UNIVERSITAT DE BARCELONA

Estudios de Fonética Experimental

Journal of Experimental Phonetics

ISSN: 1575-5533 · ISSN-e: 2385-3573

EFE
Estudios de Fonética Experimental
Journal of Experimental Phonetics

Pyrlato: A novel methodology to collect real-world acoustic data

Giuseppe Magistro [1] 0000-0002-0272-741X
Claudia Crocco [2] 0000-0003-1099-956X

[1] Ghent University (Belgium)

DOI: 10.1344/efe-2023-32-243-254

Corresponding address: giuseppe.magistro@ugent.be

Received: 12/09/2023 Accepted: 07/11/2023 Published: 28/11/2023

Magistro, G., & Crocco, C. (2023). Pyrlato: a novel methodology to collect real-world acoustic data. *Estudios de Fonética Experimental*, *32*, 243–254. https://doi.org/10.1344/efe-2023-32-243-254

Abstract

In this paper, we present *Pyrlato*, an innovative tool developed in Python for collecting acoustic data from *YouTube*. The development of this tool was motivated by the need to conveniently collect real-world spoken data. By executing this Python code, researchers can obtain a spoken corpus of specific words, syllables, constituents, and more. We illustrate the main steps of the execution to demonstrate how it works and how to use it. Additionally, we provide a complete example for reference, demonstrating how to customize *Pyrlato* according to specific requirements. Finally, we discuss the future developments we intend to cover for *Pyrlato*.

Keywords

real-world data; ecological validity; data scraping

Pyrlato: Una metodologia innovadora per recopilar dades acústiques del món real

RESUM

En aquest article presentem *Pyrlato*, una eina innovadora desenvolupada en Python per recopilar dades acústiques de *YouTube*. El desenvolupament d'aquesta eina va ser motivat per la necessitat de recopilar de manera convenient dades de parla del món real. Mitjançant l'execució d'aquest codi de Python, els investigadors poden obtenir un corpus parlat de paraules específiques, síl·labes, constituents i més. Il·lustrem els passos principals de l'execució per demostrar com funciona i com utilitzar-ho. A més, proporcionem un exemple complet de referència, demostrant com personalitzar *Pyrlato* segons requisits específics. Finalment, discutim els desenvolupaments futurs que es pretenen amb *Pyrlato*.

MOTS CLAU

dades del món real; validesa ecològica; extracció de dades

1. Introduction

In spite of the growing significance of web-based corpora for linguistic research, the internet remains largely underexploited as a resource of audio data for the linguistic research of speech. To address this gap, we developed *Pyrlato*, a prototype program designed for extracting audio data from *YouTube* speech. The data collected using *Pyrlato* present challenges in terms of the analyses they can support, but speech data sourced from *YouTube* hold considerable potential for investigating phonetic variation in spoken communication.

2. Background

Corpus-based linguistic research has received a significant boost from the use of the web as a source of linguistic data. The web-based corpora currently available probably constitute the largest collections of linguistic data ever created (cf. Stefanowitsch, 2020). Such extensive corpora enable quantitative studies with an entirely new scope compared to what was possible using traditional corpora. While linguistic data on the internet is primarily written in nature, the web also collects large amounts of different types of data, such as videos, images, and audio. Because of their naturalness, audio-visual broadcasted data from television, radio, and other media is customary in research traditions such as discourse analysis (e.g., Couper-Kuhlen, 1986; Couper-Kuhlen & Selting, 2018). In contrast, media-based audio data are normally discarded in phonetic research because of their lesser acoustic quality.

New technologies for collecting audio data for linguistic research through smartphone applications have in recent years gained significant momentum due to the needs created by the COVID-19 pandemic (Leemann et al., 2018). Phone apps enable the collection of large quantities of data suitable for acoustic analyses as well (Awan et al., in press; Magistro, 2021), representing a good compromise between the need to collect acoustically qualitative speech and the need to collect more data in a shorter time and with a low budget. However, elicited data intrinsically present ecological limitations since they are produced "on demand", i.e., outside of a natural interactional context.

At present, no tool offers the possibility of collecting large amounts of natural and yet acoustically qualitative audio data, in a short time, relying on a large informant basis. However, a further step in the quest for the ideal audio data and collection procedure for phonetic research on speech can be set by exploring the potential of web-based data for phonetic analysis. Although web-based spoken data are dirty data, they allow to collect relatively large amounts of data, from a diverse pool of speakers, in very short time and virtually with no budget. They are in this sense the ideal data for explorative and validation studies.

In the following section we discuss the theoretical value of analysing *YouTube*-based data, especially for prosodic studies, notwithstanding their lower acoustic quality. Subsequently, in section 3, we illustrate the functioning of *Pyrlato* and provide example data retrieved with *Pyrlato*-based queries.

2.1. Natural data and acoustic quality

A crucial feature of *YouTube* data is their naturalness. As emphasized by Tucker and Ernestus (2016), research on speech should adequately consider spontaneous data, i.e., specifically, non-read speech (Beckman, 1997; Warner, 2012) and "unprepared" speech (Tucker & Mukai, 2023, p. 3). One of the most relevant aspects of "non-read" and "unprepared" speech is the fact that it is produced in real, or at least more natural pragmatic and communicative conditions compared to those that are recreated in a laboratory.

Incidentally, we can observe that the use of laboratory speech, especially in phonetic studies, is closely related to a conception of accurate speech as closer to some sort of "ideal" form. Most linguistic models presuppose the existence of a canonical form of signs (Voghera, 2022, p. 31). The concept of phonetic reduction, for example, presupposes the existence of non-reduced forms from which reduced forms deviate (cf. Cangemi & Niebuhr, 2018). While this may be true in relative terms — as one form can certainly be formally richer than another — in absolute terms, it cannot be said that informal speech is "reduced" unless one has a reference form in mind. As Albano Leoni (2013, p. 133) states, by "reversing a common viewpoint, one could say that spoken language is not 'reduced', but rather that written language is 'overdetermined'." Likewise, laboratory speech could be seen as "overdetermined" compared to spontaneous and casual speech.

Of course, controlled data have several advantages: they offer researchers the opportunity to elicit exactly the desired form, despite the actual low frequency of that given phenomenon in real speech; controlled contexts allow for a close observation of subtle details, allowing for accurate measurements and a detailed phonetic analysis.

Laboratory speech, however, is produced, by definition, under non-natural pragmatic conditions. The same, although to a lesser extent, applies to speech obtained through specific elicitation techniques, such as Discourse Completion Task (Blum-Kulka et al., 1989; Vanrell et al., 2018). While certain paradigms, such as Map task (Anderson et al., 1991) and Spot the difference (Pean et al., 1993), allow to obtain dialogues that are in many ways natural, their pragmatic naturalness is severely limited by the predetermined task, with consequences on the linguistic choices made by the speakers. In fact, elicited speech presents some limitations in terms of variation representativity. In other words, the context of the laboratory or the organized and agreed-upon recording can lead speakers to avoid certain socially marked forms, for example. Therefore, certain structures and forms might never be produced, despite their frequency in everyday language. Additionally, the appropriate sociolinguistic context to trigger the production of a certain feature, may not be known, or cannot be replicated in the research setting.

In this context, collecting audio data using a social media platform like *YouTube* can present undeniable advantages, as it can provide access to a wide range of linguistic uses produced in real-world contexts.

A shortcoming of social media speech, including data from *YouTube*, is that it is often recorded in lossy formats, and the recordings are made in different conditions and frequently also in noisy contexts. Although such audio data cannot be used in the same way as laboratory data, they retain value for the examination of durational properties and fundamental frequency (see Baltazani et al., in press), since these acoustic features are relatively robust (Fuchs & Maxwell, 2016). However, *YouTube* data cannot be safely used e.g. for formant analysis due to poor audio quality (Parsa et al., 2001; Rathcke et al., 2017).

In this article, we present the features of *Pyrlato* and how it enables the extraction of audio data from *YouTube*. Additionally, we will showcase some applications to demonstrate the type of data that can be extracted, benefiting from both the absence of a researcher-collector and the fact that automated collection allows for obtaining large amounts of data in a short time. Furthermore, we will discuss the potential that *YouTube*-extracted data holds for linguistic analysis, for the study of prosody.

3. *Pyrlato*: the pipeline

Pyrlato was developed in Python 3.9. It relies on the external dependencies *pytube* (v. 12.1.0), *moviepy* (v. 1.0.3), *youtube_transcript_api* (v. 0.6.0). At the time of writing (September 2023) the code runs with updated dependencies and Python version. We briefly illustrate the algorithm behind the software for a better understanding of its potentiality and limitations. *Pyrlato* relies on Object-oriented programming, treating *YouTube* videos as object, whose subtitles are extracted and searched.

3.1. Step 1: Searching relevant videos

At first, the software searches videos using keywords inputted by the user. This function corresponds to the search bar of *YouTube*. This means that these keywords can be inputted to narrow the search (i.e. videos on specific topics, from specific speakers, in sociolinguistic contexts etc.). In the code, the variable where to specify this is Search_keyword. For example, if we are interested in the realisation of a certain syllable in read speech, as for example audiobooks, we can specify this field as in example (1).

```
(1) Search_keyword = 'audiobook'
```

Complex searches typically found in the search bar of *YouTube* are also available via shortcuts. Say, for instance, that a researcher is interested in the realisation of a certain prosodic pattern by the actress Anne Hathaway. It would be more convenient to exclude videos on the wife of William Shakespeare who was called Anne Hathaway as well. This can be done by specifying the exclusion with the '-' operator (see example 2)

```
(2) Search_keyword =
        'Anne Hathaway -Shakespeare'
```

Similarly, all other operators are available: OR for disjunctive search, *intitle:* to make sure that a certain string appears in the title of the video, *before:* to look for videos that were updated before a certain date, *after:* to search videos uploaded after the expressed date, etc. The software then starts to search for videos that match the description provided and saves the video ID in a list.

3.2. Step 2: Scraping the subtitles

At this point, the list is looped, and the subtitles are extracted and stored in a Python dictionary having the type {video_id:subtitles}. The subtitles are extracted using the automatic caption system of *YouTube*. As such, *Pyrlato* works for Dutch, English, French, German, Indonesian, Italian, Japanese, Korean, Portuguese, Russian, Spanish, Turkish, Ukrainian, and Vietnamese, for which *YouTube* provides speech recognition. Subtitles that were manually entered will also be stored. The code makes sure that the video is in the desired language. For the code in example (3), only videos in Italian will be included. It is possible, by adding a comma, to have videos from different languages (e.g. if the researcher is interested in seeing a particular phenomenon in a linguistic family).

```
(3) Language_search = ['it']
```

We refer the reader to the documentation of YouTube-transcript-api for more complications on this aspect.[1]

3.3. Step 3: Searching in the subtitles

The user is also asked to provide a string to search in the subtitles in the variable *Pattern* as exemplified in (4). This is the core aspect of *Pyrlato*, looking for a specific word, syllable, constituent etc. in the video. If a researcher, for example, is interested in the realisation of the consonant cluster *st*, the content of String will be matched in the captions.

```
(4) Pattern = ['st']
```

It is possible to indicate more alternatives by adding a comma, e.g. ['st', 'sp']. Note that the search is made possible via the Python module *re*. This means that regular expressions can be used to look for more complex patterns: for instance, to scrape data on the Italian clusters *str, spr, scr* at the beginning of a word, one could use a formula as the one reported in (5).[2]

```
(5) Pattern = ['\bs[tpc]r']
```

[1] https://pypi.org/project/youtube-transcript-api/

[2] https://docs.python.org/3/library/re.html and https://regex101.com on writing regular expressions in Python. An alternative without using regular expressions would be [' spr', ' str', ' scr'], but regular expressions are advised for more complex cases.

For languages with special characters or with non-Latin alphabet, the specific conventional writing system can be used.

The software then searches for all instances of the desired string in each video. If a string matches, the timestamp(s) of the occurring expression will be stored in a list. Each list is the value of a dictionary with the link to the video as key. Note that the timestamp that is stored in this list is not precisely aligned with the desired item, but rather with the beginning of the sentence that contains this string. This is an unavoidable limitation depending on how the encoding of the subtitles is given by *YouTube*. The results that we can get from the subtitles are like the ones reported in (6). The desired string is matched (*scremato)*, but the nearest timestamp is located at the beginning of *preferisco*. This is because, to date and to the best of our knowledge, *YouTube* does not yet provide in APIs a one-per-one timestamp.

```
(6) {'text': 'preferisco il latte scremato',
            'start': 14.06,
            'duration': 1.09}
```

Therefore, we introduce the final variable to be specified by the user, the time window in seconds to observe when cutting the sentence. This variable is named *Cut_time_fol* in our code. Example (7) shows a cutting time of 7 seconds: this is a sufficiently large time span to make sure that the desired string is found.

```
(7) Cut_time_fol = 7
```

It is possible to adjust the timespan by adding a preceding interval, this is stored in the variable *Cut_time_pre*. So, if we want a larger portion of audio with a preceding and following interval, this can be done by specifying the *cut_time_fol* and *cut_time_pre* variables.

3.4. Step 4: Getting the corpus

Finally, *Pyrlato* downloads the videos where the desired string was found and trims it following the timestamps. By default, only the audio of the video is extracted. It is possible, however, to extract the videos (i.e. for researchers working on multimodality). This can be done by changing the setting of the option *only_audio = False* in the code. The extracted fragments are saved in .mp4 in the working directory (the runtime directory in the case of Google Colab), their name has the structure in (8).

```
(8) Repetition + Ext_ + video_id.mp4
```

For instance, if we look for the instances of the German particle *doch*, the filename '*3Ext_ich stelle mich vor.mp4*' stands for that extracted audio segment where *doch* was found for the third time in the video named *ich stelle mich vor*.

The program then loops for every video and every instance found and extracts all the audio files in the working directory. Finally, a converter is also available: the user is being asked in the console if they want to convert the files in .mp3, so that they can be opened in Praat.[3]

4. Mock studies

We illustrate the application of *Pyrlato* to two possible queries for prosodic studies, by giving indications on how to run the code. An analysis of the obtained data is outside the scope of the paper, inasmuch as we intend to show the process of data collection using *Pyrlato*.

4.1. German *doch*

German modal particles are known for their expression of illocutionary force and epistemic stance (Abraham, 1991; Karagjosova, 2004). Under the assumption that modal particles interact with the intonational properties of the utterance (Prieto & Roseano, 2016, 2021, among others), it

[3] Other formats are also available. We do not encourage the conversion to lossless formats (e.g. .wav), because the audio tracks are already compressed, and the reduced information is not available anymore.

can be expected that certain prosodic properties are realised with the modal particles of German as well. This prediction is confirmed by recent work on the particle *doch* (Repp & Seeliger, 2023), which expresses a rejection of a proposition on the basis of the contradiction that arises in the speaker's knowledge. Egg and Zimmermann (2012) argue that, by virtue of its semantics, *doch* must be accented. However, they also point to the existence of an unaccented *doch*, which results from different prosodic and information-structural conditions. Their analysis is however not based on an experimental data collection. Reep and Seeliger (2023) in a production experiment with 18 speakers report that in rejection statements with *doch*, more prominent pitch accents are found in comparison to statements without the modal particle, with higher f0 maxima, greater excursion, and longer duration. To our knowledge, no bigger studies have been conducted on corpora or more naturalistic situations. We show the feasibility of such a data collection using *Pyrlato*. For ease of reproducibility, we will use the script online on *Google Colab*,[4] but the program can be used on local computers using a Python interpreter.

First, we install the required libraries by clicking on the play button at the left of the code cell (see the illustration in Figure 1). A green tick mark will appear if the required libraries are installed correctly, as in Figure 1.

```
######### Install the libraries ##############
!pip install imageio-ffmpeg
!pip install pytube
!pip install moviepy
!pip install ffmpeg
!pip install youtube-transcript-api
```

Figure 1. Installing the required libraries for *Pyrlato*.

[4] https://colab.research.google.com/drive/1zv67DpKehySkJLd0Vk7hpG7__Je4fbi4?usp=sharing

Then, we run the following code cell to import the desired libraries. We now customise the desired search. First, we set the language: in our case, this is German (*de* in the code). We decided to search the occurrences of *doch* in TV series (such as *Die Chefin* on the ZDF channel). We also set a cutting window to make sure that the entire sentence is scraped (*the cut_time* values), see Figure 2.

```
#########CUSTOM################
Language_search = ['de']
Pattern = [' doch ']
Search_keyword = 'Die Chefin'
Cut_time_fol = 7
Cut_time_pre = 5
```

Figure 2. The customized research for *doch* in German.

Next, we can run the next code cell where *Pyrlato* searches, downloads and cuts the desired occurrences. We repeated the search with other ten German TV shows and obtained a corpus of 3,277 instances (*Pyrlato* will display the number of instances by clicking on the following code cells). The scraped files are visible by clicking on the folder icon 'file' on the left of the Google Colab window, see Figures 3 and 4. A list of the extracted files is available, note that each file is named following the conventions in (5). It is possible to click and download them.

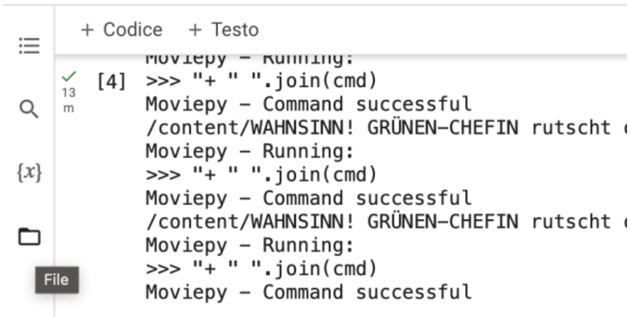

Figure 3. How to check the downloaded files in Google Colab.

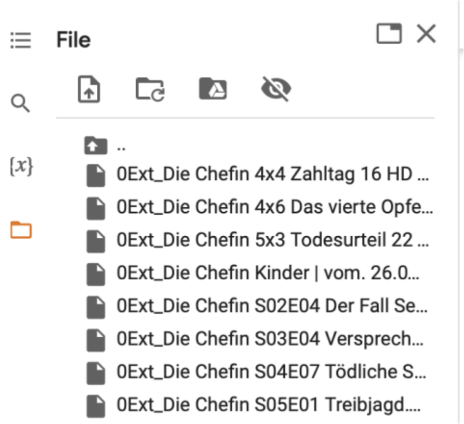

Figure 4. The download files.

As previously mentioned, there is a converter to .mp3 to ensure that the audio files can be opened in Praat. This is found in the next following cell. The last cell code will save every extracted audio in a .zip archive which is automatically downloaded from the browser to the local machine where *Pyrlato* is working. As all the recordings are saved in the .zip archive, it is simply necessary to run the cell and grant permission for the download.

We inspected a sample of 300 instances of the total corpus: 107 tokens were discarded because of poor sound quality or inaccurate subtitles. Despite this, a good number of instances are still usable for different analyses (e.g. visual and auditorily inspection of the pitch, perception of the prominence or duration values).

4.2. Gemination in Italian

We present now another application of *Pyrlato*. The difference in consonantal length is phonologically distinctive in Italian, see the geminate in e.g. *notte* [ˈnɔtːe] "night" and the short consonant in *note* [ˈnɔte] "notes". However, Italian possesses a great range of variation, especially when it comes to regional pronunciation (Crocco, 2017). In particular, it is often argued that northern speakers do not produce geminates consistently (Bertinetto & Loporcaro, 2005; Payne, 2005) or that the length of their geminates is shorter than the ones produced by central and southern speakers (Canepari, 1992). We refer the reader to Mairano and De Iacovo (2020) for a discussion on the literature. In reviewing the evidence for such claims, Mairano and De Iacovo underline that the experimental evidence for these claims (e.g. from corpora or laboratory tasks) is limited, if not lacking completely. While the previous studies that they cite are mainly based on impressionistic judgements, they use the data from the CLIPS corpus. In total, they collect a corpus of 24,242 geminated tokens coming from northern, central and southern speakers. We try to replicate a similar data collection with *Pyrlato*. For the sake of demonstration, we limit ourselves to the voiceless plosives /p/, /t/, /k/. In Figure 5 we report the customisation of this new case. Once we changed the language to 'it', we specified the strings of our interest. The use of the comma inside the square brackets allows to search for alternative realisations. In our case, we use it to list the desired strings. In the *Search_keyword* field, we have specified different local TVs and shows from each Italian region (e.g. *Telenorba* for Apulia, *Reteveneta* for Veneto, *TGR Lombardia* for Lombardy). Naturally, this specification does not guarantee the origin of the speaker, but it can work as an approximation in big data (see the discussion in the next section). For instance, it is possible to specify the names of TV presenters, interviewees, actors etc.

```
[1] #########CUSTOM################
    Language_search = ['it']
    Pattern = ['pp', 'tt', 'cc']
    Search_keyword = 'Reteveneta'
    Cut_time_fol = 7
    Cut_time_pre = 5
```

Figure 5. The customized research for the geminates /p/, /t/, /k/ in Italian.

We ran the cell containing *Pyrlato* pipeline each time with a different keyword. In our case, we used three keywords per region. In eight hours of computation (and manual input of keyword), we obtained a total of 41,893 geminates. Naturally, it is possible to download more instances with extra keywords. The corpus is also relatively balanced, given that almost 2,000 examples per region were scraped. We downloaded a small subset of 250 examples and found that in 247 instances, the presence of a geminate was correctly found. The remaining 3 cases are misinterpretations of less ‘standard words’ (e.g., the interjection *mado’* was interpreted by the *YouTube* ASR as *matto*).

As usual, the following cells can be used to convert the files in .mp3 and download the corpus in a zip file. We recommend downloading the obtained files since the runtime by Google Colab will delete them automatically. Otherwise, if the script is run locally, the files are automatically stored in the working directory.

In conclusion, Pyrlato helped us acquire the largest corpus on Italian geminates (to our knowledge). Naturally, casting such a wide net comes with the limitation of uncontrolled data: it can happen for example that in a local TV show from northern Italy, a speaker from the south has been interviewed, generating an imprecise data point in that corpus section. We believe that this is unavoidable with big data, but an accurate search and subsequent sanity check can limit such a drawback.

5. Conclusions and further improvements

The pipeline is still in its infancy, and we are working on new versions to overcome a number of limitations in the current version of the tool.

The first limitation is represented by the user’s experience. At present, *Pyrlato* requires the users to compile Python executables: it is required that the users download manually the external dependencies, correct the code to change specific settings etc. A graphic user interface (GUI) will be soon built using Tkinter and released in the upcoming versions.

The core limitation of *Pyrlato* is that it relies on the automatic captions provided by *YouTube*. Besides being a source of possible mistakes, subtitles are only supported for well-documented languages, for which the speech recognition model has sufficient training.

Finally, the absence of detailed metadata may limit the researcher's ability to contextualize and categorize the data effectively. However, this limitation can be at least partly circumvented by restricting the search field to specific keywords. A successive sanity-check is still necessary to guarantee the relevance of the data to the desired research aims.

In general, big data, characterized by their sheer volume, have often a messy nature presenting challenges in terms of quality and complexity. Despite their untidiness, big data offer a wealth of information that can reveal hidden patterns, trends, and correlations that can bring further advancement in the linguistic field. We believe that modern data analytics techniques can help researchers to have a better insight into real-world variability, uncovering patterns that might be otherwise obscured.

While *Pyrlato* may not yet possess the full range of features or capabilities that we envision, we believe that scraping data from *YouTube* and/or other social media will soon become an invaluable resource for reconstructing the different nature of linguistic forms and the intrinsic variability that is found in real-world communication.

Acknowledgements

We want to thank the audience at the conferences PaPE 2023 and “Il parlato in ambito medico: analisi linguistica, applicazioni tecnologiche e strumenti clinici” organized in Lecce.

References

Abraham, W. (1991). The grammaticization of the German modal particles. In E. Closs Traugott, & B. Heine (Eds.), *Approaches to Grammaticalization: Volume II. Types of grammatical markers* (pp. 331–380). John Benjamins. https://doi.org/10.1075/tsl.19.2.17abr

Albano Leoni, F. (2013). Il parlato e la comunicazione parlata. In G. Iannàccaro (Ed.), *Linguistica italiana all'alba del terzo millennio (1997–2010)* (pp. 129–148) [SLI 58]. Bulzoni.

Anderson, A. H., Bader, M., Gurman Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H., & Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4), 351–366. https://doi.org/10.1177/002383099103400404

Awan, S. N., Shaikh, M. A., Awan, J. A., Abdalla, I., Lim, K. O., & Misono, S. (in press). Smartphone recordings are comparable to "gold standard" recordings for acoustic measurements of voice. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2023.01.031

Baltazani, M., Coleman, J., Passoni, E., & Przedlacka, J. (in press). Echoes of Past Contact: Venetian Influence on Cretan Greek Intonation. In B. Gili Fivela B., & C. Avesani (Eds.), *Language and Speech* [PaPE 2019 Special Issue]. https://doi.org/10.1177/00238309221091939

Beckman, M. E. (1997). A typology of spontaneous speech. In Y. Sagisaka, N. Campbell, & N. Higuchi (Eds.), *Computing Prosody: Computational Models for Processing Spontaneous Speech* (pp. 7–26). Springer. http://dx.doi.org/10.1007/978-1-4612-2258-3_2

Bertinetto, P., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, *35*(2), 131–151. https://doi.org/10.1017/S0025100305002148

Blum-Kulka, S., House, J., & Kasper, G. (1989). Investigating crosscultural pragmatics: An introductory overview. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural Pragmatics. Requests and Apologies* (pp. 1–34). Ablex.

Canepari, L. (1992). *Manuale di pronuncia italiana, con un pronunciario di oltre 30000 voci e due audiocassette*. Zanichelli.

Cangemi, F., & Niebuhr, O. (2018). Rethinking reduction and canonical forms. In O. Niebuhr, F. C. Barbara Schuppler, M. Clayards, & M. Zellers (Eds.), *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation* (pp. 291–316). De Gruyter. https://doi.org/10.1515/9783110524178-009

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. Arnold & Niemeyer.

Couper-Kuhlen, E., & Selting, M. (2018). *Interactional linguistics: Studying language in social interaction*. Cambridge University Press. https://doi.org/10.1017/9781139507318

Crocco, C. (2017). Everyone has an accent: Standard Italian and regional pronunciation. (2017) In M. Cerruti, C. Crocco and S. Marzo (Eds.), *Towards a new standard: Theoretical and empirical studies on the restandardization of Italian* (pp. 89–117), Mouton de Gruyter, https://doi.org/10.1515/9781614518839-004

Egg, M., & Zimmermann, M. (2012). Stressed out! Accented discourse particles: The case of *DOCH*. In A. Aguilar Guevara, A. Chernilovskaya, & R. Nouwen (Eds.), *Proceedings of Sinn und Bedeutung 16* (pp. 225–238). MIT Press.

Fuchs, R., & Maxwell, O. (2016). The effects of MP3 compression on acoustic measurements of fundamental frequency and pitch range. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of Speech Prosody 8* (pp. 523–527). International Speech Communication Association. http://doi.org/0.21437/SpeechProsody.2016-107

Karagjosova, E. (2004). *The meaning and function of German modal particles* [Doctoral dissertation, Universität des Saarlandes]. Saarabrücken Dissertations in Computational Linguistics and Language Technology.

Leemann, A., Marie-José, K., & David, B. (2018). The English dialects app: The creation of a crowdsourced dialect corpus. *Ampersand*, *5*, 1–17. https://doi.org/10.1016/j.amper.2017.11.001

Magistro, G. (2021). Speech prosody and remote experiments: A technical report. *arXiv, 2106*, Article 10915. https://doi.org/10.48550/arXiv.2106.10915

Mairano, P., & de Iacovo, V. (2020) Gemination in Northern versus Central and Southern varieties of Italian: A corpus-based investigation.

Language and Speech, *63*(3), 608–634. https://doi.org/10.1177/0023830919875481

Parsa, V., Jamieson, D., & Pretty, B. R. (2001). Effects of microphone type on acoustic measures of voice. *Journal of Voice 15*(3), 331–343. https://doi.org/10.1016/S0892-1997(01)00035-2

Payne, E. M. (2005). Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association*, *35*(2), 153–181. https://doi.org/10.1017/S0025100305002240

Pean, V., Williams, S., & Eskenazi, M. (1993). The design and recording of icy, a corpus for the study of intraspeaker variability and the characterisation of speaking styles. In *Proceedings of the 3rd European Conference on Speech Communication and Technology (Eurospeech 1993)* (pp. 627–630). International Speech Communication Association. https://doi.org/10.21437/Eurospeech.1993-152

Prieto, P., & Roseano, P. (2016). The encoding of epistemic operations in two romance languages: Intonation and pragmatic markers. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of the 8th Speech Prosody* (pp. 888–892). International Speech Communication Association. https://doi.org/10.21437/SpeechProsody.2016-182

Prieto, P., & Roseano, P. (2021). The encoding of epistemic operations in two Romance laguages: The interplay between intonation and discourse markers. *Journal of Pragmatics*, *172*, 146–163. https://doi.org/10.1016/j.pragma.2020.11.008

Rathcke, T., Stuart-Smith, J. Torsney, B., & Harrington, J. (2017). The beauty in a beast: Minimising the effects of diverse recording quality on vowel formant measurements in sociophonetic real-time studies. *Speech Communication*, *86*, 24–41. https://doi.org/10.1016/j.specom.2016.11.001

Repp, S., & Seeliger, H. (2023) Reject?! On the prosody of non acceptance. In R. Skarnitzl, & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1355–1359). Guarant International.

Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press. https://doi.org/10.5281/zenodo.3735822

Tucker, B. V., & Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The Mental Lexicon 11*(3), 375–400. https://doi.org/10.1075/ml.11.3.03tuc

Tucker, B. V., & Mukai, Y. (2023). *Spontaneous speech*. Cambridge University Press. http://doi.org/10.1017/9781108943024

Vanrell, M. M., Feldhausen, I., & Astruc, L. (2018). The Discourse Completion Task in Romance prosody research. In I. Fedlhausen, J. Fliessbach, & M. M. Vanrell (Eds.), *Methods in prosody: A Romance language perspective* (pp. 191–227). Language Science Press. https://doi.org/10.5281/zenodo.1441345

Voghera, M. (2022) *From Speaking to Grammar*. Peter Lang. https://doi.org/10.3726/b19221

Warner, N. (2012). Methods for studying spontaneous speech. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 621–633). Oxford University Press.