Applying a fuzzy classifier to generate Sp_ToBI annotation: preliminar results

David Escudero[1], Lourdes Aguilar[2], César González[1], Valentín Cardeñoso[1], Yurena Gutiérrez[2]

[1]Department of Computer Science, University of Valladolid, Spain
[2]Department of Spanish Philology, Universitat Autònoma de Barcelona, Spain

`descuder@infor.uva.es`

Abstract

One of the goals of the Glissando research project[1] is to enrich a radio news corpus [1] with Sp_ToBI labels. In this paper we present the application of the automatic predictions of a fuzzy classifier to speed up the labeling process. The strategy is proposed after completing the following steps: a) manual annotation of a part of the Glissando corpus with Sp ToBI labels and checking of the coherence of the labels; b) training of the automatic system; c) validation or correction of the automatic system's predictions by a human expert. The automatic judgments of the classifier are enriched with confidence measures that are useful to represent uncertain situations concerning the label to be assigned. The main aim of the paper is to show that there exists a correspondence between the uncertain situations that are identified during an inter-transcriber experiment and the uncertain situations that the fuzzy classifier detects. Labeling time reduction encourages the use of this strategy.

Index Terms:Prosodic labeling, fuzzy classifier, Sp_ToBI

1. Introduction

Prosodic labeling aims to enrich spoken utterances with labels that are representative of the relationship between the prosodic form and function of the constituents of the message. Although prosodic labeling systems establish clear rules and protocols, the difficulty of the task and the inherent subjectivity of the labelers' judgments introduces a high number of inconsistencies. Prosodic labeling systems assume that uncertain situations could appear and reserve special symbols for representing them (like the symbol '?' in ToBI [2] and RaP [3] or the explicit computation of the transcriber disagreement in RPT [4]). Leaving aside these well-known difficulties of the task of manual annotation, the inter-transcriber tests of consistency have identified cases where two different transcribers decide to assign different labels to the same prosodic event. [5] suggested the use of alternative tiers for capturing ambiguities. These facts suggest the existence of an area of uncertainty across the categories, due to the perceptual and acoustic similarity of some pair of labels [6]. Fuzzy sets theory [7] has been widely used to represent those situations where it is difficult to classify a given element into the different possible categories.

It must be noted that recognizing uncertainty in the task of identification of some labels is not equivalent to saying that the prosodic categories used in the ToBI framework are fuzzy categories, since they are based on the description of the intonational phonology of the language, and, as a consequence, each of them is related to a clear phonological content. However, the process of annotation (either manually or with the aid of semi-automatic tools) has shown that the resulting labels can carry uncertain information when they have to be associated to the acoustic signal. In [8] we have shown how fuzzy sets can be used to represent situations where assigning a class to a given prosodic unit is difficult because of the high degree of uncertainty. The BURNC corpus [9] was used in the experiments, which is one of the most important references for studies on automatic ToBI prosodic labeling in English, and in this paper, we will present the application of the same fuzzy classifier to the subset of news of the Glissando corpus[1], which aims to be a reference for studies on Spanish prosody.

Since manual annotation is a time-consuming process and very costly in terms of human resources, efforts have to concentrate on developing tools for automatic prosodic labeling or, at least, to aid the experts to speed the process [10]. The state of the art on automatic prosodic labeling reports identification rates higher than 90% in binary decisions, such as the presence or absence of accent, boundary or break. However, when the system is faced with the classification of pitch accents, boundary tones or level of breaks, the rates dramatically decrease to about 70% (see [11] for a review of the state of the art). In [12], we showed that the reasons for these low accuracy rates are the high similarity among some pairs of classes and the imbalanced nature of the prosodic corpora. As expected, the difficulties of manual annotation are reflected in how successful automatic approaches to prosodic labeling are. This paper aims to show that the use of a fuzzy classifier considerably increases the performance when soft classification is performed, and that there exists a correspondence between the uncertain situations that are identified during the inter-transcriber experiments and the uncertain situations that the fuzzy classifier detects, a reason to consider the application of the tool as a good strategy to speed the process of manual annotation.

The strategy implies fulfilling the following steps: a) manual annotation of part of the Glissando corpus with Sp ToBI labels and checking of the coherence of the labels; b) training of the automatic system; c) correction of the automatic system's predictions by a human expert. Section 2 describes the process of manual labeling of the training corpus and the quality assessment procedure that has been applied. Section 3 presents the architecture and methods of the fuzzy classifier. The small number of editing operations in the revision process (as detailed in Section 4) evidences the quality of the fuzzy predictions. Additionally we show that the most uncertain predictions correspond to labels that are also problematic in the inter-transcriber consistency tests. We end with conclusions and the future work of this ongoing research.

CORPUS	L	W	S	Pitch Accents	Boundary Tones	Breaks
Sp_ToBI (this work)	4	108	2	0.68/78.35%	0.70/85.05%	0.76/88.63%
Cat_ToBI[6]	10	264	4	0.462/61.17%	0.69/86.10%	0.68/77.14%
Am_ToBI(fe)[13]	4	644	2	0.69 / 71%	0.84 / 86%	0.65 / 74%
Am_ToBI(ma)[13]	4	644	2	0.67 / 72%	0.76 / 82%	0.62 / 74%
E_ToBI[14]	26	489	4	na / 68%	na / 85%	na / 67%
E_ToBI[15]	2	1594	1	0.51 / 86.57%	0.79/ 89.33%	na / na

Table 1: Global inter-transcriber agreement results for Sp_ToBI compared with results reported for other ToBI systems. Columns labelled *Pitch Accents*, *Boundary Tones* and *Breaks* separate results according to the respective ToBI events that have been considered. The figure in the cells are the κ index and the pairwise inter-transcriber rate (as a percentage). **L** is the number of labelers, **W** is the size of the corpus in words and **S** is the number of styles. *(fe)* is female, *(ma)* is male and *(na)* means the information is not available.

2. The Sp_ToBI manually labeled subcorpus

The Glissando news subcorpus contains recordings of eight different Spanish speakers, each of them reading more than 36 news items [1]. For our purposes, two of these speakers were chosen, taking into account differences in gender (i.e. male and female) and reading style (i.e. radio speaker and advertisement actor). The labeled corpus consists of 1100 seconds of reading of news speech recorded by two professional speakers: 12 news read by a radio professional (female voice) and 12 news read by an advertising professional (male voice). These news items include a total of 3202 words (7091 syllables) labeled with 2058 pitch accents, 1115 boundary tones and 1029 breaks.

The news data-set has been annotated using the Sp_ToBI labels proposed in [16, 17], with the modifications advanced in [18] and some adjustements needed for the speaking style, contained in the guidelines distributed in http://veus.glicom.upf.edu/. The tonal inventory is adapted to the specific phenomena pertaining to declarative utterances of a news data set in terms of a reduction of the tonal inventory and the definition and representation of boundary tones. In particular, the tag =% is associated to those cases where the pitch keeps the previous tone value (i.e. sustained pitch), and the parentheses stands for allotonic variations of L+H* and L+>H* (that is, when the fall is not perceived in the pre-stressed syllable (L+)H* and (L+>)H* are used).

The procedure was perceptually based: the transcriber was encouraged to focus preferentially on perception: her task consisted in listening carefully to the utterance in order to (a) mark the subjective sense of disjuncture between each pair of words and before each pause (break tier) and (b) mark prominences and tonal events (tone tier).

Since the ToBI framework is phonologically-driven, various methods of estimating the consistency and stability of the labels assigned to the corpus were conducted: (i) periodical meetings to define guidelines to annotate read news; (ii) discussion and resolution of differences in transcription throughout a six-month period and (iii) validation of consistency among transcribers with an interreliability experiment.

In order to measure the confidence of the annotation, four experts labeled independently the same news (108 words) read by a professional speaker. Pair-wise comparisons and values of the kappa index support the stability of the labels among transcribers in the main categories, but they also show that there is confusion among others. The results of the inter-transcriber consistency test can be seen in the table 1. Values of the kappa index between 0.6 and 0.8 like the ones we obtained are com-

monly considered as substantial agreement. These consistency rates are comparable with the ones reported in similar studies for the prosodic labeling of other corpora in different languages (see table 1). Uncertainty exists, which is the main argument that supports the use of a fuzzy classifier.

3. Automatic labeling with uncertainty

We face the automatic labeling of ToBI events following the multi-class classification approach. The multi-class classification problem has the goal to assign a ToBI label to a given prosodic unit that is, typically, a word or a syllable. The multi-class classification approach contrasts with binary classification where the goal is to determine whether an accent or a boundary is present or not in the given prosodic unit.

In [11] we showed that multi-class identification of ToBI labels can be efficiently done by using *pairwise coupling classification*. The complex multi-class classification problem is divided into several simpler problems, by means of pairwise coupling. The basic idea is that it is easy for the machine to assign a label when only two classes are considered to be possible. For example it is easy to assign the label L+H* or the label L* to a prosodic unit when the only alternatives are these two classes. However, assigning the label L+H* when the alternatives are H*, !H*, H+!H*, L+!H*, L* and L*+H is a much more challenging task. Our proposal is to combine several two-class classifiers (one for every pair of possible labels) in order to achieve the multi-class classification because two-class problems provide higher accuracy results.

Furthermore, in [12] we observed that different *types of classifiers* behave differently in the classification of different tones. Thus, decision trees seem to be specialized in the identification of the most populated classes while neural networks tend to balance the number of predicted labels for each of the classes. The compromise in term of the number of samples in every class is important in automatic prosodic labeling because prosodic corpora are naturally imbalanced. For example in the BURNC corpus, 77% of the words are labeled with two out of the eight possible labels[12]. In this work, complementarity between artificial neural networks (NN), decision trees (DT) and support vector machines (SVM) classifiers has been exploited to improve the final system, combining their outputs using a fusion method.

In order to combine the decision scores that result from the three classifications modules (DT, NN and SVM), we used the comprehensive fuzzy technique proposed in [8]. The fuzzy integral technique has proven useful for combining classifiers in several contexts [19, 20, 21, 22, 23, 24]. We use the implemen-
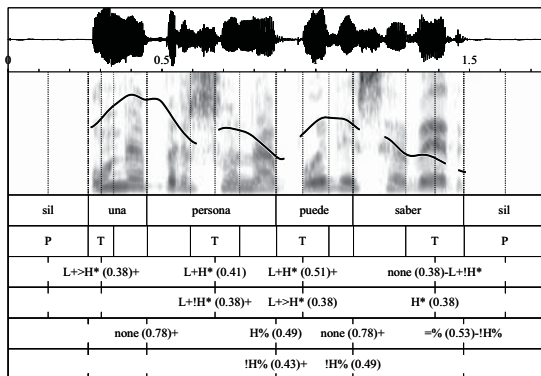
Figure 1: Multi-tier prediction interface that allows a human labeler to check the output of the automatic fuzzy classifier. Word and syllable segmentation appear in the upper tiers (T stands for lexically stressed syllable). The following tiers correspond to the predicted labels for pitch accents (aligned at the middle of the stressed syllable) and boundary tones (aligned at the end of the word).

	Label	Intertranscriber agreement per symbol	Unique label predictions
Pitch Accent	none	53.5%	50.3%
	L+H*	28.6%	37.5%
	L+!H*	6.5%	7.1%
	L+>H*	4.3%	1.7%
	H*	3.2%	0.4%
	L*	3.0%	2.9%
	L+¡H*	0.9%	0.2%
Boundary Tone	none	78.1%	77.3%
	L%	8.1%	10.4%
	H%	7.2%	3.1%
	!H%	4.6%	7.5%
	LH%	1.1%	0.5%
	=%	0.9%	1.1%

Table 2: *Inter-transcriber agreement per symbol* is the number of times (in percentage) that two of the transcribers agree assigning the same symbol to the same prosodic unit. *Unique label predictions* is the number of times (in percentage) that the fuzzy classifier predicts only one symbol per prosodic unit.

tation of the Sugeno fuzzy integral [25] as described in [26]. As result, each ToBI category is assigned to the words of the corpus with a degree of certainty. This degree of confidence is a numeric value in the [0,1] interval (the highest the value the more the certainty). The α-cut approach permits to reduce the set of candidate labels assigned to every word. The α-cut value is empirically assigned as explained in [8].

In order to train the classifiers, we have applied a multifold approach that divides the corpus into two sections: 90% training and 10% test. Some categories show a very low number of instances, so we decided to group them with similar types thereby creating particular classes. To do that, we display the inter-label distance into a *Multidimesional Scaling* (MDS) 2D plot following the perspective adopted in [27]. This MDS map is built with the confusion matrix of a decision tree classifier: the more the inter-class confusion the closer the labels in the map. This plot allows experts to make a decision regarding the different categories. The closest categories are good candidates to be collapsed into an alternative category. As result, we use the following Sp_ToBI pitch accents: H*, L* = {L* ∪ L*+H ∪ H+L* }, L+>H*, L+H* ={L+H* ∪ (L+)H*}, L+!H* ={L+!H* ∪ (L+)!H* ∪ !H*}, L+¡H* ={L+¡H* ∪ (L+)¡H* ∪ ¡H*}; and the following boundary tones: L%, H%, =%, !H%, LH% ={LH% ∪ L!H%}. Additionally, the class "none" represents the absence of tone. After performing that clustering, classification rates improved considerably.

The input of the classifier is composed of acoustic information (F0, energy and duration features) and POS tags as detailed in [11]. More details about this system can be found in [28].

4. Procedure of revision of the automatic system's predictions

The fuzzy classifier has been applied to unseen samples of the Glissando news subcorpus read by different voices than those used in the manual annotation. A total subset of 18 news (6 news read by 3 different voices, 2 female and 1 male) has been annotated by means of the predictions of the fuzzy classifier and a human expert has reviewed all the tags.

Figure 1 illustrates the graphical interface used to present the automatic system's predictions so as to a human expert can verify or correct them. The visual interface aligns the tags predicted by the fuzzy classifier with each prosodic event: pitch accents aligned with the stressed syllables and tone boundaries aligned with the end of the word. The classifier predicts presence or absence of break (with the tag "none"), and if a prosodic rupture exists, the type of boundary tone. On the contrary, information related to the levels of prosodic rupture (break indices) is not present. This information can be inferred from the type of boundary tone, since in our results, there is a statistically significant correlation between BI3 associated to H%, !H%, =%, L!H% and BI4 associated to L% (Pearson's chi squared test X-squared = 742.1301, df = 5, p-value < 2.2e-16). There are only marginal cases in which L% is associated with BI3 , and in which H%, !H%, =%, L!H% are associated with BI4.

Compared with conventional crisp classifiers, the main advantage of a fuzzy classifier is that it can provide more than one label per prosodic unit (to a maximum of three in our system), depending on the uncertainty of the predictions. Each tag is accompanied by a numerical value in the [0,1] interval, the higher the value the more the certainty, and tags are ordered from the highest to the lowest degree. At this point, it should be noted that the procedure, according to the fuzzy set theory, is not based on probabilities since the degree of certainty is independently assigned to each category. This is the reason why the values can sum up more than 1, that is, if three tags appear, we cannot infer that there are three complementary possibilities that sum up 1. On the contrary, having more than one tag in the output represents a difficult situation in which more than one label evidences a degree of certainty over the threshold set by the α-cut. Another situation that can be found is that even when only a tag is predicted, it is not necessarily accompanied by a complete confidence (marked with 1).

The task of the human expert is to evaluate the candidates: he/she checks if some of the proposed tags are the right one according to his/her perception and marks it with "+". If she/he

doesn't agree with any of the tags proposed, he/she attaches "-" and writes a new option.

Figure 1 illustrates different types of situations that the human expert can find in the process of reviewing the output of the fuzzy classifier and that reflects the uncertainty that implies using a phonologically-based system such as ToBI. With respect to tone boundaries, the most certain decision is the label "none" associated to the words "una" (a) and "puede" (can): only one label is predicted, with a high degree of certainty (0.78). In this case, the tag "none" (meaning absence of prosodic break and as a consequence, absence of tone boundary) proposed by the system is validated (+), since any tonal nor segmental cue signal a prosodic break.

On the other hand, at the end of the prosodic group "una persona" (a person) the system suggests two candidates, associated to a very similar degree of certainty: a high tone, H% (0.49) or a mid tone, !H% (0.43). Crucially, the transcriber has to decide if the difference of range is phonologically significant in this context: since the transcriber perceives that the tone decreases to a mid-tone from an L+H* nuclear pitch accent, the second option is selected.

As far as pitch accents is concerned, the difficulty in discriminating between rising accents with or without peak displacement is evidenced in the word "puede" (can). As observed in the pitchtrack, the F0 peak is situated in the syllable border, that is, both stressed and post-stressed syllables show a high tone, a fact that makes difficult the decision. The transcriber chooses the rising tone without peak displacement (L+H*), because it is generally accepted that the high tone should be completely placed in the post-stressed syllable [29].

It may also happen that the proposals of the fuzzy classifier are wrong. It is the case of the pitch accent corresponding to the word "saber" (know). The system proposes two tags: "none", meaning absence of pitch accent and a high tone (H*). Since the stressed syllable is tonally accented, but with a rising accent and not with a high tone, the transcriber dismisses both options and writes the correct one, which is L+!H*. The downstep relates to the immediately preceding high tone within the same prosodic group.

At this point, it should be said that the adequate tuning of the α-cut value (that is the number of candidates that are presented in the interface) is crucial for a correct system operation. Lowering the value of the α-cut yields a higher number of positive cases, understanding as positive those right cases found within the set of predicted labels (computed as the soft-classification rate in [8]). For the reviewer it is important to know that the probability to have the right tag in the set of candidates is really high, but on the other hand, the more the labels the harder the selection of the correct one. In our case, in the training stage we obtained soft classification rates of 82% for pitch accents and 85% for boundary tones . These rates are clearly higher than the accuracy rates that we obtain in classic non-fuzzy classification (69.2% and 81.2% respectively). The increase in the confidence rates is expected to improve the performance of the reviewing process.

In the reviewed subset of 18 news, only one label is predicted in the majority of the cases (60% for boundary tones and 45.3% for pitch accents), and few cases have three labels (3.4% for boundary tones and 2.5% for pitch accents).

5. Results

The results coming from the process of revision report that in most cases (81.8% for boundary tones and 72.6% for pitch ac-

cents) the labeler choices the first candidate of the fuzzy classifier as the right option. Only 9.2% of the boundary tones and 13.5% of the pitch accents labels needed to be edited, that means, to be corrected with a different label. In a preliminary test, the real time ratio of that labeling process has been observed to be 1:66 when the template of predictions is used. This ratio contrasts with the one obtained without any supporting template which was 1:80. These ratios have been obtained from the comparison of the time that a transcriber needs to manually label a news and the time that he/she needs to label a news of comparable size with the aid of the fuzzy classifier: in the first case, 3600 s were labeled in 44.79 s (including autochecking after the first annotation) whereas in the second case, 3000 s were labeled in 45.23 s (including as well an autochecking). The experiment has been done by the same transcriber in separate days, during a stretch of time without any interruption.

Another encouraging result is that we found a clear correspondence between the results obtained by the fuzzy automatic classifier and the results obtained in the inter-transcriber agreement tests. In the inter-transcriber test, the pair of labels that are most frequently confused by the human experts is the pair L+H* vs. L+!H* for pitch accents and the pair "none" (absence of a break) vs. !H% for boundaries (26.9% and 34.4% of the total disagreement respectively). These pairs also have the highest frequency of appearance when the fuzzy classifier predicts more than one label per word (21.8% and 29.9%, respectively). To sum up, Table 2 shows that in the range of certainty the percentages are also similar: those cases where the fuzzy classifier predicts an unique label and the frequency of intertranscriber agreement per symbol are similar for each tag.

6. Conclusions and future work

We have presented the results of an experiment in which an automatic system has been applied for the Sp_ToBI labeling of the Glissando corpus. The automatic system can generate more than one candidate label per prosodic unit according to its degree of confidence. The revision of the candidate labels provides an alternative to speed up the labeling process.

The efficiency of this strategy to speed up the labeling process is supported by the fact that only a small proportion of the predicted labels is edited. Furthermore, in most cases, the reviewer selects the first label out of the set of predictions.

The use of fuzzy labels adequately resembles the uncertainty that characterizes the human prosodic labeling process in many situations. This is evidenced by the fact that most uncertain situations for the automatic classifier correspond with labels that are the most frequently confused in manual inter-transcriber tests.

This is part of ongoing work in which an iterative training and testing process is being applied in order to improve the predictions. The reviewed labels are reintroduced in the training stage of the classifier so that the knowledge of the system increases iteratively. We are currently investigating definitions of quality metrics that measure the goodness of this iterative approach. The improvement of the revision template interface (currently in $praat$) following the suggestions of the transcribers is also current and future work.

The results of the labeling process, manual, reviewed and predicted label in the different stages, are expected to be freely available for research purposes in the web page of the project http://veus.glicom.upf.edu/.

7. References

[1] J.-M. Garrido, D. Escudero, L. Aguilar, V. Cardeñoso, E. Rodero, C. de-la Mota, C. González, C. Vivaracho, S. Rustullet, O. Larrea, Y. Laplaza, F. Vizcaíno, E. Estebas, M. Cabrera, and A. Bonafonte, "Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan," *Language Resources and Evaluation*, vol. 47, no. 4, pp. 945–971, 2013.

[2] M. Beckman, J. Hirschberg, and S. Shattuck-Hufnagel, "The original ToBI system and the evolution of the ToBI framework," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford University Press, New York, 2005, pp. 9–54.

[3] L. Dilley and M. Brown, "The rap (rythm and pitch) labeling system," Massachusets Institute of Technology, Tech. Rep. http://tedbla.mit.edu/rap.html, 2005.

[4] J. Cole, Y. Mo, and S. Baek, "The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech," *Language and Cognitive Processes*, vol. 25, no. 7-9, pp. 1141–1177, 2010.

[5] A. Brugos, N. Veilleux, M. Breen, and S. Shattuck-Hufnagel, "The alternatives (alt) tier for ToBI: Advantages of capturing prosodic ambiguity," in *Proceedings of Speech Prosody 2008*, 2008, pp. 273–276.

[6] D. Escudero, L. Aguilar, M. d. M. Vanrell, and P. Prieto, "Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labeling system," *Speech Commun.*, vol. 54, no. 4, pp. 566–582, May 2012.

[7] D. DuBois and H. Prade, *Fuzzy sets and systems: theory and applications*. Academic Pr, 1980, vol. 144.

[8] D. Escudero-Mancebo, C. González-Ferreras, C. Vivaracho-Pascual, and V. Cardeñoso Payo, "A fuzzy classifier to deal with similarity between labels on automatic prosodic labeling," *Computer Speech and Language*, vol. 28, no. 1, pp. 326 – 341, 2014.

[9] M. Ostendorf, P. Price, and S. Shattuck, "The Boston University Radio News Corpus," Boston University, Tech. Rep., 1995.

[10] A. K. Syrdal, J. Hirshberg, J. McGory, and M. Beckman, "Automatic ToBI prediction and alignment to speed manual labeling of prosody," *Speech Communication*, no. 33, pp. 135–151, 2001.

[11] C. Gonzalez-Ferreras, D. Escudero-Mancebo, C. Vivaracho-Pascual, and V. Cardeñoso Payo, "Improving automatic classification of prosodic events by pairwise coupling," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 2045 –2058, sept. 2012.

[12] C. González-Ferreras, C. Vivaracho-Pascual, D. Escudero-Mancebo, and V. Cardeñoso Payo, "On the automatic ToBI accent type identification from data," in *Proceedings Interspeech*, 2010, pp. 142–145.

[13] A. Syrdal and J. McGory, "Inter-transcriber reliability of ToBI prosodic labeling," in *Proceedings of ICSLP*, vol. 3, 2000, pp. 235–238.

[14] J. F. Pitrelli, M. E. Beckman, and J. Hirschberg, "Evaluation of prosodic transcription labeling reliability in the ToBI framework," in *Proceedings of ICSLP*, 1994, pp. 123–126.

[15] T. Yoon, S. Chavarría, J. Cole, and M. Hasegawa-Johnson, "Inter-transcriber reliability of prosodic labeling on telephone conversation using ToBI," in *Proceedings of Interspeech, Jeju*, 2004, pp. 2729–2732.

[16] E. Estebas Vilaplana and P. Prieto, "La notación prosódica en español. una revisión del sp_tobi," *Estudios de Fonética Experimental*, vol. XVIII, pp. 263–283, 2009.

[17] ——, "Castilian Spanish Intonation," in *Transcription of Intonation of the Spanish Language*, P. Prieto and P. Roseano, Eds. Lincom Europa, München, 2010, pp. 17–48.

[18] G. Elordieta, "Transcription of intonation of the Spanish language," in *Estudios de Fonética Experimental*, vol. XX, 2011, pp. 273–293.

[19] J. A. Benediktsson, J. R. Sveinsson, J. I. Ingimundarson, H. Sigurdsson, and O. K. Ersoy, "Multistage classifiers optimized by neural networks and genetic algorithms," *Nonlinear Anal., Theory, Meth., Applicat.*, vol. 30, no. 3, pp. 1323–1334, 1997.

[20] S.-B. Cho and J. H. Kim, "Combining multiple neural networks by fuzzy integral and robust classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, pp. 380–384, 1995.

[21] S. B. Cho and J. H. Kim, "Multiple network fusion using fuzzy logic," *IEEE Trans. Neural Networks*, vol. 6, pp. 497–501, 1995.

[22] P. D. Gader, M. A. Mohamed, and J. M. Keller, "Fusion of hand-written word classifiers," *Pattern Recogn. Lett.*, vol. 17, pp. 577–584, 1996.

[23] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers: A comparative study," *Pattern Recogn. Lett.*, vol. 20, pp. 429–444, 1999.

[24] D. Wang, J. M. Keller, C. A. Carson, K. K. McAdoo-Edwards, and C. W. Bailey, "Use of fuzzy-logic-inspired features to improve bacterial recognition through classifier fusion," *IEEE Trans. Syst., Man, Cybern.*, vol. 28B, pp. 583–591, 1998.

[25] M. Grabisch and M. Sugeno, "Multi-attribute classification using fuzzy integral," in *IEEE Int. Conf. Fuzzy Systems*, 1992, pp. 47–54.

[26] L. I. Kuncheva, ""fuzzy" versus "nonfuzzy" in combining classifiers designed by boosting," *Transactions on Fuzzy Systems*, vol. 11, no. 6, pp. 729–741, december 2003.

[27] D. Escudero-Mancebo and E. Estebas-Vilaplana, "Visualizing tool for evaluating inter-label similarity in prosodic labeling experiments," in *Proceedings Interspeech 2012*, 2012.

[28] D. Escudero, L. Aguilar, C. González, Y. Gutiérrez, and V. Cardeñoso, "On the use of a fuzzy classifier to speed up the Sp_ToBI labeling of the Glissando Spanish corpus," in *Proceedings of International Conference LREC*, Reykjavik, Iceland, May 2014, p. in press.

[29] T. Face and P. Prieto, "Rising accents in Castilian Spanish: a revision of Sp-ToBI," *Journal of Portuguese Linguistics*, vol. 6.1, pp. 117–146, 2007.