This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication. The final version of record is available at

http://dx.doi.org/10.1016/j.csl.2017.02.011

Identifying characteristic prosodic patterns through the analysis of the information of Sp_ToBI labels sequences.

David Escudero[a], César González[a], Yurena Gutiérrez[b], Emma Rodero[c]

[a]*Dpt. of Computer Sciences, Universidad de Valladolid, Spain*
[b]*Dpt. of Spanish Philology, Universitat Autònoma de Barcelona, Spain*
[c]*Department of Communication, Universitat Pompeu Fabra, Spain*

Abstract

This paper presents a novel methodology to characterize the style of different speakers or groups of speakers. This methodology uses sequences of prosodic labels (automatic Sp_ToBI labels) to compare and differentiate these speaking styles. A set of metrics based on conditional entropy is used to compute the distance between two speakers or group of speakers depending on the use of sequences of prosodic labels. Additionally, the most contrastive sequences of labels are identified as characteristic patterns of the speaking styles represented in a given corpus. When this methodology is applied to a corpus of radio news items, the result is that the most frequent prosodic patterns coincide with those previously characterized in studies about radio style. Finally, a perceptual test verifies that the participants attribute these characteristic patterns to the radio news style.

Keywords: Comparing prosody, prosodic labeling, radio news style, informational distance, entropy analysis, Sp_ToBI

---

1. Introduction

Prosody has an idiosyncratic value associated with some characteristic enunciations, pronunciations, inflections, pausal and other speech patterns that can be related to an individual or to a group of them. When a group of individuals adapt their manner of speaking in a similar way, in certain communicative situations, thus creating an acoustically typified image, it is generally called a *speaking style* (Goldman et al., 2009). The role of prosody on speaking style has recived some attention in the last two decades (Llisterri, 1992; Eskénazi, 1993). Regarding the analysis of individual speaking style, prosody has also been shown to be useful in speaker recognition (Adami et al., 2003; Shriberg et al., 2005). Automatic methods based on corpus analysis have proved to be effective in automatic speaking style classification, obtaining high identification rates in several applications (higher than 90% in Goldman et al., 2014 or in Rosenberg, 2013). While corpus-based *speaking style identification* seems to be an affordable problem for automatic algorithms, *speaking style characterization* is still a challenge. In fact, machine learning algorithms can be very effective in classifying and labeling speaking style, but ineffective in providing information that is useful for understanding the reasons that lead the algorithm to a given classification. The goal of this study is to outline the automatic detection of characteristic prosodic patterns by contrasting prosodic styles of two speakers or two group of speakers.

The perception and production of a given speaking style is mainly studied by analyzing acoustic factors like F0 excursion, speech rate, or prosodic variation. Nevertheless, prosodic labels can also contribute to the analysis of speaking style (Hirschberg, 2000; Rosenberg, 2013; Obin, 2011; Obin and Lanchantin, 2015). Indeed, prosodic labels code information such as prominence and phrasing, which determines the way in which speakers organize their discourse. Managing the phrasing and prominence along with the discourse in an appropriate way is crucial for being effective in communication, which is specially relevant for some type of speakers (for example, journalists). Most of the works in the literature limit the presentation of results about speaking style characterization to the description of tables and boxplots that compare the mean, standard deviation, and quartiles of the acoustic values corresponding to the analyzed styles. These statistical resources barely give information about the way speakers organize their discourse over time. As an alternative, in this paper we present a method that extracts the most informative sequences of prosodic labels or prosodic patterns that distinguish two given styles.

The identified prosodic patterns must be interpretable and recognizable, as well as to permit a discussion of how they accurately characterize the given speaker or group of speakers. ToBI is a commonly accepted annotation system in the scientific community, extended to many languages, which facilitates further comparative studies and the communication of results. The Sp_ToBI system (Beckman, 2002; Face and Prieto, 2007; Estebas Vilaplana and Prieto, 2008) will be used to describe melodic contours in the corpus analyzed in this study. The Sp_ToBI system assumes the descriptive bases of the Autosegmental-Metrical model (Pierrehumbert, 1980; Ladd, 1996, among others) in order to distinguish two phonological units: (i) pitch accents, which are related to the more promient syllables within the word and marked with an asterisk *, and (ii) boundary tones, associated with the edges of prosodic domains and marked with the symbol %. The tones in both domains are represented by the capital letters "L" for low tone, and "H" for high tone, and both can be combined to represent complex tones (e.g. L+H* or LH%). The different combinations of sequences of ToBI symbols represent the way in which speakers organize their discourse. The use of symbolic information allows us to perform a statistical analysis based on information entropy which shows that different speakers use specific ToBI symbol combinations that can contribute to the characterization of his/her speaking style.

One of the problems of the ToBI labeling is its high cost. Manual labeling is a time-consuming task (Syrdal et al., 2001), and using more than one labeler is problematic because of the potentially high intertranscriber disagreement rates (Escudero et al., 2012). However, the recent availability of tools for the automatic labeling of prosody, such as AutoBI (Rosenberg, 2010), overcomes the mentioned drawbacks. The labels predicted by an automatic system may not be as valid as the labels generated by an expert. Nevertheless, this method has the advantage of consistency, so that there is a guarantee that the same label sequences are obtained when the automatic algorithm analyzes the same or similar utterances. The output of an automatic labeling system is a sequence of ToBI labels that can be interpreted in the same way manual labels are. Consequently, the sequences of characteristic patterns that appear as a result of the speaking style analysis can be presented to a phonetician or an expert to assess the appropriateness of the patterns, or used in a perception test as described in this paper.

In order to test the methodology, we use a corpus of radio recordings. The voice of the journalists has been extensively studied in the literature (Goldman et al., 2008; Obin et al., 2008; Shriberg et al., 2009; Degand et al., 2009; Goldman et al., 2009; Shriberg et al., 2009; Roekhaut et al., 2010; Obin et al., 2010; Obin, 2011; Rosenberg, 2013; Goldman et al., 2014). The analysis of radio news items is an excellent case of study because there are abundant references about the characterization of newscasting, developed in the fields of communication and linguistics

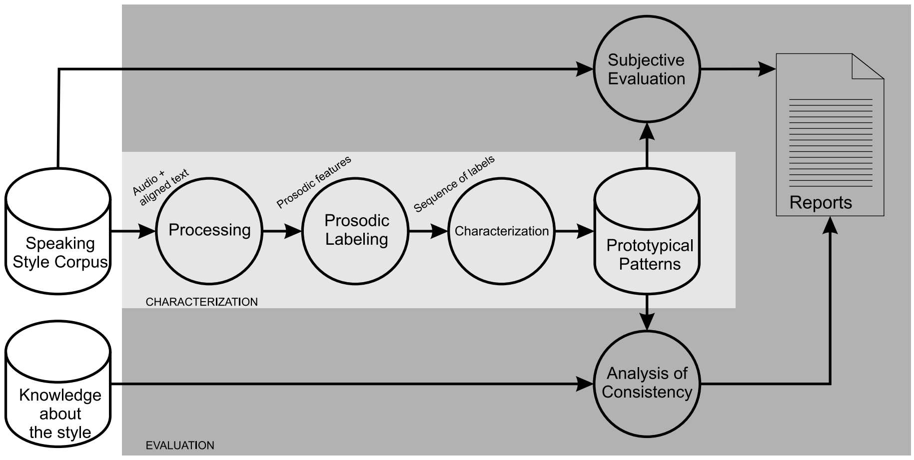

Figure 1: Scheme of the experimental procedure.

(Wheatley, 1949; Bolinger, 1998; Cotter, 1993; Medrado et al., 2005; Rodero, 2007; Price, 2008; Grawunder et al., 2008; de-la Mota and Rodero, 2012; Rodero, 2013). In this paper, apart from doing a perceptual evaluation of the characteristic prosodic patterns, we compare them with the conclusions reported in the literature about newscasting style.

Figure 1 describes the scheme proposed in this paper for characterizing speaking style and process validation. The procedure of this study is based on recordings of different speakers and styles which are then compared. Section 2.1 presents the corpus of this work and the definition of the a-priori styles to be compared. The *Processing* module computes the relevant prosodic features inside the prosodic units used as the basic reference in the labeling process (typically words or syllables). These prosodic features are the input of the *Prosodic Labeling* module. This labeling module is an automatic labeling system that has been previously trained with a manually labeled subset of the whole corpus. The output of this module is a sequence of labels linked to the prosodic units of the corpus. Section 2.2 presents the processing and labeling procedure, described in greater detail in Escudero et al. (2014b). The *Characterization* module identifies the sequences of labels that determine the style of the different speakers or groups of speakers. The metrics that identify such prototypical sequences of labels are presented in Section 2.3. In this section, we illustrate the use of the proposed metrics by comparing the style of the different individual speakers of the corpus. In section 3 we apply the methodology to identify the prosodic patterns that the newscasters of the corpus use in comparison with the patterns used by a group of actors reading the same set of news items. The *Analysis of Consistency* module permits to contrast the consistency of our results with the ones obtained in other studies (*Knowledge about the style* in figure 1). These studies are reviewed in Section 4, where we analyze the correspondence between the characteristic patterns that result from our analysis and the observations reported in the state of the art. The *Subjective Evaluation* is intended to check whether or not the listeners identify the prototypical patterns as a specific style. Section 5 details the test that has been performed for this purpose. The paper ends with the discussion and the conclusions.

Speaker ID	Speaker Type	Gender	#News Items	Duration	#words	#sentences
f11r	Newscaster	Female	36	30’53”	5,443	243
f13r	Newscaster	Female	72	1h03’56”	11,271	467
m12r	Newscaster	Male	36	32’24”	5,439	242
m14r	Newscaster	Male	71	54’32”	10,958	457
f15a	Announcer	Female	71	1h27’28”	11,126	463
f16a	Announcer	Female	72	1h07’19”	11,238	466
m09a	Announcer	Male	36	30’59”	5,431	242
m10a	Announcer	Male	36	31’07”	5,440	242
			430	6h38’38”	66,346	2,822

Table 1: Contents of the Glissando news subcorpus used in this research.

2. Experimental procedure

2.1. The speaking styles in the Glissando read news corpus

We used the Glissando corpus (Garrido et al., 2013) for our investigation. This corpus was designed to remedy the lack of corpora for prosodic studies in Catalan and Spanish. The corpus is divided into three separate parts: a news subcorpus, conversational dialogues and task-oriented dialogues. The present study only uses the Spanish news subcorpus. All the experiments reported in this paper were carried out on Peninsular Spanish.

The news subcorpus comprises a set of news items read by professional speakers. The text of the news items was obtained from real radio stations. We collected a large number of news items and then selected some of them to build the two parts of the subcorpus: A) 36 news items chosen to obtain a balanced number of different types of prosodic units; B) 36 news items selected to cover all different diphone combinations (Escudero et al., 2009). We selected four speakers per style (two male and two female newscasters and actors), resulting in a total of eight speakers. All of them were recorded reading the news items in a recording studio. Part A was recorded by eight speakers and part B was recorded by four of the speakers (two speakers per style).

Once the corpus was put together, text and audio were aligned at different levels: phonemes, syllables, words and stress groups. The sentences were analyzed to obtain the POS information associated with each of the words by using FreeLing, an open source language analysis tool (Padró and Stanilovsky, 2012). The size of the corpus (illustrated in the table 1) is larger than that of the corpora that have been used in the field. With smaller corpus in terms of samples per style and samples per speaker, other works, such as Mixdorff et al. (2005); Obin et al. (2008); Degand et al. (2009); Goldman et al. (2009), presented significant results. In this study, we show that the analysis of the Glissando corpus also yields meaningful conclusions to be drawn about radio news speaking style.

The goal of this study is to obtain characteristic prosodic patterns by contrasting the prosodic styles of two speakers or two group of speakers. First, we show that there are relevant differences between each pair of speakers. Furthermore, we contrast two different styles of professional abilities: radio news broadcasters (newscasters) – who are journalists – and advertising and dubbing broadcasters (announcers) – mostly actors – who act as voiceover on a variety of media messages (commercials, documentaries, movies, etc.). These two professional styles bear prosodic differences in terms of intonation and rhythm owing to their differing characteristics concerning speech modality, background, and speaker training.

First, in Spain, speakers presenting the news on radio are journalists. Therefore, they have pursued an education and training in journalism, covering various aspects of working with information. Since their job is not merely to present news, but also to search, select and write about it, their training in speech (how they use their voice presenting news) only covers a small part; therefore, they are less able to vary their prosody. In addition, they are conditioned by a speech modality that is both informative and persuasive. Consequently, journalists have to deal with a message that is first journalistic – seeking to be read as equally as possible avoiding prosodic prominences – and which is secondly persuasive and aims at attracting the attention of listeners. This double feature is what results in the need for it to be a short interpretive message whilst also being expressive, avoiding monotony. In conclusion, this set of factors has broadened a peculiar style of reading the news, characterized by prosodic signals of constant emphasis aimed at attracting the listener's attention, but marked by a regular melody with a fast speech rate. This style of professional talking in radio news is used by most journalists and is widely documented in various studies (Rodero, 2013).

Secondly, advertising and dubbing broadcasters (announcers) use a distinguished professional speaking style, also conditioned by their distinct speech modality, profession and training. Contrary to what is the case with newscasters, these speakers are trained in speech (voice and prosody) and in interpretation, because their professional job is based on providing the voiceover for different media messages. In this respect, most of them are actors. Unlike the journalistic message, these speakers work with more expressive messages, narrative and interpretive texts (commercials, narrations, movies, documentaries, etc.), which are conducive to a more enhanced prosodic performance. This means they are trained for narrative and dramatic reading. As a result, these conditions generate a distinguished professional speaking style more in keeping with the meaning of the message and, accordingly, one which is more expressive.

In the case of our corpus, both presenter groups read out a set of professional radio news items, but they were expected to undertake the task differently as a result of distinct targeting. During the recordings, it is important to note that none of the groups were given instructions on how they were to read the news. Thus, since the first group was formed by journalists – and considering the studies characterizing their defined prosodic style – it was expected that these professionals would read the news as they usually do in their jobs. Along these lines, an emphatic prosodic style with a more or less regular rhythm and a high speech rate was expected, portraying the same melodic patterns described in the literature. On the other hand, a different style was expected from the announcer group. Since these professionals are trained in reading various texts, it was expected that they would read the news in a narrative way, thereby adapting the prosody to the content of the text. In consequence, this style may be characterized by a greater variety of prosody, more moderate pitch movements and a more reduced speech rate. This pattern should sound more balanced with no very noticeable prosody emphasis. Ultimately, it is a style more like interpreted reading.

2.2. Prosodic labeling

The news data-set has been annotated using the Sp_ToBI labels proposed in Estebas Vilaplana and Prieto (2008, 2010), with the modifications advanced in Elordieta (2011). A phonologically-oriented prosodic annotation, such as the ToBI model, requires a wide consensus on particular aspects of a restricted speech style, such as the reading of news by professionals. Various methods of validating the consistency and stability of the labels assigned to the corpus were conducted: (i) periodic meetings to define a proposal that applies the Sp_ToBI to news reading; (ii) discussion and resolution of differences in transcription throughout a six-month period; and (iii) validation of consistency among transcribers with an inte-labeler rreliability experiment. A labeler annotated several news items from the Glissando corpus with Sp_ToBI events. An analysis of the transcriptions was carried out. We concluded that it was necessary to reduce the number of categories. Similar classes were grouped together, taking into account the consistency tests among labelers (for instance L+H* and (L+)H* form a common class). Classes with few elements were removed (for example the accent L+>¡H*, which only had 8 instances). A more detailed description of the process can be consulted in Escudero et al. (2014b).

An automatic system was trained using the news labeled with Sp_ToBI and was used to label the entire Glissando news corpus. The process was done in two phases, using a semiautomatic labeling approach. In the first phase, the human labeler annotated the Sp_ToBI events of 24 news items from the Glissando corpus (12 from a newscaster and 12 from an announcer). The automatic system was trained using these labeled news items so that it could then be used to label the rest of the corpus. In the second phase, the human transcriber reviewed and manually corrected the Sp_ToBI labels generated by the system in 36 news items (12 from a newscaster and 12 from two announcers). Finally, the system was trained again, using 60 (24+36) news items, and it was used to label all the news of the corpus. A total of 5,103 pitch accents and 2,835 boundary tones were used to train the final automatic transcription system.

The automatic labeling system applied is based on pairwise coupling (Hastie et al., 1998): the multiclass classification problem is divided into several binary classification subproblems, from which the results are combined to obtain the final classification result. To combine the results of the pairwise classifiers, the method described in (Hastie et al., 1998) is used. Moreover, different types of classifiers are used: neural networks, decision trees and support vector machines. The outputs of the different classifiers are combined using the fuzzy integral aggregation technique (Grabisch, 1995).

The word is used as the reference unit. The following features are extracted: frequency features (within-word F0 range, difference between maximum and average within-word F0, difference between average and minimum within-word F0, difference between within-word F0 average and utterance average F0); energy features (within-word energy range, difference between maximum and average within-word energy, difference between average and minimum within-word energy); maximum normalized vowel nucleus duration from all of the vowels of the word; pause duration after the word (only for boundary tones); part of speech tags (automatically obtained); Tilt parameters (Taylor, 2000); and Bézier parameters (Escudero et al., 2002) (an approximation of the pitch contours with Bézier functions, using 4 control points). Context features are also used to improve the classification results. A selection of features to model the context was achieved by using the Correlation-based Feature Selection (CFS) algorithm; features from the two previous and two following words are included.

The system is an adaptation of a system developed for English and described in Gonzalez-Ferreras et al. (2012). A classification rate of 70.8% for pitch accents and 84.2% for boundary tones was reported in the Boston Radio News Corpus. An improvement of the classifier was described in Escudero-Mancebo et al. (2014), using fuzzy logic techniques and reaching a soft classification rate of 81.8% for pitch accents. The adaptation of the classifier for its use with the Sp_ToBI labeling system is detailed in Escudero et al. (2014a). When the manual transcriber reviewed the automatic labels generated by the classifier, 72.6% of the pitch accents labels and 81.8% of the boundary tones were marked as correct. In Escudero et al. (2014b), there is a description of the revision process of the automatic labels. The suitability of using automatic labels instead of manual labels is discussed in section 6.

A summary of the labeled contents of the Glissando news corpus is shown in table 1. Table 3

	m12r	m14r	f11r	f13r	f15a	f16a	m09a	m10a
m12r		0.006	0.021	0.010	0.012	0.015	0.004	0.005
m14r	0.006		0.009	**0.003**	0.007	0.017	0.005	0.014
f11r	0.021	0.009		0.007	0.015	**0.039**	0.020	0.031
f13r	0.010	**0.003**	0.007		0.010	0.018	0.008	0.018
f15a	0.012	0.007	0.015	0.010		0.011	0.012	0.018
f16a	0.015	0.017	**0.039**	0.018	0.011		0.013	0.016
m09a	0.004	0.005	0.020	0.008	0.012	0.013		0.011
m10a	0.005	0.014	0.031	0.018	0.018	0.016	0.011	

Table 2: Symmetric distance matrix of the speakers of the Glissando corpus. The cells of the matrix represent I(T; S) with $S = \{i, j\}$ being i, the speaker labeled in the row and j the speaker labeled in the column. T are patterns of length 1. The metrics have been computed with the samples of the prosodic corpus. Maximum and mininum distances are boldfaced.

presents the count of the final set of tones and labels in the corpus. Not all the news items could be labeled because some of them are not segmented in the repository `http://veus.glicom.upf.edu`.

2.3. Contrasting sequences of prosodic labels

The output of the labeling system is a sequence of Sp_ToBI labels. Given an utterance whose transcription is: *En Bagdad, un atentado con coche bomba ha dejado al menos cuatro muertos y doce heridos* (translated as “In Baghdad, a car bomb attack has left at least four dead and twelve injured”); the labeling system produces the following sequence of ToBI labels: none, L+H*, H%, H*, L+!H*, L%, none, L+H*, L*, L%, L+>H*, L+>H*, none, L+H*, L+H*, !H%, L+H*, none, L+!H*, L*, L%. Pitch accents and boundary tones are independently predicted. The automatic labeling system assigns a pitch accent symbol (or the symbol none if the word is unaccented) to every word. In parallel, the system assigns a boundary tone symbol to the boundary words. As a result, boundary words have two labels. In the previous example, the relationship is: (*En*: none) (*Bagdad*: L+H* H%) (*un*: H*) (*atentado*: L+!H*) (*con*: none) (*coche*: L+H*) (*bomba*: L* !H%) (*ha*: L+>H*) (*dejado*: L+>H*) (*al*: none) (*menos*: L+H*) (*cuatro*: L+H*) (*muertos*: L+H* !H%) (*y*: none) (*diez*: L+!H*) (*heridos*: L* L%).

We consider the automatic labeling system as a source of information that produces symbols t of a qualitative random variable T = { H*, L+>H*, L+!H*, L+¡H*, L+H*, L*, L%, H%, !H%, LH%, =%, none }. The sequences of symbols coming from the source, belong to a speaker or a group of speakers s. We also consider s to be a value of a qualitative variable S representing a speaker or a group of speakers.

Thus, the speaking style identification problem consists in determining which speaker or group of speakers s is generating a given sequence of tone observations $t_{1..N} = t_1, \ldots, t_N$. The problem of speaking style characterization is to find the symbol sequences $t_{i..j}$ with $i \leq j$ that are the most informative to determine the style s of the source.

Information theory provides answers to this problem (Arndt, 2001). An analysis based on entropies over the tone sequences of each news item of the corpus is applied. This analysis allows us to find the sequences of tones (which we refer to as patterns) that best discriminate the speaking style. The following subsections describe the metrics that were used. In order to illustrate the operative of the metrics, we analyze the differences on the use of the tones by different pairs of speakers. Section presents 3 a more interesting study case that identifies the characteristic patterns that differentiate the styles of the two a-priori groups of speakers found in the Glissando corpus: newscasters and anouncers.

2.3.1. *Conditional entropy and mutual information*

The entropy H(T) is computed using the classic formula:

$$H(T) = -\sum_t p_t \log_2 p_t = -\sum_t \frac{n_t}{n} \log_2 \frac{n_t}{n} \tag{1}$$

n_t being the number of samples of the tone t and n the total number of samples. The relative entropy $H_s(T)$ is computed as:

$$H_s(T) = -\sum_t p_{t|s} \log_2 p_{t|s} = -\sum_t \frac{n_{ts}}{n_s} \log_2 \frac{n_{ts}}{n_s} \tag{2}$$

where n_{ts} is the number of samples of tone t and speaker s (or group of speakers), and n_s is the number of samples of s. H(T) represents the variety of T while $H_s(T)$ represents the variety of T when only the samples of s are analyzed.

Mutual information of S and T is $I(T;S) = H(T) - H_S(T)$ with $H_S(T) = \sum_s p_s H_s(T)$. Mutual information measures the relationship between S and T. It can be interpreted as a distance between the elements of S in terms of the particular use of T. If the distribution of tones is independent among the elements of S, H(T) and $H_S(T)$ will be similar, so that I(T; S) will be close to cero. On the other hand, the greater the difference in the use of the prosodic patterns T by the speakers in S, the higher the value of I(T;S).

Table 2 shows the distances between the speakers of the corpus computed in terms of I(T;S). According to the results, the most distant speakers in the corpus are *f*16*a* and *f*11*r* and the closest are *m*14*r* and *f*13*r*. Table 3 permits to analyze the reasons for these differences. The left and right part of the table constrast respectively the number of samples per tone of the closest and most distant pair of speakers in terms of mutual information. The difference between $H_t(S)$ and H(S) reveals a different use of the tone t among the different speakers in S. The values of the metric $H_t(S)$ in the left part of table 3 are similar going from 0.9859 to 0.9994. Greater variation is observed in the right part of the table with values going from 0.7467 to 0.9954. For example, the speaker *f*11*r* uses the tone L* 370 times at the time that the speaker *f*16*a* uses this tone only 100 times (remind that both speakers read the same text). This contrast is lower for the speakers *f*13*r* and *m*14*r* (204 vs 189). A higher the contrast between the values H(S) and $H_t(S)$ means that the tone t is more informative. for charactizing the differences between the speakers in S.

2.3.2. *Informational distance*

In order to analyze the situations in which a given symbol t contrasts in terms of its informative content, and taking into account the information content of the whole set of symbols, Krippendorff proposed the use of two metrics: the *informational distance* and the *informational bias* (1986).

In the first case, the objective is to compare the value of one row of the table 3 with the aggregate of the remaining rows. The amount of transmitted information between the two variables t versus not-t (referred to as $\bar{t}$) and S measures the difference between one row and the rest of them, and it is called *informational distance* $I(t\bar{t}; S)$[1]:

$$I(t\bar{t}; S) = \sum_s p_{ts} \log_2 \frac{p_{ts}}{p_t p_s} + \sum_s (p_s - p_{ts}) \log_2 \frac{p_s - p_{ts}}{(1 - p_t) p_s} \tag{3}$$

[1] Krippendorff (1986) uses $T(t\bar{t} : S)$ instead of $I(t\bar{t}; S)$ and $T(t : S)$ instead of $I(t; S)$. Our nomenclature is closer to the related metric *mutual information* which is commonly written as I(T; S) in modern texts.

Tone t	s=f13r n_{ts}	s=m14r n_{ts}	Total	Entropy (bits)	s=f11r n_{ts}	s=f16a n_{ts}	Total	Entropy (bits)
=%	**108**	**143**	**251**	**$H_{=\%}$(S)=0.986**	83	158	241	$H_{=\%}$(S)=0.929
!H%	548	619	1167	$H_{!H\%}$(S)=0.997	901	498	1399	$H_{!H\%}$(S)=0.939
H*	**266**	**338**	**604**	**H_{H*}(S)=0.990**	559	289	848	H_{H*}(S)=0.926
H%	519	556	1075	$H_{H\%}$(S)=0.999	**356**	**862**	**1218**	**$H_{H\%}$(S)=0.872**
L*	204	189	393	H_{L*}(S)=0.999	**370**	**100**	**470**	**H_{L*}(S)=0.747**
L%	**474**	**358**	**832**	**$H_{L\%}$(S)=0.986**	634	371	1005	$H_{L\%}$(S)=0.950
L+>H*	**320**	**403**	**723**	**$H_{L+>H*}$(S)=0.990**	263	354	617	$H_{L+>H*}$(S)=0.984
L+!H*	373	467	840	$H_{L+!H*}$(S)=0.991	437	495	932	$H_{L+!H*}$(S)=0.997
L+¡H*	164	205	369	$H_{L+!`H*}$(S)=0.991	286	200	486	$H_{L+!`H*}$(S)=0.977
L+H*	1902	1787	3689	H_{L+H*}(S)=0.999	1759	1961	3720	H_{L+H*}(S)=0.998
LH%	128	156	284	$H_{LH\%}$(S)=0.993	**203**	**76**	**279**	**$H_{LH\%}$(S)=0.845**
none	1931	2042	3973	H_{none}(S)=0.999	1738	2039	3777	H_{none}(S)=0.995
Total	6937	7263	14200	**H(S)=1.000**	7589	7403	14992	**H(S)=1.000**
H_s(T)	2.948	3.018		**H(T)=2.987**	3.131	2.924		**H(T)=3.068**
p_s H_s(T)	1.440	1.544		**I(T;S)=0.003**	1.585	1.444		**I(T;S)=0.039**

Table 3: Comparison of the number of tones and the conditional entropy of different speakers of the corpus. The left part of the table shows the data for the closest speakers and the right part of the table for the most distant ones. The χ-Square test applied to $n_{t|s}$ with t the patterns of the table reveals a dependence between the tones t and the type of speaker s with significant results (χ^2 = 62.4003, df = 11, p-value = 3.311e-09 for $f13r$ vs. $m14r$ and χ^2 = 782.3662, df = 11, p-value < 2.2e-16 for $f11r$ vs. $f16a$)

where $p_{ts} = n_{ts}/n$. $I(t\bar{t}; S)$ is positive or zero. It reaches a maximum when p_{ts} is zero and $p_s - p_{ts}$ is not zero or vice versa. In addition, it cannot exceed the value of 1. It becomes zero when the conditional distribution $p_{t|s}$ in t is equal to $p_{\bar{t}|s}$ in $\bar{t}$, in which case, both are equal to the marginal distribution p_s and $p_{ts} = p_t p_s$.

Table 4 shows the values of $I(t\bar{t}; S)$ computed with the same data presented in the table 3. Lower values are obtained when the samples of the speakers $f13r$ and $m14r$ are compared: the values of $I(t\bar{t}; S)$ go from 0.0000 to 0.0012 at the time that the values are in the interval 0.0003 to 0.0120 for the pair of speakers $f11r$, $f16a$. The higher value is assigned to the tone $L\%$ for the pair $f13r, m14r$ and $H\%$ for the pair $f11r$, $f16r$. These tones had been identified as highly discriminative in table 3. This metric has the drawback of being very sensitive to the size of the sample, thus penalizing the categories with fewer samples. Thus for example, in the contrast of speakers $f13r$ and $m14r$, the value of $I(t\bar{t}; S)$ for the tone LH% is lower than the value for the tone !H% (0.0028 vs 0.0057) when in reality it is not so informative, as is also shown below when we introduce the *informational bias* metric, which reduces the effect of the sample size.

2.3.3. Informational bias

The *informational bias* also compares the expected and observed probabilities, but only within each row.

$$I(t; S) = \frac{1}{p_t} \sum_s p_{ts} \log_2 \frac{p_{ts}}{p_t p_s} \tag{4}$$

The observations of the row t have the status of subsample, and $I(t; S)$ measures the degree in which this subsample is different from the rest of the samples of which it forms part.
$I(t; S)$ is related to the mutual information:

Tone t	$p_{s=f13r\|t}$	$p_{s=m14r\|t}$	$I(t\bar{t} : S)$	$I(t : S)$	#	$p_{s=f11r\|t}$	$p_{s=f16a\|t}$	$I(t\bar{t} : S)$	$I(t : S)$	#
=%	0.430	0.570	0.0002	0.0098	251	0.656	0.344	0.0013	0.0767	241
!H%	0.470	0.530	0.0001	0.0010	1167	0.356	0.644	0.0057	0.0557	1399
H*	0.440	0.560	0.0003	0.0067	604	0.341	0.659	0.0041	0.0688	848
H%	0.483	0.517	0.0000	0.0001	1075	0.708	0.292	0.0120	0.1359	1218
L*	0.519	0.481	0.0001	0.0027	393	0.213	0.787	0.0079	0.2431	470
L%	0.570	0.430	0.0012	0.0191	832	0.369	0.631	0.0033	0.0454	1005
L+>H*	0.443	0.557	0.0003	0.0061	723	0.574	0.426	0.0008	0.0185	617
L+!H*	0.444	0.556	0.0004	0.0057	840	0.531	0.469	0.0003	0.0040	932
L+¡H*	0.444	0.556	0.0002	0.0056	369	0.412	0.588	0.0007	0.0197	486
L+H*	0.516	0.484	0.0007	0.0021	3689	0.527	0.473	0.0011	0.0032	3720
LH%	0.451	0.549	0.0001	0.0041	284	0.272	0.728	0.0028	0.1471	279
none	0.486	0.514	0.0000	0.0000	3973	0.540	0.460	0.0021	0.0061	3777

Table 4: Informational distance and informational bias metrics contrast the relevance of each tone in the characterization of speaking style. The left part of the table shows the data for the closest speakers and the right part of the table for the most distant ones.

$$I(T; S) = \sum_t \sum_s p_{ts} \log_2 \frac{p_{ts}}{p_t p_s} = \sum_t p_t I(t; S) \tag{5}$$

and to $I(t\bar{t}; S)$:

$$I(t\bar{t}; S) = p_t I(t; S) + p_{\bar{t}} I(\bar{t}; S) \tag{6}$$

As this metric takes into account the number of samples in the row, it reveals information that the metric $I(t\bar{t}; S)$ can hide.

The ability of this metric to detect the capacity of tones t to characterize the members of S is evidenced in table 4. The boldface tones of table 3 are the ones that obtain higher values of $I(t; S)$ (0.0191 and 0.0098 for the pair $f13r, m14r$ and 0.2431, 0.1471 and 0.1359 for the pair $f11r$, $f16a$). Focusing on the comparison of the pair $f13r, m14r$, $I(t\bar{t}; S)$ is somewhat higher for H% than for L*. However, the value of $I(t; S)$ is substantially higher for L* than for H%. This effect is because, as shown in table 3, the number of occurrences of *H*% is noticeably higher. In a similar way, going back to the tones !H% and LH% analyzed in the previous subsection, $I(!H\%, \overline{!H\%}; S) = 0.0057$ is higher than $I(LH\%, \overline{LH\%}; S) = 0.0028$, but $I(LH\%; S)$ is higher than $I(!H\%; S)$ reflecting the fact that the impact of t=*LH*% in S is clearly higher.

In the next section, the informativeness of the tone sequences $t_{1:N}$ with N>1 is analyzed. This analysis is mainly based on I(t;S), without losing sight of the value of $I(t\bar{t}; S)$. We consider that, when characterizing style, it is not necessary that the characteristic patterns are the most frequent. In fact, in view of the information shown in table 3, it seems that the most frequent patterns are essentially the same (L+H*, none) for the four speakers analyzed, and they do not seem to be highly informative in any case. On the other hand, an infrequent tone can be very informative but difficult to find in the prosodic patterns.

3. Prototypical patterns

In this section we have grouped the speakers of the corpus according to the two preset categories $S = \{newscaster, announcer\}$. In table 5 we present the analysis of individual tones

Tone t of T	$p_{s=newscaster\|t}$	$p_{s=announcer\|t}$	$I(t\bar{t} : S)$	$I(t : S)$	#
H%	0.575	0.425	0.0014	0.0180	6369
L%	0.428	0.572	0.0009	0.0135	5606
H*	0.438	0.562	0.0005	0.0099	4421
L*	0.425	0.575	0.0004	0.0149	2329
none	0.515	0.485	0.0003	0.0010	23684
L+>H*	0.542	0.458	0.0003	0.0061	3699
!H%	0.464	0.536	0.0003	0.0031	9060
L+¡H*	0.461	0.539	0.0001	0.0036	2593
L+!H*	0.513	0.487	0.0000	0.0008	4573
L+H*	0.497	0.503	0.0000	0.0000	24945
LH%	0.473	0.527	0.0000	0.0016	1590
=%	0.520	0.480	0.0000	0.0016	1377

Table 5: Informational distance and informational bias metrics. They contrast the relevance of each tone in the characterization of speaking style.

applied to these pair of styles. Additionally, in order to find the characteristic tone sequences (or the patterns), we analyze each tone sequence of different lengths ($N = 2, \ldots, 5$) of the different styles. For each tone of the corpus, a sequence is created by concatenating the tone and the $N-1$ following tones. Then, the frequency of occurrence is calculated, along with the $I(t; S)$ and $I(t\overline{t}; S)$ metrics. Two special symbols are added to indicate the start of a news item (SON) and the end of a news item (EON). A different study is carried out considering only those patterns that end with a boundary tone. These tables are separated because Estebas Vilaplana and Prieto (2008, 2010) observed that in Spanish the combination of nuclear accent and boundary tone is usually the most informative part of the utterance in terms of prosody.

Tables 6 and 7 show the most informative sequences for speaking style characterization. The tables are created by selecting, from the whole set of patterns, the most relevant ones according to the values of the $I(t; S)$ and $I(t\overline{t}; S)$ metrics. These metrics allow two different relevance rankings of the patterns to be made. The four patterns that appear simultaneously in the highest positions of both rankings have been chosen. Appendix A shows the top positions of the individual rankings. The patterns of tables 6 and 7 are those that keep with the best balance between the two metrics. The difference between tables 6 and 7 is that the latter only includes patterns that end with a boundary tone.

The results show that the use of sequences of tones seems to provide more information to discriminate style than the use of isolated tones. Thereby, the value of $I(t\overline{t}; S)$ of the most informative isolated tone is 0.0014 for tone H% (see table 5). This value is surpassed by the sequence "L%,H*" in table 6 and by the majority of the sequences of table 7. The value of $I(t; S)$ clearly shows the greater informativeness of the sequences of tones as compared to isolated tones: the highest value of $I(t; S)$ in table 5 is 0.0180 for tone H%. This value is exceeded in one order of magnitude by most of the sequences in tables 6 and 7, reaching the value 0.3675 for the sequence "L*,L%,H*,L+H*".

Some of the tones that do not provide any information in isolation, indeed do so when they are part of a sequence. This happens with tone L+!H*, which in table 5 has $I(t\overline{t}; S) \approx 0$ and in table 6 is among the most informative when it goes before tones H% and L%.

Sequence of tones	$p_{s=newscaster\|t}$	$p_{s=announcer\|t}$	$I(t\bar{t}; S)$	$I(t ; S)$	#
L%,H*	0.248	**0.752**	0.0019	0.1860	938
L+!H*,H%	**0.698**	0.302	0.0007	0.1200	539
H*,L+¡H*	0.257	**0.743**	0.0006	0.1732	300
L+H*,L*	0.320	**0.680**	0.0005	0.0915	512
L*,L%,H*	0.183	**0.817**	0.0012	0.3065	344
L%,H*,L+H*	0.225	**0.775**	0.0009	0.2248	342
L%,H*,none	0.264	**0.736**	0.0005	0.1625	288
L+!H*,L%,H*	0.192	**0.808**	0.0004	0.2879	125
none,L*,L%,H*	0.195	**0.805**	0.0006	0.2821	195
L*,L%,H*,L+H*	0.156	**0.844**	0.0005	0.3675	128
!H%,none,L*,L%	0.223	**0.777**	0.0003	0.2282	112
none,none,L+H*,L*	0.282	**0.718**	0.0002	0.1379	103
L+H*,H%,none,L+H*,H%	**0.683**	0.317	0.0003	0.1031	221
L+H*,L%,none,L+H*,H%	**0.752**	0.248	0.0002	0.1980	101
H%,none,L+H*,H%,none	**0.695**	0.305	0.0002	0.1173	151
none,none,L+H*,L%,none	**0.638**	0.362	0.0001	0.0587	105

Table 6: Most informative sequences. The χ-Square test applied to the marginals $n_{t|s}$ with t the patterns of the table reveals a dependence between the tones t and the type of speaker s with significant results (p-value<2.2e-16 , χ^2= 774.7788, df = 15).

As the length of the patterns increases, the informativeness decreases if the metric $I(t\bar{t}; S)$ is taken into account. Nevertheless, if the metric $I(t; S)$ is taken into account, the amount of information that some sequences like "none,L*,L%,H*" (for $N = 4$) provide is higher than that provided by the patterns of shorter length. Moreover, as the length of the pattern increases, the number of occurrences of that pattern in the corpus decreases. This affects the value of $I(t\bar{t}; S)$ a lot more than the value of $I(t; S)$. This effect can be seen in the tables of the appendix. In tables 6 and 7, the patterns with the best balance between the two metrics have been selected. Choosing a pattern with a high value of $I(t; S)$ but a low value of $I(t\bar{t}; S)$ has the risk of taking as characteristic a pattern whose number is insignificant.

In the following sections, we show that these pattern sequences allow style discrimination in perceptual tests and they are consistent with the observations found in the literature about newscasting speaking style.

4. Consistency of the characteristic prosodic patterns

Radio news style is a type of professional speech used by newscasters to deliver news. Prosody in radio news is performed with very recognizable patterns (Rodero, 2013; Medrado et al., 2005). When applied to a message with a persuasive purpose, the prosody used by reporters to read radio news seeks to avoid monotony in order to capture and maintain the listener's attention. To do so, journalists highlight the information conveyed by raising the melodic contrast and stressing words not semantically relevant. This is one of the conclusions drawn by Bolinger (1998). This author characterized American radio newscasters as reading the messages "mechanically" and with a tendency to emphasize unstressed words as prepositions or auxiliary

Sequence of tones	$p_{s=newscaster\|t}$	$p_{s=announcer\|t}$	$I(t\bar{t};S)$	$I(t;S)$	#
L+!H*,H%	**0.698**	0.302	0.0028	0.1279	477
L+!H*,L%	0.364	**0.636**	0.0018	0.0468	800
L+!H*,=%	**0.659**	0.341	0.0010	0.0832	252
L*,L%	0.372	**0.628**	0.0037	0.0408	1857
L+>H*,H%	**0.636**	0.364	0.0006	0.0616	214
H*,L*,L%	0.241	**0.759**	0.0012	0.1856	116
L+H*,L*,L%	0.302	**0.698**	0.0025	0.1032	437
L+H*,L+!H*,H%	**0.720**	0.280	0.0011	0.1598	125
none,L+!H*,L%	0.328	**0.672**	0.0010	0.0761	241
none,L+H*,L*,L%	0.309	**0.691**	0.0033	0.1070	324
H*,none,L+H*,!H%	0.364	**0.636**	0.0011	0.0533	220
L+H*,none,L*,L%	0.375	**0.625**	0.0010	0.0447	248
L+H*,none,L+H*,H%	**0.585**	0.415	0.0006	0.0213	284
none,L+H*,none,L*,L%	0.385	**0.615**	0.0016	0.0521	161
none,L+H*,none,L+H*,H%	**0.631**	0.369	0.0011	0.0367	149
none,L+H*,none,L+H*,L%	0.429	**0.571**	0.0006	0.0236	140
none,none,L+H*,L+H*,!H%	0.462	**0.538**	0.0002	0.0095	117

Table 7: Most informative sequences that end with a boundary tone. The χ-Square test applied to the marginals $n_{s|t}$ with t the patterns of the table reveals a dependence between the tones t and the type of speaker s with significant results (p-value<2.2e-16 , χ^2= 416.5018, df = 15).

verbs in order "to sound impressive" (Bolinger, 1998, pg. 727). This tendency to focus on non-semantically relevant words has been observed in other studies (Rodero, 2007). As a consequence, broadcasting speaking style does not consist of neutral speech; rather it is a marked style from the standpoint of prosody. For this reason, studies relating to the reading of news on the radio characterize it as a style of speech that uses emphatic prosodic patterns, maintained in a high level of pitch, combined with a fast speech rate and few pauses providing a regular reading pace. At the same time, radio newscasters have been observed to have a higher pitch than that used in natural speech or conversation. Price (2008, pg. 305) defines this as an "overall intonation template". In this respect, Cotter's study (1993) showed that newscasters engage in a very definite style which is characterized by high pitch and high variability compared to the patterns found in conversation. Meanwhile, Grawunder et al. (2008) described German newscasters' style as having a higher pitch range than their peer reporters.

Our results confirm this thesis. Tables 6 and 7 show that the characteristic patterns of radio news style contrast with the patterns of the announcers. The low tones L* and L% are frequently used by the latter. 85% of the patterns in table 6 have a tone L%. The tone L* only appears once among the patterns that characterize radio news style. This fact is in contrast with the massive appearance of high tones (H% and H* in several configurations and alternatives) among the characteristic patterns of radio news style. Also in table 6, it can be observed that only one of the newscasting sequences does not have a high tone. In table 7, only one of the characteristic patterns of radio news style does not end in tone H%, ending in tone =%, instead a boundary tone that also contrasts with the boundary tone L%, as it usually has higher F0 values and not a descending boundary but a suspending one. These results confirm that radio news style is based on an emphatic intonation, as showed in the mentioned studies.

It should be noted that there are some patterns that appear recurrently as the length of the patterns increases. This is the case of the sequence "L%,H*", which also appears in "L*,L%,H*"; "L%,H*,L+H*"; "L%,H*,none"; "L+!H*,L%,H*"; "none,L*,L%,H*"; and "L*,L%,H*,L+H*". These patterns, along with the pattern "L*,L%", seem to be the ones that best characterize the style of announcers reading news. Something similar happens with the pattern "L+!H*,H%", which seems to characterize the radio news style. The systematic and repetitive use of long prosodic schemes, with intentionally high tones such as L+H* and H%, interspersed with unaccented words (tone "none"), configures the typical singsong of these speakers, described by several authors, as commented previously.

The systematic repetition of emphatic/rhythmic patterns is also observed in our results. Among radio newscasting patterns, it is frequent to find sequences of considerable length which combine the tones L+H*and H%. For instance, in table 6, the sequence "none, L+H*,H%" and other similar sequences, can be observed. The tone L+H* is associated with emphatic intonation. Newscasters often need to emphasize the words of the message in order to try to keep the listeners attention.

For European Spanish, Rodero (2013) confirmed the constant presence of emphatic contours, regularity in the use of pitch patterns and constant emphatic stress, mantained at a high pitch range with the aim of focusing the listener's attention on the data presented in the news. As Wheatley states, "inappropriate emotional intonation, perhaps, springs from the desire to put a great deal of expression into one's speech" (1949, pg. 213). In de-la Mota and Rodero (2012), the authors showed the presence of ascending demarcation pitches in intermediate phrases, the use of phonic resources to mark the beginning of a new prosodic unit, and various modifications in the accent pattern. This prosodic pattern is often common at the end of declarative sentences and is characterized by a strong, fast pitch rise followed by descending pitch in the same word. In our results, radio newscasters have few descending patterns. As it can be seen in table 7, the nuclear configurations that appear with a higher frequency among the characteristic patterns of the announcers are of the type "L*,L%" or "L+!H*,L%" or "L+H*,!H%". Table 6 shows that the tone following the boundary tone L% is usually tone H*. These sequences are less frequent among the patterns associated to newscasting. These configurations associated to descending boundaries - known in Spanish as "cadence" format (L* L%) or "semi-cadence" L+H* !H% - , are used by speakers to denote a more interpreted, more paused and planned speech. Among the newscasting characteristic patterns these configurations never appear. The typical boundaries of radio news style are rising patterns of the type "L+H*,H%". This way of cueing the boundary is called "anti-cadence". When a fragment presents the "anti-cadence" configuration, it indicates that the speaker has not finished talking and relevant information remains to be added to reach the full meaning of the statement. Possibly, the speaker tries to keep the attention of the listener, who will wait until the utterance is finished.

If the patterns that end with a boundary tone are compared with the rest of the patterns, the former do not seem to be more characteristic than the latter. In fact, the values of $I(t; S)$ are in general smaller in table 7 than in table 6. However, for similar values of $I(t; S)$, they usually have a higher value of $I(t\bar{t}; S)$. This is because the number of patterns ended with a boundary tone is lower than the total number of patterns, and this has a considerable impact on the computation of the metrics. For practical purposes, using this type of patterns to characterize a style can have advantages, as they are easier to locate automatically (Aguilar et al., 2009), but it is not clear that they are the most determinant to characterize style in this case.

Related to this is the fact that the combinations of a boundary tone plus initial emphatic tone

	1	2	3	4	5
Q1	9	15	25	42	133

	Positive	Negative	Both	None
Q2	141	38	36	9

Table 8: Distribution of answers in the results of the perceptual test. The sub-table entitled Q1 corresponds to the questions referring to perceptual differences between the members of the pair of utterances: 5 indicates clear differences. The sub-table entitled Q2 corresponds to the test of radio news speaker style identification. Positive is the number of times the news speaker is identified. Negative is the number of times the advertising speaker is identified as a news speaker. Both and None are the number of times that listeners assess that both or none of the utterances could belong to a radio news speaker.

seem to be a quite relevant pattern in marking style. The high presence of the sequence "L%,H*" has already been discussed. The patterns "H%,L+>H*" and "SON,L+>H*' appear in table A.9. The last one deserves special attention because it is the only time a beginning of utterance accent appears as relevant.

5. Subjective evaluation

The first goal of the perceptual test was to assess whether several pairs of utterances, with the same text content but with two different sequences of tones associated to them, were perceived differently. The second goal was to evaluate whether the sequences of tones that had been identified as characteristic patterns of the radio news style were associated with this style by listeners.

A set of sentence pairs were selected from the corpus so that, in each pair, one of the utterances belonged to the radio news style and the other to an announcer. The speakers in each pair uttered the same text. An automatic script divided the news items into sentences and checked whether the sentences contained the characteristic patterns listed in tables 6 and 7. The selected sentences had to contain characteristic prosodic patterns of the speaking style to which the speaker belonged. Additionally, this sentence could not contain any characteristic prosodic patterns of the other style to be selected.

As the goal of the test was to evaluate the representativeness of the sequences of tones, we intended to minimize the impact of speech rate and pause duration in the decisions of the listeners. The pairs were selected so that the duration of one of the sentences was not permitted to be 10% longer than the duration of the other one. The final testing set was compiled by applying a random selection from all the pairs satisfying those criteria. Sixteen pairs of sentences were selected, corresponding to the number of pairs of speakers of the different styles.

The following questionnaire was elaborated:

Q1 To what extent do you perceive differences in how these utterances are expressed regardless of the fact that they are in different voices? The possible answers are: 1 (I do not perceive any difference), 2, 3, 4, 5(the pair of utterances are clearly different).

Q2 Which of these two utterances would you most likely hear in radio news services? The possible answers are: [utterance A, utterance B, both, none of them]

The first question was included to obtain information about how the use of different prosodic patterns was perceived. The second question was included to assess whether, in the cases where the pair of sentences were differently perceived, the radio news style could be identified.

A web interface was programmed to ask the listeners about their demographics (age, sex and residence) and additional questions about profession, degree of knowledge in linguistics or communication, relationship with the media, and whether they are regular listeners of radio. The test was completed by 14 listeners, resulting in a total of 224 answers presented in table 8.

The results corresponding to question Q1 show that it was easy for the listeners to identify differences between the pairs of utterances. The mean value of Q1 is 4.2 and its standard deviation is 1.1. The t-student test indicates that this value is significantly greater than 4 with p-value =0.001495.

As for Q2, the percentage of positive answers (informants that identify the radio newscaster correctly) contrasts with the percentage of informants that judge the announcer or none of the speakers to be the radio newscaster (79.2% vs. 20.8%). The binomial test applied to these percentages, shows that most of the informants identify the radio newscasting speakers with a p-value < 2.2e-16.

The correct style identification percentages of the particular radio newscasting speakers were f11r: 87.5%; f13r: 85.7%; m14r: 73.2% and m12r: 69.6% (baseline is 50% in all cases). The lowest value corresponds to *m12r*. Going back to table 2, we observe that *m12r* behaves differently than the rest of *newscasters*: the closest speaker is the announcer *m09a* (I(T;S)=0.004) and the most distant speaker is the newscaster *f11r* (I(T;S) = 0.021) . This apparently anomalous behavior of *m12r* and the reported consistency among subjective results and objective metrics will be discussed in the following section. [2]

6. Discussion

The purpose of this study was the automatic characterization of prosodic patterns by contrasting prosodic styles of two speakers or two group of speakers in this case, radio newscasters and announcers. The use of the Autosegmental-Metrical conventions and the Sp_ToBI labels has been shown to provide an easy way to interpret the representation of the characteristic patterns of radio news style. The method presented in this work is based on the analysis of sequences of symbolic qualitative labels, such as Sp_ToBI labels. Other type of symbolic representations of the prosodic contours such as MoMEL (Mouline et al., 2004), MeLos (Obin, 2011), RaP (Dilley et al., 2006) and others could be used in future studies by following the same methodology. Nevertheless, the Autosegmental-Metrical conventions have been useful for identifying the recurrent patterns that radio news speakers use for capturing attention, as they consider the link between the sequences of symbols and the corresponding meaning or communicative function, which is a challenge for other prosodic annotation systems.

The labels generated by our automatic labeling system were validated during the automatic labeling system training stage. Once the automatic labeling system has been trained, the automatic labels generated by the system have been not manually revised as in other works like (Syrdal et al., 2001). As pointed out in the introduction, manual revision in this kind of applications, in which corpora of more than six hours long are processed, is practically unfeasible. Despite the fact that the automatic predictions have been not reviewed, the patterns that have been obtained from the application of the methodology have been shown to be informative. This result

[2] The samples are available at https://www.infor.uva.es/~descuder/testNews/. The whole Glissando corpus is available at http://veus.glicom.upf.edu/

encourages the use of automatic labels in future applications like, for example, text to speech following a similar scheme as the one following in (Obin et al., 2011; Obin and Lanchantin, 2015) with the MeLos discrete symbols.

This methodology permits to find characteristic prosodic patterns of a given speaker or of a given speaking style analyzing sequences of automatic Sp_ToBI labels. Therefore, our proposal tries to complement other methodologies that characterize speaking style by using other features, like the first and second formant of the vowels (Eskénazi, 1992) number and duration of the prosodic units and pauses (Degand et al., 2009; Goldman et al., 2009), lexical features (Adda-Decker and Lamel, 1999; Graciarena et al., 2006; Shriberg et al., 2009; Wang et al., 2013) the disfluencies (Moniz et al., 2014; Obin et al., 2008), dynamics of the acoustic features (Higuchi et al., 1997; Mixdorff et al., 2005; Tarns and Tatham, 2000; Kokenawa et al., 2005) among others. In addition to this, our proposal contributes to understand the way speakers organize their discourse by building sequences of boundaries and prominent words with a communicative purpose.

The statistical analysis methods, compared to methods based on theoretical models have a limited scope, in the sense that they allow the representation of at most as much information as there is in the training corpus. This fact, inevitably, is problematic when characterizing style from a corpus, because it runs the risk of characterizing speakers instead of style. In fact, as shown in the results, not all the speakers seem equally representative when characterizing the radio news style. This result is not surprising because, generally, in media and particularly in radio broadcasting, it is usual that the speakers particularize their utterances using a personal style. Moreover, most announcers are actors, and adapting to radio news style may not be difficult for them. The experimental procedure and results of this paper have shown that our methodology is able to detect cases in which a speaker behaves differently from the rest of speakers in their group.

7. Conclusions

In this paper we have presented an original methodology for characterizing speaking style. Our proposal is based on the calculation of the information provided by the different sequences of prosodic labels to discriminate the style. The relevance of the different sequences of prosodic labels (or patterns) for characterizing the speaking style can be ranked using these metrics of information.

The application of the methodology to a corpus of radio speech has permitted a set of prototypical prosodic patterns of news speaking style to be identified. The capabilities of the patterns for identifying speaking style have been successfully tested in subjective and objective tests.

The use of the well-known ToBI standard for defining the prosodic patterns has easily permitted the consistency of these patterns to be contrasted with respect to the ones that were expected, according to the observations reported in the state of the art about news speaking style.

The methodology has shown its capacity to obtain more information than the one that classical speaking style approaches can, mainly based on measuring the variation of acoustic features. Although the information given by the prosodic patterns can be useful in characterization activities, it can not be a substitute for the information given by the acoustic features, which seems to be more relevant in identification tasks.

8. Acknowledgments

We would like to thank Professors Carlos Vivaracho and Valentín Cardeñoso for the contributions to the programming of the automatic labeling system. Thanks also to Lourdes Aguilar, Mercedes Cabrera, Francisco Vizcaíno and Eva Estebas for the discussions about the Sp_ToBI labeling procedure. Project reference TIN2014-59852-R (MINECO/FEDER, UE).

References

Adami, A. G., Mihaescu, R., Reynolds, D., Godfrey, J. J., et al., 2003. Modeling prosodic dynamics for speaker recognition. In: Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on. Vol. 4. IEEE, pp. IV–788.

Adda-Decker, M., Lamel, L., 1999. Pronunciation variants across system configuration, language and speaking style. Speech Communication 29 (2), 83–98.

Aguilar, L., Bonafonte, A., Campillo, F., Escudero, D., Septembre 2009. Determining intonation boundaries from the acoustic signal. In: Proceedings of Interspeech, Briston UK.

Arndt, C., 2001. Information Measures: information and its description in science and engineering. Springer.

Beckman, M., January 2002. Intonation across Spanish in the Tones and Break Indices framework. Probus 14, 9–36.

Bolinger, D., 1998. The network tone of voice. Journal of Broadcasting 26, 726–728.

Cotter, C., 1993. Prosodic aspects of broadcast news register. In: Proceedings of the Annual Meeting of the Berkeley Linguistics Society. Vol. 19. pp. 90–100.

de-la Mota, C., Rodero, E., 2012. La entonación en la información radiofónica. Quaderns de filología: estudis de comunicació 45, 159–176.

Degand, L., Simon, A. C., et al., 2009. Mapping prosody and syntax as discourse strategies: how basic discourse units vary across genres. Where Prosody Meets Pragmatics 8, 79.

Dilley, L., Breen, M., Bolivar, M., Kraemer, J., Gibson, E., 2006. A comparison of inter-transcriber reliability for two systems of prosodic annotation: rap (rhythm and pitch) and tobi (tones and break indices). In: INTERSPEECH.

Elordieta, G., 2011. Transcription of intonation of the Spanish language. In: Estudios de Fonética Experimental. Vol. XX. pp. 273–293.

Escudero, D., Aguilar, L., Bonafonte, A., Garrido, J., September 2009. On the definition of a prosodically balanced copus: combining greedy algorithms with expert guided manipulation. Revista de la Sociedad Española de Procesamiento del Lenguaje Natural 43, 93–102.

Escudero, D., Aguilar, L., González, C., Cardeñoso, V., Gutiérrez, Y., May 2014a. Preliminary results on Sp_ToBI prosodic labeling assisted by an automatic fuzzy classifier. In: Proceedings of the 7th International Conference on Speech Prosody. Dublin, Ireland, pp. 457–461.

Escudero, D., Aguilar, L., González, C., Gutiérrez, Y., Cardeñoso, V., May 2014b. On the use of a fuzzy classifier to speed up the Sp_ToBI labeling of the Glissando Spanish corpus. In: Chair), N. C. C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (Eds.), Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), Reykjavik, Iceland, pp. 1962–1969.

Escudero, D., Aguilar, L., Vanrell, M. d. M., Prieto, P., May 2012. Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labeling system. Speech Commun. 54 (4), 566–582.

Escudero, D., Cardeñoso, V., Bonafonte, A., May 2002. Corpus based extraction of quantitative prosodic parameters of stress groups in Spanish. In: Proceedings of ICASSP 2002. Vol. 1. pp. 481–484.

Escudero-Mancebo, D., González-Ferreras, C., Vivaracho-Pascual, C., Cardeñoso Payo, V., 2014. A fuzzy classifier to deal with similarity between labels on automatic prosodic labeling. Computer Speech and Language 28 (1), 326 – 341.

Eskénazi, M., 1992. Changing speech styles: Strategies in read speech and casual and careful spontaneous speech. In: Proceedings ICSLP. pp. 755–758.

Eskénazi, M., 1993. Trends in speaking style research, keynote speech. In: Proceedings of Interspeech. pp. 501–509.

Estebas Vilaplana, E., Prieto, P., 2008. La notación prosódica en español. una revisión del Sp_ToBI. Estudios de Fonética Experimental XVIII, 263–283.

Estebas Vilaplana, E., Prieto, P., 2010. Castilian Spanish Intonation. In: Prieto, P., Roseano, P. (Eds.), Transcription of Intonation of the Spanish Language. Lincom Europa, München, pp. 17–48.

Face, T., Prieto, P., 2007. Rising accents in Castilian Spanish: a revision of Sp-ToBI. Journal of Portuguese Linguistics 6.1, 117–146.

Garrido, J.-M., Escudero, D., Aguilar, L., Cardeñoso, V., Rodero, E., de-la Mota, C., González, C., Vivaracho, C., Rustullet, S., Larrea, O., Laplaza, Y., Vizcaíno, F., Estebas, E., Cabrera, M., Bonafonte, A., 2013. Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan. Language Resources and Evaluation 47 (4), 945–971.

Goldman, J.-P., Auchlin, A., Avanzi, M., Simon, A. C., et al., 2008. Prosoreport: an automatic tool for prosodic description. application to a radio style. Proceedings of Speech Prosody'08.

Goldman, J.-P., Auchlin, A., Simon, A. C., 2009. Discrimination de styles de parole par analyse prosodique semi-automatique. Actes d'IDP, Paris, ISSN, 2114–7612.

Goldman, J.-P., Prisr, T., Christodoulides, G., Auchlin, A., 2014. Speaking style prosodic variation: an 8-hour 9-style corpus study. In: Proceedings of Speech Prosody.

Gonzalez-Ferreras, C., Escudero-Mancebo, D., Vivaracho-Pascual, C., Cardeñoso Payo, V., sept. 2012. Improving automatic classification of prosodic events by pairwise coupling. Audio, Speech, and Language Processing, IEEE Transactions on 20 (7), 2045 –2058.

Grabisch, M., 1995. Fuzzy integral in multicriteria decision making. Fuzzy sets and Systems 69 (3), 279–298.

Graciarena, M., Shriberg, E., Stolcke, A., Enos, F., Hirschberg, J., Kajarekar, S., 2006. Combining prosodic lexical and cepstral systems for deceptive speech detection. In: Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. Vol. 1. IEEE, pp. I–I.

Grawunder, S., Engert, H., Kaiser, H., Unger, S., Bose, I., 2008. Pitch, speech rate and rhythm as speaking style characterizing parameters in german news presentation. EASR08, UCL London.

Hastie, T., Tibshirani, R., et al., 1998. Classification by pairwise coupling. The annals of statistics 26 (2), 451–471.

Higuchi, N., Hirai, T., Sagisaka, Y., 1997. Effect of speaking style on parameters of fundamental frequency contour. In: Progress in speech synthesis. Springer, pp. 417–428.

Hirschberg, J., 2000. A corpus-based approach to the study of speaking style. In: Prosody: Theory and experiment. Springer, pp. 335–350.

Kokenawa, Y., Tsuzaki, M., Kato, H., Sagisaka, Y., 2005. F0 control characterization by perceptual impressions on speaking attitudes using multiple dimensional scaling analysis. Proc. ICASSP, SP-P1 3, 1–273.

Krippendorff, K., 1986. Information Theory, structural models for qualitative data. SAGE University Papers.

Ladd, D. R., 1996. Intonational Phonology. Cambridge University Press, Cambridge.

Llisterri, J., 1992. Speaking styles in speech research. In: Workshop on Integrating Speech and Natural Language.

Medrado, R., Ferreira, L. P., Behlau, M., 2005. Voice-over: perceptual and acoustic analysis of vocal features. Journal of Voice 19 (3), 340–349.

Mixdorff, H., Pfizinger, H., Granwinkel, K., 2005. Towards objective measures for comparing speaking styles. In: Proceedings of SPECOM. Patras, Greece, pp. 131–134.

Moniz, H., Batista, F., Mata, A. I., Trancoso, I., 2014. Speaking style effects in the production of disfluencies. Speech Communication 65 (0), 20 – 35.

Mouline, S., Boeffard, O., Bagshaw, P. C., 2004. Automatic adaptation of the momel f0 stylisation algorithm to new corpora. In: INTERSPEECH.

Obin, N., 2011. Melos: Analysis and modelling of speech prosody and speaking style. Ph.D. thesis, Université Pierre et Marie Curie-Paris VI.

Obin, N., Dellwo, V., Lacheret, A., Rodet, X., et al., 2010. Expectations for discourse genre identification: a prosodic study. In: Interspeech. pp. 3070–3073.

Obin, N., Lacheret, A., Veaux, C., Rodet, X., Simon, A.-C., et al., 2008. A method for automatic and dynamic estimation of discourse genre typology with prosodic features. In: Interspeech.

Obin, N., Lanchantin, P., March 2015. Symbolic modeling of prosody: From linguistics to statistics. Audio, Speech, and Language Processing, IEEE/ACM Transactions on 23 (3), 588–599.

Obin, N., Lanchantin, P., Lacheret, A., Rodet, X., 2011. Discrete/continuous modelling of speaking style in hmm-based speech synthesis: Design and evaluation. In: Interspeech.

Padró, L., Stanilovsky, E., 2012. Freeling 3.0: Towards wider multilinguality. In: Proceedings of the Language Resources and Evaluation Conference.

Pierrehumbert, J. B., 1980. The phonology and phonetics of English intonation. Ph.D. thesis, MIT.

Price, J., 2008. New news old news: a sociophonetic study of spoken Australian English in news broadcast speech. AAA-Arbeiten aus Anglistik und Amerikanistik 33 (2).

Rodero, E., 2007. Characterization of a proper news presentation in the audio-visuals messages. Estudios del Mensaje Periodístico 13, 523–542.

Rodero, E., 2013. Peculiar styles when narrating the news: The intonation of radio news bulletins. Estudios del Mensaje Periodístico 19 (1), –.

Roekhaut, S., Goldman, J., Simon, A., et al., 2010. A model for varying speaking style in tts systems. In: Proceedings of Speech Prosody.

Rosenberg, A., 2010. AuToBI-a tool for automatic toBI annotation. In: INTERSPEECH. pp. 146–149.

Rosenberg, A., 2013. Modeling prosodic sequences with K-Means and Dirichlet process GMMs. In: Proceedings of

Interspeech.
Shriberg, E., Favre, B., Fung, J., Hakkani-Tur, D., Cuendet, S., 2009. Prosodic similarities of dialog act boundaries across speaking styles. Linguistic Patterns in Spontaneous Speech (A25), 213–239.
Shriberg, E., Ferrer, L., Kajarekar, S., Venkataraman, A., Stolcke, A., 2005. Modeling prosodic feature sequences for speaker recognition. Speech Communication 46 (3), 455–472.
Syrdal, A. K., Hirshberg, J., McGory, J., Beckman, M., 2001. Automatic ToBI prediction and alignment to speed manual labeling of prosody. Speech Communication (33), 135–151.
Tarns, A., Tatham, M., 2000. Intonation for synthesis of speaking styles. In: IEE Seminar: State-of-the-art in speech synthesis.
Taylor, P., 2000. Analysis and synthesis of intonation using the tilt model. The Journal of the acoustical society of America 107 (3), 1697–1714.
Wang, W. Y., Biadsy, F., Rosenberg, A., Hirschberg, J., 2013. Automatic detection of speaker state: Lexical, prosodic, and phonetic approaches to level-of-interest and intoxication classification. Computer Speech & Language 27 (1), 168–189.
Wheatley, K. E., 1949. Anomalies of radio speech. American Speech, 213–215.

Sequence of tones	$p_{s=newscaster\|t}$	$p_{s=announcer\|t}$	$I(t\bar{t};S)$	$I(t;S)$	#
L*,none	**0.758**	0.242	0.0003	0.2068	128
L%,H*	0.248	**0.752**	0.0019	0.1860	938
H*,L+¡H*	0.257	**0.743**	0.0006	0.1732	300
SON,L+>H*	**0.721**	0.279	0.0002	0.1502	111
L*,L%,H*	0.183	**0.817**	0.0012	0.3065	344
L+!H*,L%,H*	0.192	**0.808**	0.0004	0.2879	125
L%,H*,L+H*	0.225	**0.775**	0.0009	0.2248	342
L%,H*,L+¡H*	0.243	**0.757**	0.0003	0.1940	152
H*,L*,L%	0.246	**0.754**	0.0003	0.1900	122
L%,H*,none	0.264	**0.736**	0.0005	0.1625	288
L+H*,H%,L+>H*	**0.726**	0.274	0.0003	0.1574	186
H*,L+¡H*,!H%	0.274	**0.726**	0.0002	0.1488	117
L+H*,L+!H*,H%	**0.714**	0.286	0.0002	0.1414	140
L*,L%,H*,L+H*	0.156	**0.844**	0.0005	0.3675	128
none,L*,L%,H*	0.195	**0.805**	0.0006	0.2821	195
!H%,none,L*,L%	0.223	**0.777**	0.0003	0.2282	112
L+H*,L*,L%,L+H*	0.256	**0.744**	0.0002	0.1736	117
L+H*,L%,H*,L+H*	0.268	**0.732**	0.0002	0.1562	123
none,none,L+H*,L*	0.282	**0.718**	0.0002	0.1379	103
none,L+!H*,H%,none	**0.706**	0.294	0.0002	0.1311	109
L+H*,L%,none,L+H*,H%	**0.752**	0.248	0.0002	0.1980	101
none,L+H*,H%,none,L+H*,H%	**0.743**	0.257	0.0002	0.1823	101
L+H*,H%,none,L+H*,H%,none	**0.717**	0.283	0.0002	0.1449	113

Table A.9: Sequences of tones sorted in terms of the metric $I(t;S)$. Only tones with $I(t;S) > 0.13$ are displayed.

Appendix A. Characteristic patterns

In this appendix four tables containing the list of the most informative patterns when characterizing the target style are presented. In all the tables, the first column contains the pattern or the sequence of tones. The second and third columns contain the relative frequency of occurrence of the pattern in each style. Boldface is used to highlight in which type of speaker the pattern is more frequent.

The fourth column is the value of the metric $I(t;S)$ corresponding to the pattern t of the row. This metric has been computed as indicated in Equation 4. The fifth column is the value of the metric $I(t\bar{t};S)$ computed as indicated in Equation 3.

Tables A.9 and A.10 differ from tables A.11 and A.12 as the latter contain only final patterns (ended by boundary tone). Tables A.9 and A.11 are sorted in terms of $I(t;S)$ while tables A.10 and A.12 are sorted $I(t\bar{t};S)$.

The Chi-Square test was applied to the marginals $n_{s|t}$ with the patterns t in the tables. Significant results with p-value<2.2e-16 that there is a dependence between the tones t and the type of speaker s: X-squared = 861.5542, df=22 for table A.9; X-squared = 1545.007, df=16 for table A.10; X-squared = 743.405, df=24 for table A.11; and X-squared = 771.705, df=16 for table A.12.

Sequence of tones	$p_{s=newscaster\|t}$	$p_{s=announcer\|t}$	$I(t\bar{t};S)$	$I(t;S)$	#
L%,H*	0.248	**0.752**	0.0019	0.1860	938
L*,L%	0.368	**0.632**	0.0010	0.0480	1926
L+H*,H%	**0.569**	0.431	0.0008	0.0151	4803
L+!H*,H%	**0.698**	0.302	0.0007	0.1200	539
L+!H*,L%	0.346	**0.654**	0.0006	0.0666	859
H%,L+H*	**0.610**	0.390	0.0006	0.0374	1340
H*,L+¡H*	0.257	**0.743**	0.0006	0.1732	300
L+H*,L*	0.320	**0.680**	0.0005	0.0915	512
H%,none	**0.567**	0.433	0.0005	0.0144	3278
L*,L%,H*	0.183	**0.817**	0.0012	0.3065	344
L%,H*,L+H*	0.225	**0.775**	0.0009	0.2248	342
none,L+H*,H%	**0.599**	0.401	0.0008	0.0308	2279
L+H*,L*,L%	0.294	**0.706**	0.0006	0.1220	456
L%,H*,none	0.264	**0.736**	0.0005	0.1625	288
L+H*,L%,H*	0.297	**0.703**	0.0005	0.1178	370
none,L*,L%,H*	0.195	**0.805**	0.0006	0.2821	195
L*,L%,H*,L+H*	0.156	**0.844**	0.0005	0.3675	128

Table A.10: Sequences of tones sorted in terms of the metric $I(t\bar{t};S)$. Only tones with $I(t\bar{t};S) > 0.0005$ are displayed.

Sequence of tones	$p_{s=newscaster|t}$	$p_{s=announcer|t}$	$I(t\bar{t} : S)$	$I(t : S)$	#
L+!H*,H%	**0.698**	0.302	0.0028	0.1279	477
L+!H*,=%	**0.659**	0.341	0.0010	0.0832	252
L+>H*,H%	**0.636**	0.364	0.0006	0.0616	214
L+¡H*,LH%	0.354	**0.646**	0.0004	0.0544	161
L+!H*,L%	0.364	**0.636**	0.0018	0.0468	800
L+!H*,LH%	**0.610**	0.390	0.0003	0.0420	154
L*,L%	0.372	**0.628**	0.0037	0.0408	1857
H*,L*,L%	0.241	**0.759**	0.0012	0.1856	116
L+H*,L+!H*,H%	**0.720**	0.280	0.0011	0.1598	125
H*,L+¡H*,!H%	0.298	**0.702**	0.0006	0.1078	104
L+H*,L*,L%	0.302	**0.698**	0.0025	0.1032	437
none,L+!H*,H%	**0.667**	0.333	0.0009	0.0934	165
none,L+!H*,L%	0.328	**0.672**	0.0010	0.0761	241
L+H*,L+!H*,L%	0.333	**0.667**	0.0009	0.0708	222
none,L+¡H*,L%	**0.640**	0.360	0.0008	0.0674	214
none,L+!H*,=%	**0.629**	0.371	0.0003	0.0573	105
none,L+H*,H%	**0.600**	0.400	0.0048	0.0364	2114
H*,L+H*,!H%	0.385	**0.615**	0.0006	0.0312	343
none,L+H*,L*,L%	0.309	**0.691**	0.0033	0.1070	324
H*,none,L+H*,!H%	0.364	**0.636**	0.0011	0.0533	220
none,L+H*,L+!H*,L%	0.368	**0.632**	0.0005	0.0496	114
L+>H*,none,L+H*,!H%	0.369	**0.631**	0.0005	0.0492	103
L+H*,none,L*,L%	0.375	**0.625**	0.0010	0.0447	248
none,L+H*,none,L*,L%	0.385	**0.615**	0.0016	0.0521	161
none,L+H*,none,L+H*,H%	**0.631**	0.369	0.0011	0.0367	149

Table A.11: Sequences of tones (ended by boundary tone) sorted in terms of the metric $I(t; S)$. Only patterns with $I(t; S) > 0.03$ are displayed.

Sequence of tones	$p_{s=newscaster\|t}$	$p_{s=announcer\|t}$	$I(t\bar{t} : S)$	$I(t : S)$	#
"L+H*,H%"	**0.571**	0.429	0.0048	0.0191	4438
"L*,L%"	0.372	**0.628**	0.0037	0.0408	1857
"L+!H*,H%"	**0.698**	0.302	0.0028	0.1279	477
"L+!H*,L%"	0.364	**0.636**	0.0018	0.0468	800
"L+H*,!H%"	**0.460**	0.540	0.0010	0.0027	6133
"L+!H*,=%"	**0.659**	0.341	0.0010	0.0832	252
"none,L+H*,H%"	**0.600**	0.400	0.0048	0.0364	2114
"L+H*,L*,L%"	0.302	**0.698**	0.0025	0.1032	437
"none,L*,L%"	0.396	**0.604**	0.0014	0.0248	942
"H*,L*,L%"	0.241	**0.759**	0.0012	0.1856	116
"L+H*,L+!H*,H%"	**0.720**	0.280	0.0011	0.1598	125
"none,L+!H*,L%"	0.328	**0.672**	0.0010	0.0761	241
"none,L+H*,L*,L%"	0.309	**0.691**	0.0033	0.1070	324
"H*,none,L+H*,!H%"	0.364	**0.636**	0.0011	0.0533	220
"L+H*,none,L*,L%"	0.375	**0.625**	0.0010	0.0447	248
"none,L+H*,none,L*,L%"	0.385	**0.615**	0.0016	0.0521	161
"none,L+H*,none,L+H*,H%"	**0.631**	0.369	0.0011	0.0367	149

Table A.12: Sequences of tones in final possition of the intonation phrase sorted by the value of the metric $I(t\bar{t}; S)$. Only tones with $I(t\bar{t}; S) > 0.001$ are displayed.