Universitat de Barcelona
Estudios de Fonética Experimental
Journal of Experimental Phonetics
LF Laboratori de Fonètica

Visualizing melody with multiple acoustic and tagging values using the visualization module of the *Oralstats* tool

Adrián Cabedo Nebot[a]

[a] Universitat de València (Spain), adrian.cabedo@uv.es

Article Info

Article history
Received: 21/01/2022
Accepted: 16/07/2022

Keywords
Visualization tools
Melody
Combined factors
R tool
Github

Abstract

This paper presents a way to visualize pitch patterns combining acoustic features (F_0, intensity or duration) with other variables, like a basic notation on ToBI (Tone and Break Indices) or the projection of acoustic transformations, following MAS (Melodic Analysis of Speech) model. This visualization and the previous data transformation leading to it have been carried out with *Oralstats*, a tool developed in *R* that is conceived to merge speech transcriptions with prosodic, linguistic, and other variables. Here, the multiple melodic visualizations available in *Oralstats* are exemplified with intonational phrases taken from a corpus of YouTubers. The complete interactive dashboard is freely available on *Github*.

1. Introduction

Researchers deal with multiple problems when exploring and analysing melodic contours in speech. Nevertheless, the action of exploring and visualizing data is indeed most relevant, since it allows to discover possible links between a rising/falling melody and a given, specific meaning, be it semantic or pragmatic. In addition, melodic patterns can be completed if more information is considered coming from the strictly linguistic material (POS tag, tone pattern assigned to a syllable, position…). As it has been pointed out by Shriberg et al. (1998, p. 43): "Instead of treating words and prosody as independent knowledge sources (…) we could provide both types of cues to a single classifier." The point is to combine the study of prosodic and linguistic information, being the latter often treated as independent. In this sense, the literature has consistently pointed out the correlation between melodic patterns or pitch contours and the expression of meanings, be they encoded in the grammars of languages (statements, questions…), or contextual/pragmatic (politeness, humour, mitigation…). For this reason, researchers could benefit from having multiple ways of pitch visualization to achieve scientific findings easier and faster. Since speech consists of multiple levels or layers, the analysis of speech should adopt such multi-layered approach, if possible.

In this sense, the present paper aims to present a dynamic way for visualizing melodic patterns along with other linguistic data. We propose a way of combining conventional methods for the labelling of the melodic contour and to project them together in a single visualization; this can be achieved by considering pitch values together with other phonetic and non-phonetic variables. To do so, the tool *Oralstats* (Cabedo, 2021) has been designed (more extensively introduced in section 3), that presents data in a dynamic, flexible, and scalable manner. *Oralstats* is dynamic in that it allows the visualization at a glance of pitch, intensity, and duration values along with tagged content, be it automatically tagged (POS, ToBI in Elvira-García et al., 2016) or with manual tags created *ad hoc*, and allows to change the visible/hidden information in the visualization chart freely. It is flexible in that it is developed with *R* programming language, thus allowing versatility and understandability for non-experts; finally, it is scalable, since more functionalities can be added just by repeating or modifying some parts of code.

2. Some pitch tagging and visualization systems

Melodic or pitch contours are incremental values of fundamental frequency (F0) values that commonly show three possible directions: rising, falling or holding (Quilis, 1999). The acoustic fluctuations of F0 can be projected onto increasingly bigger segments of the speech, like syllables, accentual groups (Garrido, 2012), intonational groups (Quilis et al., 1993), intonational clauses (Garrido, 2003; Hidalgo, 2019) or paratones (Hidalgo, 2019; Tench, 1996).

Tools like *Praat* (Boersma & Weenink, 2021) allow the researcher to view the melodic/pitch contour of any phonic segment, regardless of their length (a vowel, a syllable, a word…). By default, this visualization is carried out using absolute acoustic values, in Hertz (Hz), or relativized logarithmic values, like semitones (st). *Praat* also allows the stylization of the melodic contour, thus providing more clarity in the detection of tonal rises and falls and disregarding minor variations or micromelodies (Hirst, 2015). *Praat*'s default stylization tool has also received slight modifications to adjust to particular theoretical frameworks, as it is the case of MOMEL (Hirst, 2007).

Pitch patterns can be coded using different methods available, amongst which ToBI (Pierrehumbert, 1980). In some cases, the coding of the F0 rising or falling values is carried out sequentially, without considering time limits per unit of speech. In this line of action, one could find models like the aforementioned MOMEL/INTSINT (Bigi, 2015; Hirst et al., 2000) or the IPO model (Garrido, 2003, 2012; Hart et al., 1990). MOMEL/INTSINT and IPO contemplate a first stage of modelling and labelling, followed by a second stage in which the researcher can observe the correspondence with vowels or syllables. ToBI, in turn, begins by coding the pitch of speech units that have been segmented previously, therefore the limits of the units are established prior to the pitch analysis.

Both approaches have some advantages and disadvantages. Models that encode acoustic values first are especially useful, for instance, in discourse segmentation, but can lack precision when combining their tagging with vowels or syllables, since the tonal rises or falls can occur at any temporal point (sometimes, the coded data might appear at the beginning of the vowel or syllable, sometimes in the middle or at the end, or even at the very edges of these units).

Models that start from already segmented units (words, syllables, phones) are more homogeneous because pitch values can be taken from specific established places (first value, middle value…), but are somewhat problematic in that some variations can exceed their own limits and end themselves in next units (ToBI, for example, contemplates the option of accents displaced to the immediate contiguous syllable); in addition, vowel lengthening can make it difficult to account for an associated melodic pattern, which will be inherently more complex depending on the temporal extension.

Since *Oralstats* integrates specifically ToBI and the Melodic Analysis of Speech (MAS), in the following lines we will briefly describe these pitch coding systems, used in Section 4.

2.1. ToBI

ToBI (Pierrehumbert, 1980) is a common tonal tagging system that has been further developed or upgraded to adjust to the individual requirements of many languages (by way of example, in Spanish, ToBI includes [<], used to convey a continuous pitch rising ending in the post-tonic syllable (Estebas & Prieto, 2008).

ToBI distinguishes four levels of analysis: 1) the orthographic level, in which sentences are transcribed in words or syllables; 2) the tonal level, in which the tonal accents are associated with stressed syllables and boundary tones; 3) the level of prosodic juncture, where the presence or absence of prosodic domains is marked, from intonational phrases to prosodic words, and 4) the miscellaneous level, used to indicate paralinguistic phenomena. ToBI proposes the existence of two phonological units, *tonal accents*, which are associated with syllables with a lexical accent, and *boundary tones*, associated to the limits of the prosodic domains. Tones are represented by their initials in English: L (low tone, "low tone") and H (high tone, "high tone"). Boundary tones end with a % symbol.

In the last decades some automatic tools have been created to minimise the time needed to code a sequence of speech with ToBI. Some of them have not been entirely developed or completed, as it is the case of AuToBI (Rosenberg, 2010), but some others are fully working tools, such as Eti-ToBI (Elvira-García et al., 2016), developed for Spanish to work in *Praat*.

2.2. MAS

MAS (Melodic Analysis of Speech) is a system developed by Cantero (2002) where pitch is coded taking the pitch values of vowels as a reference and thus focusing on a percentage of rising or falling of pitch values, rather than on a nominal tag. Thus, whereas ToBI deals with semitones mainly, MAS (Cantero & Font-Rotchés, 2009) uses a pitch percentage difference between vowels measured in Hz (Cantero & Font-Rotchés, 2009):

> data in Hertz should be made relative in order to describe the contour melody. The melody that constitutes the succession of values 100Hz–200Hz is not the same as one of values 200Hz–300Hz, although the difference is exactly the same: 100Hz in both cases. In absolute terms, the difference is the same (the same number of Hz), but not in relative terms: the tonal interval is different. In the first case, there is a difference of 100% but in the second case, there is a difference of 50%. (pp. 37)

MAS transforms data in ascending or descending percentages, and it also permits to visualize the vowel pitch values standardizing them with a first value of 100. Most recent theoretical proposals for MAS have introduced the option of including not only pitch, but also intensity (or energy) and duration (Cantero, 2019).

As is the case for ToBI, also great variety of scripts have been implemented for MAS to transform and visualize data (Cantero & Mateo Ruiz, 2011; Mateo Ruiz, 2010, 2013, among others).

3. *Oralstats* design

Dealing with speech data can be overwhelming. Research usually requires more than one tool to be carried out, sometimes many of them. In addition, the epistemological approach to speech data in Phonetics, Sociolinguistics or Discourse analysis is necessarily different, and the research aim(s) determine the analyst's focus. Garrido (2018) claims that:

> Speech corpora need to include transcription and annotation to be useful for research purposes. For prosodic analysis, several types of information should ideally be available, both phonetic/phonological (phonetic or phonological transcription, prosodic phrasing) and linguistic (part-of-speech (POS), parsing, sentence type, speech acts, new/given information, focus, etc.), or paralinguistic (emotions, for example) events. (pp. 8)

In line with Garrido's (2018) claim, *Oralstats* (Cabedo, 2021) has been developed as an exploration environment combining information from speech transcripts with pitch and intensity data obtained from *Praat*. In this sense, *Oralstats* is a transformation tool: its module *Oralstats*.creation creates different data frames by crossing all the information available. Basically, it (a) creates three data frames (phones, words (POS tags included) and intonational phrases), (b) provides the prosodic information for all of them, like pitch range, mean and median, intensity mean, speech rate, or duration, for intonational phrases, and (c) assigns a ToBI tag and a declination variable. Further data frames can be generated with a slight modification of code (for instance, for wider units, like intonational clauses or paratones).

For ToBI tagging, *Oralstats* follows a set of some very basic rules, one subset of the transformation rules proposed in Eti_ToBI script (Elvira-García et al., 2016). The rules are as follow:

> if from the F0 valley to the F0 peak there is a difference greater than 1.5 semitones, a L*+ H label applies; but if there is also a difference greater than 1.5 between the start point and the end point of the stressed syllable, a L + H* label is written. If the previous two are true, but the target of the movement is in the posttonic syllable, the label will be L +>H*. (pp.778)

The basic subset of ToBI rules used in the first version of *Oralstats* considers three pitch points in vowels: the first value, the middle value, and the last value. If there is no difference greater than 1.5

semitones, the pitch value is related to the mean of the intonational group. This part of the transformation is especially weak if the word is oxytone, since sometimes the chart displays an obvious rise but a *Low* tag appears instead (this can be the case when the raise is less than 1.5 semitones with respect to the previous vowel). In its current version, *Oralstats* still needs to be upgraded in what regards this fact. At present, *Oralstats* computes more than 60 variables for each unit analyzed (intonational phrases, words, and phonemes) and the transformation code is deep and vast. In addition, Eti_ToBI is better for transformation because it takes the whole syllable as a frame, so it is still even more accurate for detecting specific pitch patterns that may exceed the space of the vowel, but that end before the next syllable.

For MAS transformation, *Oralstats* simplifies the proposal of Cantero (Cantero, 2002; Cantero & Font-Rotchés, 2009) and takes a basic approach calculating the percentage deviation between the middle points of vowels. Future versions will have to deal with micromelodic patterns within the vowels, because with the current code it would be not possible to detect circumflex patterns if these do not occur between vowels, i.e., if there is a singular melodic fluctuation within a vowel, *Oralstats* won't capture it. What it is anyway an advantage of MAS transformation method is that positive and negative deviations can be included in one of the statistical procedures available at *Oralstats,* that will allow us to observe how intonational phrases can be grouped together depending on their pitch behavior at the end.

Secondly, beyond the transformation module, *Oralstats* offers a visualization module that allows to present data in several ways, mainly with charts (some of them specific to project prosodic information), but also applying statistical procedures, like chi square test, ANOVA, decision trees or principal component analysis, among others.

Oralstats has been developed with *R* language (R Core Team, 2020); specially the visualization section includes the dynamic functionality supported by Shiny (Chang et al., 2021), a free library also developed with R. The result is an interactive dashboard that allows researchers to explore data in a synchronous mode, with the aid of successive filters and different statistical methods available. The idea of this tool is not new in *R* managing systems. As a matter of fact, although focused on business analytics, tools like Radiant (Vnijs, 2016) perform a similar task; generally, this kind of system lets researchers apply some of the statistical procedures available in *R* or *R Studio* and put them online or dynamically offline within a Shiny application; thus, it is possible (though not strictly necessary) to offer data online and analyze them dynamically.

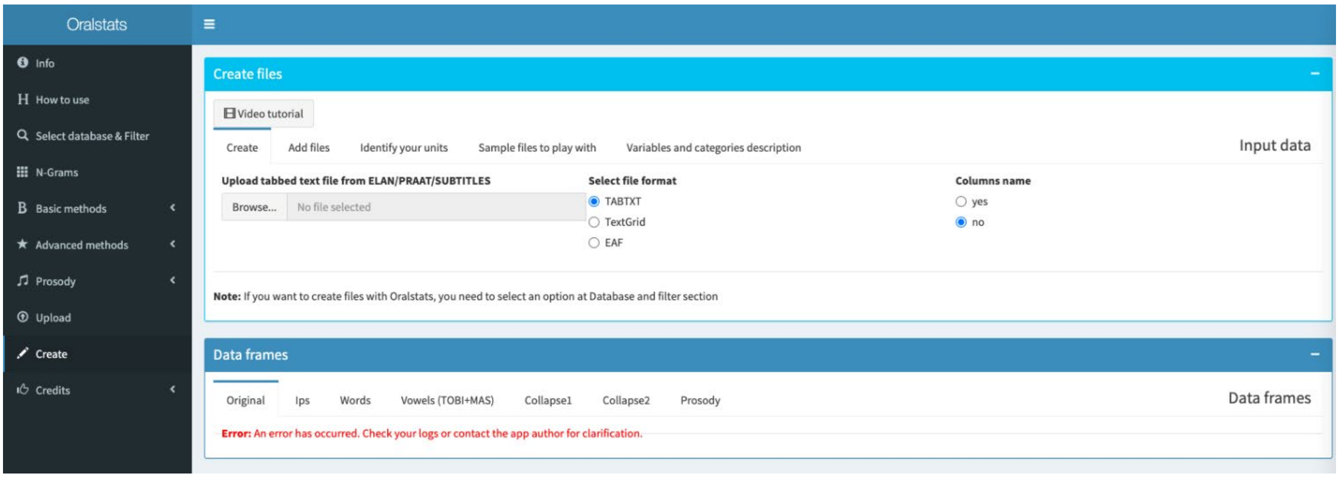
As we will see in section 4, *Oralstats* visualization module can create charts depicting multiple phonetic variables together on a same timeline. In the charts, researchers can play with the sizes of dots, line types and colors to combine several information, like pitch, intensity, duration, etc., within a given unit (for instance, words within an intonational phrase). This kind of visualization has precedents as well: Hirst (2015) establishes a similar system as a *Praat* plugin that projects melody manipulated with MOMEL algorithm along with the duration of vowels, therefore dealing with two variables only on a same chart: rhythm and pitch.

4. Case example

4.1. Variable creation and statistical visualization with *Oralstats*: a brief introduction

Before elaborating on the pitch visualization module, we will briefly introduce the creation and management of data with *Oralstats*. The program imports data directly from ELAN, *Praat* and txt files in a tabular form, crossing these files with pitch and intensity data (headerless pitch and intensity files exported from *Praat*), and generating a group of variables. Amongst others, the software considers F0 mean, F0 range, intensity mean, F0 global rising or falling; It also obtains the difference of the latter variables with the respective speaker means, and with the means of the previous unit (intonational phrase or word); it can also generate ToBI tags, and so on.

Once these variables are created, they can be explored directly in the statistical sections of *Oralstats*, obtaining correlations, chi-square test, decision trees or heatmaps, among other tests. The program also allows to filter and change data dynamically. Additionally, data generated for intonational phrases or words can be downloaded (for instance, in .xlsx) for further analysis.



Upload tabbed text file from ELAN/PRAAT/SUBTITLES: No file selected

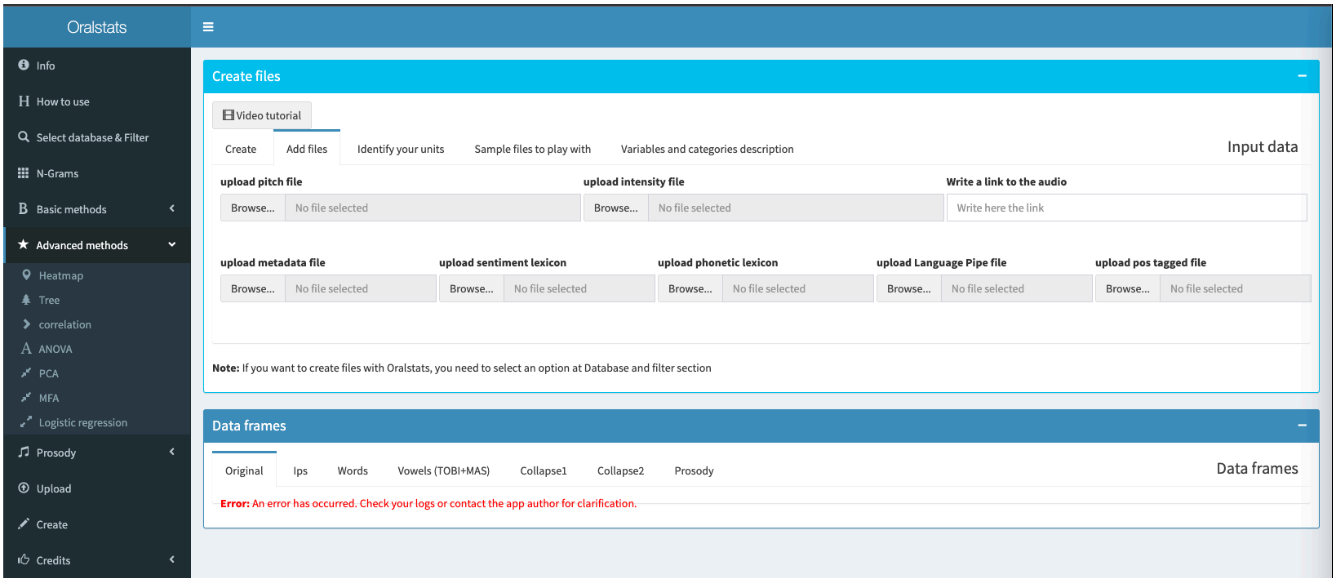
Select file format: ● TABTXT ○ TextGrid ○ EAF

Columns name: ○ yes ● no

Note: If you want to create files with Oralstats, you need to select an option at Database and filter section

Error: An error has occurred. Check your logs or contact the app author for clarification.

Figure 1. First tab in the *Create* section of Oralstats.



upload pitch file: No file selected

upload intensity file: No file selected

Write a link to the audio: Write here the link

upload metadata file: No file selected

upload sentiment lexicon: No file selected

upload phonetic lexicon: No file selected

upload Language Pipe file: No file selected

upload pos tagged file: No file selected

Note: If you want to create files with Oralstats, you need to select an option at Database and filter section

Error: An error has occurred. Check your logs or contact the app author for clarification.

Figure 2. Second tab in the *Create* section of Oralstats.

By way of example, the *Create* section in *Oralstats*, includes a tab called *Sample files to play with*. In this tab, the researcher will find a .zip file with part of the *picodeoro* database used in this paper. The .zip contains three files: one including the transcriptions of intonational phrases, words, and phonemes, and two more files, in two separate folders, including pitch and intensity values. To create and visualize data, these steps must be followed:

1. In *Select database and filter*, in the left panel, select, for example, "I will use ips from create section"
2. In *Create* section, first, the user must upload a .txt tabulated file (also a .textgrid or an .eaf file). Second, in *Add files* the user must upload the pitch and intensity files (in .txt).
3. In *Identify your units,* the user must provide some strings identifying the units to be analysed. With *picodeorodb* sample files, these strings are *ip* for intonational phrases, *word* for words, and *phon* for phonemes.
4. Lastly, in the tabs below (*original, ips, words…*) there will be options available for *Create ips, words or vowels*; These data frames, after the *Create* button has been pressed, will provide automatically generated phonetic variables (F0 mean, intensity mean, F0 range, and so on). Also, data frames from different units can be combined. For example, apart from the variables mentioned, ips data frames can include information from vowels, like ToBI or MAS tags (defining in this latter case percentages for anacrusis, body and toneme parts).

For our particular case study, we are using the interactive dashboard generated with *Oralstats*,[1] where we have access to 511 intonational phrases, divided into words and phones, from a Spanish fitness youtuber (*Pico de Oro/Padre zorro*). Figure 1 and Figure 2 are screenshots of the Shiny environment.

Figure 3 displays a decision tree containing the discourse genre as the dependent variable, and several independent variables measuring different aspects from the intonational phrases analysed.

Independent variable: discourse genre:

- twitch [*pzorrotwitch*]
- conversation [*pzorrocharla*]
- youtube show [*pzorromisa*]

Dependent variables:

- Phonic variables: F0 mean (in semitones), tonal range, intensity mean, duration, speech rate, pitch and intensity difference with the mean of the speaker, ToBI tagging.
- POS variables: quantity of verbs, nouns and adjectives.
- Sentiment analysis: quantity of positive and negative words

The total number of intonational phrases considered is 113 in *pzorrocharla*, 177 in *pzorromisa* and 221 in *pzorrotwitch*. Of these, the variables that have proven to better characterize the speaker across genres, as shown in Figure 3, are pitch mean (*PimnSt*), difference of pitch with the mean of the speaker (*PdspkSt*), intensity mean (*Imn*), and difference of intensity with the mean of the speaker (*Idspk*).

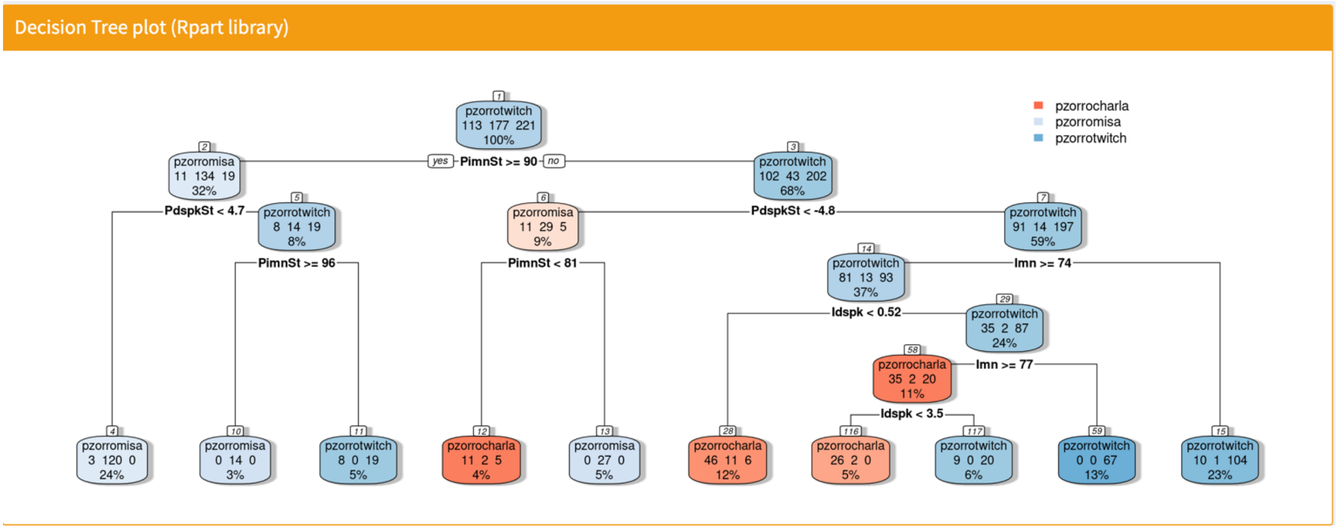


Figure 3. Decision tree classifying the prosodic properties of the speaker in different genres. The filename that appears in the corresponding node is the filename with higher frequency. For example, in the second line of nodes pzorromisa covers 32 % of total, with 134 cases and pzorrotwitch gets 68 %, with 202 cases.

Navigating the nodes, it appears that in Twitch, the speaker uses lower tone and lower intensity mean (in 104 out of 221 intonational phrases). Also an important number of intonational phrases in Twitch (67 of 221) present a high difference of intensity with the speaker mean (more than 0.52 dB), but an intensity mean between 74 and 77 dB.

In *pzorrocharla*, which is actually a debate held between two youtubers speaking about a controversial matter, the decision tree separates intonational phrases with higher intensity values, more than 77 dB (35 of 113, i.e., the conflict sequences and the overlapping excerpts), but also high intensity values with little difference with the mean of the speaker (46 of 113). Finally, *pzorromisa*, a humorous Youtube monologue, presents high variability in the pitch, with higher tones (more than 90 St) and more acute differences with the mean of the speaker, with values higher than 4.7 semitones.

[1] The case example we are using in this section is available as a free Shiny sample. The user must select *picodeorodb* at *Select database and filter* section on left panel. https://adrin-cabedo.shinyapps.io/oralstats_v_1_3/

The decision tree displayed in Figure 3 is one of the statistical tests available. Nevertheless, *Oralstats* can run multiple statistical methods with similar purposes to give researchers the option to select the statistical visualization that best fits their needs (heatmap, principal component analysis, correlation map, etc.)

4.2. Pitch visualization with *Oralstats*

As we presented at Section 4.1. for this case example speech transcripts and their prosodic values come from three different discourse scenarios: a monologue and a dialogue, both obtained from Youtube, and a discourse fragment recorded from Twitch. In the latter, the speaker interacts with followers leaving messages on the chat, so it is basically a monologue with interruptions in which the youtuber takes some time to read the messages and, if he decides so, to respond to them. The complete dataset is made up of 866 intonational phrases, 3810 words and 15085 phones.

Although we will take only one intonational phrase and its vowels as exemplification, there is a free access to all the data when *picodeorodb* sample database is loaded on the platform. The specific file from which we will select the intonational phrase to show below is the *pzorromisa* file, the one coming from Youtube monologue. We just transcribed and analysed 5 minutes of the complete recording; more specifically, we got 183 intonational phrases, 684 words and 3010 phones.[2]

ip_id	annotation	tmin	tmax	dur	trpst
ip_pzorromisa_1	señoras y señores	0.14	1.58	1.44	0.1
ip_pzorromisa_2	sean ustedes	1.67	2.70	1.03	0.1
ip_pzorromisa_3	bienvenidos	2.82	4.36	1.54	0.2
ip_pzorromisa_4	a la parroquia fitness	4.53	6.27	1.74	1.2
ip_pzorromisa_5	eterno papi zorro calentando motores para impartir	7.43	9.94	2.51	0.1
ip_pzorromisa_6	la misa	10.07	10.70	0.63	1.2
ip_pzorromisa_7	más esperada	11.93	12.80	0.87	6.7
ip_pzorromisa_8	pero antes de comenzar	19.48	20.55	1.07	0.1
ip_pzorromisa_9	pasamos lista	20.60	21.45	0.85	0.1
ip_pzorromisa_10	comprobación de zorretes	21.55	24.99	3.44	0.1
ip_pzorromisa_11	vamos equipo	25.08	26.25	1.17	0.7
ip_pzorromisa_12	activos	26.92	29.77	2.85	0.2
ip_pzorromisa_13	zorretes activados y preparados	29.98	31.88	1.90	0.1
ip_pzorromisa_14	para el despegue	31.94	32.84	0.90	0.2
ip_pzorromisa_15	bueno señores	33.09	33.95	0.86	0.1
ip_pzorromisa_16	una semana más aquí todos reunidos	34.05	36.34	2.29	0.1
ip_pzorromisa_17	con papi zorro	36.47	37.48	1.01	0.1
ip_pzorromisa_18	para comentar los horrores más horrorosos	37.58	40.38	2.80	0.4
ip_pzorromisa_19	de nuestra querida comunidad fitness	40.76	42.94	2.18	0.2

Table 1. First 19 intonational phrases of file *pzorromisa* with columns for id, annotation, beginning time, ending time, duration and posterior transition (i.e. pause duration).[3]

For illustrative purposes, we selected the intonational group with the ID ip_pzorromisa_18 and the text "para comentar los horrores más horrorosos" ('to comment on the most horrible of

[2] What it is not free available is the multimedia data. Next versions of *Oralstats* will allow to include a link to the audio.

[3] *Translation*: 'ladies and gentlemen, welcome to the fitness parrish; eternal Papi Zorro gearing up for celebrating the most expected mass. But before starting, roll call: checking my zorretes… Come on, team! Active! Zorretes activated and ready to take off! Well, gentlemen, one more week we are gathered here with Papi Zorro to comment on the most horrible of horrors in our beloved fitness community.' Note that the youtuber calls himself *Papi Zorro* ('daddy fox') and calls his followers *zorretes* ('little foxes'). The 'mass' is celebrated every Sunday and consists in Papi Zorro commenting on the most important news and social media content published by famous fitness youtubers, instagrammers, etc. He adopts a very critical, sarcastic view on these publications and personae.

horrors'). The context of this oral chunk is just at the very beginning of the speech and is provided in Table 1. The transitions between ip units are relatively fast; most of the cases ending with an [s] are followed by 0.1 pauses, which conveys an auditive impression of a longer pause:The intonational phrase *para comentar los horrores más horrorosos* contains 6 words and 14 vowels, 4 of which are stressed. The duration is 2.8 seconds, and the posterior pause is 0.4 seconds long (one of the biggest transitional timings along with ip 6 and ip 11). The pause of ip 7 (6.7 seconds) must be disregarded, since it is placed before the musical intro.

Oralstats allows researchers to get mean values for pitch and intensity, but it also allows to obtain the duration, speech rate, and other transitional variables, like before or after pauses (or fto, if the speaker before or after is different). *Oralstats* also computes the same values (mean pitch, median pitch, intensity mean, etc.) for words and phones; hence, it is also possible to access the values of the vowels in different points (beginning, middle and end) and to tag the melodic behaviour in accordance with the prosodic theoretical background selected. By default, *Oralstats* will compute a basic ToBI tagging and a basic MAS calculation. These values can be visualized as illustrated in Figures 5, 6, 7 and 8 below; Figure 4, generated with *Praat*, shows the raw pitch data in semitones:

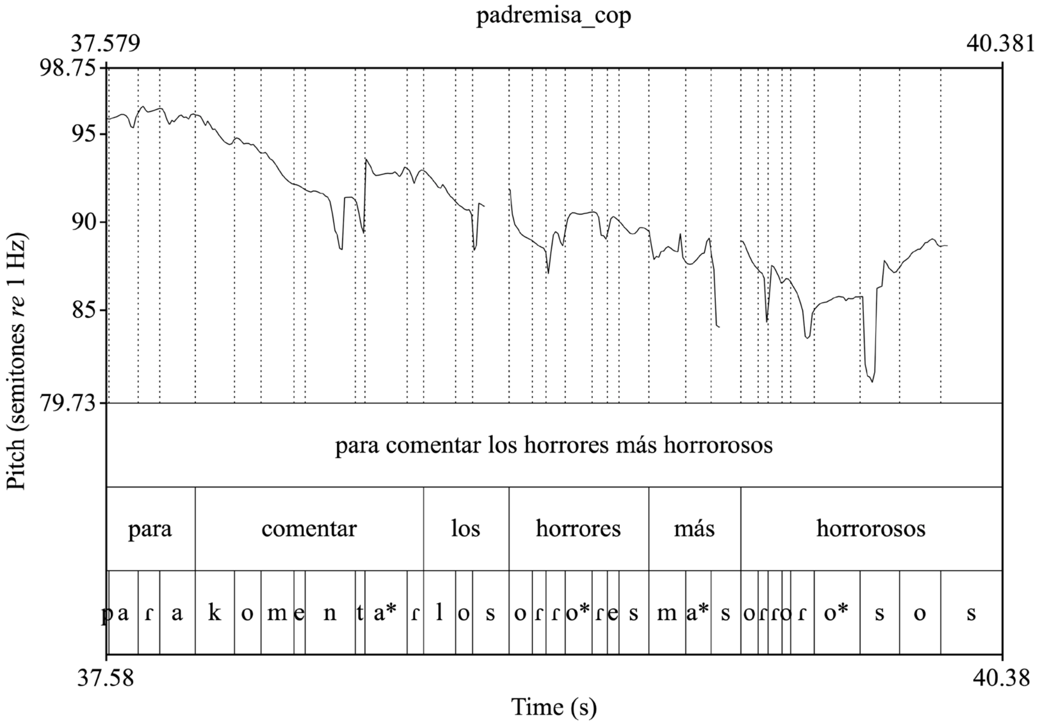


Figure 4. Pitch contour generated with *Praat*.

Figure 5 has been generated with *Oralstats* and it shows different values, combining lines, forms and color, along with numeric axes. The *x* axe includes the time where the vowels are produced, and the *y* axe presents the values of the middle portion of the vowel in semitones. The dots are sized according to the intensity mean of the vowel: the higher the value, the bigger the dot. Thus, for instance, in *para*, the intensity mean of *pa* is higher than in *ra,* and therefore the dot in *pa* is bigger. For some vowels, like the [e] in *horrores* or the second [o] in *horrorosos*, *Praat* does not register any intensity and, consequently, there is no dot shown on the chart for them. *Praat* not displaying the full extension of pitch and intensity values happens quite frequently in spontaneous speech data, a flaw that can be addressed by reducing the voice threshold in *Praat*; nevertheless, for this analysis, the common default values were selected (0.01 period times and 0.45 for voicing threshold), since the missing data does not affect the global transformation and visualization of the ip_18 under analysis.

Figure 5 also includes a continuous black line to show the ip continuity. As it can be seen in Figure 6, this trait can be modified to show other factors; for instance, researchers can associate different line displays (dotted, dashed, etc.) in different fragments of the pitch line to mark the presence of different words. However, if preferred, words can be distinguished by the color of the dots, like in Figure 5, where one dot corresponds to one vowel and dots from the same color belong to the same word. Finally, we used the ToBI generated tag to be projected along with the tonic vowels. The ending ToBI annotation, corresponding with the toneme, is L*+H; no boundary tone is showed in this chart, but it should be something like H% or even LH%.

The default chart, seen in Figure 5, can be customized to project other variables. As mentioned above, for example, Figure 6 maintains the same *x* and *y* axes, but the lines are linked to the words and the size of the dots representing intensity in Figure 2 here represents the duration of the vowels. In sum, the parameters in the axes can be accompanied with additional information just by playing with elements like the size of the dots, their colors, and the typology of lines.

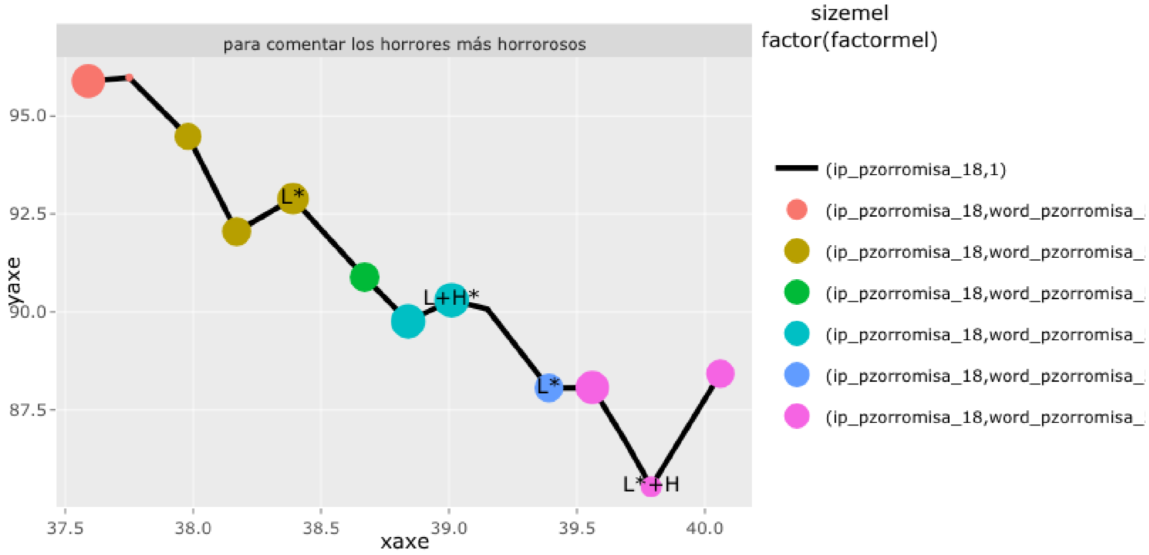


Figure 5. Stylized pitch contour generated with *Oralstats*: one dot corresponds to one vowel and dots from the same color belong to the same word; the size of dots depicts the intensity of the vowel. *Y*-axe is projecting the pitch value in semitones of the middle points of the vowels; *X*-axe is the timeline.

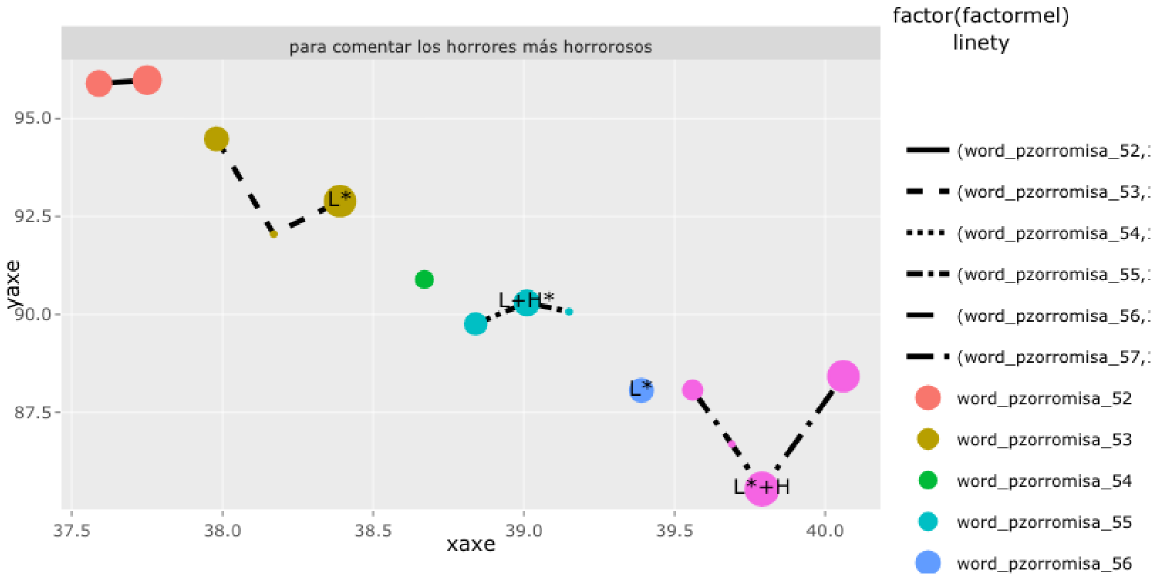


Figure 6. Stylized pitch contour generated with *Oralstats*: one point corresponds to one vowel, points size shows vowels intensity and their colors differentiate words inside the intonational phrase; the different types of lines also differentiate words. Y-axe is projecting semitones values from middle points of the vowels and X-axe is the timeline.

If the researcher wishes to shift to another visualization method, like MAS (Cantero, 2002; Cantero & Font-Rotchés, 2009), some values can be changed and therefore the melodic behaviour can be observed. Figure 7 and Figure 8 are related charts; the first shows the percentage differences between vowels, expressed here in Hz instead of semitones. Figure 7 illustrates how these differences (rising or falling) are relatively small for all the vowels, with a range that goes from -12% to 20%. A progressive fall is observed from the pitch of the first vowel (not shown on Figure 7 because it is the first value), with 255.4 Hz, to the last vowel, with 165.74 Hz; this last value is interestingly higher that the last stressed vowel [o], with 140.82 Hz.

Cantero and Font-Rotchés (2009) indicate that relating percentage increases and decreases to a first value of 100 is an efficient way to create stylized pitch contours. As in Figure 8, a first value of 100 is assigned to the first pitch value and, from that, an accumulative progression of increasing and decreasing pitch values can be computed. Comparing Figure 8 with Figure 5 (and even with Figure 4), we can see a very similar representation of the melodic contour.

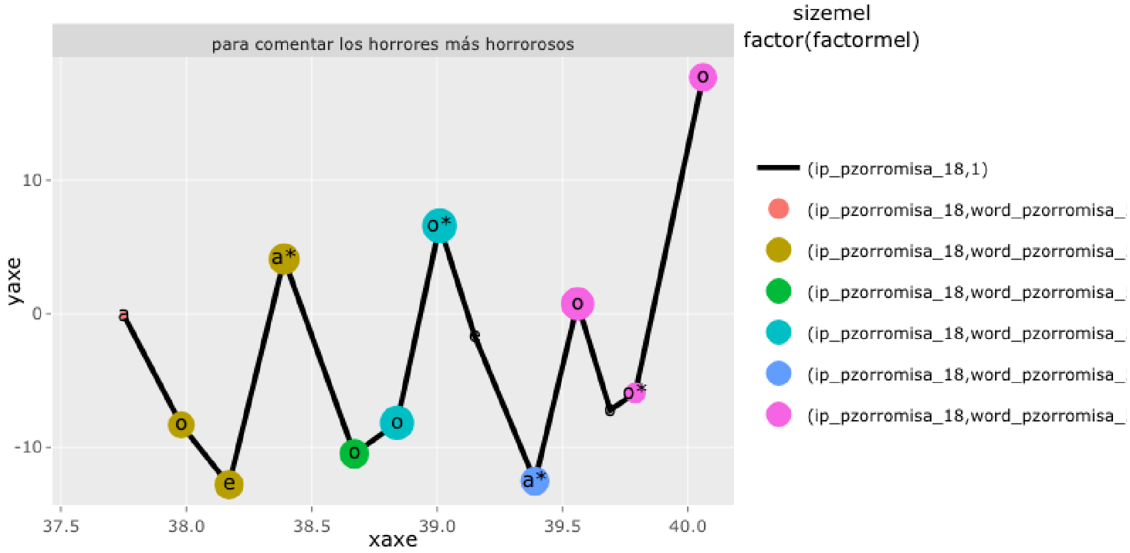


Figure 7. Stylized pitch contour generated with *Oralstats*: one dot corresponds to one vowel and dots from the same color belong to the same word; the size of dots depicts the duration of the vowel. *Y*-axe represents the percentage difference in pitch between vowels. *X*-axe is the timeline.

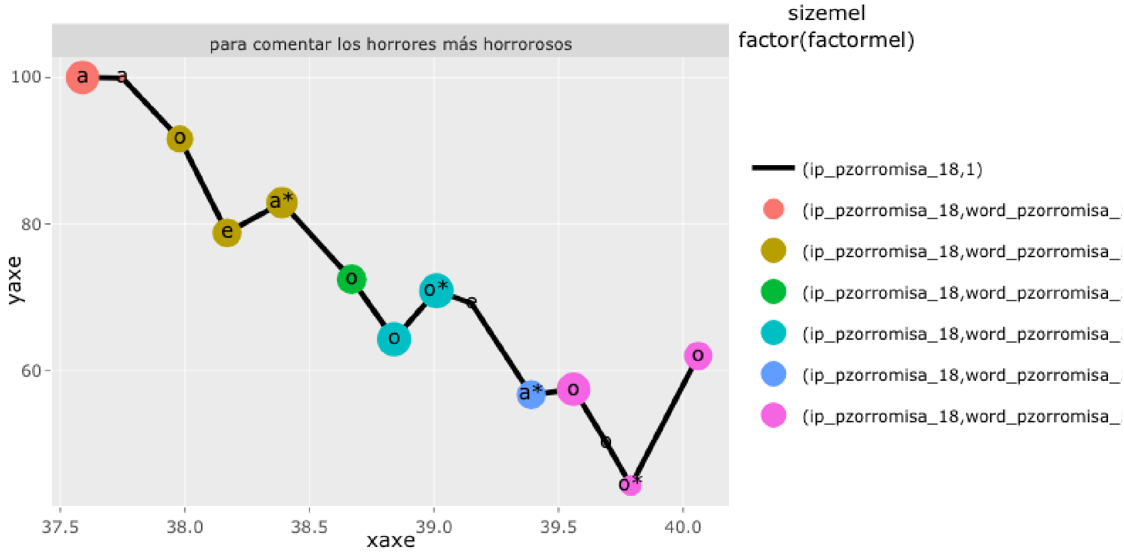


Figure 8. Stylized pitch contour generated with *Oralstats*: one dot corresponds to one vowel. Dots from the same color and line types differentiate words within the intonational phrase; the size of dots depicts the duration of the vowel. *Y*-axe is projecting pitch percentage difference in semitones of the middle points of the vowels, from one vowel to the adjacent one(s), according to the standardizing procedure of MAS; *X*-axe is the timeline.

By way of summary, *Oralstats* allows researchers to explore different ways of transforming and visualizing data, giving them flexibility and scalability on data managing. Multiple phonic information can be projected at once on the same chart (e.g., intensity, pitch, or duration), but the tool also enables researchers to explore different units, beyond the ones used in this article for illustrative purposes. For example, if intonational phrases are grouped together into bigger units, like intonational clauses (Garrido, 2003; Hidalgo, 2019) or paratones (Hidalgo, 2019; Tench, 1996), the chart can represent all these units, just by modulating the colors or line types.

Such is the case in Figure 9, where two intonational phrases are grouped within a same intonational clause (*para comentar los horrores más horrorosos / de nuestra querida comunidad fitness*). In Figure 9, the two intonational phrases are projected with two different types of line, respectively a continuous black line and a black dashed line; the size of dots represents the intensity, whereas colors discriminate words within the intonational phrase. *X* axe is again the timeline, and *y* axe represents the semitones in the middle section of the vowel.

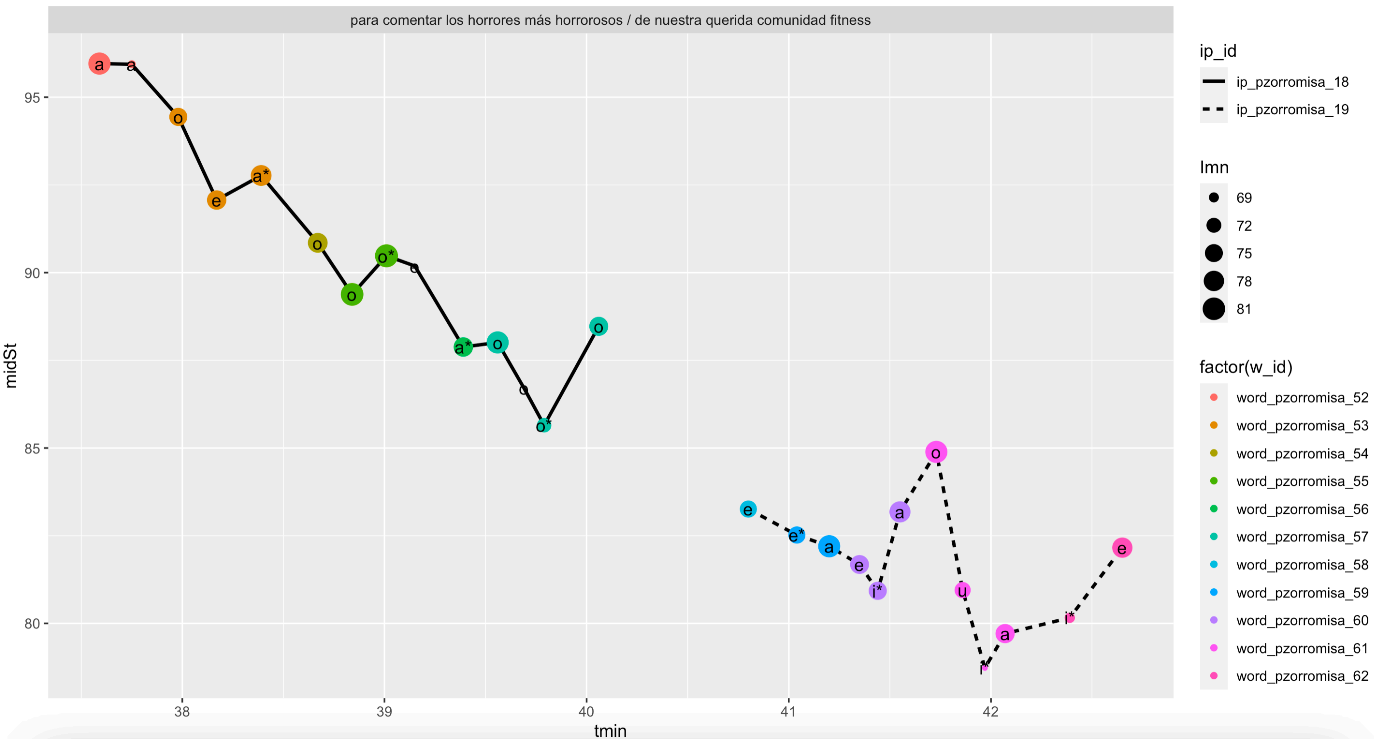


Figure 9. Stylized pitch contour generated with *Oralstats*: every point corresponds to a vowel, points size shows vowels intensity; color points differentiate words and line types differentiate intonational phrases. Y-axe shows value in semitones. X-axe is the timeline.

Besides the chart itself, one of the main advantages when visualizing data is that any researcher can change the parameters considered in a dynamic, interactive dashboard. In words of Shriberg et al.: “these and many related studies model F0, energy, and duration patterns to detect and classify accents and boundary tones; information on the location and type of prosodic events can then be used to assign or constrain meaning…” (Shriberg et al., 1998, p. 5). We followed this scientific approach to develop a tool that can help researchers when performing automatic tasks; above all, tasks related with the analysis of transcriptions and its relations with other information, like prosody.

5. Conclusion

Nowadays, researchers have moved from the stiffness of prefabricated analytics systems to the flexibility of self-developed data managing systems. In the field of phonic studies, not alien to this tendency, *Praat* has been a revolutionary tool because of its functionality and capacity, but also because researchers can create scripts to manage data in a specific way, when the default pre-set options of the tool do not suffice.

However, discourse analysts or pragmatists, who benefit greatly from the phonic approach but are not familiar with the common tools of the field, are frequently overwhelmed by *Praat*. Basically, experts from other fields trying to incorporate a phonic perspective look for the simplest application allowing them to search within strings of conversation transcripts, but also to analyse the influence on discourse of high/low values of a given phonetic factor (like pitch, intensity, or speech rate), or the meaning of a given pitch pattern associated to specific intonational units or utterances.

In this complex scenario, we have presented *Oralstats*, a free tool developed with *R* that aims to help researchers with two of the operations mentioned above: data transformation and data visualization. *Oralstats* consists of a module that transforms data combining speech transcripts with prosodic information, and including, if wished, other features like ToBI tags or even a MAS computation. The transformed data, specifically the vowels analyzed, can be visualized in customizable line charts that offer a huge variety of features to be displayed. These visualization functions enable researchers to visualize multiple data from different sources at a single glance, in one and the same chart (e.g., pitch, time, the specific phonic unit, ToBI tags, other annotations and so on).

The use of *R Shiny* environment, the main differential trait of *Oralstats* with respect to other similar solutions (Domínguez et al., 2016; Mertens, 2004; Xu, 2013) provides multiple benefits:

a) R Shiny can be used both online and offline. This functionality is useful, for instance, if a user needs to calculate a phonetic variable not covered initially by *Oralstats*, but he/she does not wish to share his/her script modification with other researchers.
b) *Oralstats* includes some of the most common statistical tests carried out by discourse phoneticians (ANOVA, chi-square, decision trees, etc.). These tests can be executed directly with *Oralstats* without using the command line of the system.
c) *Oralstats* aims to respond to the most common needs in discourse phonetics analysis, both in an acoustic or in a phonological annotated mode.
d) The programming language used is R, which, like Python, comes with a huge documentation online and allows non specialists to do statistical analysis and even low-scale programming.

More specifically, this paper has presented one functionality of *Oralstats*, the pitch contour, that allows to analyse the behaviour of vowels. This functionality takes into consideration a default *y* axe representing the pitch, but this *y* axe can be modified so as to represent other numeric values, like the MAS percentage transformation.

Beyond illustrating this functionality, we have attempted to present a small sample of the possibilities offered by *Oralstats*, which has been conceived as a new interactive way to visualize phonic data and its interaction with other speech data. Crossing all the potential interesting variables in same scenario, allowing the researcher to select or filter the desired values at any moment, and layering them in a single yet understandable visualization mode makes *Oralstats* a useful tool for analysts from different fields and with multiple research interests. At the same time, like Hirst (2015, p. 4), points out “[p]roviding linguists with better tools will surely result in the availability of more and better data on a wide variety of languages, and such data will necessarily be of considerable interest to engineers working with speech technology”, and therefore *Oralstats* uses *R* language (R Core Team, 2020), a very spread coding language around the world that has a huge community behind, making it also interesting for users with a more technical background.

In summary, *Oralstats* has been designed to be easily available, customizable, and free, hence the gratuity of the tool and its potential upgrade. With this purpose in mind, all the code can be found at *Github* (Cabedo, 2021) with a GNU license. There is yet much room for improvement and growth, but the tool has the potential to become an ally to any researcher working in fields like Forensic Linguistics, Discourse Analysis, Sociolinguistics or Pragmatics.

6. References

Boersma, P., & Weenink, D. (2021). *Praat*, version 6.1.53. Computer program. Retrieved from: http://www.*Praat*.org/

Bigi, B. (2015). SPPAS: Multi-lingual approaches to the automatic annotation of speech. *The Phonetician: International Society of Phonetic Sciences*, *111-112*, 54-69.

Cabedo, A. (2021). *Oralstats*. https://github.com/acabedo/oralstats

Cantero, F. J. (2002). *Teoría y análisis de la entonación*. Universitat de Barcelona.

Cantero, F. J. (2019). Análisis prosódico del habla: Más allá de la melodía. In M. R. Álvarez Silva, A. Muñoz Alvarado, & L. Ruiz (Eds.), *Comunicación social: Lingüística, medios masivos, arte, etnología, folclor y otras ciencias afines* (pp. 485-498). Centro de Lingüística Aplicada.

Cantero, F. J., & Font-Rotchés, D. (2009). Melodic analysis of speech method (MAS) applied to Spanish and Catalan. *Phonica*, *5*, 33-47.

Cantero, F. J., & Mateo Ruiz, M. (2011). Análisis melódico del habla: Complejidad y entonación en el discurso. *Oralia*, *14*, 105-128.

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2021). *Shiny: Web application framework for r*. https://CRAN.R-project.org/package=shiny

Domínguez, M., Latorre, I., Farrús, M., Codina-Filbà, J., & Wanner, L. (2016). Praat on the Web: an upgrade of Praat for semiautomatic speech annotation. In Y. Matsumoto, & R. Prasad (Eds.), *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016)*, Osaka, Japan (pp. 218-222). The COLING 2016 Organizing Committee.

Elvira-García, W., Roseano, P., Fernández-Planas, A. M., & Martínez-Celdrán, E. (2016). A tool for automatic transcription of intonation: Eti_ToBI; a ToBI transcriber for Spanish and Catalan. *Language Resources and Evaluation*, *50*(4), 767-792.

Estebas, E., & Prieto, P. (2008). La notación prosódica del español: Una revisión del Sp-ToBI. *Estudios de Fonética Experimental*, *17*, 263-283.

Garrido, J. M. (2003). La escuela holandesa: El modelo IPO. In P. Prieto (Ed.), *Teorías de la entonación* (pp. 97-122). Ariel.

Garrido, J. M. (2012). Análisis fonético de los patrones melódicos locales en español: Patrones entonativos. *Revista Española de Lingüística*, *42*(2), 95-126.

Garrido, J. M. (2018). Using large corpora and computational tools to describe prosody: An exciting challenge for the future with some (important) pending problems to solve. In I. Feldhausen, J. Fliessbach, & M. M. Vanrell (Eds.), *Methods in prosody: A Romance language perspective* (pp. 3-43). Language Science Press.

Hidalgo, A. (2019). *Sistema y uso de la entonación en español hablado*. Universidad Andrés Hurtado.

Hirst, D. (2007). A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation. In J. Trouvain (Ed.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, Germany. Universität des Saarlandes.

Hirst, D. (2015). ProZed: A speech prosody editor for linguists, using analysis-by-synthesis. In H. Keikichi, & T. Jianhua (Eds.), *Speech Prosody in Speech Synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis* (pp. 3-17). Springer.

Hirst, D., Di Cristo, A., & Espesser, R. (2000). Levels of Representation and Levels of Analysis for the Description of Intonation Systems. In M. Horne (Ed.), *Prosody: Theory and Experiment. Text, Speech and Language Technology*, Springer.

Mateo Ruiz, M. (2010). Protocolo para la extracción de datos tonales y curva estándar en Análisis Melódico del Habla (AMH). *Phonica*, *6*, 49-90.

Mateo Ruiz, M. (2013). De melodías y variedades del español. *Phonica*, *9*, 14-18.

Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In B. Bel, & I. Marlien (Eds.), *Proceedings of Speech Prosody 2004*, Nara, Japan. SProSIG.

Pierrehumbert, J. (1980). *The phonology and phonetics of english intonation*. Doctoral Dissertation. Massachusetts Institute of Technology, United States of America.

Quilis, A. (1999). *Tratado de fonología y fonética españolas*. Gredos.

Quilis, A., Cantarero, M., & Esgueva, M. (1993). El grupo fónico y el grupo de entonación en español hablado. *Revista de Filología Española*, *73*, 55-65.

R Core Team. (2020). *R: A language and environment for statistical computing*. R

Foundation for Statistical Computing. https://www.r-project.org/

Rosenberg, A. (2010). AuToBI: A tool for automatic ToBI annotation. In K. Hirose (Ed.), *Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010)*, Makuhari, Chiba, Japan. ISCA.

Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M., & van Ess-Dykema, C. (1998). Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech*, *41*(3), 443-492.

’t Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge University Press.

Tench, P. (1996). *The intonation systems of English*. Cassell.

Vnijs, V. (2016). *Radiant, business analytics using R and shiny*. https://vnijs.*Github*.io/radiant/

Xu, Y. (2013). ProsodyPro: A tool for large-scale systematic prosody analysis. In B. Bigi, & D. Hirst (Eds.), *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France (pp. 7-10). Laboratoire Parole et Langage.