

## Prosody and its application to forensic linguistics

Michael J. Harris

Department of Spanish and Portuguese, University of California, Santa Barbara

Stefan Th. Gries

Department of Linguistics, University of California, Santa Barbara

Viola G. Miglio

Department of Spanish and Portuguese, University of California, Santa Barbara

### Abstract

This article describes three studies in prosody and their potential application to the field of forensic linguistics. It begins with a brief introduction to prosody. It then proceeds to describe Miglio, Gries, & Harris (2014), a comparison of prosodic coding of new information by bilingual Spanish-English speakers and monolingual Spanish speakers. A description of Harris & Gries (2011) follows. This study compares the vowel duration variability of bilingual Spanish-English speakers and monolingual Spanish speakers, and touches upon corpus-based frequency effects and differences in linguistic aptitude between the two speaker groups. Finally, a portion of an ongoing study is described (Harris *in preparation*). This section describes the use of prosodic variables and ensemble methods (or methods that use multiple learning algorithms) to classify languages, even in the case of impoverished data. All three experiments have implications and applications to the field of forensic linguistics, which are touched upon in each respective section and discussed more extensively in the final section of this article. Furthermore, the applications of these methods to forensic linguistics are discussed in light of best practices for forensic linguistics, as outlined in Chaski (2013).

### Overview

Recognizing a speaker's dialect, gender, pathological conditions, native language, or socio-cultural background are skills acquired through training in different areas of linguistics, such as dialectology, acoustic/ articulatory phonetics, and sociolinguistics. They are also skills that are applicable to the field of the forensic sciences, specifically what has



Articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by the University Library System, University of Pittsburgh as part of its D-Scribe Digital Publishing Program and is cosponsored by the University of Pittsburgh Press.

come to be known as forensic linguistics, which is defined as 'an umbrella term referring to research and practice in all those areas where legal and linguistic interests converge ... it is concerned with the role, shape, and evidential value of language in legal and forensic settings.'<sup>1</sup> While the value of such skills can easily be gleaned in both criminal and civil cases, from voice identification to asylum requests,<sup>2</sup> it is not entirely uncontroversial who can be an expert witness in such cases and what methodology can be used that would comply with the modern scientific standards of objectivity and replicability. Relying on expert linguists, i.e. individuals who know the dialects/languages (near-)natively, have acceptable linguistic credentials, or have superior skills in discourse analysis has been an accepted practice in court both in the USA and Europe. However, dialectal differences are being eroded by exposure to the mass media world-wide, and as a result the evidence based on them is less reliable in court; moreover, such finely-honed expertise is very time-consuming and experts are few and far between, especially if they need to know exotic or non-widely spoken languages. Even more relevant is the criticism raised by Chaski over a decade ago that relying on this type of expertise would violate the Daubert standard of empirical reliability, because techniques so heavily centered on the individual linguistic expert do not provide falsifiable criteria, known error rate, or even standard operating procedures to perform the analysis technique (2001:1351). Although relying on the individual's expertise is still accepted in legal practice, specialized journals lay more and more emphasis on experimental techniques, with known error rates and falsifiable methods, in order to establish standards to be followed in forensic linguistics, such as which phonetic parameters are important for speaker identification and what statistic method best isolates them (Jessen 2007, Becker et al. 2008), or how reliable is the human ear in distinguishing speech from non-speech sounds.<sup>3</sup>

Jessen (2007) makes a clear argument for the use of acoustic analysis to complement auditory analysis for forensic speaker classification, and Becker et al. (2008) provides statistic models of formant features that represent the vocal tract characteristics of speakers, and are therefore capable of accounting for between-speaker and within-speaker variability. In the same vein as these studies, this article reviews the use of statistically-informed empirical methods used to analyze the features of spoken speech. We would like to go even further, however, and move beyond the acoustic phonetic analysis of speakers' voice characteristics. The novelty of this paper, in fact, lies in applying empirical methods to speakers' characteristics associated with the grammar of their dialect/language, in these cases different aspects of pronunciation that can be theoretically informed and empirically verified via scientific methodology. This is to say that the methodologies described in the current paper investigate acoustic properties beyond the phonemes, or the 'baseline acoustic system' to use Hansen, Slyh, & Anderson's (2004) term. Instead, it uses high-level phonological cues to successfully profile speakers; high-level phonological cues are those associated with the manner in which a word is pronounced, such as volume or melodic tone. Past studies have considered prosodic cues in speech recognition. Lea (1973, 1976) advocates prosodic data for speech recognition algorithms. Lea (1973) suggests that stressed syllables contain more salient information for speech recognition systems as they tend not to be reduced and are pronounced with more volume (as compared to unstressed syllables) and Lea (1976) suggests that intonational curves can be used to automatically determine prosodic units, which are helpful in the parsing of information by speech recognition technology. However, the results of more recent research using high-level acoustic

---

<sup>1</sup> Taken from the webpage of the International Summer School of Forensic Linguistic Analysis, organized and taught by Malcolm Coulthard, founder of the International Association of Forensic Linguists <<http://www.iafl.org>>.

<sup>2</sup> See for instance the applications discussed in Peter Tiersma's notes on forensic linguistics (<http://www.languageandlaw.org/FORENSIC.HTM> (accessed 08/31/14)).

<sup>3</sup> Experimental topics such as these are in fact becoming increasingly common in the *International Journal of Speech, Language and the Law*, the journal published by the International Association of Forensic Linguists and the International Association for Forensic Phonetics and Acoustics. <http://www.iafl.org/journal.php> (08.31.14).

cues for speaker are associated with “quite remarkable error rates” (Hansen, Slyh, & Anderson 2004:1). In comparison, the data described in the current paper are empirically verified as significant predictors of the dependent variable according to accepted statistical methodology. This is important, as forensic computational linguistics “is oriented primarily to research- and empirically-driven protocols...” (Chaski 2013:335). This is to say that the current paper shows that we can go beyond the acoustics of the voice and tap information related to different languages or dialects of the same language, without sacrificing the use of those scientific methods that guarantee the replicability of results, while at the same time minimizing dependence on an individual researcher's (or witness's) expertise.

This article explores how prosody informs us about a speaker's native language or dialect. The application of these experiments to the field of forensic linguistics will be explained through the exemplification of modern prosodic research and how it can be applied to distinguish bilingual versus monolingual speech and to discriminate between languages even on the basis of a diminished or impoverished speech signal. Prosodic research is applicable to forensic linguistics because it can be applied in speaker or language identification and linguistic profiling (for instance assessing a speaker's dialect or native language). Identification of speaker, language, and author is one of the four categories of answers provided by forensic linguistics in law enforcement investigation and/or legal settings, and linguistic profiling is another (Chaski 2013). In a broad linguistic sense, prosody refers to rhythm, stress, and intonation in language. Stepping away from the linguistic meaning of these elements, it is equally important to note that they contribute to what is commonly referred to as *accent* (the difference between, say, a native, regional or foreign accent, rather than the linguistic meaning, which is akin to stress). While these prosodic elements are certainly not the only phonetic elements that contribute to a perceived difference in accents (vowel quality, for instance has been shown to vary between dialects of the same language, see for English Fox & Jacewicz 2009), they do contribute to the audible differences between two dialects of the same language, once again what is commonly known as an individual's accent.

Below, we show how prosodic elements allow the distinction between two dialects of the same language, as well as the distinction between three different languages even in the case of an impoverished (or distorted) speech signal. Furthermore, rather than relying upon anecdotal evidence, these results will be empirically verified via statistical modeling, providing an empirically-founded perspective that is also more informative due to the methodologies' precision and ability to account for interactions between two variables. ‘Interactions between two variables’ refers to cases where a dependent variable (or the phenomenon being investigated) behaves in a particular way in the presence of factors X and Y that is different/unpredictable from the behavior of the dependent variable in the presence of just either X or Y. For instance, if women spoke faster than men and young people spoke faster than older people, but at the same time younger women spoke slowest (rather than fastest), one would consider this an interaction of the independent variables (‘causes’) Sex (*male vs. female*) and Age (*young vs. old*).

The remainder of this article will be structured as follows: a brief introduction to prosody and some current methods of quantifying prosodic features are presented. This section is followed by an overview of three different studies that have used prosodic features for dialect or language distinction. The first of these sections reviews Miglio, Gries, & Harris (2014) and will discuss intonation as it varies between monolingual Mexican Spanish speakers and bilingual American Spanish-English speakers, when used to mark new information in the speech signal. The second will describe portions of Harris & Gries (2011); specifically it will discuss how vowel duration variability differs in the Spanish of monolingual vs. bilingual speakers; it is generally accepted that vowel duration variability is related to differences in speech rhythms (as in Low & Grabe 1995), and it is thus related to the following, third section of this paper. The third section, in fact, will discuss a portion of Harris (*in preparation*), a study that uses a combination of prosody-based variables to distinguish between utterances of Spanish, English, and Portuguese sentences, even after the quality of these recordings has been so degraded as to make them indistinguishable (in terms of language) to a trained phonetician's ear. These three sections are followed by a general conclusion. Once again, the ability to use linguistic features, and specifically prosody, to distinguish between a bilingual and a monolingual speaker has

potentially very important applications in terms of speaker identification; finally, the practical application of this research to the field of forensic linguistics will be discussed in the conclusion of this paper.

## Introduction

Speaker recognition largely based upon basic acoustic features of a speaker's voice has achieved some considerable success in the past; for instance, machine-learning algorithms have been successfully trained to match recordings of a speaker with other recordings of the same speaker (e.g. Schötz 2002). The studies described in the current paper are novel in investigating prosody via modern computational methodology. As mentioned above, prosody is a facet of phonology that is concerned with rhythm, stress, and intonation. This is important because these properties differ across languages (and, at times, dialects), and the nature of the research described is applicable to best practices for forensic linguistics (discussed in the conclusion to this paper). Prosody is often described as concerning itself with the so-called *suprasegmental* features of the speech signal; this is to say that rather than be confined to a specific segment such as a syllable, it can span across several segments. This distinction does not always prove easy to define, however;<sup>4</sup> rhythm is often measured by the duration of vowels, consonants, or syllables, for instance. In this paper, we will simply define prosody as including the aforementioned rhythm, stress, and intonation, and any measurements or metrics used to quantify these properties.

Before defining the meaning of rhythm, stress, and intonation, it is worthwhile to discuss some of the uses of prosody in terms of communication. That is, apart from linguistic definitions of prosody, what do speakers use prosody for? Prosody is in fact used for many aspects of human communication. It can indicate the mental or emotional state of a speaker, the meaning of a sentence (e.g., a prosodic representation of irony can in fact invert the truth value of the literal interpretation of an utterance), the importance or novelty of a concept or character within a narrative, emphasis or contrast in a phrase, etc. This is to highlight that the import and utility of prosody in the linguistic signal is not negligible. Furthermore, prosody is particularly appealing for use in the forensic sciences for several reasons concerning the identification of a speaker's linguistic or dialectal characteristic features that will be outlined below. However, it is first advisable to define and discuss three main facets of prosody.

The first facet of prosody to be discussed is intonation. *Intonation* refers to changes in pitch (or melodic tone) across a phrase. In tone languages, pitch has grammatical or lexical meaning. While languages such as English or Spanish are not tone languages, this does not mean that intonation does not play an important role in their use. Pitch plays a discursive role, helping to distinguish elements of a conversation and indicate their purpose in discourse or to clarify the meaning of phrases. For example, in English (as indeed, in many languages), rising pitch indicates an interrogative sentence, rather than a declarative sentence. Intonation also plays an important role by focusing the listener's attention on new or novel information via a rising pitch in a number of languages, including English (Vallduví 1992).

*Rhythm* is concerned with the relative variability of segment durations between languages or dialects. That is, some languages sound like they are spoken with a more steady or predictable rhythm while others seem to be spoken with more variation in rhythm. Linguists originally hypothesized that languages were spoken with syllable-timed or stress-timed rhythm (e.g. Pike 1945). In stress-timed speech, syllable (and vowel) durations vary according to the placement of stress and seem to be less uniform throughout the entire phrase. Dutch and English are traditionally classified as

---

<sup>4</sup> The exact nature of those phonological elements that define the units of prosody, while not completely controversial, can be subject to debate. For differing prosodic constituents, for instance, see Selkirk (1978) and Beckman and Ayers Alam (1997).

stress-timed languages. In syllable-timed speech, syllable (and vowel) durations are relatively uniform within a phrase. According to traditional speech rhythm typology, French and Spanish are syllable-timed. This distinction was originally described as a dichotomy, with all languages being characterized by either one or the other rhythm. However, the identifying of ‘intermediate’ rhythm languages, such as Catalan, which has some characteristics of a stress-timed languages and others of syllable-timed languages, indicates that speech rhythms should be conceived of as existing on a continuum between the two extreme poles of stress-timing and syllable-timing (Dauer 1987).

*Stress* refers to the placement of emphasis on a syllable within a word or phrase. The most common example of stress is *lexical stress*, or the prominence of one syllable within a word; this would be the difference between the verb (*to record* (emphasis on the second syllable, *recórd*) vs. the noun *record* (emphasis on the first syllable, *récord*). As evidenced by the preceding example, English has phonemic stress; this means that there are some words that are pronounced with identical sounds and only differ in the location of the main emphasis. Stress can be also be used to give prominence to one word within a sentence or phrase; this is commonly known as *prosodic stress*. Prosodic stress is used to indicate contrast or emphasis in a phrase. While correlates differ across languages, it is generally accepted that languages use some combination of intensity, pitch, and duration to indicate stress. This is to say that the syllable bearing stress in a word will be pronounced more loudly, and/or have higher pitch (or a higher/ rising melodic tone), and/or have relatively longer segment durations (Fry 1958).

Having defined prosody, it is also important to note its importance to speaker identification, dialect distinctions, and the forensic sciences in general. Although prosody is not the only facet of phonetics/ phonology important to speaker identification, several facts about prosody suggest that it may be important and/or beneficial when dealing with speaker or accent identification. Firstly, it appears that some prosodic elements are acquired at a very young age. In the case of rhythm, various studies have shown that neonates and infants can distinguish between languages of two different rhythm classes, but not necessarily between two languages of the same rhythm class (e.g. Nazzi, Bertoni, & Mehler 1998), suggesting that even newborns are sensitive to rhythmic typology. Furthermore, some prosodic elements do not appear to vary greatly after their acquisition, even if the speaker learns a second language: Harris & Gries (2011) show that vowel duration variability between monolingual Spanish speakers and bilingual Spanish-English speakers do not vary in high frequency words. Prosodic patterns also appear difficult to mimic: Wretling & Eriksson show that speech impersonators could alter spectral characteristics, but less so timing at the phoneme and word levels in imitating speech, suggesting that rhythm is somehow “hard coded” in the adult speaker (1998:1).

In addition to these practical advantages of the use of prosodic cues for use in the forensic sciences, it is worthwhile to add that many prosodic metrics are relatively uncomplicated to measure; they are often unambiguous, and, in some cases, the process can be automated (e.g. Loukina *et al.* 2009), effectively helping to remove researcher bias. The analyses in the current paper take advantage of two linguistics tools that can be downloaded free: PRAAT (Boersma & Weenink 2010) and the R programming language for statistics (R Core Development Team 2013). PRAAT is a program for phonetic analysis that provides a visual representation of recorded speech. This allows the measurement of various phonetic elements. In the context of this article, this software was used to measure cues such as pitch (F0), intensity, and segment duration. These phonetic cues were then analyzed using R. R is an open-source programming language that allows for great flexibility in the analysis of data and the graphical representation of these results. This article highlights three different analyses performed upon prosodic data, which also demonstrates the capabilities of R in handling varied data sets. Thus, the three subsections that follow will present mixed effects modeling, multifactorial linear regression, and an ensemble method called random forest. Following these three sections, a general conclusion and discussion will further expand the applicability of this research for the forensic science community.

## Prosodic Analyses

### Intonation<sup>5</sup>

In conversation, a speaker uses cues to signal new information to the listener, making the parsing of information easier. These cues can be both prosodic and syntactic, and the cues to signal new information vary cross-linguistically. Specifically, so-called *plastic languages* such as Dutch or English (Vallduví 1992) tend to use changes in pitch to signal new information, while *non-plastic languages*, such as Spanish or Italian (Zubizarreta & Nava 2010; Swerts, Kraemer, & Avesani 2002) use word order instead. That is, where word order can vary more freely in a language (as in Spanish, for instance), a change of syntactic structure is often used to draw the listener's focus to information that is new to a conversation. Meanwhile, speakers of languages, which have more fixed word order such as English, use prosodic information, specifically changes in pitch, to signal new information (see Figure 1). The current study concerns itself with how different speakers (Spanish monolinguals vs. English-Spanish bilingual speakers) use prosodic cues to signal that information is novel to the discourse (*new information*) as compared to information that has already been introduced to the conversation (*given information*).

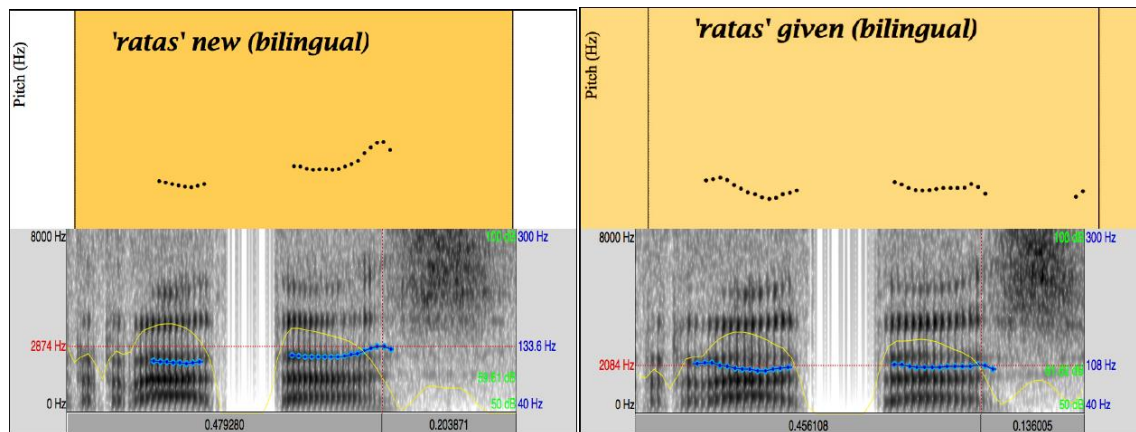


Figure 1: Pitch contours of naturalistic bilingual Spanish speaker from California; the left panel displays new information while the right panel is given information.

Figure 1 shows that speakers do use pitch to signal new information in naturalistic speech. In the left panel, the word *ratas* ('rats') is pronounced with a rising pitch contour. This is the first time it is mentioned in the narrative. The right hand panel shows the second mention of the same word; here notice that the pitch is flat<sup>6</sup>.

Miglio, Gries, & Harris (2014) investigate the differences between Mexican monolingual Spanish speakers and Chicano bilingual Spanish speakers in the signaling of new and given information; the authors are particularly

<sup>5</sup> This is based on Miglio, Gries, & Harris (2014).

<sup>6</sup> Notice that Figure 1 represents the pitch contours of a bilingual Spanish speaker from California. However, even in monolingual Spanish speakers there may be changes in pitch contour to signal new or given information. Just because speakers of a non-plastic language (such as Spanish) can use word order to signal new information, this does not mean that they do so exclusively, neglecting prosodic cues completely.

interested in their use of pitch to signal new information. As mentioned above, monolingual Spanish speakers would be expected to employ word order more than prosodic cues in the signaling of new information, according to the traditional plastic versus non-plastic language typology. Of course, the bilingual speakers mentioned above also speak a plastic language, English, in addition to Spanish. Thus, the authors hypothesize that bilingual speakers will employ more prosodic strategies for new information as compared to monolingual Spanish speakers – i.e. their knowledge/use of English (a plastic language) makes their Spanish display more pitch movement for new information.

## **Variables**

In order to compare the use of prosodic cues in the signaling of new information, Miglio, Gries, & Harris (2014) measured and recorded the following variables:

Dependent Variable:

- PITCH\_MOVEMENT: if speaker displayed pitch movement across the word *yes* vs. *no*.

Independent variables / Fixed Effects:

- SPEAKER\_TYPE: *monolingual* vs. *bilingual*;
- SPEAKER\_SEX: *male* vs. *female*;
- NEW\_GIVEN: if information is *new* vs. *given*;
- DURATION\_LOG: the log of the length of the stressed vowel in ms;
- PHRASE\_FINAL: if the word is in final position in the IU, *yes* vs. *no*;
- INTENSITY: peak intensity of the word in dB.

Random Effects (i.e., variables that address the fact that data points by specific speakers and for specific words are interrelated, which allows one to filter out idiosyncrasies; cf. below):

- INFORMANT: makes adjustments to intercepts according to the speaker to account for speaker-specific variation;
- WORD: makes adjustments to intercepts according to each word to account for word-specific variation.

## **Statistical Analysis**

A linear mixed-effects model selection process was performed using R (R Development Core Team 2013). The notion of ‘model selection’ means that a (within reasonable limits) maximally complex model was generated including all the fixed effects and their two-way interactions, from which interactions and then main effects were eliminated if they did not significantly improve the performance of the model. This process essentially evaluates if the variables substantially account for variability of the dependent variable, PITCH\_MOVEMENT or not. When all non-significant predictors have been eliminated, the resulting minimal adequate model contains only those variables and interactions that significantly contribute to the model.

As characterized above, *random effects* are effects that often do not contribute much to the interpretation of the results but whose variation must not be accounted for by the fixed effects in the model because the fact that there are multiple

data observations of the dependent variable for each speaker and each word violates the assumption of independence of data points that underlies the significance tests of the fixed effects. Since a certain speaker may favor pitch excursions regardless of information structure or since certain words may be more commonly pronounced with pitch excursions than others, the random-effects structure include an adjustment along the intercept of the model for each speaker and each word. These adjustments improve the performance of the model and make the results more representative of a wider population.

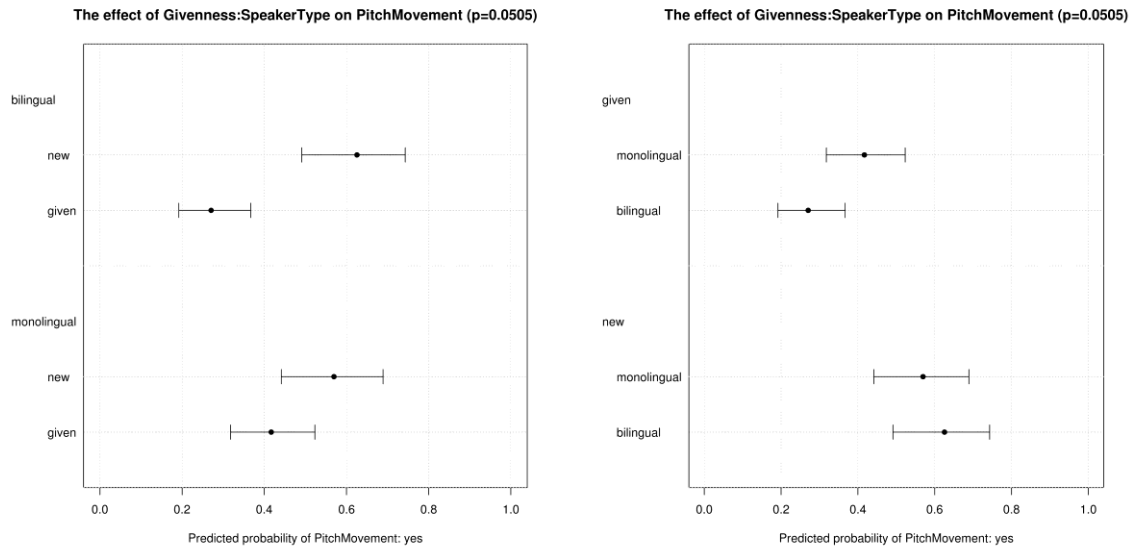


Figure 2: Interaction NEW\_GIVEN : SPEAKER\_TYPE. Both panels display alternate views of the same interaction.

## Results and discussion

Miglio, Gries, & Harris (2014) show that Chicano bilingual and Mexican monolingual Spanish speakers both use pitch movement to signal new information. Specifically, their analysis returned a marginally significant two-tailed  $p$ -value of 0.0505 for the interaction NEW\_GIVEN : SPEAKER\_TYPE; since their hypothesis about the Chicano speakers was directional/one-tailed, however, this results reflects a significant one-tailed  $p$ -value of 0.02503 and shows that Chicano speakers use pitch movement more often as compared to Mexican monolingual speakers. Thus, the Spanish of the Spanish-English bilingual participants is more plastic as compared to that of the monolingual participants, which suggests that the English prosodic system influences the Spanish prosodic system of bilingual speakers. The ability to identify differences between bilingual and monolingual prosodic systems presents a promising avenue for speaker profiling or speaker identification.

## Vowel duration

There have been several attempts to empirically quantify differences in speech rhythms between dialects or languages. One phonetic feature that has been suggested as a major contributor to perceived rhythmic variation is vowel duration. More specifically, vowel duration variability has been measured with metrics in an attempt to distinguish between rhythm classes (e.g. Low & Grabe 1995). As mentioned above, the distinction between stress-timed and syllable-timed speech hinges on the fact that in stress-timed languages, syllable (and vowel) duration theoretically vary according to the placement of stress and is less uniform throughout the entire phrase. Meanwhile, in syllable-timed languages, syllable (and vowel) durations are theoretically more uniform within a phrase.



Harris & Gries (2011) compare the vowel duration variability in the Spanish of bilingual Spanish-English speakers and monolingual Mexican Spanish speakers. This proves an interesting comparison, since the monolingual speakers speak a syllable-timed language, while the bilingual speakers also speak a stress-timed language; thus, this comparison affords a perspective of a prosodic system in bilingual speech. Harris & Gries (2011) hypothesize that bilingual speakers would exhibit a more stress-timed (or English-like) rhythm due to their bilingualism as compared to monolingual speakers. Furthermore, the metrics utilized also provide valuable insight into the way that speech rhythms and other suprasegmental features may be affected by differences in linguistic abilities; that is, prosodic features appear to vary depending on which of the two languages is dominant for any individual bilingual speaker. This approach is also novel in the inclusion of corpus-based frequency effects. As Harris & Gries note: "... it seems plausible, and even likely that duration units are affected by word frequencies" (2011:5).

### **Variables**

For the sake of space, only those variables relevant to the current discussion will be listed:

Dependent Variable:

- SPEAKER\_TYPE: *monolingual* vs. *bilingual*;

Independent variables:

- SD\_LOG: the log of the standard deviation of  $\text{duration}_{\text{syllable } x}$  and  $\text{duration}_{\text{syllable } x+1}$  plus 1 within the IU;
- VAR\_COEFF\_LOG: the log of the variation coefficient of  $\text{duration}_{\text{syllable } x}$  and  $\text{duration}_{\text{syllable } x+1}$  plus 1 within the IU;<sup>7</sup>
- LEMMA\_FREQ: the log of the frequency of the lemma in which the segment occurred in the Corpus del Español (Davies 2002-).

### **Statistical analysis**

Harris & Gries (2011) use an automatic stepwise bidirectional logistic regression model selection process, trying to predict SPEAKER\_TYPE: monolingual using the R functions `glm` and `stepAIC` (R Development Core Team 2013 and Ripley 2011). This is to say that the model selection process was automated, removing and adding predictors and two-way interactions in order to obtain a parsimonious model that predicted the dependent variable, namely monolingual versus bilingual speaker and could not be significantly improved by the addition or subtraction of another predictor.

---

<sup>7</sup> The variation coefficient of a vector of numbers  $x_1, 2, \dots, n-1, n$  is defined as the standard deviation of that vector divided by its mean.

## Results

Figure 3 represents the effect of VAR\_COEFF\_LOG: on the  $x$ -axis, we represent VAR\_COEFF\_LOG, on the  $y$ -axis we represent the predicted probability of the speaker being monolingual, and the point cloud is summarized by a locally-weighted smoother. The plot shows that the overall trend of VAR\_COEFF\_LOG is positively correlated with the prediction of monolingual. That is, the overall slope of this predictor is the opposite of what was originally hypothesized. However, it is important to view the implications of this non-linear trend. The prediction is most strongly 'bilingual' in the small range of intermediate variability while the extreme ranges of variability largely lead to the prediction of 'monolingual'. This suggests that monolingual speakers are able to employ a full range of variable vowel durations, whereas bilingual speakers tend to display an intermediate level of variability.

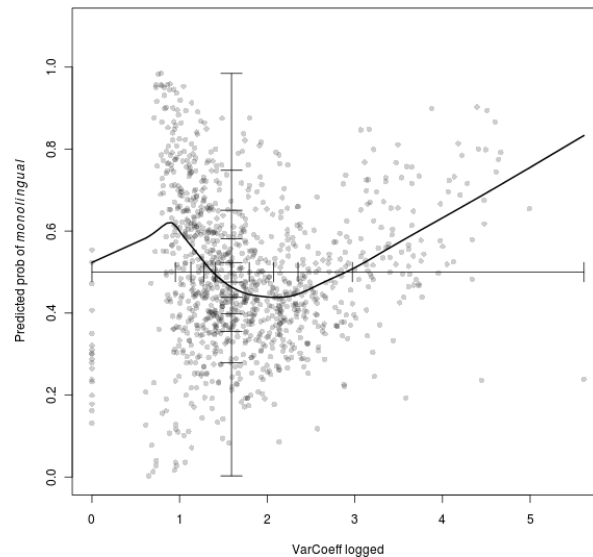


Figure 3: The main effect VAR\_COEFF\_LOG.

Figure 4 represents the interaction SD\_LOG : LEMMA\_FREQ. It shows that with high-frequency lemmas, the variability values of monolingual and bilingual speakers do not differ (because on the right sides of the two plots, the regression lines are at similar SD\_LOG values (of around 3). However, with words whose lemma frequency is below 9, monolingual speakers have higher SD\_LOG values (because on the left sides of the two plots, the regression line in the left panel is below the one in the right panel). That is, with frequent words, bilingual speakers behave like monolingual ones, but with less frequent words, bilingual speakers' vowel durations are less homogeneous. This may be explained by linguistic aptitude. The bilinguals' lesser proficiency in their less dominant language manifests itself in slower speech and less heterogeneous production in those low-frequency words to which they are less exposed.

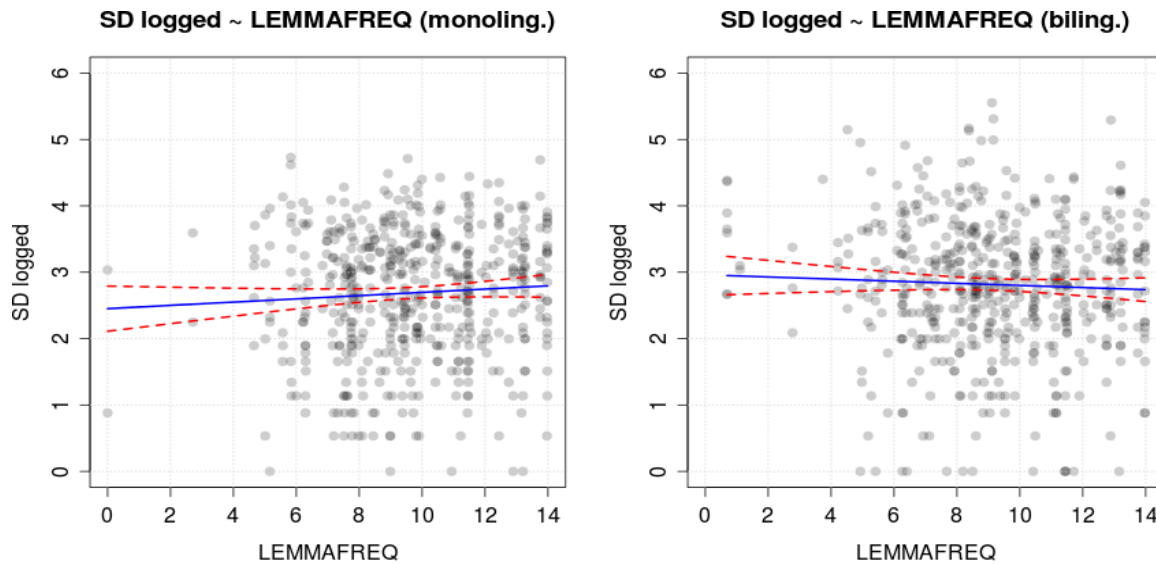


Figure 4: the interaction SD\_LOG : LEMMA\_FREQ. The left panel represents *monolingual* speakers and the right panel represents *bilingual* speakers.

## Discussion

Several variables shed light on the varying linguistic aptitudes of monolingual and bilingual speakers. VAR\_COEFF\_LOG suggests that monolingual speakers are able to employ a far greater range of vowel duration variability, ranging from extremely similar pairs of syllables in terms of vowel durations to extremely dissimilar pairs. Meanwhile, bilingual speakers seem to display a much flatter range in their vowel duration variability, grouping around the mean of vowel duration variability in terms of VAR\_COEFF\_LOG. The interaction LEMMA\_FREQ : SD\_LOG suggests that bilingual speakers display and more heterogeneous production of words, particularly in words that have low frequencies. Vowel duration variability metrics provide interesting insights into prosody, and allow a partial distinction in speech rhythm typologies. However, these metrics also provide a unique perspective upon many other facets of monolingual and bilingual speech, particularly when corpus-based frequency effects are taken into account.

In terms of forensic linguistics, these data provide an interesting perspective on the influence of a first language on a second language, and especially on a speakers' linguistic aptitude manifested in terms of rhythm. The potential application of these methods to dialect identification or speaker profiling are certainly promising. One particularly appealing factor is that these metrics are based largely on vowel durations. In terms of prosodic factors, vowel durations prove relatively simple to measure and there is well-established methodology for their measurement (e.g. White & Mattys 2001). Furthermore, their measurement can be automated with Praat scripts (Boersma & Weenink 2010) or the methodology of Loukina, Kochanski, Shih, Keane, & Watson (2009), who use algorithms to automatically segment speech into vocalic and intervocalic units.

## A prosodic approach to language discrimination

The data in the current section comprise a portion of an experiment designed to investigate what prosodic features contribute to perceived differences between three languages; this study used the (reportedly) syllable-timed English, (reportedly) stress-timed Spanish (see above), and (reportedly) intermediately timed Portuguese (e.g. Frota & Vigário 2001). A variety of variables were coded in order to differentiate between these three languages; variables related to

syllabic rhythm and non-rhythmic variables were included. Certain variables were coded within two different prosodic constituents: the *intonational phrase*<sup>8</sup> and the *phonological phrase*<sup>9</sup>. Essentially these are two different units that are smaller than a sentence yet larger than a word, with *intonational phrase* being larger than *phonological phrase*. These units were determined according to the prosodic characteristics that indicate a pause or break in an utterance (e.g. Beckman & Ayers Alam 1997 & Pierrehumbert 1980).

It is important to note that the variables coded in this section are based largely upon an impoverished speech signal (although it was verified with the full speech signal in order to avoid coding errors). The reason for this is that these data form part of a larger study examining the perception as well as the production of speech rhythms. For this reason, utterances of English, Portuguese, and Spanish were low-pass filtered at 450 Hz; this process mutes all sonic energy below 450 Hz, which removes non-syllabic information from the utterances (Arvaniti 2012). This is to say that these variables were derived from a speech signal that was greatly impoverished, providing implications as to the efficacy of these methodologies even in the case of less-than-ideal raw data.<sup>10</sup> In this perception experiment, participants rated the similarity of the 6 utterances; a single speaker of each language provided two utterances. The two Portuguese utterances were the only utterances consistently rated as maximally similar by participants, thus they were combined into a single level: *Portuguese*. Meanwhile, the remaining 4 utterances are *English 1*, *English 2*, *Spanish 1*, and *Spanish 2*.

## Variables

The following variables were coded in order to determine their importance in distinguishing between the utterances of the various languages:

Dependent variable:

- UTTERANCE: a categorical variable specifying the utterance from which the metrics were derived:  
*English 1*, *English 2*, *Portuguese*, *Spanish 1*, *Spanish 2*

Independent variables / Fixed effects:

- Vowel duration variables: DURATION\_V (the length of the vowel in ms); SDV\_INT\_PH and SDV\_PHON\_PH (the standard deviation of the vowel durations within the *intonational phrase* and the

---

<sup>8</sup> Equivalent to a break index of 4 in the ToBI system, or a full intonational phrase (e.g. Beckman & Ayers Alam 1997). This is represented by an edge tone, to use the AM framework terminology (e.g. Pierrehumbert 1980).

<sup>9</sup> Equivalent to a break index of 3 in the break tier of the ToBI system (e.g. Beckman & Ayers Alam 1997), or an intermediate (intonation) phrase.

<sup>10</sup> Note again that the variable coding was verified using the full (unaltered) speech signal; this is due to the fact that the main purpose of the study was the evaluation of speech rhythms, not the ability to distinguish language with impoverished speech signals. Thus, the current results do not purport to be conclusive evidence, but only preliminary evidence of language distinction with sub-optimal data.

*phonological phrase*); PVIV (the PVI of the duration of the current and the next vowel within the IU (if there was one), computed as in (1));

- Syllable duration variables: DURATION\_S (the length of the syllable in ms); SDS\_INT\_PH and SDS\_PHON\_PH (the standard deviation of the syllable durations within the *intonational phrase* and the *phonological phrase*); PVIS (the PVI of the duration of the current and the next syllable within the IU (if there was one), computed as follows);

$$PVI = \frac{|segment_a - segment_b|}{mean(segment_a, segment_b)}$$

- Pitch variables: MIN\_PITCH (the minimum pitch across the segment); MAX\_PITCH (the maximum pitch across the segment); SDPITCH\_INT\_PH and SDPITCH\_PHON\_PH (the standard deviation of the mean pitch in Hz across each vowel within the *intonational phrase* and the *phonological phrase*); PITCH\_PRESENT\_PHON and PITCH\_PRESENT\_INT (the percentage of the *phonological/intonational phrase* where F0 is present divided by the length of the entire *phonological/intonational phrase*);
- Intensity variables: MEAN\_INTENSITY (the mean intensity across the vowel); SDINTENSITY\_INT\_PH and SDINTENSITY\_PHON\_PH (the standard deviation of the mean intensity in decibels across each vowel within the *intonational phrase* and the *phonological phrase*).

### Statistical Analysis

A random forest ensemble was generated using the previously mentioned variables in order to predict the dependent variable, UTTERANCE; permutation accuracy variable importance was then determined. Random forests are a series of decision trees that are each based upon a random subset of the available predictors and available data (Breiman 2001). One particular advantage of random forests is that they “are able to better examine the contribution and behavior that each predictor has, even when one predictor’s effect would usually be overshadowed by more significant competitors in simpler models” (Shih 2011:1). Permutation accuracy variable importance determines the amount that a variable contributes to the prediction of the dependent variable, UTTERANCE in this case.

### Results

Table 1 shows the percentage of correctly classified segments as compared to the observed values, as classified by a random forest ensemble; the bold cells across the main diagonal of the table represent those segments correctly identified by the random forest.

predicted	observed				
	English 1	English 2	Portuguese	Spanish 1	Spanish 2
English 1	<b>44</b>	0	0	0	0
English 2	0	<b>23</b>	5	0	0
Portuguese	0	0	<b>56</b>	0	0
Spanish 1	0	0	3	<b>31</b>	0
Spanish 2	0	4	0	0	<b>45</b>

**Table 1:** Classification accuracy for random forest ensemble.

The classification accuracy of this model is 94%; this model performs highly significantly better than the most stringent baseline for classification accuracy (namely a model that chooses the most common level of the dependent variable);  $p_{binomial} < 1e-85$ . Furthermore, the permutation accuracy affords a perspective of those variables that are informative; variables can be considered informative and important if their variable importance value is above the absolute value of the lowest negative-scoring variable (e.g. Strobl et al. 2009, see [Figure 5](#)).

## Discussion

While the theoretical implications of this data set still require an inspection of trees in the random forest ensemble in order to determine the directional behavior of these main effects and potential interactions, it does raise an interesting question on the application of highly automated methods to linguistic data set; implementing forensic linguistic research through highly-automated computerized software increases consistency and objectivity; this ties with Chaski's concept of replicability as a best practice for forensic linguistics (2013). [Figure 5](#) makes it clear that at least 4 variables and potentially a 5<sup>th</sup> could be removed from this model as they are not significant according to prescribed criteria for random forest ensembles (e.g. Strobl et al. 2009).

This possibility of deleting data following Occam's razor brings up the question of model simplicity vs. model classification accuracy. When interpretation of the data for theoretical implications is the goal of a statistical analysis, criteria for the inclusion of variables normally weigh the contribution to model accuracy against model complexity (e.g., minimizing *AIC*) and against other criteria (e.g., avoiding collinearity [i.e. the presence of highly correlated predictors which makes statistical coefficients and tests highly unstable] and overfitting [i.e. the fact that a statistical model fits one particular data set so well that it does not generalize]).

Thus, for the purposes of linguistic research, one would interpret a minimal adequate model, or a model that is comprised of only those variables that contribute 'substantially' to the classification of the dependent variable ('substantially' here according to the model evaluation criterion chosen, e.g. significance value or *AIC*). However, in the case of forensic evidence, where classification or prediction accuracy is the goal, additional variables may be included as they improve this accuracy. Of course, when classification accuracy is the goal, the inclusion of additional variables will have to be weighed more by practical concerns than theoretical motivation, such as the above-mentioned issue of overfitting, but also practical concerns such as admissibility in court, especially in forensic linguistic use. In particular, these variables should be motivated by linguistic or forensic theory (as even the inclusion of random

variation as a variable improves classification accuracy marginally). Furthermore, they should be able to be measured consistently and without undue complication, as to minimize researcher bias and permit their measurement in a reasonably timely manner. Finally, it is vital to consider model assumptions and data distribution issues, avoiding, for instance, collinear predictors; in fact, the interdependence of data points is particularly problematic for forensic science use, since 'correlated evidence has to be avoided' (Becker et al. 2008). In practice, this means that a higher classification accuracy may have to be sacrificed in order to have fewer interdependent data points for evidence to be admitted in court.

The other conclusion that is immediately relevant is that language distinction is certainly possible even in the case of a speech signal of diminished quality. In the case of forensic linguistics, they may prove important as laboratory-quality data are not always available. In fact, returning to Chaski's best practices for forensic linguistics (2013), this ties to the concept of forensically viable data, as it appears that this method would be effective even in the case of data with sub-optimal recording quality.

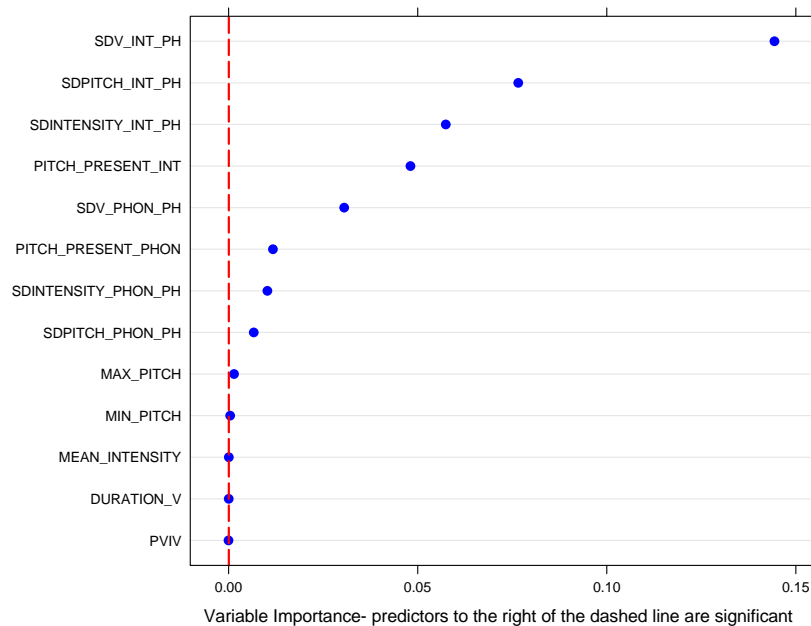


Figure 5: Permutation accuracy variable importance for random forest ensemble.

## Conclusions

As researchers and practitioners move increasingly towards finding common, scientifically acceptable standards for forensic linguistic methodology, we believe that the various processes described in this article lead to several general conclusions that may prove to be highly beneficial to the forensic sciences in the future, and especially in forensic linguistics. Firstly, we can see evidence of dialectal variation in many prosodic cues. Miglio, Gries, & Harris (2014) show that intonational encoding of information structure vary across two closely related dialects. It is important to note that the bilingual participants were raised speaking Spanish in the home by Mexican parents. Given their common linguistic roots with the monolingual Mexican speakers, the variation that occurs appears to be solely due to the influence of bilingualism in their prosodic system. Meanwhile, Harris & Gries (2011) show that not only bilingualism

but also linguistic ability and corpus-based word frequencies contribute to vowel duration variability in bilingual speakers as compared to monolingual speakers. This suggests several methods of speaker identification. For instance, a comparison of high-frequency and low-frequency words could help determine whether a speaker is a monolingual or bilingual speaker. Finally, the third data set discussed suggests that prosodic variables are highly accurate in the classification of languages, and can therefore contribute to their classification, as well as being potentially very useful for the identification of individual speakers, even when only small amounts of data or low-quality data are available. This is particularly true when more complex multivariate methods are used, whose sampling approaches deal well with data problems (e.g., collinearity) and which offer forensic linguists classifiers that can be largely untainted by individual researchers' or expert witnesses' (unconscious) biases.

Let us turn our attention to the application of the current methods in terms of best practices for forensic linguistics, as outlined by Chaski (2013) in order to establish linguistic evidence that is “scientifically respectable and judicially acceptable.” (2013:334). Although these criteria are presented specifically for author identification, they apply to all types of linguistic evidence, including prosodic data. To this end, the current article presents how the methods currently described adhere to several of these best practices. Specifically, the following best practices have been touched upon (from Chaski 2013:334-335):

- a) developed independent of any litigation;
- b) tested for accuracy outside of any litigation;
- c) tested for known limits correlated to specific accuracy levels;
- d) related to standard (“generally accepted”) techniques within the specific expertise and academic training;
- e) related to a specific expertise and academic training;
- f) able to work reliably on “forensically feasible” data;
- g) replicable.

All the methods described in the current paper were developed and tested independently of litigation. In fact, having been performed in academic settings, the current data and methods are based upon accepted scientific and statistical methodology, and rooted in linguistic theory. This means that they adhere to specific accuracy levels accepted according to current literature for each respective method (e.g. for random forest see Strobl, Malley, & Tutz 2009). Linguistic expertise based upon academic training in linguistics gave rise to the methodologies described herein. Both hypotheses and interpretations of results are rooted in linguistic theory (e.g. for speech rhythms see Harris & Gries 2011).

As mentioned above, in terms of being forensically feasible, the results of the random forest methodology are particularly promising. Forensic data are often sub-optimal, having not been collected in a laboratory setting (e.g. Chaski 2013). The random forest method achieved high classification accuracy, despite the fact that the variables were largely derived from an impoverished speech signal. This indicates that prosodic data may be beneficial to the forensic linguist even in the case that the speech signal is distorted in some manner. Finally, the methods applied to forensic linguistics should be replicable, allowing for the same results, regardless of the researcher, thus guaranteeing unbiased results.

While a fair amount of exploration was involved in the analysis of these data sets, their results are largely intended to be exploratory in terms of forensic linguistics. A more systematic protocol for the analysis of prosodic data would have to be developed to completely eliminate researcher bias. This said, the random forest method is particularly promising in terms of replicability, as it is a highly automated computerized method. This is to say that the algorithms of the ensemble methods are programmed in the function `randomforest` from the library `party` (Hothorn, Hornik, Strobl,



& Achim Zeileis 2006).<sup>11</sup> While all of these described methods are not currently automated, a protocol or program to make these techniques replicable by any researcher is certainly possible. Thus, in addition to the theoretical advantages afforded by the application of prosodic variables to the forensic sciences enumerated in the introduction, this article affords several practical examples of the application of these methods to forensic linguistics. The empirical methods applied in the various studies analyzed in this paper, while they do not intend to substitute the linguist testifying as an expert witness in court, are a means to limit subjectivity in forensic linguistic practice. While it is often impossible to maintain in front of a judge that, say, the voice heard on the tape is definitely that of a specific subject, or represents without a doubt a speaker of a certain dialect, it is often easier (and more easily admissible in court) to state that X is *not* the voice of a specific subject, or that it represents a certain dialect speaker, as that implies a simple binary choice in demonstrating that two samples are sufficiently different one from the other (see Tiersma (no date), and Tiersma and Solan 2002 on these points). Semi-automated, statistically vetted methods as advocated in this paper are ways of increasing objectivity and replicability of the analysis, and we have shown here that they work also in cases of a degraded acoustic signal. Our article shares these points with recent forensic acoustic articles (Jessen 2007, Becker et al. 2008, the 'degraded' signal in this last paper, is limited to F1, F2, F3 transmitted through phone lines). It should be noticed that even these acoustic phonetic studies admit that purely linguistic features of speech, such as differences in speaking styles must also be carefully analyzed and addressed in forensic applications.

The novelty of this paper, in fact, lies in harnessing (some of the) speakers' characteristics associated with the grammar of their dialect/language, in our cases prosody, i.e. grammar beyond the segment, which is hard-wired from a very early age and therefore more difficult to imitate or modify. An advantage of both automated methodology (for instance to separate vowels from consonants in the signal), or from concentrating on suprasegmentals (as in our different studies) means that for the technique to be successful, it does not require an expert on the specific language analyzed, which would, on the other hand, be needed for non-automated segmentation and/or segmental analysis.

We hope to have pointed out that prosody and intonation are important linguistic features that can be used for language and speaker classification, with many potential applications in forensic contexts. We also believe that we have shown how linguistic research of this type can avail itself of the same scientific methods that guarantee the replicability of results in the hard sciences, minimizing dependence on an individual's expertise, thus contributing to provide the standards that forensic linguistics is aiming to attain as a scientific discipline.

## References

- Arvaniti, A. 2009. Rhythm, Timing, and the Timing of Rhythm. *Phonetica* 66(1-2). 46-63.
- Beckman, M. E. & G. A. Elam. 1997. Guidelines for ToBI Labelling. Unpublished manuscript, Ohio State University. Version 3.0 March 1997. [Online] Available from <[http://ling.ohio-state.edu/Phonetics/etobi\\_homepage.html](http://ling.ohio-state.edu/Phonetics/etobi_homepage.html)>.
- Becker, T., M. Jessen, & C. Grigoras. 2008. Forensic speaker verification using formant features and Gaussian mixture models. Interspeech 2008 Special Session: Forensic Speaker Recognition - Traditional and Automatic Approaches. Brisbane, Queensland, Australia. September 2008. 1505-1508. [Online] Available from

---

<sup>11</sup> Although random forest is an automated method, there are several settings that can be adjusted that can, on occasion, affect the results of the ensemble. However, a protocol for optimizing these settings can be found in (Strobl, Malley, & Tutz 2006).

- <[http://interspeech2008.forensic-voice-comparison.net/papers/Becker,%20Jessen,%20Grigoras%20\(2008\)%20Forensic%20speaker%20verification%20using%20formant%20features.pdf](http://interspeech2008.forensic-voice-comparison.net/papers/Becker,%20Jessen,%20Grigoras%20(2008)%20Forensic%20speaker%20verification%20using%20formant%20features.pdf)>
- Boersma, P. & D. Weenink. 2010. Praat: doing phonetics by computer (Version 5.1.29 [Computer program]. Retrieved June 21, 2009, from <<http://www.praat.org/>>.
- Breiman, L. 2001. Random Forests. *Machine Learning* 45(1): 5-32.
- Chaski, C. 2001. Empirical evaluations of language-based author identification techniques *Forensic Linguistics* 8(1) 1350-1771.
- Chaski, Carole. 2013. Best practices and admissibility of forensic author identification. *Journal of Law and Policy* 21(2). 333-376.
- Davies, M. 2002-. Corpus del Español: 100 million words, 1200s-1900s. <<http://www.corpusdelespanol.org>>.
- Dauer, R. M. 1987. Phonetic and phonological components of language rhythm. *Proceedings of the 11th International Congress of Phonetic Sciences* 5. 447-450.
- Fox, R. & E. Jacewicz. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America* 126(5). 2603-2618.
- Frota, S. & M. Vigário. 2001. On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus* 13(2). 247-275.
- Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech* 1(2), 120-152.
- Hansen, E., R. Slyh, & T. Anderson. 2004. Speaker recognition using phoneme-specific GMMs. ODYSSEY04 -- The Speaker and Language Recognition Workshop. Toledo, Spain May 31- June 3, 2004. [Online] Available from <[http://www.isca-speech.org/archive\\_open/archive\\_papers/odyssey\\_04/ody4\\_179.pdf](http://www.isca-speech.org/archive_open/archive_papers/odyssey_04/ody4_179.pdf) >.
- Harris, M. & St. Th. Gries. 2011. Measures of speech rhythms and the role of corpus-based word frequency: a multifactorial comparison of Spanish(-English) speakers. 2011. *International Journal of English Studies* 11(2). 1-22.
- Harris, M. *In preparation*. Quantifying speech rhythms: Perception and production data in the case of Spanish, Portuguese, and English. Doctoral dissertation. University of California, Santa Barbara. USA.
- Lea, W. 1973. Evidence that Stressed Syllables are the Most Readily Decoded Portions of Continuous Speech. 2013 Archival Reprint of presentation at the *86th Meeting, Acoustical Society of America*, Los Angeles, CA.
- Lea, W. 1976. The Importance of Prosodic Analysis in Speech Understanding Systems. 2013 Archival Reprint of research conducted by the author under the DARPA Contract DAHC 15-73-0-0310.
- Loukina, A., Kochanski, G., Shih, C., Keane, E., & Watson, I. 2009. Rhythm measures with language-independent segmentation, 1531–1534.
- Low, E. & E. Grabe. 1995. Prosodic patterns in Singapore English. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, Vol 3, 636-639.
- Jessen, M. 2007. Speaker classification in forensic phonetics and acoustics. In C. Müller (ed.) *Speaker Classification I - Fundamentals, Features, and Methods*. Lecture Notes in Computer Science Volume 4343, Berlin/Heidelberg: Springer Verlag. 180-204.

- Miglio, V., S. Th. Gries, & M. Harris. 2014. Intonation signaling new & given information in English-Spanish bilinguals and Mexican Spanish speakers. Poster presented at *Current Approaches to Spanish and Portuguese Second Language Phonology*. March 14-16, 2014, Georgetown University, Washington DC.
- Nazzi, T., J. Bertoncini, & J. Mehler. 1998. Language discrimination by newborns: Towards understanding the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24(3). 756-766.
- Pierrehumbert, J.B. 1980. The phonology and phonetics of English intonation. PhD dissertation, MIT. Published 1988 by Indiana University Linguistics Club.
- Pike, K. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press.
- R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <<http://www.R-project.org/>>.
- Ripley, B. 2011. MASS. Version 7.3-13 Package for R.
- Schötz, S. 2002. Linguistic & Paralinguistic Phonetic Variation in Speaker Recognition & Text-to-Speech Synthesis. [Online] Available from <[http://www.speech.kth.se/~rolf/gslt\\_papers/SusanneSchotz.pdf](http://www.speech.kth.se/~rolf/gslt_papers/SusanneSchotz.pdf)>.
- Selkirk, E. O. 1978. On prosodic structure and its relation to syntactic structure, in T. Fretheim, ed., *Nordic Prosody II*. Trondheim: Tapir, 111-140.
- Shih, S. 2011. Random forests for categorical dependent variables: an informal quick start R guide. [Online] Available from <<http://www.stanford.edu/~stephsus/R-randomforest-guide.pdf>>.
- Strobl, C., J. Malley, & G. Tutz. 2009. An Introduction to Recursive Partitioning: Rational, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Psychological Methods*. 14(4). 323-348.
- Swerts, M., Kraemer, E. & Avesani, C. 2002. Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics* 30(4). 629-654.
- Tiersma, P. (No date). *What is Forensic Linguistics?* [Online] Available from <<http://www.languageandlaw.org/FORENSIC.HTM>>
- Tiersma P. & L. Solan. 2002. The linguist on the witness stand: forensic linguistics in American courts. *Language* 78(2). 221-239.
- Vallduví, E. 1992. *The informational component*. Garland Publishers: New York/London.
- White, L. & Sven L. Mattys. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35(4). 501-522.
- Wretling, P., & A. Eriksson. 1998. Is articulatory timing speaker specific? — Evidence from imitated voices. In Branderud, P. and H. Traunmüller (eds.) *Proceedings of FONETIK 98*. Stockholm: Dept. Linguistics, Stockholm University.
- Zubizarreta, M.L. & E. Nava. 2010. Encoding discourse-based meaning: prosody vs. syntax. Implications for Second Language Acquisition. *Lingua* 121(4). 652-669.