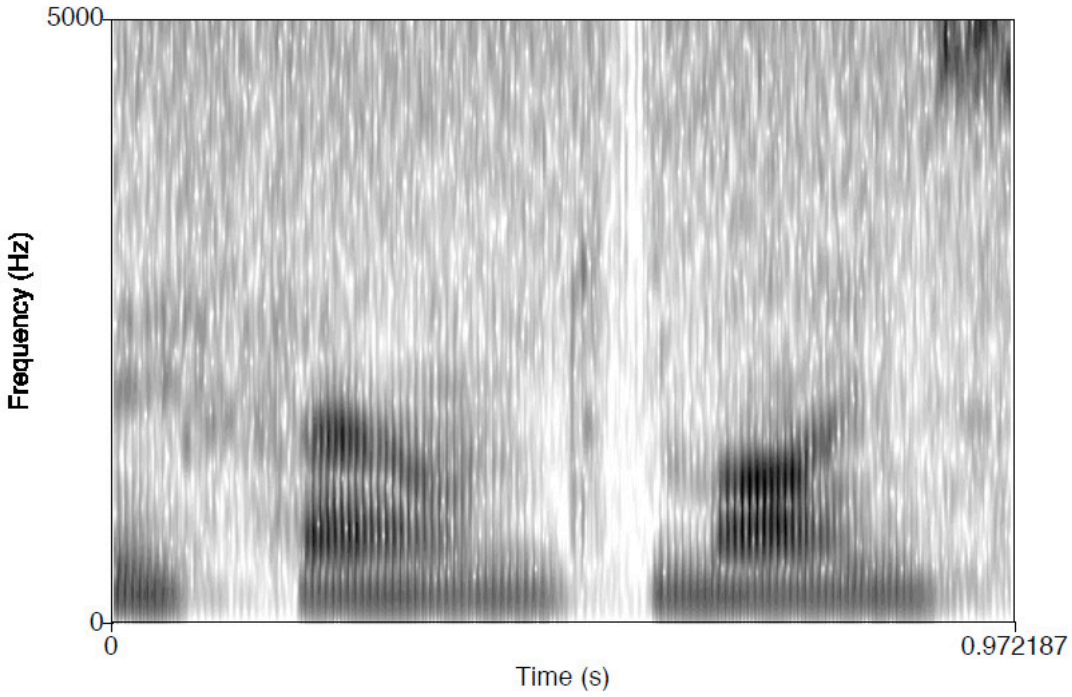


Figure 5. Spectrographic representation of "<u>found</u> mines[weeper]" (11); male speaker.

Examples (8) (Fig.4) and (11) (Fig. 5) were produced by different speakers and the durations are different, but the spectral pattern in *found* (11) is almost identical, *mutatis mutandis*, to that in *trained* (8): in both cases there is clear formant structure throughout the voiced portion of the closure for [n(d)] and no voicing bar without it, as there would be in a canonical voiced [d]. The plosive release in Figure 5 is completely voiceless, though not aspirated. This is again quite normal in English and it is difficult to see on what grounds one could possibly state definitively whether or not the stop in (8) (Fig. 4) has been deleted. In that case, even techniques like palatography would not disambiguate the token. It is thus hard to see the justification for extrapolating a phonological rule of deletion from these and the other examples in this section, and even if deletion could be demonstrated, it is hard to see how to justify the claim that it is governed by the same rule that deletes, say, the final /t/ of "I've never seen the film Gorillas in the <u>Mist</u> [mɪs]."[9] The latter would be

[9] An invented example is given here, since there is not a single example of a sentence-final coronal stop cluster with deletion in the data set analysed in T&T.

marked for speakers of York English and one would expect it to behave quite differently from the examples which are governed by their normal CSPs, yet the same variable rule is purported to apply to all these cases.

3.3 Assimilation

The problem of masking is compounded in cases of assimilation across the (t,d) token. Again, this is particularly a problem with nasals, which frequently assimilate to the place of articulation of a consonant following (t,d). When the underlying token is voiceless, it is sometimes possible still to detect a glottalised reflex of it, as in (12):

(12) *she's on a <u>different</u> plane* [dɪf̬ɪm̬ ˀp̬lɛˑn].

Reflexes of /d/ are, however, much harder to detect, as in (13), where the speaker is describing an early record player, and (14), which is shown in Figure 6.

(13) *a a a <u>sound</u> box* [saʊmbɔks] *was only a diaphragm*

(14) *we built, um, Bradford <u>combined</u> court* [kʰəmbaĩŋkʰɔːʔ] *centre*.

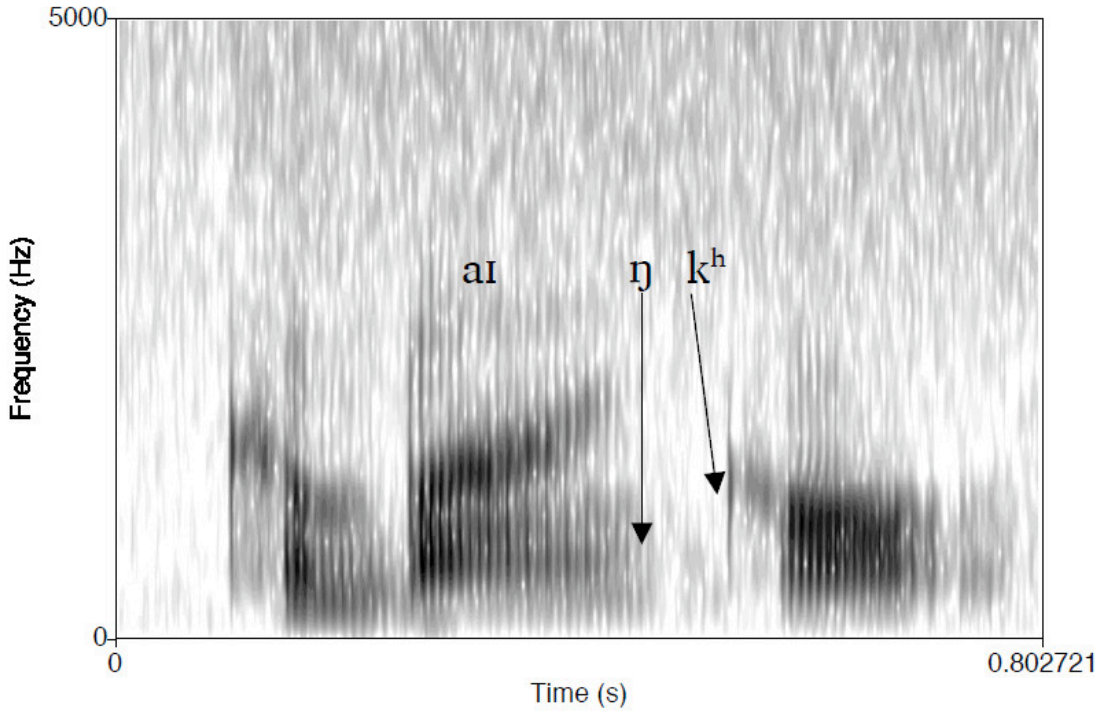


Figure 6. Spectrographic representation of "<u>combined</u> court" (14); male speaker.

It could be argued that these assimilation cases constitute evidence in support of a lexical rule of word-final coronal stop deletion: the assimilation in (14) can only occur because the /d/ between the nasal of *combined* and the velar plosive of *court* has been deleted before the postlexical rule of assimilation across word boundary applies. However, examples like (12) show that deletion is not a prerequisite for assimilation, since assimilation of the /n/ in *different* to the place of articulation of /p/ in *plane* occurs across the glottal reflex of the word-final stop, showing that segmental adjacency is not a prerequisite for assimilation.

3.4 Sequentiality

Example (4) above raises a further question, albeit one which is partly bound up with masking and assimilation, that is the possibility that a phonetic reflex of (t,d) might not occur sequentially between its "preceding" and "following" segments. The spectrogram in Figure 2 shows the audible glottalisation on the vowel of *kept* and into the [p] closure. It is well known that the phonetic cues to segmental identity are not restricted to the temporal slot implied by phonemic (or indeed generative) representations. The cueing of coda voicing by the duration of the preceding vowel is a commonplace. So it might be argued that there is a reflex of /t/ present in the *kept* of (4), although it is not sequentially aligned in the word-final position. Again, this is a topic which merits further experimental exploration, into both perception and production, beyond the scope of the present paper, but again the problem is raised of how to classify such tokens for variable rule analysis. T&T decided to classify them, not without some misgivings, as having undergone deletion because they were trying to replicate Guy (1991) and so far as they could ascertain, this would have been Guy's practice.

In (4), there is clear oral articulation of the [pˀ] of *kept* as well as the glottalisation. By contrast, voiceless velar stops immediately followed by another stop in York English (and many British varieties) are frequently realised as glottals without any velar articulation[10]. These tokens pose a different problem for classifying segments in sequence: in (15) the [t^h] of *worked* is released so [ʔ] and [t^h] can be taken as sequential reflexes of /k/ and /t/ respectively:

[10] Very occasionally, preceding /p/ is also realised as a glottal, as in *the whole place <u>except</u> us* [iʔsɛ̱ʔˈt^hʊs].

(15) *and that was where my dad worked and* [wɜ̃ ʔtʰən] *where the Barbican...*

However, this is not possible in (16) to (18), which are all from different speakers:

(16) *I w- worked part-time* [wəʔäʔtaɪm] *in funerals*

(17) *She knocked straight* [nɒʔstɻɛɪ̃] *into us yeah*

(18) *being an infant teacher was helpful in that respect because* [ɹɪsbɛ̰ʔˈbɪkʊz̥].

The preceding segment in each case is realised as a glottal stop, and it appears that the (t,d) token is absent. A parallel example, (7), was discussed under *Masking* above, but even if there were no masked alveolar gesture, [ʔ] is also a possible pronunciation of (t,d) in this variety, as shown in (19), so an alternative (or concurrent) interpretation of the problem is that it is impossible to disambiguate whether [ʔ] is a reflex of /k/ or /t/ or both.

(19) *you felt as* [fɛlʔəz] *if you moved you'd fall off*

It would be necessary to do detailed phonetic comparisons of a number of tokens with potential sequences of glottals to establish whether there is, for example, a regular pattern of variation between a lengthened [ʔ] in *worked* versus a shorter glottal reflex of /k/ in *(I) work*, which would indicate (although not conclusively) that there was an undeleted /t/ in this token of *worked*.

In their replication study, T&T again opted to code tokens such as (4) and (16) to (18) as deleted because that appeared to be the North American practice, but this is a rather problematic strategy. The problems are further complicated by the fact that preceding /k/ is very unevenly distributed across the data, as shown in Table 2 above: whereas 23% of regular past tense forms have preceding /k/ only 3% of monomorphemes and none of the semi-weak forms do. Since ambiguous glottals are overwhelmingly produced in tokens with preceding /k/ and following consonants this could be further skewing the findings for morphological class.

4.0 Variable (Lexical) Phonological Rules and (t,d)

Having addressed some of the problems of method and interpretation posed by the phonetic and statistical analysis of (t,d) data, we now turn to their theoretical implications. Although variable rules have their roots in generative grammar and specifically generative phonology, their ontological status has been a matter of debate (see, for example, Fasold (1991) or the brief overview in Mendoza-Denton, Hay and Jannedy (2003)): do they represent a convenient statistical tool for measuring variation or are they an albeit imperfect model of speakers' competence[11]? Whatever the general answer to this question, the linguistic characterisation of (t,d) in terms of the generative Lexical Phonology (henceforth LP) model, which drives the predictions concerning morphological class tested in T&T, entails that the rule be a phonological rule, at least so far as morphological class and preceding context are concerned, that is, it applies during the derivation of the word (as well as post-lexically). The question thus arises of how this particular rule fits into the phonology as a whole. It is unproblematic for processes strictly associated with the derivation of verbal forms, such as the deletion of the suffix vowel of {*-ed*} and voicing agreement of the final consonant, to occur before the variable deletion rule applies. However, the timing of the application of the rule with respect to processes affecting preceding and following consonantal segments does have direct bearing on the analysis. This is perhaps best examined with reference to further examples from T&T's data.

In (20) there is a clear release of the [t^h] accompanied by a short aspiration burst, so the token is an unambiguous example of non-application of the rule:

(20) *he was a bit wet when it comes to <u>contact</u> sports* -
[k^hɒnt^{h}a̰ʔt͡^{h}spɔːʔs]

[11] Notwithstanding the problems outlined in this paper, (t,d) is an interesting example of how the statistical model of a variable rule can differ from the linguistic variable rule being modelled: morphological category is an independent factor group in the statistical analysis whose function is to model the consequences of the iterative application of the linguistic variable rule, which in the LP view has no need of the input of an independent variable of morphological category, since it falls out of the structure of the phonological component of the grammar. This mismatch between a putative linguistic variable rule and the statistical modelling of its behaviour is not in itself problematic.

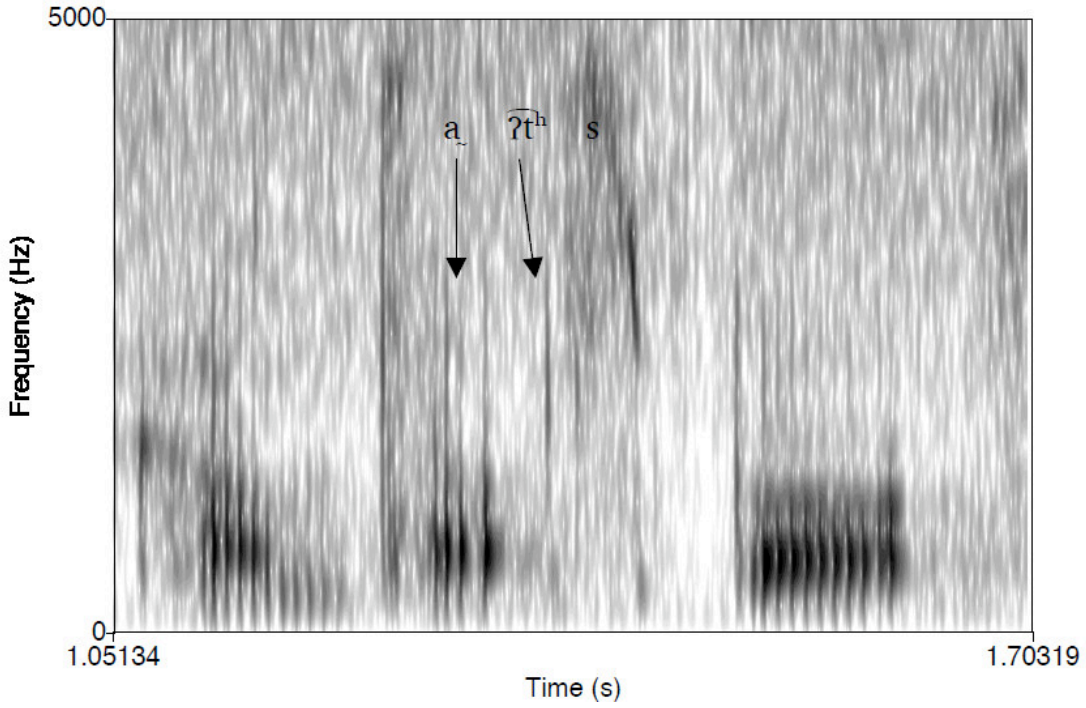


Figure 7. Spectrographic representation of "contact sports" (19); male speaker.

The following context is unproblematically [s]. However, the preceding context is less straightforward: /k/ is realised as a glottal, which raises the question of what exactly the preceding context was when the rule applied, [k] or [ʔ][12]. It might be argued that what matters for the rule is that [ʔ] is a stop, and its place of articulation is not important, but phonetically it is realised as creak on the /a/ vowel (see Figure 7), as arguably something which is qualitatively very different from [k]. Of the 1118 tokens in Table 1, 71 preceding /k/s are phonetically glottal stops and 5 are glottalised; glottals thus represent nearly 7% of the data set and 45% of preceding stops, so this is far from a trivial question.

A similar problem occurs with vocalised /l/, as in (21):

(21) *So she told me off* [tʰɐʊmiɔf] *for shouting at her*

York English is not known as a strongly /l/-vocalising variety, but there are ten such tokens in the data set and one where there is no obvious sequential reflex of /l/:

[12] Since the rule applies iteratively, the answer to this question may actually be different at different stages in the derivation, thus introducing a further complicating element.

(22) *my friend told me right* [tʰəmɹ̩ɹ̩aɪ] *yesterday*

In these and other cases of the absence of a preceding phonetic consonant, the question arises of how long in the derivation the underlying cluster remained a cluster and so subject to the (t,d) rule. Whereas tokens with preceding phonetic laterals have a mean rule application rate of 19%, of the ten tokens[13] where the word-final consonant is preceded by a phonetic vowel in the surface form, six (60%) have the final consonant deleted. This may be simply due to the small number of tokens, but it is interesting that syllabic phonetic laterals, also few in number, pattern in the same way as the non-syllabics which surface phonetically (25% deletion, N=8).

Questions of rule ordering also affect the following phonological context. In cases like (23), where the /t/ coarticulates with the following /j/, the same question arises: what is the following context when the rule applies, in this case postlexically?

(23) *like [the baby] kept you up* [kʰɛpˀt͡ʃʲʊpˀ] *24 hours a night*

Following /h/ is particularly problematic in this respect. In (24) the following context is phonetically a vowel, but underlyingly it is consonantal. What, then, is the following context when the rule applies?

(24) *Yeah that that was it we was walking down Micklegate and we grabbed him* [gɹabdɪm]

These problems are compounded when the processes affecting adjacent consonants also affect (t,d), as illustrated by (16) above, reproduced here:

(16)… *I w- worked part-time* [wəʔäʔtaɪm] *in funerals*

Here, [ʔ] is a perfectly normal reflex of both coda /t/ and /k/ in many varieties of British English so it is not only the preceding consonant whose identity is in question at the point of application of the rule, but the surface (t,d) token itself: is it deleted or not? If not, has /t/-glottalisation occurred before or /k/ glottalisation and/or (t,d)?

[13] There were in fact 18 tokens in the whole data set, but some were excluded on other grounds for the analysis shown in Table 1. The problem would, of course, be more serious in other varieties of British English where /l/-vocalisation is more common.

The questions raised here cannot be dismissed by saying the rule relates to abstract phonological units or categories of sonority, major class features etc: in order to carry out variable rule analysis, the analyst has to code each token for preceding context, and it is crucial to know what that context is. This is particularly important in cases where the preceding context could be a vowel, which means the cluster may not actually be a consonant cluster when the rule applies, and equally so where the following context may be a vowel, given that following consonant versus following vowel has been known (unsurprisingly) to have the most robust effect on (t,d) since the very earliest studies. With an iterative rule, such problems are intractable. It is difficult to see how to determine whether the chicken of rule application came before or after the egg of, say, /l/-vocalisation.

5.0 Discussion and conclusions

This survey of a range of problems which came to light during T&T's attempts to replicate North-American studies of (t,d) with data from northern England has been somewhat brief, due to space constraints, and apparently rather eclectic. However, as already indicated, many of the issues are inter-related and all raise questions not only about (t,d) as a linguistic variable analysable in terms of Lexical Phonology but also about the nature of variable rules in general and indeed about the relationship more broadly between phonetic output and phonological analysis.

The phenomenon of masking might seem to pose purely practical problems, and the argument could be adduced from the point of view of perception that the masking causes the hearer not to hear a reflex of /t,d/ and it is thus reasonable to model its perceived absence as a result of deletion. However, the generally accepted treatment of "neutralisation" in (t,d) by excluding tokens in neutralising (following) contexts on the grounds that it is impossible to perceive whether the (t,d) token is deleted or not, demonstrates that (t,d) is modelled on the basis of production rather than perception. Since masking and neutralisation introduce the same uncertainty in the first step of the analysis, that is deciding whether a token is realised or not, they should at the very least be treated in the same way: either neutralised tokens should be included in the analysis because they form part of what the hearer hears (and presumably recognises as (t,d) sites), or masked tokens should be excluded because, as with neutralisation, it is impossible for the analyst or the hearer to detect whether deletion has occurred. Given that production and perception must ultimately be linked, this decision might still be construed as merely a practical, operational one, but it must nevertheless

be addressed and it cannot be given proper consideration without also considering the abstract model of the behaviour of (t,d), to which we shall return below.

Assimilation was presented in §3.3 above as compounding the problem of masking. Could it be the case, on the other hand, that it confirms that deletion has taken place? In this view, deletion would lead to, e.g., an underlying /n/ and /b/ being adjacent in *sound box* (13), making the assimilation of place of articulation unsurprising. However, the problem of undetectable gestures for [t,d] remains, and the evidence of *different plane* (12), pronounced [dɪfɹ̥ɪm̥˕˜ʔp̥lɛˑn], shows clearly that assimilation can still take place when the intervening segment is not deleted, so its usefulness as a diagnostic is rather doubtful. Moreover, assimilation and the other processes affecting preceding and following consonants raise the question, addressed in §4, of how (t,d) relates to other processes affecting its conditioning: does it apply before or after /l/ vocalisation, /h/ deletion or indeed assimilation? Does it perhaps feed any of those processes? So far as T&T could ascertain, the assumption in the literature seems to be that (t,d) takes underlying phonological units as its input. This assumption has to be justified, however: on what basis can it be argued that (t,d) belongs in the (lexical) phonology whereas those other processes are either phonetic or post-lexical or even lexical but applying after (t,d)?

This brings us to the fundamental problem of the nature of (t,d), its relation to phonology and phonetics, and the nature of variable rules. Why, one might ask, should deletion be a phonological rule at all? The original conception of variable rules was a part of a Generative Phonology-type rule. As I have acknowledged, variable rules have evolved into more of an analytic construct than a theoretical one, but they nevertheless retain their claim to model, albeit at some remove, how speakers produce and perceive variable patterns of speech. (t,d), as I have also acknowledged, goes further than this, working backwards from the observation that the variable appears to be conditioned by the morphological class of words to the assumption that it really is a phonological rule operating both lexically and post-lexically. It behoves the advocates of this view of (t,d) not only to demonstrate that the patterns of variability are consistent with the predictions of LP (which T&T were unable to do), but just as importantly, to demonstrate the compatibility of the variable rule with the model in other respects, in other words to demonstrate that this is a (lexical) phonological rule. In its lexical component, LP deals with contrastive phonological units and their morphophonological alternations. There is no reason why lexical LP rules should not be variable, but that does not of itself make (t,d) a candidate to

be a lexical rule any more than l-vocalisation or the glottalisation of /k/ in *worked* (15) or *knocked* (17) would be. The conditions for (t,d) are introduced by the morphology (except, of course in the case of monomorphemes) but there is no phonological contrast between /t,d/ and zero (except in the trivial sense that anything might be said to contrast with zero) and no morphophonological alternation involved.

An alternative analysis might be that (t,d) is a phonetic Continuous Speech Process. Being phonetic does not preclude being variable and structured, but as well as allowing a more holistic approach in the light of what is known of other CSPs in English, viewing it this way obviates the need to justify a more abstract phonological analysis. It does not, of course, mean that issues like masking, the ordering of processes and assimilation disappear, nor does it obviate the need to make a reasoned case for such an analysis, but that analysis will have to await a further, fuller treatment.

Acknowledgements

My gratitude continues for Sali Tagliamonte's generosity in inviting me to collaborate with her after I expressed an interest in her preliminary findings, as presented in Tagliamonte (2000), and for the stimulating exchanges we have had since then. Although we have discussed many of the questions in the present paper, the analyses and opinions expressed here are my own and not all shared by her, and she should not be called to account for them. I am also grateful to Sali Tagliamonte for access to data collected with the support of the Economic and Social Research Council of the United Kingdom (the ESRC) under Research Grant #R000238287, Grammatical Variation and Change in British English: Perspectives from York. Many other colleagues have provided encouragement and stimulating discussion, particularly Paul Foulkes and audiences at the First International Conference on the Linguistics of Contemporary English in Edinburgh and at UKLVC 5 in Aberdeen, both in 2005.

References

Bayley, R. (1995). Consonant cluster reduction in Chicano English. *Language Variation and Change*, *6*, 303-326.

Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents.* Berlin: Foris Publications.

Docherty, G. J., Foulkes, P., Milroy, J., Milroy, L., & Walshaw, D. (1997). Descriptive adequacy in phonology: a variationist perspective. *Journal of Linguistics*, *33*, 275-310.

Fasold, R. (1972). *Tense Marking in Black English.* Arlington, VA: Center for Applied Linguistics.

Fasold, R. (2003). The quiet demise of variable rules. *American Speech*, *66*, 3-21.

Guy, G. (1980). Variation in the group and the individual: the case of final stop deletion. In W. Labov (Ed.), *Locating Language in Time and Space* (pp. 1-36). New York: Academic Press.

Guy, G. (1991). Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change*, *3*, 1-22.

Guy, G. (Forthcoming). *Language Variation and Linguistic Theory.* Oxford: Blackwell.

Guy, G. & Boberg, C. (1997). Inherent variability and the obligatory contour principle. *Language Variation and Change*, *9*, 149-164.

Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, *1*, 85-98.

Labov, W. (1997). Resyllabification. In H., Frans, R. van Hout, & L. Wetzels (Eds). *Language Variation and Phonological Theory* (pp. 145-180). Amsterdam: John Benjamins.

Labov, W., P. Cohen, C. Robins, & J. Lewis. (1968). *A Study of the Nonstandard English of Black and Puerto Rican Speakers in New York City.* (Cooperative Research Report no. 3288). Washington DC: U. S. Office of Education.

Mendoza-Denton, N., Hay, J., & Jannedy, J. (2003). Probabilistic sociolinguistics: beyond variable rules. In R. Bod, J. Hay, & S. Jannedy (Eds). *Probabilistic Linguistics*. (pp. 97-138). Cambridge, MA: MIT Press.

Rand, D., & Sankoff, D. (1990). *GoldVarb: A Variable Rule Application for the Macintosh.* Montréal: Canada. Centres de recherches mathématiques, Université de Montréal. Version 2.

Santa Ana, O. (1992). Chicano English evidence for the exponential hypothesis: a variable rule pervades lexical phonology. *Language Variation and Change*, *4*, 275-288.

Sigley, R. (2003). The importance of interaction effects. *Language Variation and Change*, *15*, 227-253.

Tagliamonte, S. A. (1998). Was/were variation across the generations: View from the city of York. *Language Variation and Change*, *10*, 153-91.

Tagliamonte, S. A. & Temple, R. A. M. (2005). New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change*, *17*, 281-302.

Temple, R. A. M. (2000). Now and then: the evolution of male-female differences in the voicing of consonants in two varieties of French. *Leeds Working Papers in Linguistics and Phonetics*, *8*, 193-204.

Wolfram, W. (1969). *A Sociolinguistic Description of Detroit Negro Speech*. Washington, D.C.: Center for Applied Linguistics.

Wolfram, W. (1993). Identifying and interpreting variables. In D. Preston (Ed.). *American Dialect Research*. (193-221). Amsterdam and Philadelphia: John Benjamins.

Accentual Patterns in the Spoken French of the Early 20th Century

Ian Watson

Christ Church, University of Oxford

Abstract

Three of the earliest recordings of spoken French were analysed prosodically to determine whether they showed evidence of accents early in APs, such as are found in current French, or rather an absence of such accents as suggested by contemporary early-20th century accounts. A considerable proportion of APs had early accents, although their f0 contours were not always akin to those reported in current forms of the language.

Keywords

French, Accent, Language Change, Historical Laboratory Phonology

1.0 Introduction

The student of French language who compares traditional models of French prosody (such as are found in most textbook accounts) to recent research and overview papers (di Cristo, 1999, Post, 2000, Jun & Fougeron, 2000, Gussenhoven 2004) can easily be plunged into a state of Orwellian doublethink. According to the traditional approach, French has either no accents at all (see e.g. Rossi, 1980), or at best very limited accentuation on the final syllable of rhythmic groups/intonational phrases. According to recent studies, French has accents every 1.74 syllables (Gussenhoven, 2004, Post, 2000) with lexical words bearing clear accentual marking, possibly even in post-focal positions (Di Cristo & Jankowski ,1999) and with a large number of word-initial secondary accents (Fónagy, 1980, 1989, Astésano et al., 1995).

Four potential explanations of this discrepancy are:

(i) one of the two accounts is fundamentally flawed;
(ii) the two accounts are based on different notions of what is meant by 'accent';
(iii) the two accounts are based on different notions of what is meant by 'French';

(iv) there has been a significant recent change in the prosodic structure of French.

Explanation (iv) has been advocated in a number of recent studies (Fónagy, 1980, 1989, di Cristo 1999). Fónagy was the first in recent times to investigate the use in French of secondary accents; whereas primary accents are found on phrase-final syllables, secondary accents occur on earlier syllables, notably on the first syllable of polysyllabic words. Fónagy found evidence of an increase in the prevalence of such accents across the twentieth century.

The present study reports on initial results from a project designed to re-assess the diachronic explanation using a wider range of early recordings than were available to Fónagy. In the last ten years, numerous recordings of spoken French from the period 1911-1920 have been made available either commercially or through electronic publication by the Bibliothèque Nationale de France. Analyses of extracts from three of these recordings are presented here. These are used to test the hypothesis that secondary accentual marking early in rhythmic groups was already a common feature of the spoken French of the early 20th century and that the language of that era did not therefore differ substantially in this respect from that of the early 21st century.

The focus on the diachronic explanation (iv, above) does not exclude consideration of explanations (i) – (iii), indeed it arguably implies that all will be taken into account (they are in any case not mutually exclusive). In comparing accentual patterns in French across two eras, a stable definition is needed of both 'accent' and 'French'; these definitions may then be compared to those of the traditional and modern approaches to determine the degree of correspondence between them (explanations (ii) and (iii)). Concerning explanation (i), the proposition that one of the two accounts is flawed, recent studies have a strong empirical and theoretical basis. They are solidly underpinned by laboratory studies employing acoustic analysis; with the proviso that (as their authors indicate) there may be stylistic limitations to the applicability of their findings (see, e.g. Lucci, 1983), there are no apparent grounds to challenge them. The traditional account is thus certainly flawed as a description of current French. The possibility remains, however, that it is an accurate description of a slightly earlier form of the language, an account which simply failed to be modified as diachronic changes occurred. In this case, the data from the recordings analysed here should match the traditional analysis better than does 21st century French, thereby disproving the hypothesis above.

2.0 The traditional view of French accentuation

2.1 Origins and main features

Reference to *a* traditional approach to French prosody necessarily involves an amalgam of analyses ranging across the 20th century which are not totally identical. Nonetheless, since Fónagy's influential article (1980) it has become accepted practice in the literature (cf. Astesano et al., 1995, di Cristo, 1999) to conflate views which share the attribution to French prosody of four essential features:

(i) all words have an oyxtonic (final syllable accented) rhythm when spoken in isolation or when the accent is realised in continuous speech (Grammont, 1913, 1963, Dauzat & Fouché, 1935);
(ii) in continuous speech, most words lose their accent, leaving the only accent in the intonation phrase on the *phrase*-final syllable (Pulgram 1965, 1967);
(iii) the most consistent acoustic marker of this phrase-final accent is duration, as phrase-final syllables are often low in intensity and may not be clearly pitch prominent (Delattre, 1966a);
(iv) special emphasis may also be marked by a distinct 'accent d'insistance' which is applied to the first or second syllable of a word carrying special emotional importance.

(iv) is the only exception to the oxytonic pattern and its use is rare, especially compared to that of analogous procedures in other languages; this infrequency is emphasised in a number of papers (Grammont, 1913, Marouzeau, 1924).

2.2 Prosodic structure

From the four features listed above, a fifth follows logically: (v) there is only one type of unit in French prosody, variously called the breath group, tone unit, or sense group; here, following current usage, it will be referred to as the Intonational Phrase (IP). French prosody is thus distinct from language with a richer accentual structure, such as English or German. In these, there may be several accented syllables in each IP, and sub-sets of these accented syllables may be grouped with their unaccented counterparts to form smaller units such as the foot or the accentual phrase. The traditional view sees no reason to hypothesize such

units in French, as they would in effect be coextensive with the IP (cf. di Cristo, 1999).

2.3 The definition of French

This traditional view is derived from articles by leading French phoneticians of the first half of the 20th century. Such articles rarely cite any data and they inevitably adduce no instrumental evidence. Few define 'French' with any sociolinguistic or stylistic details, the implication being that there is a single coherent code denoted by this term. Others are explicitly normative, indicating that they are describing *'français correct'* (Fouché, 1933, 1936), a form which is more easily identified by what it is not than by any positive properties. Fouché (1936) contrasts 'le français correct' with all of the following:

> the language of peasants;
> the language of the provinces (everything except Paris);
> the language of the people ("peuple") of Paris (as opposed to some sections of the Bourgeoisie);
> the language of that section of the Paris Bourgeoisie which has moved to Paris from the provinces.

Less negatively, the object language is found, he claims, in certain Parisian families; 'des familles où depuis trois generations au moins, il n'y a pas eu d'alliances provinciales'[1] (given that these are not families of the 'peuple'). Should this seem too restrictive, he also adds that children who move to Paris may learn to speak like such families; '*l'enfant qui y arrive [à Paris] à la condition qu'il fréquente une école'*. [2]

The identification of 'French' with a (relatively small) group of people is further circumscribed in Fouché's earlier (1933) article, in which he specifies, with explicit reference to prosody, that his descriptions apply only to spontaneous conversation, not to a higher, more careful register. This is a surprising limitation as it suggests that the educated bourgeoisie are more likely to use a low-status feature of speech in a higher than a lower register (see further discussion in section 4). Fouché seems here to hark back to Vaugelas' 17th century admonition that good French should be sought in the speech of a selected section of the court 'la plus saine partie de la cour,' with one circumscribed in-group, the erstwhile

[1] "Families where for at least three generations there have been no marriages with provincials."
[2] "The child who arrives there [Paris] provided he attends school."

courtiers, replaced by another, namely approved 20th century bourgeois families.

2.4 The notion of accent

The notion of accent in the traditional account of French is a broad one: accent is said to involves the perceptual prominence of a syllable which may in theory be marked by any of the three acoustic features (length, intensity, pitch prominence) which are commonly found cross-linguistically (Lehiste, 1970). In practice, the (predictable) phrase-final accent and the accent d'insistance differ in their commonly observed main correlates (Delattre, 1966a). The former is marked above all by phrase-final lengthening. It may also be marked by pitch prominence, but this is not criterial, and it is often markedly *lower* in intensity than surrounding syllables. The *accent d'insistance* is marked by pitch prominence and intensity, although it may also be associated with increased duration. As French lacks both vowel reduction and distinctions in vowel length, it follows that syllables which are neither given an *accent d'insistance* nor are in phrase-final position should be of approximately equal duration.

2.5 Predictions of the traditional approach

The predictions of the traditional approach may thus be summarized as follows:

(i) only phrase-final syllables will normally be accented
(ii) other syllables will normally be of approximately equal length
(iii) those other syllables will not be prominent in terms of their pitch or intensity
(iv) there will be occasional emphatic marking of non-final syllables (the *accent d'insistance*)
(v) this marking will involve intensity and pitch excursions, possibly with increased duration
(vi) there is only one level of prosodic structure in the language.

3.0 The modern approach

3.1 Main features

Research on French prosody in the last 30 years has produced a number of competing models. Although their assumptions and analyses

are too different to allow for conflation, all make use of an autosegmental-metrical approach (e.g., Ladd 1996), and share sufficient features distinguishing them from the traditional approach for it to be legitimate, for the purposes of this study, to treat them together. The main features of modern approaches may be summarized in four points modelled on and contrasting with those in section 2.1:

(i) words are oxytonic when phrase-final, but not necessarily elsewhere in continuous speech; accents may be earlier in the word, notably on the first syllable;
(ii) most content words, and under some circumstances even clitics, bear at least one accent in continuous speech; words of three or more syllables may bear more than one accent;
(iii) phrase-final accents are marked by duration but are also generally pitch-prominent; other accents are marked by pitch (although probably also with some lengthening: see below);
(iv) special emphasis may also be marked by an '*accent d'insistance*'; this occurs in the same position as non-emphatic early accents, typically on the first syllable of the relevant word. Autosegmental-metrical models have not generally treated this *accent d'insistance* as phonologically distinct from other initial accents. However, it has been claimed that the f0 peak associated with it has a sufficiently different shape from that of other initial accents to warrant such a separate treatment (Astésano et al., 1995, Jankowski et al., 1999). On this view, the *accent d'insistance* is marked by a particularly sharp rise and following fall, other initial accents having a gentler rise and being followed by a less pronounced dip.

3.2 Prosodic structure

It follows from the above that in this approach, there are considerably more accents in continuous speech than there are Intonation Phrases. This allows for the possibility, *contra* the traditional account, that French might have more than one level of prosodic phrasing, as within an IP there may be several sub-groupings of an accented syllable with unaccented syllables. Different autosegmental accounts differ as to how many levels they recognise; in particular some recognise a level equivalent to the foot in English (e.g. Di Cristo, 2000, Gussenhoven, 2004), while others (Verluyten, 1984, Jun & Fougeron, 2000) do not. All, however, make a distinction between the overall IP and at least one smaller unit, variously called the Accentual Phrase, Tonal Unit, Prosodic Phrase. In what follows, the term Accentual Phrase (AP) will be retained

for this purpose, while the term 'prosodic phrase' will be used as a cover term to denote both units of prosodic phrasing, i.e. APs and IPs.

IPs are thus composed of one or more APs. Both IP and AP are marked by final lengthening, this being greater at the end of the former than the latter. Fundamental frequency patterns are also associated with the boundaries of prosodic phrases, notably the beginning of the AP. A typical pattern involves an early low tone followed immediately by a rise, often reaching a peak on the first syllable of the first content word in the AP. Although there are variants to this sequence, it can often be used as a criterion, along with final lengthening, for establishing the boundaries between APs (Jun & Fougeron, 2000).

3.3 The view of French

The experimental evidence underpinning the modern approach has predominantly come from highly controlled read speech. There has been little overt control for the social background of speakers, although subjects have generally been drawn from educated populations of university or advanced school students. compared to the target group of the traditional approach, therefore, social cohesion has been replaced by a degree of educational consistency. Attention has also been paid to stylistic variation, in that comparisons have been made between read speech, retellings of stories, interviews, lectures and conversations (Lucci, 1983, Post 2000). Overall, the main finding of these studies has been one of variability in the usage of non-final accents. These are most prevalent in 'didactic' speech and least common in conversation. They are nonetheless attested across the range of styles examined and are generally treated as a central part of French prosody (Astésano, 1995, di Cristo, 2000, Post, 2000, Jun & Fougeron, 2000, Gussenhoven, 2004), all styles taken together. The notion of 'French' prevalent in the modern approach is thus less exclusive but less precise than that found in the writings of the earlier generation, reflecting in part a move away from the ideal of a single 'correct' form of the language with a social elite.

3.4 The notion of accent

The definition of accent in autosegmental accounts has varied from model to model, but all emphasise the role of f0 as an accentual marker. This reflects the central interest of many authors in developing overall compositional models of intonation based on the concatenation of f0 movements associated with accented syllables (pitch accents, in the standard autosegmental terminology introduced by Pierrehumbert, 1980,

and derived from Bolinger, 1958; see Ladd, 1996, for discussion). Duration is also seen as a criterial marker of primary, phrase-final accent, the degree of lengthening being greater at the end of an IP than of an AP. For secondary non-emphatic accents there is probably also a small effect of lengthening (Astésano et al., 1995). With different degrees of lengthening for accents in different positions, the modern approach thus predicts that there will be considerably more variability in syllable duration within an IP than the traditional approach.

Apart from the phrase final accent, further accents are typically (but not obligatorily) found on early syllables in lexical words, e.g. ˌinterˈdit, ˌrenouˈveau, ˌimpréviˈsible. The theoretic status of these early accents varies across different models; in what follows, for terminological simplicity, the term 'nucleus' will be borrowed from the British tradition of intonation analysis (Cruttenden, 1997) to denote the phrase final accent in both APs and IPs, while the term 'secondary accent' or 'early accent' will be used to denote others.

A principle of stress clash prevents the attribution of accents to adjacent syllables within an AP, except, rarely, in the case of emphatic accents. Thus a two-syllable word in phrase-final position, which must receive the primary accent on its second syllable cannot normally receive an earlier accent:

1. C'est un garˈçon
2. *C'est un ˌgarˈçon

Two-syllable words earlier in a prosodic phrase may be accented on either (but not both) of their syllables. Longer lexical words most frequently receive a secondary accent on their 1st syllable, but in those of 4 or more syllables (a small fraction of the French lexicon, see di Cristo, 1999), a secondary accent may occur on the first or on a later syllable, typically the second (Verluten, 1982, Gussenhoven, 2004). Clitics and form words in general are not usually accented, although in spontaneous speech this limitation is far from always observed. With the exception of the prohibition on stress clash, the principles governing the distribution of non-final accents allow for considerable stylistic and individual variability. However, one pattern, identified by Fónagy (1980) has been observed repeatedly in a range of studies. Christened by Fónagy the '*arc accentuel*' (accentual arch), it involves the gathering together of a semantically important and coherent group of words so that the first syllable of the first content word and the last syllable of the last content word are accented:

1. Le ˌprofesseur de linguiˈstique
2. Le ˌPrésident de la RuˈssIe.

3.5 Predictions derived from the modern approach

The modern approach suggests that:

(i) within an Intonational Phrase there may be a number of accents;
(ii) most lexical words, and in certain circumstances, clitics, bear at least one accent; polysyllabic words may have more than one (but see [v], below);
(iii) prosody is organized hierarchically. An intonation phrase may consist of several smaller accentual phrases. Each of these will be marked by a degree of lengthening of the phrase-final syllable, the lengthening of the last syllable of the whole IP being greater than that of the internal APs;
(iv) there will always be an accent on the phrase-final syllable in APs and IPs;
(v) stress clashes (accents on successive syllables) are avoided within APs;
(vi) otherwise accent placement is not limited to final syllables;
(vii) there is a tendency to place an accent early in a phrase, most often on the first syllable of the first content word;
(viii) there is considerable variability as to when/how frequently non-final accents are realised;
(ix) some of this variability is related to stylistic variation; but
(x) there is a tendency for semantically important groups of words to be gathered together rhythmically such that they are contained in an 'accentual arch' formed by accents on the first syllable of the first content word of the group and on the last syllable of the last such word, there being no other intervening accents;
(xi) there is, as in the traditional account, the possibility of the 'accent d'insistance'. Modern accounts vary as to whether this should be considered a phonologically separate entity from the much more frequent normal initial accent.

4.0 The development of early accentuation

4.1 The picture emerging from the literature

Fónagy's (1980, 1989) identification of the role of an early accent in French APs was based on his analysis of radio broadcasts from the 1940s. The observed frequency of the phenomenon disqualifies it from being dismissed as the (allegedly rare) 'accent d'insistance'. His subsequent investigations suggested that the tendency to early accentuation varied according to speaking style, and that his initial observations based on news reports had probably caused him to overestimate it. Nonetheless, he judged that the tendency was probably spreading to other styles of the language, a claim that is now generally accepted (cf. di Cristo, 1998). Based both on contemporary comments and on its early detection by foreign scholars (Schuchardt, 1880, Meyer-Lübke, 1890, Scherk, 1912), Fónagy proposed that the origin of the phenomenon probably lay in the latter half of the 19th century, but that its generalization gained pace through the 20th century.

Other studies pose problems for this proposal and tend to suggest both an earlier origin and a possible earlier spreading of the phenomenon. An origin at least as early as the 18th century is proposed by Carton (1971, see also di Cristo, 1999) who quotes remarks to this effect by Voltaire and Rousseau. However, studies of French poetry suggest a still earlier date. The classical French alexandrine, with twelve syllables divided into two hemistiches of six syllables allowing accents only at the end of each hemistich appears to correspond ideally with the traditional, phrase-final accentuation described by the traditional account of French rhythm. Yet as early as 1912, a study by Lote demonstrated that this metrical pattern was already being disrupted in the mid- (and perhaps the early-) 17th century by the incursion of further accents into the 12 syllable line. More modern work by Pensom (1993, 1998) traces the distribution of accents implied by metrical patterns back from the 20th century, through classical 17th century practice to the medieval period, and suggests that there may never have been a period in which early accents were totally absent from the language.

On this view, then, rather than disappearing from the language for several centuries then being re-introduced (Lyche & Girard, 1995), early accents may have been a constant feature of French, albeit one that was for a period stylistically marked. Exactly who did and did not use this feature, in which styles and when, remains unclear. As noted above, a greater usage was observed in 'didactic' styles, notably news-reporting and lecturing, than in conversation at various points in the mid- and late 20th century (Fónagy 1980, 1989, Lucci, 1983). We have seen (section

2.3) that French phoneticians of the 1930s denied the presence of early accents (which for them could only be the *accent d'insistance*) in what they defined as standard French, while recognising that it was found in the speech of provincials, the lower classes and even the bourgeoisie when speaking in higher registers.

Even this claim has to be treated with some suspicion, however. Various foreign observers of the 1920s and 1930s refer to early accents as being a notable feature of Parisian speech, without any suggestion of a limitation to particular social groups (Schwartz, 1930, Gill, 1936). This suggestion, fiercely rebuffed by Dauzat (1936) in a comment directly following Gill's paper was, in contrast, accepted by some phoneticians both of his time and earlier. It is mentioned by Passy in 1890. In 1930, Schwartz reports learning in courses given at the Insitut de Phonétique at the Sorbonne under its then director, Hubert Pernot, that 'due to the mere effort of beginning to speak, some stress may be noticed at the beginning of a phrase: "*très souvent*, *au commencement d'une phrase.*"[3] Pernot describes this initial accent in his own work on prosody (Pernot, 1929-30) and states explicitly that it is found in a range of registers: '*ce phénomène est sensible dans la conversation; il l'est beaucoup plus encore dans la lecture, la diction ou quand on parle en public. On pourra s'en assurer en écoutant le premier conférencier venu.*'[4] Two aspects of this claim call seriously into question the assertions of other contemporary phoneticians that early accents were not a feature of the conversational French of this era. One is its timing, six years before Dauzat's (1936) claim that early accentuation was shocking to French ears. The second is the observation by Schwartz (1930) that Pernot urged his students to pay particular attention to the speech of young Parisians; we may thus assume that the speech of this group influenced Pernot's own descriptions heavily. If this group was indeed producing early accentuation regularly in conversation, then Dauzat's claim that 'any child coming to Paris … who attends school' would learn '*français correct*' as he defined it, without such initial accents, loses much of its force.

There is thus no coherent and generally accepted history of the development of early accents in French prosodic units. On the one hand are studies that offer some evidence of diachronic development and of the style-shifting often associated with it (Fónagy, 1989, Lucci, 1983). On the other hand, the case for diachronic change is partly built on traditional descriptions of accentuation which the weight of evidence warns us not to

[3] "… very often at the beginning of a phrase."

[4] "This phenomenon is audible in conversation; it is much more so in reading, elocution, or public speaking. To convince oneself of this, one needs only to listen to the first lecturer who comes along."

accept uncritically as an accurate representation of the spoken French of their time.

4.2 Analysis of early recordings

As part of his 1980 study, Fónagy compared the accentual properties of three political speeches made in 1914-15 to three dating from 1974. Fónagy concludes that the comparison lends weight to the claim of a diachronic development in the role of early accents having taken place. Both Fónagy's experimental method and the precise questions he asks make his findings difficult to assess in the terms of the present article, however. Fónagy's methodology involves perceptual judgements of degree of accentuation, unsupported by the sort of acoustic analyses that are now the common currency of prosodic studies. There is no doubt that for a language in which the position of accent is controversial, native speaker judgements form a valuable source of information. On the other hand, judgements made about a prior state of the language when the purpose of the enquiry is precisely to see if the language has changed in the meantime raise the problem that the participants cannot be assumed to be native speakers of the relevant (perhaps now defunct) variety. The present study therefore focuses on the objectively definable acoustic criterion of fundamental frequency patterning.

Fónagy also considers accentual patterns only as realised in individual words, rather than looking within an AP or IP. His two statistically significant findings concern the proportion of words having a more marked accent on their first than on their last syllable and the number of accented clitics (greater in 1974). Although these findings *could* be evidence for an overall increase in the use of early accent, they are not necessarily conclusive, as there could be a countervailing higher proportion of words with primary final stress but a secondary (as opposed to no) accent on their first syllables in 1914-15.

Some evidence of an overall increase in the use of early accent can indeed be derived from the raw figures Fónagy presents (Fónagy, 1980, table 68); in 1914-15, 27% of words had such an accent, as opposed to 36% in 1974. Thus, already in the recordings from 1914/15 there was, in political discourse, significant use of early accent and this usage appears to have increased by 1974. The latter conclusion can only be tentative; the figures might be skewed, for example, by the relative numbers of disyllabic and polysyllabic words (the former, in phrase final position, not being normally able to bear an accent on their first syllable, see section 3, above) and this information is not given. Furthermore, the same raw figures reveal that the proportion of words with an accent both on the

final syllable and earlier in the word was *greater* in 1915/15 than in 1974 (26% vs. 22%).

As with the literature survey in 4.1, the assessment of Fónagy's study of the three early political speeches leaves a suggestive but unclear picture of the true nature of the usage of early accent in the first decades of the 20th century. The project presented in preliminary form here aims to improve this picture both methodologically, by appealing to modern instrumental phonetic techniques and by adding to the number of genres and to the number of early recordings analysed.

5.0 The recordings

From sections 2 and 3 above, it will be clear that it would be most valuable to have recordings from the early 20th century of spontaneous conversations, preferably between members of those Parisian bourgeois groups alleged by the traditional account not to use early accents. Such recordings do not exist. The '*Archives de la Parole*', founded by Ferdinand Brunot, was highly active in the relevant period in collecting examples of regional forms of speech (*patois*) but preserved what was seen as standard French exclusively in the speech of famous men uttering largely pre-prepared texts. Three of these have been chosen for partial analysis for the current paper. They offer slight differences in style:

1) Alfred Dreyfus reads, not always quite accurately, a section from his memoirs (1912);

2) Ferdinand Brunot's speech (1911) at the opening of the '*Archives de la Parole*' was captured live. Although pre-scripted, his production of it is animated with a degree of theatricality. A recent stylistic study of the recording by Freyermuth and Bonnot (2007) described it as '*jouissant à la fois de la spontanéité de l'oral et de la rigueur d'un écrit très travaillé et construit.*';[5]

3) Emile Durkheim delivers part of a lecture (1913). Again recorded live, this is the text which probably most closely approximates spontaneous speech.

Although these recordings contain formal, rather than conversational speech, they offer a wider range of styles than the political speeches analysed by Fónagy (1980). They offer the possibility of assessing how

[5] 'enjoying both the spontaneity of the spoken word and the rigour of carefully constructed written text'

far early stress was a general feature of public speaking in this period. For each text, a section from the beginning lasting from just under two to two and a half minutes, was chosen, to form a coherent sub-section of the overall recording (actual lengths: Durkheim 1' 54"; Brunot 2' 30"; Dreyfus 1' 56").

6.0 Method

6.1 Parameters

Although durational information probably plays a role in early accent (Astésano et al., 1995), the main parameter has been shown by numerous studies to be f0 movement (see di Cristo, 1999 for an overview), there being an f0 peak followed by a fall on the accented syllable. For the purposes of this preliminary report, only f0 patterns have been investigated, although both duration and intensity information will be examined at a later stage of the project.

6.2 Prosodic phrasing

Each recording was analysed into syllables, words, Accentual Phrases and Intonational Phrases, using auditory information along with waveforms, spectrograms, pitch traces and intensity traces produced using the PRAAT program (Boersma, 2001). Much of the time, the prosodic phrasing follows syntactic structure quite clearly (recall that these were pre-prepared texts). Alongside the syntax, a number of prosodic phenomena indicated boundaries, above all the presence of pauses and phrase-final lengthening. Tonal information was also sometimes used, notably the presence of an AP initial low tone.

6.3 f0 peaks

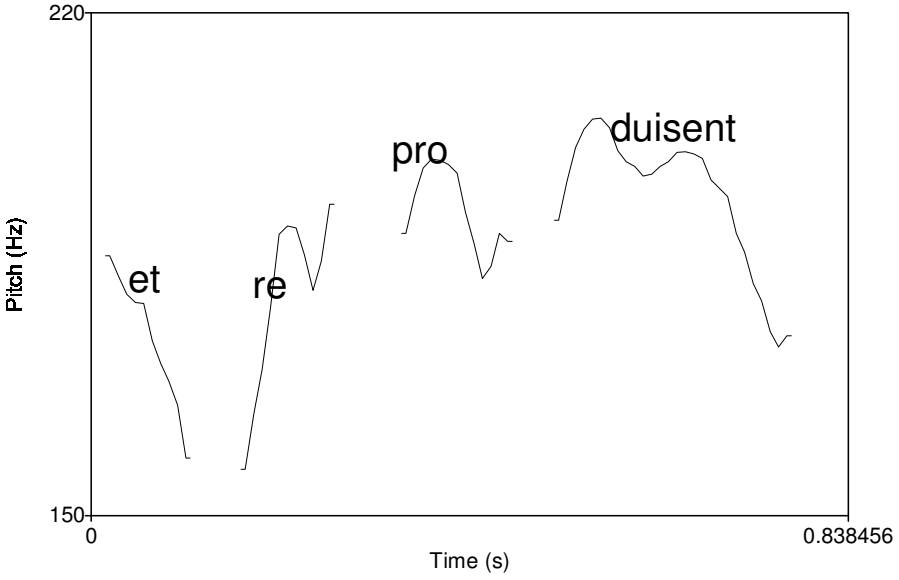
Each AP was then examined for the presence of f0 peaks on early syllables. By 'early syllable' is meant a syllable at least two syllables before the final accented syllable in the AP. Peaks on the final or pre-final syllable were ignored as these typically form part of the phrase-final pitch movement. These early peaks were recorded, along with the lexical nature of the word in which they occurred and the syllable on which they appeared in the case of a polysyllabic word. From this information was worked out; (i) the number of such peaks, expressed as a proportion of all APs; (ii) the number of peaks appearing on lexical words vs. clitics; (iii) on which syllables of polysyllabic words the peaks tend to occur. The question arises as to whether an attempt should be made to distinguish

initial emphatic accents (*accents d'insistance*) from others. It will be recalled that according to the traditional view, all pre-final accents are *accents d'insistance*, whereas according to the modern approach the majority are not and it is not clear that a phonological distinction can be made between the two cases. It was therefore decided to note all early f0 peaks together at this stage of the project, leaving the shape of the peaks, should these be found to be prevalent, to be analysed at a later stage

6.4 Rapid rises

Initial analysis of f0 patterns revealed that a number of APs contained an early rapid rise in f0 which rather than leading to an f0 peak, was followed by a fairly flat f0 pattern before the phrase-final accent. In the majority of these cases, at least half of the overall pitch rise in the AP was concentrated on a single syllable, which therefore had a much steeper f0 slope than either the AP as a whole or any other syllable. Nonetheless, the actual highest frequency in the AP (prior to the final accent) was on a later syllable than that with the steep slope, generally a syllable which itself had a relatively flat f0 pattern and carried no other evidence of being accented (see example below, Figure 1). In a small subset of examples, the rapid rise spanned two very short syllables, these consisting of clitics which cannot normally be accented.

Figure 1: f0 trace for AP 'et reproduisent', showing rapid rise on 're'.



This pattern matches neither the steady rise typically referred to by the traditional analysis, nor the 'accent d'insistance' pattern, nor the peak-dip-peak pattern reported in many modern studies. However, it is in

complementary distribution with the last of these and resembles it more closely than it does the steady rise; it could therefore be an alternative realisation of, or a precursor to the early peak-f0. Cases of the rapid rise without associated peak (henceforth 'rapid rise') were therefore recorded and analysed. The analysis involved determining, as for the early peak measurements (section 5.3, above): the proportion of APs with a rapid rise; the number of rapid rises appearing on lexical words vs. clitics; which syllables of polysyllabic words were marked by the rapid rise. There was also a supplementary statistical analysis (paired *t*-tests) to establish whether the slope of the rapid rise section was significantly different from that from the entire AP, excluding the phrase final accent; the slope was calculated as f0 rise over time for each of these units.

6.5 Even rises

The traditional account suggests that within a rhythmic group, pitch typically rises evenly from the beginning of that group till the nucleus. The number of APs in which this pattern was observed in the three recordings was noted. The AP, rather than the IP, was chosen, because it is the minimal accentual unit examined here and thus corresponds closely to the rhythmic group in the traditional approach. The lexical structure of APs with even rises was examined, to establish whether they contained lexical patterns that, according to the modern approach, would allow for accents other than that on the final syllable. An AP consisting of a single mono- or disyllabic lexical word preceded by clitics would not allow for such accents, as clitics cannot usually be accented (although exceptions to this have been observed, cf. di Cristo 1999) and the proscription of stress clashes prevents the first syllable of the disyllable from being accented, given that this syllable carries the obligatory phrase-final stress.

7.0 Results

7.1 Prosodic structure

The numbers of IPs and APs in each text are summarized in Table 1.

Table 1. Number of IPs and APs per speaker.

Speaker	**IPs**	**APs**
Durkheim	51	115
Brunot	65	163
Dreyfus	47	124

A small proportion of these APs and IPs are entirely falling in pitch, or consist solely of a vocative or short exclamation and are thus uninformative for the present study. Only those showing one of the target patterns are discussed below.

7.2 Early f0 peaks

The number of early f0 peaks observed is given in table 2, both as a raw figure and as a percentage of APs seen to contain one; none of the APs in the sample had more than one peak before the final accent. The percentages are low; overall only 12.1% of APs have the early f0 peak that is considered typical of modern French.

Table 2: Number of early f0 peaks.

Speaker	**No. of f0 peaks**	**No. of APs**	**% of APs with early peak**
Durkheim	19	115	16.5
Brunot	23	163	14.1
Dreyfus	9	124	7.25

The difference between the read text (Dreyfus) and those spoken with a degree of spontaneity is marked, but the proportion of APs with an early peak is very low across all three texts.

7.3 Early Rapid Rises

The figures for these are tabulated in table 3. It will be noted that in contrast to the early peaks, this feature does not differentiate Dreyfus from the others. Rather, Durkheim, the most spontaneous sounding of the speakers, stands out from the others as having a greater number of rapid rises.

Table 3: Number of early rapid rises.

Speaker	**No. of early rapid rises**	**No. of APs**	**% of APs with rapid rise**
Durkheim	32	115	27.8
Brunot	34	163	20.9
Dreyfus	27	124	21.8

Statistical tests were carried out to establish whether the slope of the rapid rise section was significantly different for that from the entire pre-nuclear AP; slopes were calculated for the relevant syllable and the overall prenuclear AP. These were then compared using paired *t*-tests. For all three recordings, these were highly significant (see table 4, below).

Table 4: Paired t-tests for significance of rapid rise feature.

Speaker	***t value***	***df***	***p*** **<**
Durkheim	-6.082	31	.001
Brunot	-6.219	33	.001
Dreyfus	-8.88	26	.001

Table 5: Number of early f0 peaks and rapid rises combined.

Speaker	**No. of early peaks & rapid rises combined**	**No. of APs**	**% of APs with the two features combined**
Durkheim	51	115	44.3
Brunot	57	163	35
Dreyfus	36	124	29.0

The rapid rise is thus shown to be acoustically distinct from the remainder of the AP containing it, although whether the feature is perceptually salient remains to be established by further tests at a later stage of the project.

The rapid rise and the early peak-f0 were by definition mutually exclusive in the APs observed in this study; recall that rapid rises involve the absence of a following pitch fall, the latter being a criterion for the early peak-f0. In other respects, the two features resemble each other. It was hypothesized in section 5.4 that the rapid rise might thus be an alternative form of, or historical precursor to, the early peak-f0. Further work is needed to corroborate this hypothesis, but it is worth noting that if it is correct, then the combined percentage of APs with these features, while low compared to modern norms (see Table 5), is large enough to make it difficult for advocates of the traditional account to explain it away by appealing to the notion of '*accent d'insistance*'. In any case, the rapid rise pattern, which forms the majority of the cases in table 5 does not resemble at all that for the *accent d'insistance*.

7.4 Even rises

The numbers of even rises observed for each recording are listed in table 6. For none of the recordings do more than about a quarter of the APs have this pattern, the number being considerably less for the Brunot recording.

Table 6: Number of even rises: number in parenthesis shows how many could in principle have received a secondary accent.

Speaker	**No. of even rises**	**No. of APs**	**% of APs with even rises**
Durkheim	30 (10)	115	26.1 (8.7)
Brunot	26 (13)	163	16 (8)
Dreyfus	33 (24)	124	26.6 (19.4)

Table 6 also shows, in parenthesis, the number of APs with even accent that could in principle have received an additional early accent, and therefore on which a pitch peak or early rapid rise could have occurred; these are APs containing more than one lexical word, or a single lexical word with at least three syllables, so that a secondary accent could occur without producing a stress clash. For the Durkheim and Brunot texts, the percentage of such APs is in single figures; even for Dreyfus, it is less than one fifth of the total.

8.0 Discussion

The initial hypothesis tested here was that there is no difference between early 20th and early 21st century French with respect to the presence of early accents; this is not supported by the results. There is not a predominant tendency in these readings to have an early f0 peak, corresponding to a pre-nuclear accent, as described in current analyses of modern French. Neither the early f0-peak feature nor the early rapid rise, proposed here as an alternative or precursor form of that peak is found in more than approximately 25% of the APs examined here, and even when combined, these two features are found in considerably fewer than half of the APs. In the current language, in contrast, the presence of a pre-nuclear accent is treated as being the default case.

However, the alternative description, offered by the traditional account and still being proposed more than twenty years after these recordings were made, corresponds still less well to the data. That account predicts a general pattern devoid of pre-nuclear accents of any sort and specifically

a slow even increase in pitch until the nucleus. Only a minority of APs showed this pattern and nearly half of these APs involved combinations of lexical items that do not allow for an early accent.

What these recordings seem to suggest is thus a form of French that is prosodically between the extremes of the modern and traditional descriptions. This form already has the f0 patterns that underlie descriptions of French accentuation today (and thus is phonologically similar to today's variety) but also has a majority of APs without an early rise or f0 peak, and is thus statistically different from that suggested by modern descriptions.

This is not to suggest that the early 20th century French described here literally stands diachronically at some mid-point between forms of the language corresponding to the traditional and modern accounts. The present research leaves open the possibility that early accents have always been a part of the prosodic phonology of the language (cf. Pensom 1998 and section 4.1, above) but have become more common across the 20th century. It also remains possible that when other acoustic parameters are included in the investigation, notably duration, evidence for a ,greater number of early accents will be found. Indeed, this was, impressionistically, the judgement of the author when performing the f0 analyses. In later parts of the present project, these possibilities will be examined using recordings of different styles of speech from later decades.

Thus of the four explanations proposed in section 1 for the disparity between the traditional and modern approaches, three none is shown to be solely valid, but three contribute to the observed disparity between traditional and more recent descriptions of French prosody. There is evidence of a diachronic shift at least in the frequency of the usage of early pitch movements in French. The traditional account of French accentuation does not seem an adequate description of the formal use of the language even at the beginning of the 20th century. However, some of this inadequacy might be explicable through the difference in the definition of 'French' used to support the traditional account and that used in contemporary research in that the recordings examined in this paper could not exactly match that traditional definition.

9.0 Summary and Conclusions

This preliminary report demonstrates that highly educated French speakers of the early 20th century produced spoken French whose prosodic patterns do not match those suggested by traditional accounts of French prosody, and which were being advocated by phoneticians throughout the early decades of the 20th century and which still have

currency in a number of textbooks. The French produced by Brunot, Durkheim and, to a lesser extent, Dreyfus, has few examples of the typical even pitch rise described by the traditional account and rather makes use of f0 patterns associated with an early accent in modern French. Nonetheless, such f0 patterns occur only in a minority of APs; statistically, at least, there is thus a difference between the French of the early 20th and 21st centuries. These differences will be further examined in the remaineder of this project, taking account of a wider range of recordings and of further acoustic correlates of accent.

References

Astesano , C., Di Cristo, A. & Hirst, D. (1995). Discourse-based empirical evidence for a multi-class accent system in French. *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Stockholm) *4*, 630-3.

Boersma, P. (2001) PRAAT, a system for doing phonetics by computer. *Glot International* ***5*** *(9/10)*, 341-345.

Bolinger, D. (1958) A theory of pitch-accent in English. *Word, 14*, 109-49.

Carton, F. (1971) L'accent d'insistance en français contemporain. *Actes du XIIIe Congrès International de Linguistique Romane (Québec)* 205-19.

Cruttenden, A. 1997 *Intonation*. (*2nd ed*.). Cambridge, C.U.P.

Dauzat, A. (1936) Comment on Gill, "Remarques sur l'accent tonique en français contemporain". *Le Français Moderne, 4*, 318-9.

Dauzat, A. & P. Fouché (1935) *Où en sont les études du français (*phonétique et orthographe). Paris: d'Artrey.

Delattre, P. (1966a) A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics, 4***,** 183-198.

Delattre, P. (1966b) Les dix intonations de base du français. *The French Review, 40 (1),* 1-14.

Di Cristo (1998) Intonation in French In D. Hirst & A. di Cristo (Eds.), *Intonation systems* (pp. 195-218). Cambridge: Cambridge University Press,.

Di Cristo, A. (1999) Vers une modélisation de l'accentuation du français: première partie. *Journal of French Language Studies, 9,* 143-179.

Di Cristo, A. & L. Jankowski (1999) Prosodic organisation and phrasing after focus in French. *Proceedings of the International Congress of the Phonetic Sciences 14 (2)*, 1565-1568.

Fónagy, I. (1980) L'accent en français: accent probilitaire. In I. Fónagy & P. Léon (Eds). *L'accent en français contemporain* (pp. 123-233). Studia Phonetica 15.

Fónagy, I. (1989) Le français change de visage? *Revue Romane, 23(2),* 225-54.

Fouché, P. (1933) La pronunciation actuelle du français. *Le Français Moderne, 1*, 43-67.

Fouché, P. (1933-4) L'évolution phonétique du français du XVIe siècle à nos jours. *Le Français Moderne, 1-2*, 217-36.

Fouché, P. (1936) Les diverses sortes de français au point de vue phonétique. *Le Français Moderne, 4*, 199-216.

Freyermuth, S. & J.-F. Bonnot (2007) Ferdinand Brunot entre académisme et innovation: analyse phonostylistique et rhétorique du *Discours d'inauguration des Archives de la parole.* In *Colloque international: Le français parlé des médias,* Stokholms Universitet, 203-219.

Gill, A. (1936) Remarques sur l'accent tonique en français contemporain. *Le Français Moderne, 4,* 311-18.

Grammont, M. (1913) *Traité pratique de pronunciation français.* Paris: Delagrave.

Grammont, M. (1963) *Traité de phonétique.* Paris: Delagrave.

Gussenhoven, C. (2004) *Phonology of tone and intonation.* Cambridge: Cambridge University Press.

Jankowski, L, C. Astésano & A. Di Cristo (1999) The initial rhythmic accent in French: acoustic data and perceptual investigation. *Proceedings of the International Congress of the Phonetic Sciences, 1***,** 257-260. San Francisco.

Jun, S.-A.& C. Fougeron (2000) A Phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 209-242). Dordrecht: Kluwer.

Ladd, D.R. (1996) *Intonational Phonology.* Cambridge: Cambridge University Press.

Lehiste, I. (1970) *Suprasegmentals.* Boston: MIT Press.

Lote, G. (1912) La déclamation du vers français à la fin du XVII^{e} siècle. *Revue de Phonétique, 2*, 313-364.

Lucci, V. (1983) *Etude phonétique du français contemporain à travers la variation situationnelle.* Publications de l'Université de Grenoble.

Lyche, C. & F. Girard (1995) Le mot retrouvé. *Lingua,* 95 *(1-3,:* 205-21.

Marouzeau, J. (1924) Accent affectif et accent intellectual. *Bulletin de la Société de Linguistique de Paris*, *XXV*, 79-86.

Meyer-Lübke, W. (1890) *Grammatik der romanischen Sprachen, 1: Lautlehre.* Leipzig: Riesland.

Passy, P. (1890) *Etudes sur les changements phonétiques.* Paris:Didot.

Pensom, R. (1993) Accent and metre in French. *French Language Studies, 3*, 19-37.

Pensom, R. (1998) *Accent and metre in French: a theory of the relation between linguistic accent and metrical practice in French, 1100-1900.* Bern: Peter Lang.

Pernot, H. (1929-30) L'Intonation. *Revue de Phonétique, 6*, 273-289.

Pierrehumbert, J. (1980) The phonetics and phonology of English intonation. Ph.D. thesis, MIT.

Post, B. M. (2000) *Tonal and Phrasal Structures in French Intonation.* Ph.D. Thesis. The Hague: Thesus.

Pulgram, E. (1965) Prosodic systems: French. *Lingua, 13*: 125-144.

Pulgram, E. (1967) Trends and predictions. *In Honor of Roman Jakobson*: 1641.

Pulgram, E. (1970) *Syllable, Word, Nexus, Cursus.* La Haye: Mouton.

Rossi, M. (1980) Le français, langue sans accent? In *L'accent en français contemporain,* 12-51. Ottawa: Didier.

Schuchardt, (1880) Revue dritique de Windisch: Irische Grammatik. *Zeitschrift für Romanische Philologie, 4***,** 124-155.

Scherk (1912) *Über den französischen Akzent.* Doctoral dissertation, Berlin: Schmersow, Kirchhain.

Schwartz, W. (1930) The Parisian accent according to the Institut de Phonétique. *The French Review 4 (3)*, 233-242.

Verluyten, P. (1982) *Recherches sur la prosodie et la métrique du français.* Doctoral Thesis, Antwerpen Univeristeit.