UNIVERSITAT DE BARCELONA

Estudios de Fonética Experimental
Journal of Experimental Phonetics

LF Laboratori de Fonètica

CALLIOPE: A multi-dimensional model for the prosodic characterization of Information Units

Sonia Cenceschi[a], Licia Sbattella[b], Roberto Tedesco[c]

[a] Politecnico di Milano (Italy), sonia.cenceschi@supsi.ch
[b] Politecnico di Milano (Italy), licia.sbattella@polimi.it
[c] Politecnico di Milano (Italy), roberto.tedesco@polimi.it

ARTICLE INFO

Article history
Received: 14/01/2021
Accepted: 22/08/2021

Keywords
Prosody
Prosodic model
Information unit
Calliope

ABSTRACT

CALLIOPE is a conceptual multi-dimensional model that aims at approximating and categorizing the prosodic phenomena taking into account of all possible independent factors affecting the sound of so-called Information Units (IUs). In CALLIOPE, each IU is associated with a tuple composed of 12 labels, each belonging to a different dimension representing a characteristic influencing the prosodic behaviour. Its ultimate aim is creating well-defined corpora suitable for linguistic and engineering research.

1. Introduction

CALLIOPE (Combined and Assessed List of Latent Influences On Prosodic Expressivity) is a conceptual multi-dimensional model that aims at approximating and categorizing the prosodic phenomena taking into account of all possible independent factors affecting the sound of so-called Information Units (IUs). It starts from the need to characterise the acoustic variability of human speech, with the ultimate aim of creating corpora suitable for the analysis of dialogue and dialogic interaction, and the development of deep learning algorithms. In fact, creating an appropriate and well-defined prosodic corpus is a substantial condition for the correct functioning of automatic algorithms, which otherwise risk to provide unbalanced results. CALLIOPE focuses on the relationship between linguistic, phonotactical and intonational parts of Information Units (IU), and includes the linguistic variability in its categories in order to provide to provide a framework adaptable to any research. Following Cresti (2000), a given IU is composed by a *textual realisation* (i.e., a written phrase) and an *acoustic realisation*, and conveys an *illocutionary act* (Austin, 1975). The acoustic realisation carries both phonotactics and prosody. Phonotactics is the possible arrangement of phones in words of a given language while, according to Fujisaki (1997), prosody is the systematic organisation of various linguistic units into an utterance in the process of speech production, whose realisation involves both segmental and suprasegmental features.

Humans establish a dialogue relationship with others through many different paralinguistic clues: kinesthetics, environmental context, and other exogenous complex linguistics and spoken signals. In this perspective, the acoustic component (in other words, the prosody) plays a crucial role. Consider,

for example, *irony* and *pragmatic focus*: with the first, the messages conveyed by the sentence and its acoustic realisation are in contradiction creating a refined pragmatic game, while the second can be used to introduce a new element in the discussion, or to underline the conclusion of an intervention within the conversation. Moreover, the understanding of a spoken message does not always coincide with the information conveyed by the linguistic content (Cole, 2015) and, at the same time, it can be decomposed into many aspects not limited to single word processing. In light of this it becomes necessary to accurately describe what IUs samples we are going to analyse (e.g. specifying who/where we register, under what conditions, if spoken read/elicited/spontaneous speech, etc.), trying as much as possible to approximate the complexity of

To our knowledge, no model allowing IUs' prosody conceptual categorisation has been proposed, in particular taking into account the comprehensive set of variables involved in the communication process. This proposal fits into this perspective, trying to draw from different theories belonging to the most varied fields of study. In the next section, we will mention the models that come closest to our proposal, and the most relevant theories that inspired and provided insights for this work. Following sections will describe the model's dimensions, the methodology and the validation. A conclusive part will underline strengths and weaknesses of CALLIOPE, and future perspectives.

2. State of the art

CALLIOPE can be placed in the same field of scope of Calliope & Fant (1989) that provided a multidisciplinary documentation to address speech data processing at the intersection of speech technologies, speech production & perception, linguistics, and phonetics. On the methodological front it is worth mentioning Leoni (2001, 2017), and Leoni & Giordano (2005), a project including contributions that have been an inspiration for this work, although strictly focused on the Italian language. Looking at the pertaining works, many of them are insightful but often represent different or partial points of view on IU's categorization. To date, the most similar approach and "conceptual attitude" to CALLIOPE can be found in the *Amper* project[1] (Romano et al., 2014), and De Iacovo (2019) for Italo-romance speech. *Amper* is a wide-ranging and participatory project that starts from audio data collection to propose a classification of the geo-prosodic variation of speech relying on quantitative vector and clustering evaluations. However, although the IU is the fundamental unit for both the projects, CALLIOPE does not provide quantitative descriptions, but rather qualitative ones. For this reason, while *Amper* characterize the IU relying on acoustic features, our model defines a tuple[2] of high-level conceptual descriptors related to the phenomena influencing its prosody.

Some researches that could remind CALLIOPE are DIT++ (Bunt, 2009) and the ISO 24617-2 (Bunt et al., 2017), but they are focused on spoken, written and multimodal dialogue annotation. Moreover, they adopt a communication perspective, marking up each unit (dialog turns, and not single IUs) with one or more labels in order to model the dialog over time. Instead, widening the perspective, a wide range of research studies and projects focuses on the influence of the linguistic meaning on human understanding processes, narrowing to the neurological functioning, such as Davis & Johnsrude (2007), Kazanina et al. (2006), Baker et al. (2009). Kompe & Kompe (1997) describes algorithms and statistical models leveraging prosodic information on various levels of speech understanding, while Noth et al. (2000) investigate about how prosody can be used in automatic speech understanding systems.

2.1. Speech Act theory

In the Speech Act theory "To say something is to do something" (Austin, 1975, p. 12) and an utterance is considered as an action. In this perspective, the issuing of an utterance changes with the intention of the speaker, corresponding to his/her type of attitude,

[1] https://www.lfsag.unito.it/ricerca/amper-ita/#/

[2] A tuple is here intended in informatics sense: a list of elements characterizing a particular type of data, and distinguished by an identifier value. It should not be confused with the mathematical meaning, where the list is ordered and delineates a vector in space.

called illocutionary act (e.g. *advice*, *suggestion*, or *opinion*). The illocutionary ict refers to a IU used with a specific communicative intention (whether it is spontaneous or not), where prosody *is the necessary means of transducing the pragmatic conception in a concrete and audible entity* (Cresti, 2020). In this perspective, every illocutionary act influences differently the sound of a given IU: for this reason CALLIOPE considers the illocutionary act a dimension of its conceptual space, relying on the classification provided by Kratzer (2012), useful for describing both spontaneous and non-spontaneous speech.

However, this is just one of many different categorizations, and it has been chose because already assimilated in different disciplines (see section 5 for insights regarding labels' numbers and their validation). For example, Cresti et al. (1998) extend the Speech Act Theory, to the Theory of Language into Act (L-Act) defining over 90 acts for spontaneous speech (Cresti, 2014).

2.2. Intonation and Intonation Units

CALLIOPE has been inspired by The *Interactive Atlas of Romance Intonation* (IARI) (Prieto et al., 2010), which is focused on the study of intonation in different Romance languages. IARI leverages on ToBi (Beckman et al., 2004) which is not of interest for our purpose being a tagging and transcription system and not a conceptual theory aimed at a general IUs' classification. Instead, we adopted as a dimension the general intonation units classification used in IARI: statements, yes-no questions, wh-questions, echo questions, imperatives and vocatives. Each of these labels will cross with other CALLIOPE dimensions to generate a different prosodic realization. For example, the statement "*Domani è bel tempo*" 'Tomorrow the weather will be fine' will sound differently when the speaker changes emotion.

2.3. Emotions and psychological aspects

Prosody is also a mean for communicating emotions, and so we need to take into account that the acoustic realisation of the IU could carry an affective valence (Zentner et al., 2008). Emotion categorisations are innumerable: we were initially inspired by Plutchik (1991), and Tomkins (1984) to finally head towards Douglas-Cowie et al. (2007), but the topic is deepened in paragraph 2.1. Emotions-speech relationship has been deeply investigated to reproduce the neurological detecting processes, and links between acoustic features and emotional states are widely studied (Williams & Stevens, 1972; Nicholson et al., 2000; Vogt, 2010). Emotions are strictly connected with the dialogue according to speakers' goals and mutual interactions. This topic has been largely studied in the Interpersonal Motivational Systems (SMI) theory (Liotti & Monticelli, 2008) from a psychological point of view. Detection of SMIs has been initially based on the analysis of textual contents (Fassone et al., 2012); nevertheless, sound plays a very important role. For example, the same request is pronounced differently if we want to induce the interlocutor to trust us, or if we want to assert some kind of social superiority (but not necessarily changing the IU's associated illocutionary act). This variability in prosody is proved by Sbattella et al. (2014), where an original multilevel model of verbal interaction combines textual and acoustic components of the speech to automatic extract SMIs in forensic interrogations.

2.4. Focus and pragmatics

Another useful point of view for acoustic IUs' classification, is the placement of pragmatic stresses. CALLIOPE includes a dimension for the *focus*, which is a particular acoustic emphasis placed in order to attract the listeners attention to a specific part of the IU. The way we intend the focus comes from the L-AcT theory, which underlines that the sound is a direct consequence of a pragmatic act. In our classification, we considered categories proposed by Gussenhoven (2008), but many other theories exist, such as those proposed by Büring (2009), Domínguez (2004), or Fujisaki (1997). Speech styles and context change prosody as well: read, recited and spontaneous speech lead to different prosodies (Nencioni, 1983; Llisterri, 1992), and also position in space and context (Harris, 1997; Ghaffarzadegan et al., 2014).

2.5 Phonotactics and other factors affecting prosody

Each language defines restrictions on the feasible combination of phonemes leading to different sound and segmental or suprasegmental combinations (phonotactics), and therefore the language variation deeply influences prosody (Booij, 1999; Bennett, 2012). For this reason, in CALLIOPE, we conceptually consider the language as a space where other parameters are varying (Figure 6). Moreover, only works considering a direct connection with the prosodic realisation have been used in CALLIOPE. However, if we widen the horizon to theories regarding phenomena indirectly influencing IU acoustic realisations, we find, for example, the linguistic and expressive abilities, their typology and nature, and the social context (Cole, 2015). Many of these phenomena are essentially not enumerable or difficult to quantify, but they do affect IU's acoustic realisations and must be taken into account.

	Dimensions	Field of study
Dialogic dimensions	D_1 Structure	Grammar & phonotactics
	D_2 Illocutionary Act	Pragmatics
	D_3 Intonational Focus	Pragmatics
	D_4 Rhetorical Form	Pragmatics
	D_5 Motivational State	Psychology
	D_6 Speech Loudness	Spatial dislocation
	D_7 Spontaneity	Sociology
	D_8 Acoustic Pause	Unforeseeable factors
	D_9 Emotions	Psychology
Background dimensions	D_{10} Subjective Expressiveness Skills	Clinics
	D_{11} Social Context	Sociology
	D_{12} Language, Dialect or Local Variety	Grammar & phonotactics

Table 1. CALLIOPE dimensions and the research fields they derive from

3. The CALLIOPE model

CALLIOPE is a conceptual model aiming at categorising IUs according to their prosodic form, relying on a list of labels linked to factors influencing the acoustic components of the spoken message.

It must be underlined that CALLIOPE is not a geometrical space or a taxonomy, but a label system, potentially expandable if each label is branched. Therefore, terms like "dimension", "space", and "point" are used for synthesis purposes, with a conceptual rather than a geometric meaning. CALLIOPE defines a multidimensional "space", where each *dimension* represents a characteristic influencing the acoustic paralinguistic components of IUs; each dimension is actually a categorical variable, assuming values in a set of labels called *elements*. Each IU is thus associated to a "point" into this space; more formally, a generic IU is associated to a tuple T(IU) composed of 12 elements:

$$T(IU) = (l_1,l_2,...,l_{12}) : l_i \in \quad D_i, 1 \leq i \leq 12$$

Thus, a corpus can be described in CALLIOPE by one or more tuples (see § 3), where we define a label for each dimension. As shown in Figure 1, more IUs can be associated to the same tuple, but a single IU cannot be associated to more than one tuple.

A new corpus composed by a set of IUs can be created to be described by one or more tuples in order, for example, to allow reproducibility or data integration, while the attempt to describe an existing corpus with CALLIOPE, can help in highlighting any limits due to prosodic variability. For example, in AI research, the definition of a tuple permits to define well the corpus with respect to the scope of network, and understand which factors of variability can affect its performance with respect to the training data.

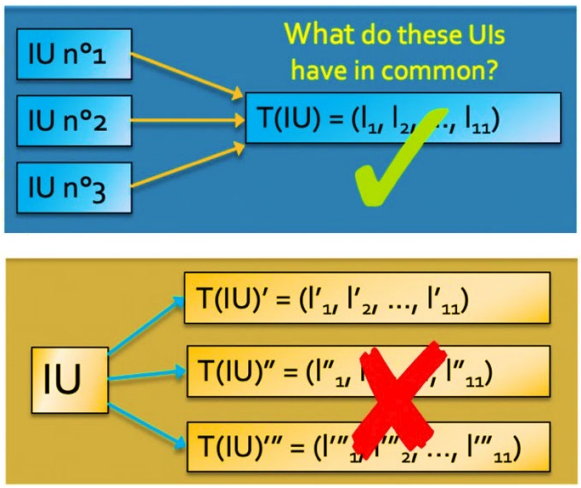

Figure 1. Surjective IU-tuple correspondence.

Some combinations are not possible in CALLIOPE, or commonly used: for example, it is not possible to utter a IU like *"Giovanni ama Maria!"* ("Giovanni loves Maria!") as exclamatory (D_1) and ironic (D_4) at the same time (D'Imperio, 2002); this means that several tuples are never observed or (which is the same) our space contains several points where no IUs can be associated. On the other hand, we argue that, for a generic IU, it is possible to select a precise element for each dimension of our model; this means that every IU is truly a point in our "space".

CALLIOPE dimensions are divided into two groups. The first one contains *Dialogic dimensions*: characteristics directly related to the communication context; the corresponding D sets are fully defined. The second group contains *Background dimensions*: characteristics existing regardless of the presence of interaction; the corresponding D sets are finite, but so large that we prefer to consider them as open sets.

Table 1 lists the dimensions for each group and shows the research field each dimension derives from. Each dimension is describe in detail in the following sections, while the complete list of CALLIOPE labels is provided in Appendix A.

3.1. Dialogic dimensions

This section details the CALLIOPE *dialogic dimensions*:

- D_1 Structure.
- D_2 Illocutionary Act.
- D_3 Intonational Focus.
- D_4 Rhetorical Form.
- D_5 Motivational State.
- D_6 Speech Loudness.
- D_7 Spontaneity.
- D_8 Acoustic Pause.
- D_9 Emotion.

All images have been generated with *Praat* (Boersma & Weenink, 2017).

3.1.1. D_1 (Structure)

This dimension refers to the sentence typology, as determined by both a specific punctuation mark and a peculiar prosodic characteristic. We adopted the following typologies: declarative, interrogative with 1 tonal unit, interrogative with 2 or more tonal units, interrogative disjunctive, echo question, exclamative, and vocative. Such categories have been inspired by the *Interactive Atlas of Romance Intonation* (Prieto et al., 2010)

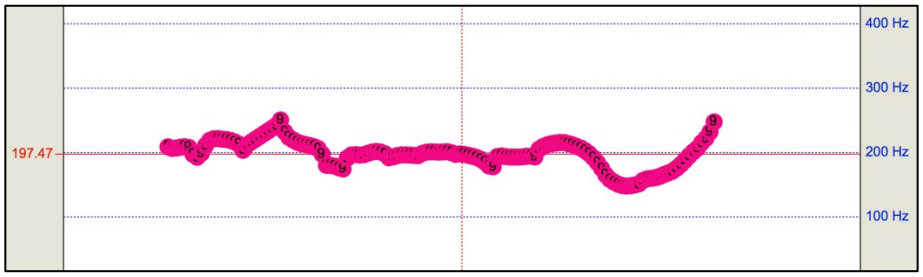

Figure 2. Italian intonation (Milanese speaker) of a Question with 1 tonal unit, obtained with the *Praat* Smooth command.

It is necessary to emphasize that intonation is one of the main components of prosody, and is often used to easily describe different structures (see Figure 1), but this is not sufficient to characterise and discriminate the *Structure* dimension.

The Pitch and Intensity envelopes, together, can provide a first approximated representation of the IU *Structure*. See figure 2, where the two contours suggest the presence of a repetition. However, other spectral features contribute to prosodic variations and perception. Furthermore, note that a *Structure* typology does not correspond to the same prosodic realisation in all languages; for example, not all languages use final intonation rising to indicate a question. This characteristic (which is actually present in several dimensions) forced us to add to CALLIOPE a specific dimension defining the current language: D_{13} (*Language*, *Dialect* or *Local Variety*).

It should also be noted that examples reported in Figures 2 and 3 are related to *read speech* from a Milanese speaker, but they will present a different sound (and different specific intonation) with respect to the same IU pronounced by the same person in spontaneous conditions. This aspect is taken into account in the *Spontaneity* dimension (D_7). This concept is valid for all the examples related to figure 3-5: their *acoustic realization* descends from a text. Instead it would be worthwhile valid the reverse for spontaneous speech.

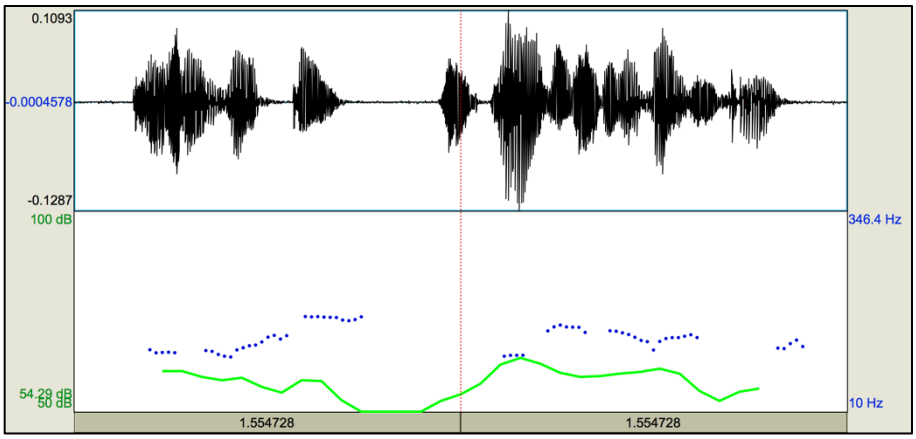

Figure 3. Italian Echo question (Milanese speaker): intensity (green) and pitch (blue dots) envelopes.

3.1.2. D_2 (Illocutionary Act)

This dimension derives from the concept of Speech Act, where "to say something is to do something" (Austin, 1975, p. 12), and an utterance is considered as an action. In this perspective, the issuing of an utterance in a speech situation changes with the intention of the speaker, corresponding to his/her type of attitude, called *illocutionary act* (examples are: advice, suggestion, and opinion). Hacquard (2011, p. 1484) define it as "the category of meaning used to talk about possibilities and necessities,

essentially, states of affairs beyond the actual". In this work, we rely on 26 illocutionary acts, following the categorisation explained in Kratzer (2012) as an expandable starting point. Notice that this is just one of many different categorisations, which can define as many as 90 acts (Cresti et al., 2000; Firenzuoli, 2003, extending the Speech Act theory to the theory of Language into Act (L-Act) for spontaneous speech (Cresti, 2014).

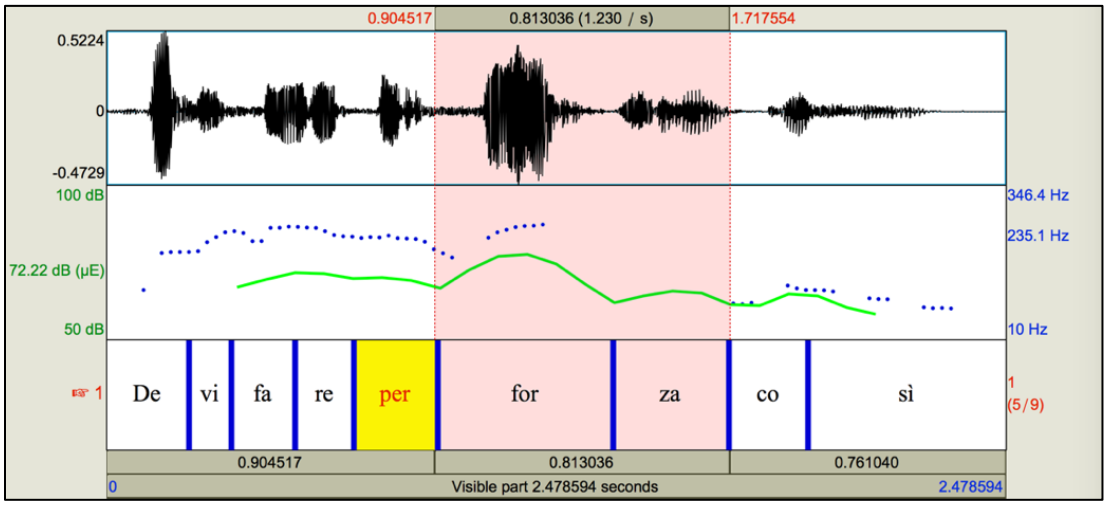

Fig. 4. Italian Corrective focus (Milanese speaker) in elicited speech. The IU is: *"Devi fare per forza così!"* 'You must do that!' with the focus on the syllable *for-*.

3.1.3. D_3 (Intonational Focus)

This dimension highlights important elements of the spoken message. The acoustic realisation of the focus depends on speaker's culture and language (Goldrick, 2004). For example, the Italian language is strongly syllable-timed: syllables take approximately an equal amount of time to be pronounced, and they are temporally stretched by speakers when they intend to underline a word. An example is shown in figure 4, where the focused word *"forza"* (in this context it means "you must...") is the longest one and presents higher pitch and intensity. The focused syllable is characterised by changes in fundamental frequency excursion and intensity-related parameters, with respect to their average values.

Listeners' expectations affects how prominence is perceived (Tamburini et al., 2014). In our model, we adopted *elements* like Presentational, Corrective, Counter-presuppositional, etc. These elements have been extracted from Gussenhoven (2008) and they carry a clear pragmatic function. The full list is shown in the Appendix. We also included Non-focused, for IUs lacking any *Intonational Focus*.

3.1.4. D_4 (Rhetorical Form)

We consider only the rhetorical forms (Harris, 1997; Wallace, 1970) that are detectable in acoustic realisation of IUs: Irony, Aposiopesis, Prepetition, Anacoluthon, etc. We also included the Null element for IUs lacking any rhetorical form. The full list, as well as the precise semantics, is shown in Appendix. Simple examples are the Enumeratio (e.g. "1, 2, 3,. . . "), where commas insert temporal pauses, Irony, where the acoustic realisation can communicate an opposite effect with respect to the linguistic part of the IU, and Aposiopesis, where a sentence is deliberately broken off and left unfinished (e.g. "Get out, otherwise…!").

The Dialisis and Parenthesis elements are two particular cases: both of them are made of two different sentences joined together and considered as a single IU (*"Ho scordato, ma chi se ne importa,*

l'ombrello", "I forgot, but who cares, the umbrella").

3.1.5. D_5 (Motivational State)

In the Interpersonal Motivational Systems (SMI) Theory (Liotti & Monticelli, 2008), motivations within interpersonal exchanges are analysed in an evolutionary perspective. The AIMIT manual defines the following SMIs: Care-seeking, Sexual bonding, Caregiving, Attachment, Competitive, Peer-cooperative and Social rank. We also included the Neutral element, for IUs where no SMI is detectable. It is important to underline that this dimension differs both from *Social Context* and *Illocutionary Act*; in fact, SMIs are related to the relationship between the interlocutors and their mutual role. In each different *Social Context* speakers can independently play different interpersonal roles (e.g. the judging and the judged) and inside this SMI they can choose one of the 26 *Illocutionary Act* elements (e.g. both the judging person and the judged person can use IUs to convince, suggest, ask, etc).

At the same time, the SMI modifies the acoustic realisation of the IU because prosody is influenced by the role we are playing with respect to our interlocutors (e.g. the judging person will be more secure and less undecided than the judged person, pronouncing the same IU). The SMI theory is admittedly very abstract; notice, however, that SMIs can be automatically extracted from IUs. As an example, see the DIKE project (Sbattella et al., 2014), a pioneer work regarding automatic dialogue analysis exploiting an original multilevel model of verbal interaction taking into account of the mutual indivisible relationship between SMI and prosody.

3.1.6. D_6 (Speech Loudness)

This dimension depends on the intensity with which a person is speaking due, for example, to physical distance from the interlocutor, or specific needs. According to Zhang & Hansen (2007) we set its label to: whispered, soft, neutral, loud and shouted. In fact, *Speech Loudness* affects several acoustic features, generating different prosodic realisations (e.g., Jovičić, 1998; Fux et al., 2011; Zhang & Hansen, 2007; Hansen et al., 2017). For example, think about talking with someone positioned quite far apart or close to you: it is immediately clear that a same IU is expressed by very different acoustic realisations also if the communicative intention and linguistic contents are maintained. In Figure 5 (on the left), the intonation and the intensity envelopes of the same sentence pronounced by the same speaker, screaming (above) and with a normal loudness (below) are shown. The changes are therefore evident, even remaining at a macroscopic level of analysis.

3.1.7. D_7 (Spontaneity)

This dimension reflects the speech typologies found in corpora (Nencioni, 1983), the three main types of enunciative styles: Spoken, Read, Recited. We chose to consider also the Elicited speech, and the inclusion of Social Media Speech could be interesting in the future (Cenceschi et al., 2021). Like in Emotion, such element set is potentially expandable but we think that, together with *Illocutionary Act* and *Social Context*, we can provide a good approximation for all the IUs.

3.1.8. D_8 (Acoustic Pause)

This dimension describes the presence of purely acoustic pauses inside an IU. This kind of pauses lack any pragmatic valence and is due to extemporaneous phenomena (or, like in poetry, for creating particular effects). Often, at the textual realization level, such pause is represented by a comma or three dots (e.g., *"Vieni, a bere un caffè."*, "Come, to drink a coffee").

Elements are Present and Not Present. Notice that bracketing commas and the incidental sentence are not a pause; instead, they are considered by the Parenthesis and Dialysis elements (in *Rhetorical Form*).

3.1.9. D_9 (Emotions)

We are aware that classifying emotions is controversial and no definite list exists (Schuller et al., 2011; Cowie & Cornelius, 2003; Vasco et al., 2010). Origlia et al. (2014) provide a comprehensive overview of the classification and features extraction in emotions. A wide range of studies address the quantitative characterization of emotions based on the extraction of acoustic features (e.g. Carbone & Petrone, 2020; Likitha et al., 2017) but interesting reviews summarize these approaches, such as Swain et al. (2018), or Akçay & Oğuz (2020). However, being CALLIOPE a qualitative classification, we just need a set of discrete categories in order to describe the vast majority of prosodic cases, while the purpose of these works is to distinguish emotions acoustically. For this reason, we choose not to limit this dimension to primary emotions (e.g. Tomkins, 1984), but to extend labels to a wider set with the 48 categories proposed by the Human-Machine Interaction Network on Emotion – HUMAINE (Schröder et al., 2006) as reported in the Appendix. We also included the Neutral element for IUs lacking a particular emotional state. Figure 5 (right) shows an example limited, as for *Speech Loudness*, to intonation and intensity envelopes for a same speaker speaking simulating sadness (above) and anger (below). The fact this example is made by recited speech is not to be considered a limit, but only that these samples are labeled in CALLIOPE as *recited speech* for D_7 (*Spontaneity*).

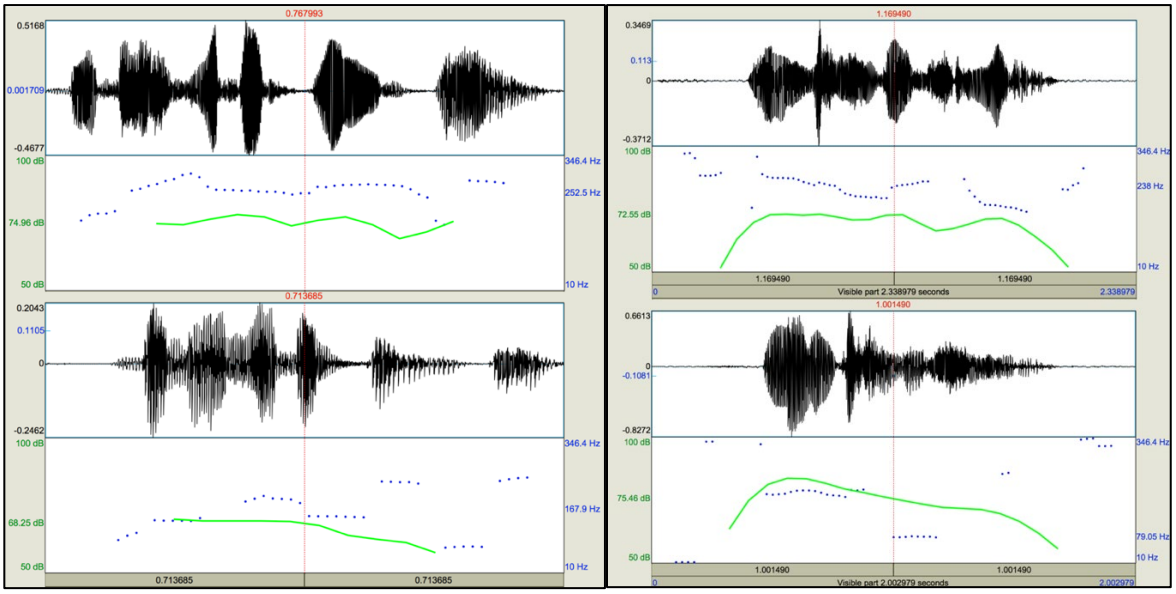

Fig.5. The same speaker pronouncing the same IU: (1) On the left, screaming (above) and with a normal loudness (below): *"Domani è bel tempo!"* 'Tomorrow the weather is fine!'; (2) On the right: sadness (above) and anger (below): *"Non voglio uscire!"* 'I don't want to go out!'.

3.2. Background dimensions

This section details the CALLIOPE *background dimensions*. The Appendix shows some examples: as they are open sets, they have not been marked with numbers. These dimensions can be described by countless labels, and it is considered useful to leave the researcher the freedom to choose the label and the degree of specificity.

3.2.1. D_{10} (Subjective Expressiveness Skills)

This dimension describes the expressiveness skills of the speakers. For example, a speaker with cognitive problems influencing the expressiveness level may not complete the words, or have a flat and monotonous intonation. It is not possible to define a comprehensive list of elements because there are

endless variations and subjective nuances: that is why D_{10} is a *Background dimensions* with an open set of elements.

3.2.2. D_{11} (Social Context)

This dimension can be intended as a component of the Speaking Model developed by Hymes for the classification of interactions, where the *Setting* and *speaker and the audience* (*Participants*) play a crucial role in modelling the interaction. The Social Context is then to be intended similar to the *Setting*: it "refers to the time and place of a speech act and, in general, to the physical circumstances" (Hymes, 2013, p. 55). Instead, we do not consider *Participants* because the CALLIOPE perspective focuses on IU's sound realization rather than interaction, and speaker's characteristics are considered in other dimensions of the model. *Social contexts* affect speakers' prosody (Klasen et al., 2018). For example, if we imagine a university lecture or a political meeting, the same IU (e.g. greeting) will have a different prosodic realization. Indicating the speaker's *Social Context* helps in ensuring the repeatability of experiments, and allows to contextualise accurately the research: in line with the previous example, another case is an experimental research concerning the persuasive ability of the political speech. It is not possible to provide a comprehensive list of elements, and this is the reason why D_{11} is a *Background dimensions* with an open set of elements for which we propose some examples such as Political context (debate, meeting), Teaching activities, Informal situation, Ceremony, etc.

3.2.3. D_{12} (Language, Dialect or Local Variety)

We added to our model an explicit dimension where we represent each language variety. In this sense, each language variety can be imagined as a space "slice", where the phonotactics (Goldrick, 2004), and the expressiveness modalities of the specific community are applied. Figure 6 has been inserted to clarify how the linguistic variety is positioned in the Calliope model (it is a conceptual representation without value outside the model).

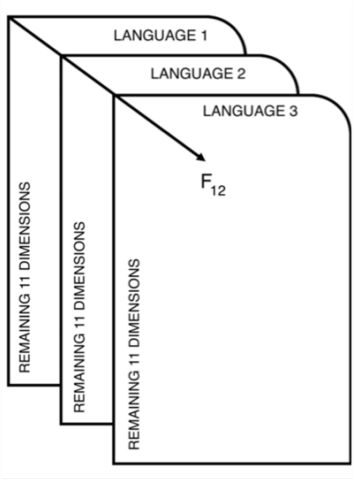

Fig. 6. How to visualise *Language, Dialect or Local Variety* dimension in CALLIOPE.

Indicating the speaker's *Language, Dialect or Local variety* helps to ensure the repeatability of experiments, and allows to contextualise accurately the research (e.g. in a research concerning how local accents affect the acoustic realisation of the yes/no question). In order to overcome misunderstandings (Haugen, 1966), we define *Language, Dialect or Local variety* as reported by Berruto & Cerruti (2015) and Coseriu (1980), considering them as an open set of elements, to be defined according to the purpose of the research. For example, a broad-spectrum corpus for gender recognition may have a national language as a label, plus subsets regarding local varieties, while sociolinguistics research can apply labels such as *Palermo dialect*, or *Dublin variety of Irish English.*

4. Methodology

In order to apply this categorization to a series of audio samples, it is necessary to define a label for each dimension of CALLIOPE before collecting the audio, taking into account the purpose of the research we are dealing with.

As an example, we can consider López Zorrilla et al. (2018), where a neural network was trained to automatically recognize a corrective pragmatic accent in a set of IUs. Starting from the background dimension, the audio samples are contextualized

defining from D12 to D1, as shown in table 2: both sentences have been recorded by professional actors, and the corpus is described by two tuples, corresponding to the two columns.

This process, which can seem simple and/or obvious, actually makes it possible to address systematically any weak point of the specific research as reported in the introduction, and (if possible) to remedy issues concerning the prosodic variability by integrating or redefining the corpus.

A corpus having more than one label for several dimensions will be described by a series of tuples, one for each characterizing combination. As a consequence, corpora described by a single tuple will be less variable on the prosodic level than those described by more than one (for example, sociophonetic researches such as the study of the rhotic for a specific dialect in a selected speakers' community).

DIMENSION	"Domani è bel tempo."	"DoMAni è bel tempo."
12. Language	IT Milanese variation	IT Milanese variation
11. Context	Experimental recording	Experimental recording
10. Personal skills	Standard (able-bodied speaker)	Standard (able-bodied speaker)
9. Emotion	Null	Null
8. Spontaneity	Recited speech	Recited speech
7. Punctuation form	Null	Null
6. Speech mood	Normal	Normal
5. Interpersonal motivational system	Neutral	Neutral
4. Rhetorical form	None	None
3. Focus	**Presentational**	**Corrective**
2. Illocutionary act	**Descriptive**	**Objection**
1. Structure	Declarative	Declarative

Table 2. Example of tuples for the Italian sentence "Tomorrow the weather is going to be fine." with/without pragmatic corrective focus on the second syllable of "tomorrow" (*domani*).

5. Validation

CALLIOPE is a conceptual categorization system, whose validation depends on the mutual independence of its dimensions and, consequently, of each label. In this perspective, a possible methodology could exploit perception tests focused on single labels taking into account mutual inferences between dimensions. If a label can be perceived/recognized by listeners, then the related IU will contain textual and/or acoustic clues also exploitable in the field of automatic recognition. A first step has been made on SI-CALLIOPE, a corpus based on the *Interactive Atlas of Romance Intonation* Italian script for elicited speech (Prieto et al., 2010) used for automatic prosodic features recognition (Cenceschi et al., 2019, 2018b, 2018c; López Zorrilla et al., 2018). It is characterized by a high

prosodic variability, but 6 dimensions can be fixed as follows:

- D_{12} = IT Milanese variation
- D_{11} = Daily situations
- D_{10} = Able-bodied speakers
- D_9 = Null (emotionally neutral)
- D_8 = Null (no pauses)
- D_7 = Elicited speech

The following labels have been tested (numbering according to the annex).

For D_1 dimension:

(1) Declarative
(2) Interrogative with 1 tonal unit
(3) Interrogative with 2 or more tonal units
(4) Interrogative disjunctive
(5) Echo questions
(6) Exclamative
(7) Vocative

For D_3 dimension:

(2) Corrective focus

Each listener was asked to recognize the presence of each one of these labels in a set of randomized samples[3] (both in real IUs, pseudo-words IUs, and IU's pitch envelopes) in order to investigate the influence of semantics, phonotaxis and intonation on detecting processes.

Results show that these labels can be independently recognized by listeners relying both on text and audio clues, with some interesting exception for which the text is almost useless (e.g. the pragmatic focus detailed in Cenceschi, 2019). Then, the characterizing features were investigated in four of the eight labels: Question (different typologies), Exclamation, Statement, and Corrective Focus. Results have been useful in training automatic recognition algorithms relying on embedding techniques and acoustic features (Cenceschi et al., 2018a; López Zorrilla et al., 2018).

The results suggest that these 4 labels can be independently recognized both at perceptive and digital level. At the same time, it is clear that the entire conceptual model validation will require a very extensive work, which may lead to several updates over time as happened for the L-AcT theory. Nothing will prevent to expand the labels set for one or more dimensions: for example, *Emotions*, *Illocutionary Act* or *Speech Mood* could be tuned using more labels. In this sense, it is essential to apply CALLIOPE to different sets of data, in order to test and refine the model and include as many prosodic nuances as possible. Moreover, as labels are validated, it will be of great interest to study their mutual inferences and relationships.

6. Discussion and future perspective

Even if CALLIOPE is at an early stage, we believe that its application for corpora description can however be extremely helpful in:

- Defining the prosodic context of a research
- Ensuring recordings' repeatability
- Highlighting the criticalities and strengths of a prosodic research
- Providing interesting clues for the tuning and enrichment of the model
- Increasing the dialogue between linguistic and engineering disciplines

As underlined in Section 2, this proposal collects and attempts to integrate theories belonging to different fields of the prosodic research, but unlike the majority of previous approaches, it attempts to categorize the IU's prosody basing on multiple set of qualitative labels, and not to quantify one or more acoustic behaviours (e.g. dialect or language variety characterization), or a specific aspect of the human communication (e.g. dialogue modelling). This makes it complementary (and not comparable) to the

[3] http://calliope.deib.polimi.it.

models mentioned in the state of the art, because it can be applied before performing a quantitative analysis, in order to define the prosodic variation of its samples.

Moreover, CALLIOPE does not catalogue the IU in a dialogue perspective: speech productions preceding and following a specific IU influences its acoustic and textual realization, but we chose to not consider temporal and dialogic perspectives at this stage. An expanded model should to consider existing theories and trends in this field (Wilks et al., 2010; Arora et al., 2013; Ferrari, 2004) causing a further huge increase in complexity, especially for validation.

The complexity of the prosodic phenomenon is a general problem in research, and in particular when attempting to measure a variability or, for example, to characterize a universe of speakers.

CALLIOPE attempts to describe IUs' prosodic variability by means of tuples with an interdisciplinary approach, and urging to reflect on factors not necessarily belonging to one's own field of expertise. Moreover, the validation proposal described in Section 5 could provide interesting clues regarding the features related to psychoacoustic processes applicable, for example, in automatic recognition algorithms, NLP systems, sociophonetic or pragmatic research. Further steps will include the analysis of the remaining labels of Section 5, plus the developing of further algorithms for their automatic recognition. However, the possible developments are many, but the necessary step in order to validate CALLIOPE is to start building corpora by applying it. It will allow, for example, to highlight the dimensions that need greater accuracy and greater number of labels, and to precisely identify non-existent tuples, or non-independent labels.

References

Akçay, M. B., & Oğuz, K. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers, *Speech Communication*, *116*, 56-76.

Arora, S., Batra, K., & Singh, S. (2013). Dialogue system: A brief review, *arXiv:1306.4134*.

Austin, J. L. (1975). *How to do things with words*. Oxford University Press.

Baker, J. M., Deng, L., Glass, J., Khudanpur, S., Lee, C. H., Morgan, N., & O'Shaughnessy, D. (2009). Developments and directions in speech recognition and understanding, Part 1 [DSP Education], *IEEE Signal processing magazine*, *26*(3), 75-80.

Beckman, M. E., Hirschberg, J. B., & Shattuck-Hufnagel, S. (2004). The original ToBI system and the evolution of the ToBI framework. In S-A. Jun (Ed.). *Prosodic typology: The phonology of intonation and phrasing* (pp. 9-54). Oxford University Press.

Bennett, R. T. (2012). *Foot-conditioned phonotactics and prosodic constituency* (Unpublished doctoral dissertation). UC Santa Cruz, United States of America.

Berruto, G., & Cerruti, M. S. (2015). *Manuale di sociolinguistica*. Utet Università.

Boersma, P., & Weenink, D. (2017). Praat, a system for doing phonetics by computer (version 6.0. 28). *Institute of Phonetic Sciences University of Amsterdam.*

Booij, G. (1999). The role of the prosodic word in phonotactic generalizations, *Amsterdam studies in the theory and history of linguistic science series*, *4*, 47-72.

Bunt, H. (2009). The DIT++ taxonomy for functional dialogue markup. *Proceedings of the AAMAS 2009 Workshop: Towards a Standard Markup Language for Embodied Dialogue Acts* (pp. 13-24), Bucarest*, Romania (AAMAS2009).*

Bunt, H., Petukhova, V., Traum, D., & Alexandersson, J. (2017). Dialogue act annotation with the ISO 24617-2 standard. In Dahl A. D. (Ed.). *Multimodal interaction with W3C standards* (pp. 109-135). Springer.

Büring, D. (2009). *Towards a typology of focus realization*. In M. Zimmermann & C. Féry (Eds.). *Information Structure* (pp. 177-205). Oxford University Press.

Calliope, L., & Fant, G. (1989). *La parole et son traitement automatique*. Masson.

Carbone, F., & Petrone, C. (2020). L'impact de la prosodie et du lexique des émotions sur

l'activité électrodermale en français, *Colloque Langage et éMOTions*, Bordeaux, France.

Cenceschi, S. (2019). *Speech analysis for automatic prosody recognition* (Unpublished doctoral dissertation). Politecnico di Milano, Italy.

Cenceschi S., Meluzzi C., & Nese, N. (2021). Speaker's identification across recording modalities: a preliminary phonetic experiment, Proceedings of the 16th AISV national conference, *Studi AISV* (Vol. 7), Officinaventuno, in print.

Cenceschi, S., Tedesco, R., Sbattella, L., Losio, D., & Luchetti, M. (2019). PESInet: Automatic Recognition of Italian Statements, Questions, and Exclamations With Neural Networks, *Proceedings of the Sixth Italian Conference on Computational Linguistics 2019, Bari, Italy, (CLiC-it19).*

Cenceschi, S., Sbattella, L., & Tedesco, R. (2018a). Influence of semantics on the perception of corrective focus in spoken Italian. In Botinis A. (Ed.). *Proceedings of 9th Tutorial and Research Workshop on Experimental Linguistics (pp. 21-24), Paris, France (Exling18)*.

Cenceschi, S., Sbattella, L., & Tedesco, R. (2018b). Towards automatic recognition of prosody. In Klessa K., Bachan J., Wagner A., Karpiński M. & Śledziński D. (Eds.). *Proceedings of the 9th International Conference on Speech Prosody, (pp. 319-323). Poznan, Poland (SpeechProsody2018).*

Cenceschi, S., Sbattella, L., & Tedesco, R. (2018c). Verso il riconoscimento automatico della prosodia. In Avesani C. (Ed). Proceedings of the 15th AISV national conference, *Studi AISV* (Vol. 3), (pp. 433-440) Officinaventuno.

Cole, J. (2015). Prosody in context: A review, *Language, Cognition and Neuroscience, 30*(1-2), 1-31.

Coseriu, E. (1980). "*Historische Sprache" und "Dialekt"* (pp. 45-61). Franz Steiner Verlag.

Cresti, E. (2000). *Corpus di italiano parlato* (Vol. 1). Accademia della Crusca.

Cresti, E. (2014). Syntactic properties of spontaneous speech in the Language into Act Theory. In Raso T. & Mello H. (Eds). *Spoken Corpora and Linguistic Studies* (pp. 365-410). John Benjamins.

Cresti, E. (2020). The pragmatic analysis of speech and its illocutionary classification according to the Language into Act Theory. In S. Izre´el, Mello H., Panunzi A. & Raso T. (Eds.). *In search of basic units of spoken language: A corpus-driven approach* (pp. 181-219). John Benjamins.

Cresti, E., Martin, P., & Moneglia, M. (1998). L'intonazione delle illocuzioni naturali rappresentative: analisi e validazione percettiva. *Atti delle IX giornate del gruppo di fonetica sperimentale, AIA* (pp. 51-63). Unipress.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing research, 229*(1-2), 132-147.

De Iacovo, V. (2019). *Intonation analysis on some samples of Italian dialects: An instrumental approach* (Vol. 3). Edizioni dell'Orso.

D'Imperio, M. (2002). Italian intonation: An overview and some questions. *Probus, 14*(1), 37-69.

Domínguez, L. (2004). *Mapping focus: The syntax and prosody of focus in Spanish* (Unpublished doctoral dissertation). Boston University, United States of America.

Douglas-Cowie E. et al. (2007). The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data. In Paiva A., Prada R., & Picard R.W. (Eds). *Proceedings of the Second International Conference ACII 2007* (pp.488-500). Springer.

Fassone, G., Valcella, F., Pallini, S., Scarcella, F., Tombolini, L., Ivaldi, A., & Liotti, G. (2012). Assessment of Interpersonal Motivation in Transcripts (AIMIT): An inter and intra rater reliability study of a new method of detection of interpersonal motivational systems in psychotherapy, *Clinical psychology & psychotherapy*, *19*(3), 224-234.

Ferrari, G. (2004). State of the art in Computational Linguistics. In Van Sterkenburg P. (Ed). *Linguistics today: Facing a greater challenge*, (pp. 163-186), John Benjamins Publishing Company.

Firenzuoli, V. (2003). *Le forme intonative di valore illocutivo dell'italiano parlato. Analisi sperimentale di un corpus di parlato spontaneo (LABLITA)* (Unpublished doctoral dissertation), Università di Firenze, Italy.

Fujisaki, H. (1997). Prosody, models, and spontaneous speech. In Sagisaka Y., Campbell N., & Higuchi N. (Eds). *Computing prosody: Computational models for processing spontaneous speech* (pp. 27-42). Springer.

Fux, T., Feng, G., & Zimpfer, V. (2011). Relevant acoustic features of speech signals for natural-to-shouted voice transformation, *Proceedings of the 6th European Congress on Acoustics, Forum Acusticum* (Vol. 97, Suppl. 1). *Acta Acustica & Acustica (EAA). Aalborg, Denmark.*

Ghaffarzadegan, S., Bořil, H., & Hansen, J. H. (2014). Model and feature based compensation for whispered speech recognition. In Li, H. et al. (Eds.). *Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association, Singapore (Interspeech2014)* (pp. 2420-2424).

Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production, *Journal of Memory and Language, 51*(4), 586-603.

Gussenhoven, C. (2008). Types of focus in English. In *Topic and focus* (pp. 83-100). Springer.

Hansen, J. H., Nandwana, M. K., & Shokouhi, N. (2017). Analysis of human scream and its impact on text-independent speaker verification, *The Journal of the Acoustical Society of America, 141*(4), 2957-2967.

Harris, R. A. (1997). *A handbook of rhetorical devices*. Retrieved from https://hellesdon.org/documents/Advanced%20Rhetoric.pdf

Hacquard (2011). Modality. In Maienborn C., von Heusinger K., & Portner P. (Eds.), *Semantics: An international handbook of natural language meaning* (pp.1484–1515). Mouton de Gruyter.

Haugen, E. (1966). Dialect, Language, Nation 1. *American anthropologist*, *68*(4), 922-935.

Hymes, D. (2013). *Foundations in sociolinguistics: An ethnographic approach*. Routledge.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(30), 11381-11386, National Academy of Sciences.

Klasen, M., von Marschall, C., Isman, G., Zvyagintsev, M., Gur, R. C., & Mathiak, K. (2018). Prosody production networks are modulated by sensory cues and social context, *Social cognitive and affective neuroscience, 13*(4), 418-429.

Kompe, R., & Kompe, R. (1997). *Prosody in speech understanding systems* (Vol. 1307). Springer.

Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives* (Vol. 36). Oxford University Press.

Jovičić, S. T. (1998). Formant feature differences between whispered and voiced sustained vowels, *Acta Acustica united with Acustica*, *84*(4), 739-743.

Leoni, F. A. (2001). Il ruolo dell'udito nella comunicazione linguistica. Il caso della prosodia, *Italian Journal of Linguistics*, *13*, 45-68.

Leoni, F. A. (2017). *Lingua e patologia: Le frontiere interdisciplinari del linguaggio.* Aracne.

Leoni, F. A., & Giordano, F. (2005) (a cura di), *Italiano parlato. Analisi di un dialogo.* Liguori, Napoli.

Likitha, M. S., Gupta, S. R. R., Hasitha, K., & Raju, A. U. (2017, March). Speech based human emotion recognition using MFCC. In *2017 international conference on wireless communications, signal processing and networking, Chennai, India (WiSPNET)* (pp. 2257-2260).

Liotti, G., & Monticelli, F. (2008). *I sistemi motivazionali nel dialogo clinico*. Raffaello Cortina.

Llisterri, J. (1992, July). Speaking styles in speech research, *ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language, Dublin, Irland.*

López Zorrilla, A., De Velasco Vázquez, M., Cenceschi, S., & Torres Barañano, M. I. (2018). Corrective focus detection in Italian speech using neural networks, *Acta Polytechnica Hungarica, 15*(5), 109-127.

Nencioni, G. (1983). *Di scritto e di parlato: Discorsi linguistici* (Vol. 6). Zanichelli.

Nicholson, J., Takahashi, K., & Nakatsu, R. (2000). Emotion recognition in speech using neural networks, *Neural computing & applications, 9*(4), 290-296, Springer.

Noth, E., Batliner, A., Kießling, A., Kompe, R., & Niemann, H. (2000). Verbmobil: The use of prosody in the linguistic components of a speech understanding system. *IEEE Transactions on Speech and Audio Processing, 8*(5), 519-532.

Origlia, A., Cutugno, F., & Galatà, V. (2014). Continuous emotion recognition with phonetic syllables. *Speech Communication*, *57*, 155-169.

Plutchik, R. (1991). *The emotions*. University Press of America.

Prieto, P., Borràs-Comes, J., & Roseano, P. (2010). Interactive atlas of Romance intonation. Retreived from http://prosodia. upf. edu/iari.

Romano, A., Contini, M., & Lai, J. P. (2014). *L'Atlas Multimédia Prosodique de l'Espace Roman: uno strumento per lo studio della variazione geoprosodica*. in Tosques F. (Ed.), 20 Jahre digitale Sprachgeographie, Berlin: Humboldt-Universität - Institut für Romanistik, 27-51.

Sbattella, L., Tedesco, R., & Trivilini, A. (2014). Forensic examinations: Computational analysis and information extraction. *Proceedings of the International Conference on Forensic Science-Criminalistics Research Singapore (FSCR)* (pp. 1-10).

Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (2011). Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge, *Speech Communication*, *53*(9-10), 1062-1087.

Schröder, M., Pirker, H., & Lamolle, M. (2006, May). First suggestions for an emotion annotation and representation language, *Proceedings of The International Conference on Language Resources and Evaluation, Genoa, Italy (LREC2006).*

Swain, M., Routray, A., & Kabisatpathy, P. (2018). Databases, features and classifiers for speech emotion recognition: a review, *International Journal of Speech Technology, 21*(1), 93-120.

Tamburini, F., Bertini, C., & Bertinetto, P. M. (2014). Prosodic prominence detection in Italian continuous speech using probabilistic graphical models. In Campbell N., Gibbon D., & Hirst D. (Eds). *Proceedings of the 5th International Conference on Speech Prosody Dublin, Ireland (SpeechProsody2014)* (pp. 285-289).

Tomkins, S. S. (1984). Affect theory. In Scherer K. R. & Ekma P. (Eds.), *Approaches to emotion* (pp. 163-195). Psychology Press.

Vasco, V., Gensini, S., & Leoni, F. A. (2010). *Tu chiamale se vuoi emozioni": Espressione e riconoscimento degli stati d'animo nel parlato* (Unpublished doctoral dissertation). University La Sapienza, Italy.

Vogt, T. (2010). *Real-time automatic emotion recognition from speech* (Unpublished doctoral dissertation). Universität Bielefeld, Germany.

Wallace, K. R. (1970). *Understanding Discourse: The Speech Act and Rhetorical Action.* Louisiana State University Press.

Wilks, Y., Catizone, R., Worgan, S., & Turunen, M. (2010). Some background on dialogue management and conversational speech for dialogue systems, *Computer Speech and Language, 25*(2), 128.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates, *The Journal of the Acoustical Society of America, 52*(4B), 1238-1250.

Zhang, C., & Hansen, J. H. (2007). Analysis and classification of speech mode: whispered through shouted, *Eighth Annual Conference of the International Speech Communication Association, Antwerp, Belgium (Interspeech2007)* (pp. 2289-2292).

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion, 8*(4), 494-521.

Appendix: CALLIOPE's labels

D_1 - Structure

(1) Declarative
(2) Interrogative with 1 tonal unit
(3) Interrogative with 2 or more tonal units
(4) Interrogative disjunctive
(5) Echo questions
(6) Exclamative
(7) Vocative

D_2 – Illocutionary Act

25 labels divided into five main groups.

Alethic:
(1) Assumption
(2) Confirmation
(3) Objection
(4) Admission
(5) Ascertainment
(6) Description
(7) Explanation
(8) Clarification
(9) Inference

Epistemic:
(10) Intuition
(11) Conjecture
(12) Inference
(13) Doubt
(14) Supposition
(15) Prediction
(16) Query

Appreciative:
(17) Opinion
(18) Judgement

Volitive:
(19) Desire
(20) Decision

Deontic:
(21) Advice
(22) Permission
(23) Request
(24) Exhortation
(25) Admonition
(26) Instruction

D_3 - Intonational focus

(1) Presentational
(2) Corrective
(3) Counter-presuppositional
(4) Definitional
(5) Contingency
(6) Reactivating
(7) Identificational
(8) Non-focused

D_4 - Rhetorical form

(1) Irony
(2) Aposiopesis
(3) Prepetition
(4) Anacoluthon
(5) Expeditio
(6) Eutrepismus
(7) Dialysis
(8) Sentential Adverb
(9) Polysyndeton
(10) Asyndeton
(11) Rhetorical question
(12) Parenthesis
(13) Epizeuxis
(14) Enumeratio
(15) Neutral

D_5 - Interpersonal Motivational State

(1) Attachment
(2) Caregiving
(3) Rank
(4) Sexual
(5) Peer cooperation
(6) Neutral

D_6 - Speech mood

(1) Whispered
(2) Soft
(3) Neutral

(4) Loud
(5) Shouted

D_7 - Spontaneity

(1) Spoken
(2) Read
(3) Recited
(4) Elicited

D_8 - Punctuation form

(1) Single commas
(2) Null

D_9 - Emotions

Negative and forceful:
(1) Anger
(2) Annoyance
(3) Contempt
(4) Disgust
(5) Irritation

Negative and not in control:
(6) Anxiety
(7) Embarrassment
(8) Fear
(9) Helplessness
(10) Powerlessness
(11) Worry

Negative thoughts:
(12) Pride
(13) Doubt
(14) Envy
(15) Frustration
(16) Guilt
(17) Shame

Negative and passive:
(18) Boredom
(19) Despair
(20) Disappointment
(21) Hurt
(22) Sadness

Agitation:
(23) Stress
(24) Shock
(25) Tension

Positive and lively:
(26) Amusement
(27) Delight
(28) Elation
(29) Excitement
(30) Happiness
(31) Joy
(32) Pleasure

Caring:
(33) Affection
(34) Empathy
(35) Friendliness
(36) Love

Positive thoughts:
(37) Courage
(38) Hope
(39) Humility
(40) Satisfaction
(41) Trust

Quiet positive:
(42) Calmness
(43) Contentment
(44) Relaxation
(45) Relief
(46) Serenity

Reactive:
(47) Interest
(48) Politeness
(49) Surprise

Neutral label:
(50) Null

D_{10} - Subjective expressiveness skills (Open set)

- Able-bodied speaker
- Verbal dyspraxia
- Aphasia
- ...

D_{11} - Social context (Open set)

- Thesis dissertation
- Political debate
- Religious ritual
- Teaching activity
- Informal situation
- Ceremony
- …

D_{12} - Language, dialect or local variety (Open set)

- Genoese variety of the Ligurian dialect
- Dublin variety of Irish English
- …