Universidade Estadual de Campinas
Instituto de Estudos da Linguagem

Gabriel Catani

Newscasting Prosody:

a Forensic Sociophonetic Approach

Prosódia telejornalística:
uma abordagem sociofonético-forense

Campinas
2023

Gabriel Catani

Newscasting Prosody:
a Forensic Sociophonetic Approach

Prosódia telejornalística:
uma abordagem sociofonético-forense

Master Thesis presented to the Institute of Language Studies of the University of Campinas in partial fulfillment of the requirements for the degree of Master of Linguistics.

Dissertação de Mestrado apresentada ao Instituto de Estudos da Linguagem da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestre em Linguística.

Advisor: Prof. PhD. Livia Oushiro
Co-advisor: PhD. Renata Regina Passetti

Este arquivo digital corresponde à versão final da dissertação de mestrado defendida pelo aluno Gabriel Catani sob orientação da Prof.ª Dr.ª Livia Oushiro e coorientação da Dr.ª Renata Regina Passetti.

Campinas
2023

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Estudos da Linguagem
Leandro dos Santos Nascimento - CRB 8/8343

C28n

Catani, Gabriel, 1996-
Newscasting prosody : a forensic sociophonetic approach / Gabriel Catani. – Campinas, SP : [s.n.], 2023.

Orientador: Livia Oushiro.
Coorientador: Renata Regina Passetti.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

1. Sociofonética. 2. Fonética forense. 3. Prosódia (Linguística). 4. Variação linguística. 5. Envelhecimento. I. Oushiro, Livia, 1980-. II. Passetti, Renata Regina, 1981-. III. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. IV. Título.

Informações Complementares

Título em outro idioma: Prosódia telejornalística : uma abordagem sociofonética forense
Palavras-chave em inglês:
Sociophonetics
Forensic phonetics
Prosody (Linguistics)
Language variation
Aging
Área de concentração: Linguística
Titulação: Mestre em Linguística
Banca examinadora:
Livia Oushiro [Orientador]
Plínio Almeida Barbosa
Ronald Beline Mendes
Data de defesa: 12-05-2023
Programa de Pós-Graduação: Linguística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-9010-7113
- Currículo Lattes do autor: http://lattes.cnpq.br/4220291161726827

BANCA EXAMINADORA:

Livia Oushiro

Ronald Beline Mendes

Plínio Almeida Barbosa

IEL/UNICAMP
2023

Ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós Graduação do IEL.

To the memory of the ones
who first explored these paths

ACKNOWLEDGMENTS

I am grateful to my family and especially to Marilia, for her patience and affection.

I am grateful to my advisors Livia and Renata, whose dedication and wisdom made this work possible.

In this sense, I am also grateful to Professor Plinio for his support and care.

I am grateful to Professor Ronald, for the valuable comments and for the time and attention provided.

I am grateful to my colleagues and friends from VARIEM, with whom I've continuously had the opportunity to learn.

From those, I am especially grateful to Gustavo and Julia, for all the help and for the companionship.

I am grateful to the professors of IEL, for their good will, for the teachings and for promoting a healthy learning environment.

I am grateful to CAPES and its contributors, for the support to this research. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

ABSTRACT

This thesis presents an analysis of intraspeaker prosodic variation in newscasting speech, a relevant topic for the development of sociophonetic analyses with potential forensic applications. Through the systematic acoustic analyses of a 19 year speech sample of a middle-aged Brazilian journalist, this research sought to investigate the behaviour of intonational, vocal quality, and intensity parameters in this instance of non-vernacular speech.

Moreover, the research also investigates the correlation of news themes and content valence on the analysed prosodic parameters. Each audio sample was manually classified in regard to its theme and its content valence. All the acoustic parameters were extracted via Praat and statistical analyses were done using the R programming language.

Results showed an overall great variability in most of the prosodic measures between the sampled news editions. There was an observable trend of change in some of fundamental frequency parameters, going towards a lower pitch and a less dynamic speech. In addition, these results highlight age-related changes in vocal behaviour and vocal characteristics before senescence, even in the case of a speaker subjected to vocal training and long-term speech therapy.

The comparison between the themes suggests that some news subjects follow similar acoustic patterns, tending towards a more dynamic vocal behaviour. On the other hand, the greatest differences in vocal behaviour were observed in a specific set of news in which the anchor clearly displayed sadness. The valence of the utterances showed a low correlation with the acoustic parameters.

The findings of this research highlight the importance of social factors in forensic phonetics practices, indicating the relevance of adopting a sociophonetic approach to forensic linguistics. Furthermore, the

parameters reported here may serve as a reference for further forensic analysis, given that there is little data regarding intraspeaker prosodic variation in non-vernacular speech.

RESUMO

Esta dissertação apresenta uma análise da variação prosódica intrafalante na fala telejornalística, tópico relevante para o desenvolvimento de análises sociofonéticas passíveis de aplicação no contexto forense. Através de análises acústicas sistemáticas de uma amostra com 19 anos de fala de um jornalista brasileiro de meia-idade, buscou-se observar o comportamento de parâmetros associados à entoação, qualidade vocal e intensidade nessa modalidade de fala não-vernácula.

Além disso, a pesquisa também investiga a influência dos temas das notícias e da valência do conteúdo dos enunciados sobre os parâmetros prosódicos analisados. Todos os enunciados do *corpus* foram classificados manualmente em relação ao seu tema e à valência do seu conteúdo. Os parâmetros acústicos foram extraídos através do Praat e as análises estatísticas foram realizadas utilizando a linguagem de programação R.

Os resultados mostraram grande variabilidade entre os programas na maior parte dos parâmetros investigados. Foi observada tendência de mudança em alguns parâmetros da frequência fundamental, em direção a frequências mais baixas e a uma fala menos dinâmica. Além disso, os resultados ressaltam alterações no comportamento vocal associadas ao envelhecimento antes da senescência, mesmo no caso de um falante com treinamento vocal e acompanhamento fonoaudiológico de longo prazo.

A comparação entre os temas das notícias sugere que a apresentação de alguns assuntos seguem padrões acústicos semelhantes, tendendo a um comportamento vocal mais dinâmico. Por outro lado, as maiores diferenças na fala do âncora foram observadas em notícias nas quais o falante demonstrava tristeza. A valência dos enunciados mostrou correlação fraca com os parâmetros acústicos analisados.

Os achados deste estudo salientam a importância da consideração de fatores sociais nas práticas da fonética forense, indicando a relevân-

cia da adoção de uma abordagem sociofonética na linguística forense. Ademais, os valores associados aos parâmetros aqui reportados podem servir de referência para futuras análises forenses, dada a escassez de dados relativos à variação prosódica intrafalante em fala não-vernácula.

LIST OF FIGURES

3.1 Means and standard deviations of f_0 derivatives by year . . . 66
3.2 Linear regression plots of intonational parameters 67
3.3 Linear regression of harmonics-to-noise ratio 69
3.4 Linear regression of long term average spectra 70
3.5 Stability in intonational and vocal effort parameters 72
3.6 Stability across voice quality measures 75
3.7 Dunn's multiple comparison of f_0 measures 81
3.8 Dunn's multiple comparison of f_0 1^{st} derivative measures . . 82
3.9 Dunn's multiple comparison of diverse measures 83
3.10 Dunn's comparison of voice quality and intensity measures . 84
3.11 Regression of manually categorised valence by year 86

LIST OF TABLES

2.1 Levels used in the coding of the news 55
2.2 Prosody Descriptor Extractor parameters used and associated information 57
2.3 Prosody Descriptor Extractor hyperparameters used for extraction 58

3.1 Linear models results 65
3.2 Kruskal-Wallis test results (n = 1969, df = 8). 77
3.3 Linear regressions for valence codification 87

CONTENTS

List of Figures

List of Tables

Introduction **15**

1 Background **20**
1.1 Sound and Society 20
1.1.1 Sociolinguistics 21
1.1.2 Phonetics 25
1.1.3 Sociophonetics 26
1.1.4 Prosody 28
1.2 Forensic Phonetics and the emergence of Forensic Sociophonetics 31
1.2.1 Non-contemporary speech samples 34
1.2.2 Intraspeaker variation 38
1.3 Newscasting speech 45

2 Methods **52**
2.1 Corpus 53
2.2 Transcription and coding 53
2.3 Data extraction 55
2.3.1 Extraction from ELAN 55
2.3.2 Extraction from Praat 56
2.4 Analyses 58
2.4.1 Change and continuity through time 58
2.4.2 News themes 59
2.4.3 Valence analyses 60

3 Results **63**
3.1 Change and continuity in the acoustic parameters 63
3.1.1 Change 64
3.1.2 Continuity 71
3.2 News themes 76
3.3 Valence analyses 85

4 Discussion **88**

5 Conclusion **105**

References **111**

INTRODUCTION

A voice can tell us much more than just the content spoken. For instance, our intonation is capable of expressing a myriad of meanings, and characteristics associated with our vocal quality can indicate group membership and alter the way we are understood by others (Bolinger, 1985, 1989; Ladd, 2008; Laver, 1968, 2000; Madureira, 2016). Such linguistic functioning is constituent of language and inextricable from speech. However, most sociophonetic research leave aside the prosodic dimension of speech in favour of the analysis of segmental features (Thomas, 2010).

Intraspeaker variation, also known as stylistic variation, refers to linguistic variation observed within the speech of a single subject. Although observed in the speech of single speakers, this dimension of variation is directly associated to variation between speakers (Bell, 1984, 2001). Thus, it can regarded as one key topic for a better comprehension of language variation and change (Labov, 2001b, p. 85).

The present research aims to explore both of these subjects: prosodic variation and intraspeaker variation. Because studies investigating stylistic prosodic variation are scarce, approaching speech by this angle may grant novel and valuable insights about language functioning as a whole.

Namely, this work investigates prosodic variation in the speech of a single middle-aged Brazilian newscaster throughout 19 years. The corpus consists of 50 audio recordings of a TV news program presented by the journalist in Brazilian Portuguese (BP), ranging from year 2000 to 2019. Acoustic parameters related to intonation, voice quality and intensity were extracted from these recordings and statistically analysed.

In addition to investigating the subject from a longitudinal perspective, this work also analyses the role of the news themes and the positive–negative valence of the spoken content on the speaker's linguistic beha-

viour. Hence, the results of this inquiry are expected to provide an useful description about intraspeaker variation in monitored situations.

The analysed corpus can be taken as an example of very careful and controlled speech uttered by a highly trained speaker. Consequently, changes associated with age observed in this particularly contained environment can serve as reference of the maximum control that a typical speaker can achieve over long-term voice changes. Thus, while analysing sociophonetic aspects of speech production in monitored non-contemporaneous speech samples, this research seeks to contribute to the growing body of forensic linguistics literature.

On the other hand, considering the performative characteristic of the newscasting genre, the analysed data may also be taken as a reference of contemporaneous speech elasticity. Thus, the findings of this research can potentially act as beacon for forensic comparison involving the analysed acoustic parameters. Moreover, they may also offer a good reference point for high-profile suspects, such as politicians and other public figures that may be subject to media training and speech therapy accompaniment.

This research is thus set to explore the following specific questions:

- How variable is the prosody of a newscasting anchor while presenting the news?
- Did the newscaster prosody change throughout the course of 19 years? If so, which acoustic parameters changed? How did they change? Which parameters remained stable?
- Do the theme of the news and the positive–negative valence of the utterances affect the speech prosody of the newscaster? If so, which parameters are affected? How are they affected?

The application of linguistics in the forensic context is at least as old as the use of vocal recognition being as forensic evidence. Deffenbacher et al. (1989) state that the oldest record known dates back to the 17th century. Indirectly, it may be considered as old as criminal procedure

itself, since much of the judicial functioning relies on vocal performance, rhetoric and intonation as resources to elucidate the situations and to persuade the judges into giving a favourable ruling.[1]

Given the transiency of speech and the crucial need of material evidence in forensic context, the usage of voice recordings has become increasingly popular in forensic cases. After all, it would be arguably much simpler to present a misleading statement regarding authorship of a questioned voice than to fake a recording of the voice in question.

An important source of voice recordings used in forensics is intercepted speech, which is defined as the recording or apprehension of speech without the knowledge of the speakers involved in the interaction (Orletti & Mariottini, 2017). In older methods of vocal communication interception, such as mechanical telephone tapping, the interceptor would have to deal with complicated logistics of installing the necessary hardware before the speech event. Nowadays, voice data is frequently self-recorded by the users of many widespread communication services in which the users can send and receive recorded audio messages to each other via computer and smartphones.

In this messaging applications, audio data is commonly stored indefinitely in both the sender's and the receiver's devices. The transmitted audio data also often remain archived in one or more servers of the service providers. Due to these characteristics, access to voice recordings with fair audio quality has become increasingly easier. Consequently, there has been an increasing usage of these recordings as audio evidence in legal cases. Therefore, the development of theories and methods associated with forensic phonetic analysis is essential for the provisioning of adequate analysis and treatment of this type of data.

One of the main concerns of forensic linguistics is how to deal with variation in speech (Foulkes, Scobbie & Watt, 2010). Differences between what are conventionally called languages or dialects are mostly

[1] On this account, McMenamin (2002) enumerates a series of early works about language and law.

recognised by all. Nevertheless, the simplest observation of language in use shows us that there are many different ways of saying basically the same thing. More precisely, these differences can be observed at all linguistic levels, ranging from phonetic minutiae to the use of very dissimilar syntactic structures.

In this sense, different meanings can be associated with the usage of different linguistic forms. These meanings do not necessarily coincide among different speakers of the same community. What sounds relaxed and friendly to one person, may sound disrespectful and impolite to another, while it may simply not sound marked at all to a third person. Thus, the use of different linguistic forms constitutes a complex sociolinguistic structure. Not only is there variation in the forms used, but also in the meanings associated with the forms.

Linguistic variation is not limited to heterogeneity of forms and meanings within a dialect. The usage of different linguistic variants can also be observed in the speech of the same speaker. This can be more or less conscious for the speaker, and can be associated with different factors such as the speech situation (Labov, 1972), the interlocutor (Bell, 1984) and the communication of a certain stance (Eckert, 2003).

This level of variation is not independent of the social context in which it occurs. The value of their usages comes from the association of the relationship between form and meaning assumed by the speaker with the relationships between form and meaning at play in the interaction. Access to potentially more accurate interpretation of these relationships depends on the subject's contact with specific social loci or communities (Eckert, 2008; Labov, 1972). Accordingly, intraspeaker variation is not unconditioned, as it is interrelated to the variation and form–meaning associations in society (Bell, 1984).

In Chapter 1 I provide some background to these topics based on the existing literature. I start by discussing the relationship between Sociolinguistics and Phonetics and the influence of these fields into the development of Sociophonetics. Subsequently, I present a brief introduction on prosody and its acoustic investigation. I then proceed to

review some studies about newscasting speech. And in the last section I deal with questions related to forensic sociophonetics: the analysis of non-contemporaneous speech recordings and the theory of intraspeaker variation.

In Chapter 2, I present some information about the newscaster and the corpus, coupled with the steps taken for data extraction. I further explain the three axes of analysis employed in the data: longitudinal analysis, news themes analysis and valence analysis.

Next, in Chapter 3, I proceed to display and describe the results of the quantitative analyses of the data. Firstly, I present results concerning the longitudinal analysis, dealing with change and continuity in the anchor's speech. I then present the results of the analysis of the news themes. Lastly, I present the results about the valence analysis, regarding both the manual classification as well as the ones done via Sentiment Analysis lexicons.

These results are discussed in Chapter 4. In this chapter, I aim to interpret the quantitative results based on the previous sociolinguistic and phonetic literature. Lastly, in Chapter 5, I draw conclusive remarks associated with the development of forensic sociophonetics based on the research findings. I also suggest some directions for further research.

CHAPTER 1

BACKGROUND

This chapter presents a review of different topics pertinent to the development of this work. It starts by covering theoretical and practical contributions from Sociolinguistics (Section 1.1.1) and Phonetics (Section 1.1.2) to the study of variation of speech sounds. Afterwards, it discusses the conjunction of the interests of both areas, shaping Sociophonetics (Section 1.1.3). It subsequently introduces the study of speech prosody and some of its acoustic characteristics (Section 1.1.4). Thereafter, it presents Forensic Sociophonetics (Section 1.2) as an application of Sociophonetics in the forensic field and discusses associated concepts, namely non-contemporary speech samples (Section 1.2.1) and intraspeaker variation (Section 1.2.2). Lastly, it discusses newscasting speech (Section 1.3), given that this topic comprehends the corpus of the research.

1.1 Sound and Society

As modern Linguistics developed, many fields sought to explore speech in its multiple dimensions and with various interests. In present time, the conjunction of mutual interests in empirical speech analysis crystallised a dynamic interface between two discretely consolidated disciplines: Phonetics and Sociolinguistics. This interface has been referred to as "Sociophonetics" (Foulkes, Scobbie & Watt, 2010). In order to comprehend this relatively new field, it is relevant to understand both of its structuring disciplines.

1.1.1 Sociolinguistics

Sociolinguistics is generally accepted as the study of the relationship between language and society. There are, however, many specific interpretations of what this term represents, as many kinds of study may be referred to through this usage (Labov, 2006).

For instance, talking about the social role of language, Wolfram (n.d.) characterises various types of sociolinguistic investigations, such as the study of language attitudes on a macro level, the study of language contact, the studies of patterns in conversations and interaction, and so on. Furthermore, the use of this term alone can be somewhat broad, notwithstanding the fact that research practices within each of these types of investigation associated with language and society are mostly stable and specific.

Labov's famous opening words in *Sociolinguistic Patterns* states:

> I have resisted the term *sociolinguistics* for many years, since it implies that there can be a successful linguistic theory or practice which is not social. (Labov, 1972, p. xiii)

Regardless of a theory being able to treat language without much consideration to what is traditionally viewed as "social" (which might be seen as communal; supraindividual; external; and/or performatic), in the bigger picture, the linguistics object is still attached to a social dimension, since language is intertwined to the only circumstance in which we are known to exist, in society – or more simply, circumscribed by those we call "others". Therefore, even if some linguists feel compelled to argue that *there are* "successful theories" in which social factors are (apparently) not present, recognising the fact that language is an immanently social phenomenon, as all language is situated in society, makes this assertion inaccurate.

A judgement about the grammaticality of a given sentence, for instance, could superficially appear to be disconnected from social factors, as it could be implied that the judgements are solely dependent on the

single judge. But this subject's own language, and therefore the associated judgements on this matter, are resultant from a language acquisition process that is determined by social interactions and shaped by socio-historic processes. Consequently, the symbolic associations in the judge's language are inevitably socially mediated, since the existence of this speaker is socially constrained.

As researchers, it is not necessary to study what is superficially regarded as "social" or to study language in its interactive dimension. Nevertheless, it is certainly important to recognise the role of this dimension on the linguistic functionings that could appear socially unchained, such as subjective judgements and behaviour or internal language structure.

While Labov (1972, p. xiii) argues that linguists "ha[d] not yet emerged from the shadows of [their] intuition", perhaps one of the main problems of working with these intuitions is not recognising them as socially dependent. This realisation, in turn, leads to better practices that may solve the problems that Labov (1972) points out: the lack of empirical rigour in linguistics and the disregard of the importance of "real world" language data.

In what concerns the naming of the field, if it wasn't necessary to state the importance of the social aspects of linguistics, we certainly would not have the several aforementioned fields of linguistic studies united by the "Sociolinguistics" hyperonym. Given the variety of these studies, it is of interest to define more specifically which of these sociolinguistic traditions better describe this work, in order to avoid overgeneralisation: this research take on sociolinguistics is closely related with what became known as Variationist Sociolinguistics, a field traditionally characterised by the usage of quantitative research methods in order to analyse linguistic variation in the linguistic system.

The notion that variation is part of linguistic system is one of the corollaries of variationist studies. One of the structuring literature and one of the founding works of the area is Weinreich, Labov and Herzog (1968), presented at the symposium *Directions for Historical Linguistics*. The authors review the notion of sound change developed by the Neogrammari-

ans, specially debating and building upon Paul's (1891 [1889]) *Principles of the History of Language*. They progressively visit subsequent research that inherited its views on language change from these scholars, commenting on Saussure (1916), Bloomfield (1927) and Chomsky (1965).

Mainly, Weinreich, Labov and Herzog (1968) argue that the antinomies between structure and history, stemming from Paul's (1891 [1889]) theory and reappearing in further works, lead to paradoxes that precludes a cohesive treatment of language change. Those problems first arise from the idea that the "individual languages" (concept akin to "idiolects") should be the central object of linguistics studies, while *Sprachusus* (or dialects) are considered analytic abstractions based on the idiolects. This notion establishes an "irreconcilable opposition between the individual and society" (Weinreich, Labov and Herzog, 1968, p. 106).

The authors further propose a new perspective of variation, asserting that the equation of structuredness to homogeneity, consequent of the centrality of the idiolect, was detrimental, establishing the idea of orderly heterogeneity. In the process, Weinreich, Labov and Herzog (1968) also present what would become central questions or problems for the variationists, regarding language change. In general terms, this agenda consisted in exploring what constrains language change (the constraints problem); how language transitions from one stage to another (the transition problem); how these changes are embedded in the social and linguistic systems (the embedding problem); how these changes are subjectively evaluated by the speakers (the evaluation problem); and what makes a given change occur in a given time (the actuation problem).

Nevertheless, sound was one of the main concerns of variationist theorisation even before the proposal of these problems. As the notion that everyday speech was the most productive place to study language developed, "sound change" was a frequent term in the associated literature, such as in Labov (1963).

This interest in vernacular speech derived from the assumption that this was the most systematic linguistic source, as it was less subject to

conscious modifications resulting from societal pressures (Labov, 1972). The objective of systematically obtaining such unmonitored and undisturbed speech, however, implied an investigative paradox. How could an outside observer witness speech typical of moments when the speakers are not being observed?

Labov's (1972) strategy to overcome "The Observer's Paradox" was the elaboration of a method that sought to account for multiple formality levels within the situation of interview recording. This method, known as *sociolinguistic interview*, became the main data source for sociolinguistic studies.

Labov (2006) notes that the beginning of the usage of acoustic analysis in sociolinguistic speech recordings was influenced by discussions about the usage of formants for vowel discrimination in English made by Cooper et al. (1952), Cooper, Liberman and Borst (1951) and Delattre et al. (1952). Labov, Yaeger and Steiner (1972) also acknowledge the studies by Peterson and Barney (1952), and Fant (1958). In their work, Labov, Yaeger and Steiner (1972) analysed English vowels from interviews recorded in multiple locations, mapping their first and second formants using spectrograms.

Most sociolinguistic work, nevertheless, adopted a different strategy regarding sound analysis. Aiming to investigate large samples of vernacular speech, the classification of phonetic variants has often been implemented via auditory methods, resulting in less detailed, categorical data (Silveira & Oushiro, 2022; Zimman, 2020). One example of this type of approach is the classical department store survey by Labov (2006 [1966]) where he classified, by ear, hundreds of tokens of post-vocalic /r/ sampled in three stores with different target public.

Subsequent research in the field gradually incorporated contemporary techniques to its framework, following the wide dissemination of computational resources. Yet, some of these traditional practices, such as the usage of categorical variables, still remain in the field.

1.1.2 Phonetics

The main enterprise of Phonetics is the description and analysis of speech. Given the complexity of its main research object, phonetic studies can make use of a variety of approaches.

As the mechanisms of sound generation and the characteristics of speech sounds can be thoroughly analysed mobilising a series of techniques and resources, phonetic research can be associated with many objectives and serve many purposes, ranging from bio-physiological concerns to speech production, speech-related pathologies, vocal synthesis, computer recognition, and so on (Ladefoged & Johnson, 2014).

One of the main branches of phonetics deals with the acoustic analysis of speech sounds. Broadly speaking, acoustics consists of the study of sounds and its related physical phenomena. Subsequently, Acoustic Phonetics' interests revolve around the sounds of speech.

Regarding data collection, differently from Sociolinguistics, the most traditional method of data gathering for phonetic research is laboratory recordings. This type of recording is typically set in controlled environments crafted with the goal of providing high-quality audio registers favouring subsequent analyses.

Although in both fields the speech data are usually elicited by the researcher, in sociolinguistic interviews the target linguistic phenomena are generally only loosely evoked by the researcher, who incites speech via semi-open questions stimulating a conversational behaviour on the speaker. This is done mostly as an attempt to keep interviewee's attention away from the studied phenomenon, prompting for a less self-conscious speech. Although this scenario is changing, Phonetics' recordings are generally not as constrained by this concern with obtaining vernacular speech.

In this sense, experimental phonetic research usually shows greater interest in maintaining parity of the procedures and the circumstances of data collection. Researchers strive for greater experimental control by reducing parallel sources of variation and trying to isolate the target phenomenon.

Regardless of this interest, absolute parity of conditions is unfeasible. The control of the researcher over the circumstances of the experiments can only go so far. And although this limitation is usually at least tacitly acknowledged, ideally being explicitly informed, social variables that may correlate with observed linguistic behaviour may be not recognised in the experiments and, thus, may not be taken into account in the analyses.

1.1.3 Sociophonetics

As a diverse set of works are produced under the term "Sociophonetics", the nature of the interface between Phonetics and Sociolinguistics can entail diverging interpretations, since it could be seen as a subfield of either area, or as a new field of research altogether. Labov (2006) considers the emergence of Sociophonetics as a separate discipline as a potential distraction from what he regards as the most important fact: the approximation of the areas of research. He argues that although the methods and analyses under this name exceeds those of contemporary Sociolinguistics, those "studies are not disjoined from the broader field of sociolinguistics" (p. 501). Foulkes (2015) considers that although the integration of practices from these two parent fields promotes gains for both, "trully sociophonetic research offers more than the sum of its parts" (p. 1). Given that the methods as well as the objectives of sociophonetics are not particularly constrained by its predecessors, and for the sake of convenience, I henceforth treat sociophonetics as a specific field.

As do the results of the approximation of these two different starting points, the definition of the field also varies. In this direction, Foulkes, Scobbie and Watt (2010) suggest that the variability of the usage of "sociophonetics" is associated with the particular research concerns of its adopters. The authors situate the field by naming some of its characteristics: "the integration of the principles, techniques, and theoretical frameworks of phonetics with those of sociolinguistics" (Foulkes, Scob-

bie & Watt, 2010, p. 703). Nevertheless, they further highlight the permeable borders of the area due to its growth, involving theory and practices associated with psycholinguistics, language acquisition, computational linguistics and so on. This visible expansion of the field can perhaps support the classification of sociophonetics as an independent field, as sociophonetics seems to grow through the interfaces with fields in which it was not theoretically contained.

Many authors seem to agree that a great influx of works associated with sociophonetics started around 1990 (Foulkes, Scobbie & Watt, 2010; Silveira & Oushiro, 2022; Thomas, 2019; Zimman, 2020, inter alia). In fact, the development of more easily accessible computational resources for acoustic analysis appears to be one crucial factor for the increasing adherence to the field (Silveira & Oushiro, 2022).

Regardless of the increase in popularity, most sociophonetics research seems to be associated with data and methods most traditionally used in the preceding areas. Hay and Drager (2007), for instance, draw attention to the importance of employing qualitative approaches such as ethnographic method in conjunction with the phonetic techniques. Foulkes, Scobbie and Watt (2010), by their turn, attest that not many longitudinal studies have been carried out with adult speech due to logistical complications and to a lingering assumption of stability of linguistic usage during adulthood. This may be partially attributed to the wide adoption of what is known as apparent time hypothesis, in which this presumed stability of adult speakers' speech is used to assume changes in progress based on the disparities between different aged speakers (Hay & Drager, 2007).

While the speech analysis through modern phonetic techniques seems to be the tonic, there is a large influence from sociolinguistics data collection strategies, such as the recording of conversational interviews. In an endeavour to obtain data with a higher ecological validity, that is, data with higher proximity of what would more commonly happen in real life, researchers have been steering away from the more controlled yet more artificial techniques of data collection, such as

laboratory data (Silveira & Oushiro, 2022; Zimman, 2020). Nevertheless, accompanying sociolinguistics historical development, there is an apparent bias towards the studies of vernacular speech, in detriment of the equally systematic non-vernacular speech. In a similar fashion, there is a scarcity of studies dealing with intraspeaker variation vis-à-vis those comparing characteristics of multiple speakers.

Although the interest prosodic features has been increasing, it is still a minor subject in sociophonetic studies, that are traditionally more focused on vowel production and perception (Silveira & Oushiro, 2022; Thomas, 2019). Research on other prosodic domains, such as voice quality, in the field is even rarer (Thomas, 2019).

In this sense, while it is important to highlight the theoretical potential and the rich contributions of sociophonetics to the field of Linguistics, approaching speech in a comprehensive manner is paramount for the development of the area. To be able to address language variation from its finest acoustic details, articulating these analyses with their social context is sociophonetics greatest strength.

1.1.4 Prosody

Prosody, be it intonational, rhythmic or associated with intensity and phonatory quality, is an inextricable part of speech. It often serves a purpose of modalising and modifying the contents of an utterance. Even outside the constraints of the sound-based communication, prosodic aspects are made present in communication taking different forms. Through this dimension, we are able to compose and try to express the meaning of sentences in such a way that the same written statements, that is, apparently similar lexical forms and morphosyntactic structures, can be realised and interpreted in completely different ways. One of its main functions relates to the emotional, affective and attitudinal dimension of communication and expression (Madureira, 2016).

Considering this, we may say that regardless of its name, from Greek “προσῳδία”, with “ᾠδή” meaning “chant” (Bailly, 1935), “song”, “ode”

(Liddell & Scott, 1940), prosody is not only an element of speech sounds but of language as a whole. Such is the necessity of this expression in our communication that as early as computers began serving as an interactive medium, we started filling the gap of prosodic information with graphical representations of facial expressions to convey emotional information with the so called "emoticons". Those further evolved into a myriad of multimedia expressive resources, in many cases seemingly trying to increase the similarity between the representation and the represented, adding specific glyphs, colors, movements, sounds and even the possibility of customisation of the facial traits so they could match those of the interactant.

Bolinger (1985, p. 195) argues that signaling emotion is the primitive of intonation. Regarding this prosodic dimension, the author mentions the idea from Tronick, Als and Brazelton (1980), that "opposite emotions are expressed by opposite behaviors" (Bolinger, 1985, p. 194), in this case, a higher pitch would be correlated to more active and potent emotions in opposition to lower pitch. His biological interpretation of the workings of intonation and voice quality, understood as a specialisation of the ability of reading "symptoms" (Bolinger, 1985, pp. 194–195), may be considered very accurate, specially by those that notice their own reliance on such information as a strategy to guide their actions.

The author illustrates this symptomatic signaling with an analogy with pressure gauges from boilers. Instead of relying on the observation of the bulge in the seams of the boiler, we would access the state of the boiler through a specialised mechanism.[1] Thinking of language, the main point is to be able to synchronically access another persons's emotion or stance in an interaction allowing us to adjust our behaviour without the need, for instance, of the verbal expression of discordance and a subsequent argument, while bearing non-confrontational inten-

[1]For those not well acquainted with boilers, thinking of a tire pressure gauge can yield a similar effect. While both cases might not be equally dangerous, it's surely easier to access the information about the pressure with the specialised resource – maybe with the exception of extreme and undesirable cases of a flat tire or a blown-up boiler.

tions. One important aspect of this characterisation is that these expressive mechanisms are parts of the system. In Bolinger's (1985, p. 195) words: "symptomatic communication of this sort is PRESENTATIVE and not REPRESENTATIVE" (emphasis in the original).

The relationship between specific acoustic characteristics to the perception of a linguistic sound is not exclusive to vowels or phonetic segments. As part of speech, prosodic information is also acoustically structured. Furthermore, a set of specific physico-acoustic information about speech sounds are closely correlated with our corresponding perception of prosody. These can be classified as prosodic parameters, as they are able to describe prosodic information from speech (Barbosa, 2019).

Based on a conventional division of prosody, we can distinguish four of these relationships (Barbosa, 2019; Hirst & di Cristo, 1998). The fundamental frequency (f_0), measured in Hertz or semitones, is an acoustic parameter associated to the frequency of vibration of our vocal folds. Perceptually, it correlates with pitch, the sensation of a high or low (deep) sound. Duration is related to the time between two specific points, delimiting an analytic units as a syllable, usually measured in milliseconds (ms). It is associated with the perception of a sound being long or short and with speech being perceived as fast or slow. Intensity is associated with sound pressure and thus with sound strength. It is usually measured in decibel (dB). Subjectively, it is associated with loudness or volume, that is, how loud or quiet a given sound is. Voice quality is associated with spectral characteristics stemming from speech apparatus settings and vocal folds behaviour. It can correlate perceptually with timbre and may confer the sensation of creakiness, harshness and breathiness, among others.

It is important to notice that the relationship between these acoustic parameters and their perceptual correlates are not linear. Moreover these perceptual correlates may also be dependent on factors other than their main acoustic correlate alone (Barbosa, 2019).

This research analyses parameters concerning all of the above dimensions, except for duration. Fundamental frequency (f_0) is analysed

via summary statistics (such maximum, minimum and median values, standard-deviation and interquartile semi-amplitude); f_0 peak parameters, that are associated with peak values of the fundamental frequency in a given time interval; and first derivative measures, associated with the rate of f_0 change. Intensity is measured via its coefficient of variation and spectral emphasis. Voice quality is analysed via measures of fundamental frequency perturbation (jitter, shimmer); the harmonics-to-noise ratio; soft phonation index, which is associated with vocal folds abduction; and long term average spectrum (LTAS) slopes, associated with perceived vocal effort. These measures will be presented with their analysis on Chapter 3 and Chapter 4.

The main takeaway from these remarks is that there are many possible ways of analysing prosody involving acoustic measures that bear close association with our abstract perception of prosodic information in speech. As these prosodic parameters can be systematically extracted and analysed, they constitute a prominent resource of phonetic data. Exploring this dimension of phonetic variation can be valuable both from a theoretical standpoint and for practical applications of linguistics such as its usage in the forensic context.

1.2 Forensic Phonetics and the emergence of Forensic Sociophonetics

Forensic Sciences is the generic name of a field that comprises the application of scientific knowledge in the legal context. Subsequently, Forensic Linguistics can be generally seen as the application or usage of Linguistics' knowledge and practices to the forensic context, that is, to the legal and/or criminal setting, generally providing support in trials and judicial processes (McMenamin, 2002).

As Linguistics itself encompasses a wide array of subjects, interests and proficiencies, we can separate the application of Forensic Linguistics in relation to Linguistic disciplines. Perhaps the most widespread

application of Linguistics in the forensic context comes in the shape on Forensic Phonetics.

Forensic phonetics deals with analyses of speech sounds within the forensic environment. Given its practical nature, one straightforward way of providing a glimpse of this branch of phonetics is by describing some of the tasks it comprises.

One of the most prominent tasks of the field is Speaker Comparison (Foulkes & French, 2012). This task commonly consists of opposing intercepted or questioned audio samples from a person of interest against samples provided by one or more suspects ("reference samples"), seeking to arrive at a probabilistic result regarding the authorship of the questioned sample, that is, if the audios were produced by the same speaker or by different ones. Another task is Speaker Profiling, in which suspects are potentially outlined based on the characteristics of the investigated voice, which might be associated with a given regional dialect or to the speech of specific groups (Rose, 2002).

In most cases, two relevant concepts for the tasks performed in the area are *similarity* and *typicality* (Jessen, 2008). *Similarity* corresponds to the differences or similarities between two different voices, given their characteristics. *Typicality*, on its turn, deals with how typical a voice or a specific characteristic in the population of speakers is. The conjunction of both these factors helps the analyst to rate the strength of a given evidence.

As a hypothetical example, we may think about two scenarios: in the first, we are comparing two recordings in which the voices have a median f_0 of about 120 Hz, a fairly common value for males. In the second, the two voices have respectively a median f_0 of 80 and 85, a less typical value for the male population. Although in the second scenario the two voices are a little less similar, the fact that their median f_0 are in a far less typical range than the ones in the first scenario can constitute a stronger evidence of the two voices being spoken by the same speaker in the second scenario than in the first scenario, given that many different speakers have a median f_0 around 120 Hz but only a few have

a median f_0 below 90 Hz. The example above makes it clear that an evaluation of typicality calls for reference data about the population of interest in order to evaluate the cases in real scenarios. It is only possible to know about the distribution of median f_0 for a given population through phonetic research.

Regarding methods, forensic phoneticians make use of several techniques of data analysis. Approaches range from auditory analysis to automatic data extraction and speaker recognition. In this sense, the usage of acoustic phonetics methods plays an important role in the field. Given the serious implications that the results of such analyses may hold, striving for objective and well-founded practices is paramount. A conjunction of this approach with the perceptive-auditory methods is often used, granting a more thorough analysis. Moreover, it is increasingly common that the case analysis proceeds to be validated by another expert before it is finished (Gold & French, 2019).

One of the factors that greatly determines the affinity between forensic phonetics and sociolinguistics is data collection. Both areas greatly make use of interviewing types of elicitation, generally with a similar methodological goal of obtaining a more informal type of spoken data. While the underlying motivations can slightly differ, as sociolinguistics usually aims to avoid the interference of supravernacular influence in the data, with the underlying assumption that vernacular speech is the most systematic, both share the preference to work with more casual types of data.[2] Another prolific similarity relates to the study of properties of the speech of a single subject, which is further discussed in Section 1.2.2.

While the term Forensic Phonetics is somewhat well established and so is becoming the usage of Sociophonetics, given the fruitful contributions that sociophonetics can provide to forensic sciences, it can be useful to use a specific term for referencing the advances in the integra-

[2]Of course this is not set in stone, since Sociolinguistics' interest in non-vernacular speech has increased and the goal of forensic analysis can, in some cases, benefit from the elicitation of more formal speech, e.g.: in a case where the questioned voice is situated in a formal setting.

tion of Sociophonetics' theories and methods into the Forensic practices. Given its specific application and methodological concerns, it seems appropriate to delimit a section of this field as Forensic Sociophonetics. In the following subsections I elaborate on questions concerning the employment of Linguistics into legal matters.

1.2.1 Non-contemporary speech samples

Non-contemporary speech samples refer to speech samples collected in different periods of time, that is, speech samples that are not contemporary in relation to each other. This type of samples bear special relevance in the forensic context, as comparison of voices of interest is usually carried out by using samples that were produced at different times in which the evidence/questioned samples come before the reference samples (Hollien & Schwartz, 2001). The interest regarding the contemporaneity of the samples, thus, is to be able to account for potential longitudinal changes that may occur during this gap between the recordings (Künzel, 2007).

It is relevant to state, from the start, that physiological changes are not disconnected from social changes, as social behaviour is able to mould our bodies in such a way that behavioural factors may slowly modify what could be expected simply from a biological standpoint of ageing. Speaking of more fine grained influences of our social behaviour, common recent practices, like holding and typing on a mobile device can contribute to physical conditions such as the "smartphone pinkie", a deformation of the fifth digit of the hand, due to its usage as a support for holding mobile devices (Özdil, Çatıker & Bulucu Büyüksoy, 2022). We also have culturally dependent ways of behaving and carrying our bodies, such as how to sit or how to stand – which could, themselves, compete against our tendency to adapt to our surroundings, as in the mundane case of sitting on a badly designed chair. Even though most of these age-related changes are interconnected and interdependent, it seems reasonable, for our purposes here, to distinguish between two

large categories of changes, generally speaking: physiological changes and social changes.

Some of these physiological changes, for instance, appear to be more inevitable. Some appear to be related to our ageing body itself and not directly from any specific behaviour – except, perhaps, breathing.

Ageing can be understood as the process of becoming older, specially for humans. In many cultures people keep track of their age, the time passed since they were born. Thus, the relevance of age is directly related to the importance of the passage of time.

In Forensic Sociophonetics, the age factor is of great relevance. Given the importance of time in speech, it's also important to consider potential vocal changes related to age, envisioning the comparison of data collected within a long time gap. Not only is it important to obtain a comparison threshold for different speakers of different ages, but also to encompass different conditions, whether these conditions are related to the medium, such as mobile phone communication (Passetti & Barbosa, 2015; Passetti, Madureira & Barbosa, 2022), or to specific groups of speakers.

Despite different theories about why and how we get old, the fact is that we do. And as we do, a series of alterations in our bodies take place. Since voice production is dependent on other, more general, bodily functions such as musculoskeletal, respiratory and cardiovascular systems, it's fair to infer that changes in these systems would result in vocal change (Ramig & Ringel, 1983). Some examples of changes that can directly affect the way we speak and sound are alterations in our respiratory system, atrophy of vocal folds, calcification of larynx, *inter alia* (Leeuw & Mahieu, 2004; Sataloff, Kost & Linville, 2017). Among perceptual-acoustic parameters that deteriorate with older age, there is an overall increase in instability measures of f_0 related do voice quality (such as jitter and shimmer), decrease of overall sound pressure and increase in breathiness, which may be measured via harmonics-to-noise ratio and Soft Phonation Index (Leeuw & Mahieu, 2004).

As we get older, we are also subjected to different social expectations.

In a generalised take on Western culture, a child is expected to go to school and an adult is expected to work. Adults are, most of the time, also expected to become more responsible than kids. With the approach of senescence, there also come associated ideals of maturity related to ideas of experience and accomplishment (Mirowsky & Ross, 2010). This conception can lead to the overall evaluation of older people as more prestigious or respectable than the young. As such, these types of social values can spread over to behaviours associated with the older, including linguistic behaviour.

While age based on our date of birth may look as an objective way of assessing the impact of time on human bodies, changes and deterioration of the body can occur in a different pace for different subjects due to genetic and social factors, such as heart disease tendency or smoking habit. Ramig and Ringel (1983) note that there is a remarkable amount of variability across traditional age categorised data, that is, a great amount of variation within the same age or age group. Since older persons have been alive for a longer time, their speech is more likely to be affected by their different lifestyles, leaving more room for the emergence of latent physiological conditions. Consequently, differences within an age group tend to increase over time, leading to smaller intragroup cohesion and, hence, to a research challenge.

Addressing this complexity, Ramig and Ringel (1983) analysed the speech of 48 male subjects within three age groups (25–35; 45–55; 65–75 years old) divided by their overall health condition. The participants' condition was classified into good or poor based on measures such as heart rate, blood pressure and forced vital capacity. The results showed a high influence of their health conditions, as the healthier produced voice with less jitter and shimmer, that is, perceptually less hoarse and whispery, and within a higher frequency range than their less healthy counterparts of similar age. Chronological age alone showed difference only for shimmer between the elders and the others, and mean f_0 was not different between the groups. As the authors suggest, this shows not only the impact of age in voice production, but the importance of

better methods for estimating physiological age of the speakers.

On the other hand, healthy elders were affected in more demanding but also part of day-to-day activities. On a longitudinal study, Leeuw and Mahieu (2004) saw that nonpathological vocal changes due to ageing in a time span of 5 years can be observed in older males between 50 and 81 years old. As time passed, the voices of the subjects became increasingly perceived as rough by expert judges and less stable regarding frequency and breathiness. They also questioned the subjects about their judgement of their own voices and its capacities, such as maintaining a normal conversation and making telephone calls. The self-evaluation by the 11 subjects that participated in both occasions showed an increase in the reports of avoidance of large groups of people due to voice concerns and increased perception of day-to-day changes in their voices.

Considering the above discussion about the interconnectedness of social and physiological changes, this is an interesting result, as it shows another type of relationship between physiological changes and social behaviour. Here, the social practice of avoidance of large groups appears to come as a consequence of the physiological condition of self-perceived vocal degradation.

Also in relation to ageing and health, Orlikoff (1990) studied acoustic parameters on male voices by comparing old healthy (68–80 years old), old cardiopath subjects (60–79 years old), and healthy younger adults (26–33 years old). Looking into fundamental frequency and amplitude measures, he analysed mean absolute jitter and shimmer, mean relative jitter and shimmer, sound pressure level and its standard deviation, peak to peak amplitude, and f_0 mean and standard deviation. The main differences were between the old and the young, even though atherosclerosis had a significant impact on vocal parameters.

While he saw no significant difference for mean f_0, both older groups differed from the young in the standard deviation of f_0. The overall variability concerning all observed parameters between the 4 members of each group was also compared, and the old healthy speakers vocal parameters were almost twice more variable than the young. The unhealthy

older speakers were even more variable than their healthy counterparts. Variability of jitter measures was higher for the healthy old men compared to the young and, again, even higher on the unhealthy older speakers. Both jitter measures were significantly different between the cardiopaths and the young and both shimmer measures were different between the young and the two older groups. When grouping the older groups, the only parameter that showed no significant difference was the mean f_0.

One interesting aspect of this health conditioned differences in the vocal ageing process is the impact of vocal training, as it is common practice in voice professionals. This is somehow assumed by Ramig and Ringel (1983), as they specifically note the fact that people with professional voice training were excluded from their sample.

1.2.2 Intraspeaker variation

Another relevant topic for forensic linguistics concerns the variability of phenomena among the utterances of a single speaker. This matter is widely discussed in the Sociolinguistics literature, which can provide interesting contributions to forensic analysis. In Sociolinguistics, it is often considered that there are three main models that treat intraspeaker variation, also called stylistic variation (Hernández-Campoy & Cutillas-Espinosa, 2012; Schilling, 2013; Schilling-Estes, 2002). These models are known as Attention Paid to Speech (Labov, 1972); Audience Design (Bell, 1984, 2001); and Speaker Design (Coupland, 2007; Eckert, 2003).

As Coupland (2007) notes, the terms *style* and *design* have a degree of proximity in regard to planning and purposefully creating and acting. Irvine (2001) introduces a similar perspective, aiming to reconcile the everyday usage of the term with its more technical usage inside Socioliguistics arguing in favour of style as distinctiveness based on the actions of individuals. This agency-based definitions, nevertheless, directly involve epistemological assumptions associated to the nature of human behaviour. Since this greater philosophical discussion falls bey-

ond the scope of this work, my usage of the *style* here does not bear this specific connotation, serving only as a synonym of intraspeaker/within-speaker variation in a more traditional sense.

The first prominent theorisation in the field came from Labov (1972) in which the comprehension of the variation in a single speaker linguistic usage was associated to the overall level of attention paid to their speech. Within this framework, situations that represented a higher level of formality, due to social expectations, would be likely to be matched with a higher level of attention paid to linguistic usage. This more cautious way of speaking would tend to be more affected by normative ideals and by the speakers' conception of the societal prestige forms, somewhat extraneous to their own way of speaking. With the heightening of attention paid to speech thus would come an increase in misuse of some of these constructions, sometimes overgeneralising specific rules in an attempt to adapt to the seemingly correct way of speaking.

Although fitting intraspeaker variation in this single formality axis may sound simplistic, the point of this theorisation was not to elaborate on all the mechanisms that lead to stylistic variation, but to find a way to operationalise this important dimension while dealing with language variation and change in a systematic manner. This way, researchers would be able to account for this dimension in their collected data based on the evaluation of the context in which the data were produced, instead of assuming that variation in this dimension was simply a matter of chance. More specifically, this conceptualisation was designed and has been used mainly in the context of traditional sociolinguistic interviews, where speakers are recorded throughout different linguistic tasks with different levels of control, such as answering to a more or less structured set of open questions about their lives, telling an emotional story about themselves, reading different types of texts, reading word lists and lists of minimal pairs.[3] Attributing different formality status

[3]Words that have only one different sound between each other, such as *pie* and *tie* (Ladefoged & Johnson, 2014)

for each section of the interview indicates to the researcher what forms are part of the speakers' vernacular language – their everyday speech – which is assumed to have less attention paid to and, thus, assumed to be more systematic due to less interference from other complex and potentially harder to account social factors.

Although the approach of formality and attention paid to speech conditioning variation is intuitive and has led to relevant considerations about linguistic behaviour, it is admittedly a hypothesis regarding a efficient way to systematise intraspeaker variation – and not a final theory about the nature of this phenomenon itself. Labov's (1972) approach was later considered to be simplistic due to its single dimensionality, but we can say that this was, in a way, an intentional feature of his conceptualisation. Regarding attention as an element of a methodological paradox[4], he states:

> There are a great many styles and stylistic dimensions that can be isolated by an analyst. But we find that *styles can be ranged along a single dimension, measured by the amount of attention paid to speech.* The most important way in which this attention is exerted is in audio-monitoring one's own speech, though other forms of monitoring also take place. This axiom (really an hypothesis) receives strong support from the fact that speakers show the same level for many important linguistic variables in casual speech, when they are least involved, and excited speech, when they are deeply involved emotionally. The common factor for both styles is that the minimum attention is available for monitoring one's own speech.
>
> (Labov, 1972, p. 208, emphasis in the original)

Furthering the discussions, Bell (1984) asserts the importance of

[4]Which could be partly described by the following quotation: "We must somehow become witness to the everyday speech which the informant will use as soon as the door is closed behind us [...]" (Labov, 1972, p. 85).

thinking of style not as any other independent variable, but as a distinct axis of sociolinguistic variation. Judging Labov's (1972) views on style as "narrow and mechanistic" (Bell, 1984, p. 146), he outlines his own conceptualisation of style named "Audience Design". Through his study of the speech of a New Zealander radio broadcaster in two different news program recorded in similar conditions, he found that the broadcaster had a different linguistic behaviour depending on which program/station he was presenting on. He argues that the content in both cases was read and spoken with the same amount of attention paid to speech, highlighting the shortcomings of Labov's model, and critiques his association of specific tasks to specific styles (such as "minimal-pair style"). While Bell (1984) doesn't deny the impact of attention in intraspeaker variation, he conceives it as "[...] a mechanism, through which other factors can affect style" (Bell, 1984, p. 150).

Moving onto his proposal, he reaffirms the importance of the interrelationship between interspeaker and intraspeaker dimensions, establishing the "Style Axiom" as follows:

> Variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the "social" dimension.
>
> (Bell, 1984, p. 151)

Considering the dependence on the interspeaker variation and redirecting the emphasis of stylistic variation from the sole speaker, Bell (1984) posits that intraspeaker variation relies on the attributes of the audience, and their influence over the speaker. He proposes that the speaker mainly shifts their style based on the interlocutors involved directly and indirectly in the interaction, designing their speech to these subjects. Further, he ranks different roles of the audience regarding their distance from the speaker: addressees, auditors (who are not directly addressed), overhearers (who are not addressed and not ratified) and eavesdroppers (who aren't known by the speaker).

Moreover, Bell (1984) also notes the importance of setting and topic. In his later work, he hypothesises that "Style-shifting according to topic or setting derives its meaning and direction of shift from the underlying association of topics or settings with typical audience members" (Bell, 2001, p. 146), since these presuppose addressee related variation. Also in later considerations about his early work, he comments that the model should expand to account for the initiative dimension, under the name of "referee design" but without losing a reasonable level of generalisation and falsifiability (Bell, 2001). Nevertheless, his first version of the model already accounts for some of these factors.

Though his critique of the notion of style as attention paid to speech is valid, it is not necessarily a critique of Labov's own views on style. As stated by Bell (1984, p. 150): "[...] attention is better regarded as a factor in the linguistic interview,* rather than the all-embracing dimension of style". The asterisk on this quotation refers to a note stating that this was suggested by Labov on personal communication. Interestingly, the above quote is preceded by the following: "Although Labov clearly intended it as a theoretical and not just a methodological construct". Before the above quote, though, Bell (1984) cites the following passage by Labov (1972, p. 99):

> We can therefore put forward the hypothesis that the various styles of speech we are considering are all ranged along a single dimension of attention paid to speech, with casual speech at one end of the continuum and minimal pairs at the other. If future research succeeds in confirming this hypothesis, and quantifying attention paid to speech, we will then have a firmer foundation for the study of style shifting, and more precise relations can be established in the study of sociolinguistic structures as a whole.

Albeit one might agree with Bell's interpretation while reading this excerpt, Labov's (1972, p. 208) quote regarding multiple stylistic dimensions, coupled with the following two, can lead to an opposing view:

> We may define a sociolinguistic variable as one which is correlated with some non-linguistic variable of the social context: of the speaker, the addressee, the audience, the setting, etc. (Labov, 1972, p. 237)

> [...] every speaker we have encountered shows a shift of some linguistic variables as the social context and topic change [...] (Labov, 1972, p. 208)

In fact, Labov (2001a, p. 87) recognises that audience-based adaptation is part of style-shifting. In the occasion, he also addresses the misunderstanding about his founding proposal:

> The organization of contextual styles along the axis of attention paid to speech (Labov 1966a) was not intended as a general description of how style-shifting is produced and organized in every-day speech, but rather as a way of organizing and using the intraspeaker variation that occurs in the interview. (Labov, 2001a, p. 87)

A third interpretation of the intraspeaker variation phenomenon is known as Speaker Design (Hernández-Campoy & Cutillas-Espinosa, 2012; Schilling, 2013; Schilling-Estes, 2002). This view criticises the late-acknowledged Audience Design question of not dealing enough with the speaker's initiative and creativity. For the adepts of this view, stylistic variation could be described as "[...] the activity in which people create social meaning, as style is the visible manifestation of social meaning" (Eckert, 2003, p. 43).

Eckert (2003) argues that Bell's (2001) structuring style as a function of a present or imagined audience limits what she conceives as "speaker agency". Coupland (2007) also brings into discussion the importance of perception in the evaluation of stylistic shifts. He suggests that presenting a higher percentage of usage of specific features not necessarily means that the speaker would be differently perceived by the interlocutor, or by anyone, in that matter. Interpreting frequency of

usage in such a manner would potentially clash with the notion of sociolinguistic indexicality, in which contextualisation is prioritised over frequency inasmuch similar forms could index very different meanings (Eckert, 2008).

An example of this is presented by Eckert (2008). She shows that the hyperaticulated pronunciation of (-t) in English was shown to be employed by and correlated with fairly distinct groups, bearing different meanings for each. This pronunciation was adopted by Orthodox Jewish boys, nerd high-school girls, and gay men. In this case, according to Coupland's (2007) reasoning, simply accounting for the frequency of this feature would not only be unsuccessful but erroneous.

Another example, from São Paulo Portuguese, is the rhotic approximant pronunciation of coda (-r), that is associated both with rural values, related to *caipira* and with urban lower-class periphery-living speakers (Oushiro, 2014).

One important notion that perpasses the discussion about language variation and style is prestige. Although this view is more prominent on Labov's attention-paid-to-speech model in which he argues that speakers adopt more prestigious forms as a consequence of higher attention to speech, prestige is also at play in both Audience and Speaker Design.

In Audience Design, the forms employed by the speaker can be thought of as those that would be more highly regarded by the audience. A potential adoption of a less standard speech, in this case, can be seen as a shift towards covert prestige, an opposing force to the traditional normative values in the community (Trudgill, 1972).

On Speaker Design, the overt–covert prestige relationship could work in a similar manner, as the speakers would also need to deal with the perceived linguistic values at play in order to use linguistic resources to shape their persona. Another possible interpretation, perhaps more in line with the notion of power relations and local meaning negotiation, would be thinking of multiple axes of prestige associated to a multitude of values for different groups.

Regardless of how we structure prestige in relation to these views on

intraspeaker variation, the fact is that societal pressures affect the way people speak in different situations. These situations can be defined as a function of different factors: the speaker, the listeners, the environment and so on. Nevertheless, defining one of these pivoting points as the main conditioner of intraspeaker variation prior to the analysis can be a source of bias – which is potentially avoidable. In this sense, adopting a more holistic point of view, recognising that many different factors can be relevant for explaining this type of variation can lead to a better understanding of general linguistic behaviour.

1.3 Newscasting speech

One particular vector of language prestige is the television. In this medium, newscasting speech is a frequent source of normative linguistic reference (Carvalho, 2004; Catani, 2021). Although most sociolinguistic studies have focused on the study of vernacular speech, Johnstone and Bean (1997) point out that studying public speech, in its various forms, is the best way to investigate speaker's competence more comprehensively.

One particular characteristic associated to newscasting speech, however, is that the social meanings associated to the employed linguistic forms seem to work differently for segmental and suprasegmental features. While selected segmental variants are commonly taken as the "correct way" of talking, the prosodic behaviour of the news presenters is not usually seen as such (Catani, 2021). This may indicate that the meanings are independently associated with these phonetic dimensions, which may be due to profound results of normative processes in language. Furthermore, the study of this particular kind of speech can be fruitful not only as a way to explore prosodic behaviour and language variation and change, but also to better understand the relationship between phonetic forms and social meaning.

A relevant reference for the present research is the work by Valente et al. (2021) about age-related prosodic differences in the speech of a

male European Portuguese public figure. Their corpus consisted of 30 audio samples from TV interviews, each around 4.5 seconds long, for each of the three analysed ages: 51, 74 and 82 years old (total span of 31 years). From these data, they extracted 17 acoustic parameters using Prosody Descriptor Extractor (Barbosa, 2021), the same script used in the present study, as described in Chapter 2.

The authors observed an overall increase over time in the f_0 measures, such as minimum, median and maximum values. Fundamental frequency variability (standard deviation; interquartile semiamplitude) also increased as well as the fundamental frequency range of the subject. Nevertheless, all of those measures show a nonlinear trend, increasing from ages 51 to 74 but decreasing from ages 74 to 82 while still being at higher values than those seen at 51. The results concerning the first derivative measures of f_0, which are associated to the rate of change of f_0, were significant only in the comparisons within the first age group, with an overall decrease of the rate of change, corresponding to less steep rises and falls, and decrease of the standard deviation of the change rate, that is, a decrease in its variability. Regarding intensity, they found that the speech in the older samples were less variable, even though a similar inverted V shape was also present in the coefficient of variation, due to an increase in variability between ages 74 and 82. It is worth noting, though, that the age difference between the aforementioned samples is of only about 8 years, while the first two analysed samples are about 23 year apart. As for the measured spectral emphasis, an overall decrease was also observed. Both of these measures correspond to decrease in vocal effort.

Valente et al. (2021) raise important considerations related to the need of longitudinal works about age-related prosody modification with more speakers. In agreement with the authors I emphasise the need for more research of this kind involving young to middle aged adults, which should help delimit when some of the age effects prevalent on the literature starts to take place or which are the turning points of prosodic variables. These works should allow researchers and forensic phoneticians

to focus on particular speech features given the potential age of the subject, improving the Speaker Comparison process. For this, we are also certainly in need of more prosodic studies in Portuguese, which should map typical phenomena to specific populations and linguistic communities.

Specifically about newscasting speech, Gasser et al. (2019) ran both production and perception experiments. In the former, they compared recordings of news speech from the Boston University Radio News Corpus (Ostendorf, Price & Shattuck-Hufnagel, 1995) with non-newscasting reading of the same content. In the latter, they applied surveys to test whether the participants could distinguish newscasting speech from "everyday speech". For that they used a delexicalised[5] version of the data from the first experiment. Their results show that the newscasters employed a less variable intensity. Male newscasters also had lower values of minimum f_0 than in the non-newscasting readers, which was correlated to the evaluation of the speaker as a newscaster. Moreover, for the male speakers there was no difference between newscasters and non-newscasters in f_0 means, range, standard deviation and maxima.

The most distinct aspect of the tasks described in this paper is a survey applied to twelve radio newscasters. The authors asked the newscasters to rate their agreement to a set of completing options for the sentence "When delivering the news on air, how important is it to you to sound..." (Gasser et al., 2019, p. 23). The following options were given: trustworthy, charismatic, objective, persuasive, authoritative, enthusiastic, likable, engaging. The highest rate, on a scale of 1 to 5, was attributed to trustworthiness ($\mu = 4.83$) followed by engagement ($\mu = 4.75$), and persuasiveness ($\mu = 2.83$) was rated the lowest.

Some respondents also answered brief open questions about potential differences between their professional versus everyday voice, what they tried to convey while broadcasting and how. In those answers the

[5]In this case, using a low-pass filter, a filter that attenuates frequencies higher than a given threshold. In the case of delexicalisation experiments, it is used as an attempt to obfuscate lexical content of utterances, which may not be fully effective.

authors note the ubiquitous desire to "sound relaxed, conversational and natural and to avoid affectation" while presenting the news (Gasser et al., 2019, p. 24). Remarkably, they also report that several of the informants alleged that there was "no difference between their on-air and conversational voices" (Gasser et al., 2019, p. 24). On this account, they also quote the following typical answer: "I work very hard in not sounding like I'm reading the news. I imagine telling my story to a friend or family member in a conversational and colloquial manner."

While this quote gives us an insight about the potentially not self-evident[6] nature of the modifications in the speech characteristics as commented on the paper, it also suggests that the observed shift in usage could be more directly explained by stylistic conditioning than by the speakers' conformity to the newscasting genre itself. According to the participant, in order to try to control his way of speaking, seeking to sound more colloquial, he tries to portray a different audience – one to which he would speak in a more colloquial manner. This is particularly interesting considering Bell's (1984, 2001) remarks on variation conditioned by Audience Design. Nevertheless, this shows one of many apparent complexities in the understanding of public speech. The participants' comments show two opposing pressures: the socially imposed forces which directs the speaker to talk differently than he usually does and the subjective perceived intention to speak the same. Playfully put, we have to work hard to sound like we are not working hard, and we hardly sound so.

On the other hand, it is worth further exploring the professionals' statement on the similarity between their regular and professional speech. Experiments with professional speakers in reading tasks elicited through different prompts may shed some light on the contrasts between different usages by voice professionals and their characteristics. Barbosa and Boula de Mareüil (2018) analysed data from different reading tasks assigned to 8 news announcers, male and female native speak-

[6] In the authors terms, "unconsciously".

ers of Brazilian Portuguese or French. The participants were asked to read *The North Wind and The Sun*, an Aesop's fable often used in phonetic experiments, in three different registers: neutral, broadcast news style, and consecutive imitation of a female newscaster of the participants' language. Data were divided according to differences in the position of f_0 peaks in relation to the syllables from Accentual Phrases. For the Brazilian Portuguese speakers, the authors found no difference between f_0 medians of the neutral and broadcast style, but observed an increase in the range of f_0 in relation to non-neutral readings when the accentual phrase was anchored in the initial syllable of the word.

These results for the Brazilian Portuguese newscasters may in principle be interpreted in the same direction as those presented by Gasser et al. (2019). The difference in f_0 range seems to be conditioned by particular realisations of the Accentual Phrases that were not controlled in Gasser et al. (2019), and no difference was observed regarding the means. Nevertheless, at least for the male professionals, the overall results by Barbosa and Boula de Mareüil (2018) seem to sustain the distinctions between the neutral and non-neutral tasks regarding aspects such as stress positioning, spectral emphasis, as well as some f_0 measures. Such elasticity, as conclude the authors, "allows prosody to operate as a socioprofessional marker, indexical of a particular speaking style" (Barbosa & Boula de Mareüil, 2018, p. 426).

Tielen (1990), in turn, analysed listener's categorisations of Dutch voices regarding the speaker's profession. Forty listeners performed the evaluation of multiple texts read by 60 male and female speakers of three professions: nurses, managers, and agents at information offices. The listeners had to identify the speaker's sex, age group and profession among 6 given categories; in addition to those mentioned above, they also had the following options: teacher, shop assistant, and radio reporter. Finally, they were also asked to rate the characteristics they associated with the typical speech of the 6 professions in a set of bipolar scales.

Listeners were quite successful in guessing the speakers' sex and

age. Although less precise, the results for occupational guesses were somewhat successful. Nonetheless, this classification showed a sex bias, where females tended to be more classified as nurses and males as teachers. Due to the bias, an overall correct classification of the nurses was observed. An apparent success was also achieved in the overall classification of the managers, whereas the information agents had more evenly distributed answers. It's worth noting that the radio reporter category was the overall less chosen occupation.

While the aforesaid bias might be explained by the uneven number of workers in each of these professions, the descriptors associated with their voices allude to a perhaps deeper social dimension of voice perception and subsequent speaker evaluation. This could even contribute to the asymmetric distribution of workers itself. Despite the apparent male tendency to rate nurses as having a more sweet, vulgar and emotional voice than female judges, there were even greater differences for all the respondents in regard to the professions, specially between managers and nurses. In this case, managers were rated as much more polished, business-like, and severe than the nurses. Taking this into account, it seems safe to assume that a clear association of voice characteristics and occupation was observed.

In tandem with the newscaster survey results, Strangert (1991, p. 39) affirms that their style of speaking is "guided by demands for effectiveness and intelligibility". Strangert (1991) further presents a comparison between a Swedish professional and a non-professional female newsreader in which the latter was asked to read a transcript of an authentic broadcast delivered by the former. The author states that there were different usages of stress and different pause strategies, with the professional favouring semantically governed less marked filled pauses while the non-professional tended to employ more syntactically dependent and prominent pauses. According to her, the professional conferred a more emphatic characteristic to the speech. Similarly to Barbosa and Boula de Mareüil (2018), the author noted a greater tonal range in the professional reading, with distinctive f_0 peak patterns.

Castro, Serridge et al. (2010) compared parameters related to pauses and f_0 on the speech of five TV anchors and five TV interviewees, whose speech didn't appear to be associated with any particular profession (Castro, Freitas et al., 2010, p. 2). Aside from the notable absence of filled pauses on the speech of the anchors in comparison to the interviewees, which strikingly differs from Strangert's (1991) observations in Swedish, they saw no other significant difference regarding pauses in the compared samples. The f_0 measures were also very similar between the two television types of speech: the f_0 mean, standard deviation, and interquartile range were all statistically similar, as were the percentage of static, rising and falling tones. Only the intravocalic mean melodic movement, defined as "the sum of the absolute value of changes in fundamental frequency from one frame to next, considering the values within vowels" (Castro, Serridge et al., 2010, p. 18), yielded a significant difference. Their interpretation was that this difference could indicate a higher usage of short-range pitch movements by the anchors, but they recognise that the tone percentage results don't support this hypothesis.

On the other hand, although Castro, Serridge et al.'s (2010) results suggest a high degree of similarity between news anchors and interviewees, in a different study, Castro, Freitas et al. (2010) ran delexicalised (low-pass filtered) perception experiments using the same recorded data in conjunction with religious and political voice samples recorded from television. Twenty Brazilian Portuguese native speakers were asked to classify three stimuli for each of the television speaking genres throughout two tasks. In the first task, they had to choose between two options shown on a screen, with one of them being the correct one. In the second task, they had to choose between the remaining three categories that the participant had not picked in the first task. All of them were well acquainted with the aforementioned speaking styles. The overall accuracy of the answers was 90%, a significant result considering the null hypothesis of 50% of accuracy, given the experimental design. Highlighting only the TV news style, the accuracy level was the same: out of the 60 stimuli only 6 of the answers given wrong.

CHAPTER 2

METHODS

Considering the previous discussion in Chapter 1 this research seeks to explore the interface between sociolinguistics and phonetics with the goal of broadening the scope of forensic linguistics supported by the sociolinguistics developments about intraspeaker linguistic variation. More specifically, with acoustic measures associated with speech prosody, I employ quantitative methods to analyse change and continuity in the voice of a Brazilian newscaster. The results of these analyses are expected to shed light on intraspeaker prosodic variation in non-vernacular speech – a modality that is specially relevant for forensic analysis since many of the voice samples involved in forensic practice are collected in somewhat monitored settings and, thus, non-vernacular. By sampling data produced over a two-decade span, this research also seeks to better understand the effects of the ageing process in the speech of a trained voice professional, assisted by medical voice accompaniment. Furthermore, the research aims at evaluating the viability of the usage of publicly available data, not recorded for the purpose of acoustic analysis.

Since news speech is seen as one of the most standard ways of speaking in the Brazilian context, the results of this research may also assist further works, not only via the acquired acoustic data, but also as an instance of analysis of a type of speech highly resistant to language change. If linguistic change is observed in this highly controlled situation, one could expect more change over the lifespan in other contexts. The following chapter details the data and methods used in this investigation.

2.1 Corpus

The corpus selected for the analyses consists of 50 recordings of a television news program broadcast between the years 2000 and 2019, with a maximum of 3 recordings per year. Each iteration of the program have around 35 minutes and is broadcast during prime time, usually when the main programs of the TV networks are exhibited. The speech of one single anchor was transcribed and each news segment was coded according to its theme (see Section 2.2).

The subject is a male native Brazilian Portuguese speaker, who is considered a successful news anchor working at a large Brazilian broadcasting network. His voice is apparently healthy, with no known associated problems and with long-term speech therapy accompaniment. His accent sounds similarly to that of speakers from the city of Rio de Janeiro, where he has lived most of his life, but his usage shifts to a neutralised dialect when presenting the news, with markedly less salient features. The journalist's age ranges from his mid-30's to mid-50's in the collected data.

After a preliminary review, part of the data was deemed unfit for the research purposes due to low signal-to-noise ratio and thus discarded. Non-news segments, such as announcement of other programs from the network, were also removed. After the removal of these data, there remained a total of 1,989 utterances corresponding to about 155 minutes of speech data.

2.2 Transcription and coding

The content of the newscaster's speech was entirely transcribed in ELAN (Brugman & Russel, 2004). ELAN is a free[1] multimedia annotation tool intended for linguistic analyses and documentation (Hellwig et al., 2022). Its workflow is based on alignment between media and an-

[1]Available under the GNU General Public License v3.

notation, coupled with editable shortcuts and wide file format support, which makes it a great alternative to other transcription software.

The annotation style was based on the conventions developed by Projeto SP2010 (Mendes & Oushiro, 2013), a set of rules that aimed to represent spoken data in a faithful but understandable way, striving for the standardisation of the transcriptions in order to allow easier manipulation of large quantities of data (Mendes & Oushiro, 2012). The words of the corpus were transcribed in its traditional orthographic form, even when pronounced in a distinct manner. On the other hand, syntactic modifications were not made, maintaining the proffered structure even in the rare cases of unorthodox sentences.

The annotations were delimited considering perceived silent pauses, semantically complete sentences, and duration (by convention, under 8 seconds per annotation). Orthographic punctuation signs such as commas (,) and periods (.) are omitted and ellipses (...) can be employed to represent micro pauses. Proper nouns of non-public persons were anonymised, as were other related personal pieces of information.

In some moments of the news program, the subject interacts verbally with another speaker (e.g.: news reporter, weather forecaster). When this happened, the interactants' speech was also transcribed. Even though these occurrences have increased with ongoing changes in the news broadcasting format (as noted in Catani, 2021), they were still very infrequent in the present corpus. Thus, due to the lack of representativeness, the newscaster's speech data under this circumstances was discarded.

Adding to the transcripts, the news were manually coded according to their main topic. In the process, the annotations pertaining to each different news were read in its entirety. Although this is ultimately a subjective classification, not directly representative of how the program structures its news, most of the news fell within main topics that are customarily present in the organisational schemes not only on television news but on general journalistic practice, e.g.: *economics*, *politics*, *sports*, and so on. Furthermore, the classification was done solely by

the author, granting stability across the categories. The categories used are: politics news, violence news, economy news, sports news, diverse news, election news, disaster news, science news and emotional news. A brief description of these categories and their shorthand are presented on Table 2.1.

Level	Theme	Description
pol	Politics	National and international politics
viol	Violence	Police, violent crimes, war
econ	Economy	Market indexes, quotations etc.
spo	Sports	Football and sports in general
div	Diverse	Behaviour and everyday life
ele	Election	Election coverage
dis	Disaster	Disasters (hurricanes, landslides etc.)
sci	Science	Science, technology, research, nature
sad	Emotional	Visibly emotionally affected news

Table 2.1: Levels used in the coding of the news

2.3 Data extraction

2.3.1 Extraction from ELAN

The resulting ELAN (.eaf) files were further converted into the default format used by Praat (.TextGrid files) (Boersma, 2001; Boersma & Weenink, 2022). For this, a modified script from treinaPB (Silveira, 2022) was used. This script converts annotations from .eaf files into .TextGrid files. The text and the audio file are split into single independent pieces, removing unannotated streches for better performance on Praat. In addition to this conversion, an R script by Oushiro (unpublished) was also used for extracting the news topic and detailed time information from ELAN directly into a .csv file.

2.3.2 Extraction from Praat

Before the main extraction, the data was again manually reviewed. In this occasion, the data quality was reassessed and potential sources of error in the acoustic analysis were removed. These included segments with ambiental noise, music, voice overlap, recording artefacts and such.

prosody descriptor extractor

Prosody Descriptor Extractor (Barbosa, 2021) is a Praat script that enables the user to extract multiple prosodic-acoustic measures of different kinds: intonational, voice quality, vocal effort and rhythmic. The only input needed is an audio file coupled with a TextGrid with a chunk tier, that is, a tier with an arbitrary length segmentation of the analysed audio.

The user can also independently provide a tier of Vowel-to-Vowel units, a measure delimited by two consecutive vowel onsets (Barbosa, 2009), as well as the segmentation of sustained vowels, pauses and tones for additional extractions. In total, the script can extract 38 acoustic measures, 24 of which via the chunk tier (on version 3.1). These parameters are independently extracted for each segmented chunk. The acoustic measures used in the present analyses are reported on Table 2.2.

When running the script through Praat, the user is presented with a graphical window in which the existing tiers and input hyperparameters for the acoustic extractions, such as the upper and lower limits of accepted f_0 measures, are selected and defined. These threshold values for f_0 calculations were set after manually checking the subject's f_0 measures in Praat. The other following hyperparameters were left at their default values, shown on Table 2.3.

Smoothness threshold is the cut-off frequency of the smoothing filter used in the f_0 peaks calculation. The f_0 step corresponds to the time step for calculating the f_0 derivatives. Time window is the window used to make spectral calculations. The spectral emphasis threshold is used solely for the extraction spectral emphasis. The language file contains

Parameter	Description	Unit
Intonation		
f0med	Median of f_0	Hz/st re 1 Hz
f0sd	Standard deviation of f_0	Hz/st re 1 Hz
f0SAQ	Interquantile semi-amplitude f_0	Hz/st re 1 Hz
f0min	Minimum of f_0	Hz/st re 1 Hz
f0max	Maximum of f_0	Hz/st re 1 Hz
f0base	Baseline of f_0	Hz/st re 1 Hz
f0peakwidth	Peakness of f_0	Hz/st re 1 Hz
f0peakrate	Rate of f_0 peaks	Peaks/s
sdf0peak	Standard deviation of f_0 peaks	Peaks/s
df0posmean	Mean positive 1^{st} derivatives of f_0	Hz/frame
df0negmean	Mean negative 1^{st} derivatives of f_0	Hz/frame
df0sdpos	Standard deviation of positive 1^{st} derivatives of f_0	Hz/frame
df0sdneg	Standard deviation of negative 1^{st} derivatives of f_0	Hz/frame
Intensity		
emph	Spectral emphasis	dB
cvint	Coefficient of variation of intensity	NA
Voice Quality		
hnr	Harmonic to noise ratio	dB
slltas-med	Slope of LTAS in medium frequencies	dB
slltas-high	Slope of LTAS in high frequencies	dB
spi	Soft phonation index	dB
jitter	Local jitter	%
shimmer	Local shimmer	%

Table 2.2: Prosody Descriptor Extractor parameters used and associated information

data regarding phones duration for the given language and it is used to normalise raw syllable-sized duration. Tables for other languages are available.

After running it, the script outputs a comma-separated .txt file with the acoustic information corresponding to each chunk. A step-by-step

Hyperparameter	Value
f_0 smoothness threshold	5 Hz
f_0 step	0.05 s
Time window	0.03 s
Spectral emphasis threshold	400 Hz
f_0 minimum threshold	60 Hz
f_0 maximum threshold	295 Hz
Language reference	BP

Table 2.3: Prosody Descriptor Extractor hyperparameters used for extraction

guide of the script minutiae is available at the documentation provided by the author on the script repository.[2]

2.4 Analyses

After manual inspection of the extracted data, a set of outliers was removed. The analyses were divided in the three axes described bellow: change and continuity through time, news themes analysis and valence analyses.

2.4.1 Change and continuity through time

Considering the longitudinal characteristic of the sampled data, one of this research goals was to assess the role of time in the vocal behaviour and characteristics of the subject's voice. This question is specially relevant since most of the longitudinal analyses of prosodic vocal character-

[2]Available at https://github.com/pabarbosa/prosody-scripts/blob/master/ProsodyDescriptorExtractor/Documents/ProsodyDescriptorExtractor_Manual.md

istics seem to deal with the voice of older speakers, leaving a gap about how these acoustic parameters act during middle age. In a more technical sense, it is also relevant to observe how some of the less analysed acoustic parameters extracted behave in general, such as those related to f_0 derivatives. From a sociolinguistics standpoint, there is interest in both the single speaker long-term changes and the general characteristics of newscasting speech – commonly equated to a beautiful and correct way of speaking in the Brazilian context.

With these considerations in mind, linear regressions were run for each variable presented on Table 2.2 with the goal of checking potential changes in the acoustic parameters in the sampled period. These regressions were done using the ordinary least squares method and the date of the news transmission in days was used as the predictor. The alpha value was $\alpha = 0.01$.

2.4.2 News themes

A factor that can be associated with phonetic variation in broadcast speech is the themes of the news. Although this could appear to be particular to newscasting, the news themes are directly related with topic. These themes could each be associated with a specific target audience, they can be guided by different sets of behavioural rules and can also be a locus for the expression of the newscaster positions or the expression of the network stances in relation to a subject. Thus the themes can be associated to many of the concerns of sociolinguistics intraspeaker variation research.

Kruskal-Wallis tests were used to check for differences between news with differently categorised themes. These non-parametric tests were used because the assumptions required by its parametric correspondent, one-way ANOVA, were not met.

For each acoustic variable that presented a significant difference between news groups in the Kruskal-Wallis test, Dunn's (1964) multiple comparison tests were run to check which groups differed from the oth-

ers for each of the parameters analysed. The usage of this multiple comparison test follows Dolgun and Demirhan's (2017) findings using Monte Carlo simulations in multiple comparison scenarios. In their research, the robustness of Dunn's (1964) test was demonstrated when comparing variables with higher number of unbalanced groups. Although the sample has a relatively large amount of data, the groups are unbalanced – mostly due to the category "Politics" (*pol*), which has 832 observations, against a mean number of 142 observations in each of the other 8 groups and "sad" (n = 98) as the category with the smallest number of observations.

The Dunn's (1964) tests were applied using the Holm-Bonferroni method for controlling family-wise error rate (Holm, 1979), avoiding false positives. Both tests used $\alpha = 0.01$.

2.4.3 Valence analyses

A more generalised approach to account for topic and more specifically to the emotional dimension of the speech is the analysis of valence. In this type of analysis one can classify an arbitrary speech unit based on an abstract axis delimited by opposing positive and negative poles, respectively associated with subjectively "good" and "bad" values. These characterisation can be done in many ways, from a binary categorisation to a more fine grained approach considering nuances in the classified elements. Two different techniques were used with this purpose: Sentiment Analysis and manual valence classification.

sentiment analyses

Sentiment analysis is a method for systematic evaluation and classification of subjective information associated with a specific content. These pieces of information can be categorised in relation to multiple axes of emotions, stances, and others, through the usage of various sets of resources.

In the current research, Sentiment Analysis was used to evaluate the valence (negative, neutral, positive) of word-tokens from the news corpus using the following pre-built Brazilian Portuguese lexicons – groups of word-value pairs:

- NRC Word-Emotion Association Lexicon (also known as EmoLex) (Mohammad & Turney, 2010, 2013)
- SentiLex (M. J. Silva et al., 2010; M. J. Silva, Carvalho & Sarmento, 2012)
- oplexicon v3.0 (Souza & Vieira, 2012; Souza et al., 2011)

This resource was experimentally employed in the current research, testing the viability of the usage of some previously constructed models as a replacement of manual classification for sociophonetics research purposes. The analysis was conducted using the R Language (R Core Team, 2020) with via the Syuzhet package (Jockers, 2015).

EmoLex was used directly from Syuzhet and the other two lexicons were downloaded[3] and their field names were modified in accordance to Syuzhet's expected format. The outputs of these analyses are integers, with the sign corresponding to the sum of token valence of each analysed chunk. These results were later transformed into binary data (positive and negative) allowing a comparison with the manual codification described below.

MANUAL VALENCE ANALYSIS

The valence of the contents of each chunk was manually coded into negative, neutral or positive. For this step, only the transcripts were used since the auditory classification of the samples could be influenced by

[3]Respectively retrieved from https://b2find.eudat.eu/dataset/b6bd16c2-a8ab-598f-be41-1e7aeecd60d3 and https://www.inf.pucrs.br/linatural/wordpress/recursos-e-ferramentas/oplexicon/

the acoustic characteristics of the observations. This could lead to a circular relationship such as data being classified as positive due to high median f_0, since high f_0 median is associated with positiveness.

One potential caveat of this method of categorisation is that the content of the news does not always match the newscaster intonation. News which the content might not be clearly negative might sound very negative due to the newscaster intonation. The newscaster can also give a positive spin in sad news and vice-versa.

Nevertheless, intersubjective agreement between ratings of the overall valence would require ratings by different listeners, perhaps even a larger scale perception experiment, since the researchers' opinions on the valence of previously repeatedly listened news may not align with the judgement of naive and prototypical listeners. Although valence classification also involves subjectivity, content valence rating is adopted here as a more objective approach to valence than what we could call "performative valence". This method also provide results using similar data of those used in the Sentiment Analysis classification and it is assumed to be less time consuming than auditory approaches. Future studies are needed to evaluate the relationship between the content valence and the performative valence of news and other scripted media.

Data were categorised with numbers corresponding to the valence levels: 1 = negative; 2 = neutral; 3 = positive. Information was directly inserted in a spreadsheet column that was later appended to the main .csv file. Each chunk was individually ranked but information of the surrounding annotations were used for contextualisation. News about negative topics such as crime were classified as negative, including the news about a police case being solved. Unclear and ambiguous cases were classified as neutral, which became akin to a default/unmarked category. This neutral category was later merged with the positive values, since only a few of the news were classified as markedly positive.

After the classification was done, their results were contrasted. Linear regressions were run for each of the four sets of categorised data checking the trend of negative news chunks proportion over time.

CHAPTER 3

RESULTS

This chapter presents the main results of the analyses of the newscaster's speech data. While this chapter bears a more descriptive approach to the quantitative results, a more in-depth and qualitative discussion of these findings is presented in Chapter 4.

I begin with results of the acoustic parameters analysed across time, describing whether features have remained stable along his lifespan or not. Ordinary least squares linear regressions were used for this specific analyses. In Section 3.1.1, I present the results of the variables that have changed over time. In Section 3.1.2, I present the results of the variables that showed negligible change or that have maintained an overall stability throughout the sampled years, despite the observed variation.

Further, in Section 3.2, I analyse the relationship between the acoustic parameters and the news themes. These analyses were carried using Kruskal-Wallis tests in conjunction with Dunn's tests.

Lastly, I present the analyses of valence of the narrated contents through two different approaches: sentiment analysis and manual content analysis. The Sentiment Analysis was applied using three third-party Brazilian Portuguese lexicons, whereas the manual content analysis was done by manually classifying the contents of each analysed chunk.

3.1 Change and continuity in the acoustic parameters

Given the extended time interval (almost 20 years) that delimits the corpus, we start with the results related to the time variable. Time was measured in days, starting by April 03, 2000, date of the first edition in

the corpus as "day 1" until the most recent edition, from September 02, 2019, 7092 days later.

The results of each test are shown in Table 3.1. The first column present the parameters' names as they are on the script; the second column presents the degrees of freedom and the F-statistic; the third and fourth columns respectively show the probability value for each test and the adjusted R^2 (the coefficient of determination) that indicates the amount of explained variation (between 0 and 1). I start by presenting the variables associated with f_0 derivatives, followed by other intonational measures and then by voice quality and intensity measures.

3.1.1 Change

means and standard deviations of the first derivatives of f_0

Figure 3.1 shows the linear regression of the absolute value of the means (left) and the standard deviations (right) of the positive (blue) and negative (orange) derivatives of f_0 (Hz) by time. These means and standard deviations respectively correspond to the rate of f_0 change through time and the standard deviation of this rate. Although all calculations were based on days, the data is arranged on the horizontal axis according to the year of the transmission for ease of reading.

The negative values were converted to their absolute value for the sake of graphical comparison. Each observation is represented by the plus sign "+" for the positive derivatives and the "x" sign for the originally negative values. The greater the neighbouring number of observations, the bluer the corresponding sign.

Both measures of derivatives on both plots show a downwards trend. That is, as time passes, the absolute means and standard deviations of the aforementioned derivatives decreases, which corresponds to smoother f_0 rises and falls.

Variable	(df) F-statistic	Significance	$\bar{R}^2$
slLTAShigh	(1, 1967) = 381.95	< .001	.16
df0posmean	(1, 1962) = 135.57	< .001	.06
df0sdpos	(1, 1967) = 130.58	< .001	.06
slLTASmed	(1, 1965) = 98.48	< .001	.05
hnr	(1, 1967) = 83.04	< .001	.04
f0max	(1, 1967) = 71.51	< .001	.03
f0sd	(1, 1967) = 67.27	< .001	.03
df0negmean	(1, 1967) = 63.16	< .001	.03
df0sdneg	(1, 1967) = 47.14	< .001	.02
sdf0peak	(1, 1967) = 45.26	< .001	.02
f0SAQ	(1, 1945) = 37.14	< .001	.02
f0med	(1, 1967) = 31.08	< .001	.02
shimmer	(1, 1967) = 10.30	< .01	< .01
cvint	(1, 1967) = 10.05	< .01	< .01
f0peakrate	(1, 1899) = 5.48	< .05	< .01
jitter	(1, 1945) = 1.26	> .10	< .001
emph	(1, 1967) = 0.88	> .10	< .001
f0base	(1, 1967) = 0.39	> .10	< .001
f0peakwidth	(1, 1967) = 0.35	> .10	< .001
SPI	(1, 1962) = 0.07	> .10	< .001
f0min	(1, 1967) < 0.01	> .10	< .001

Table 3.1: Linear models results

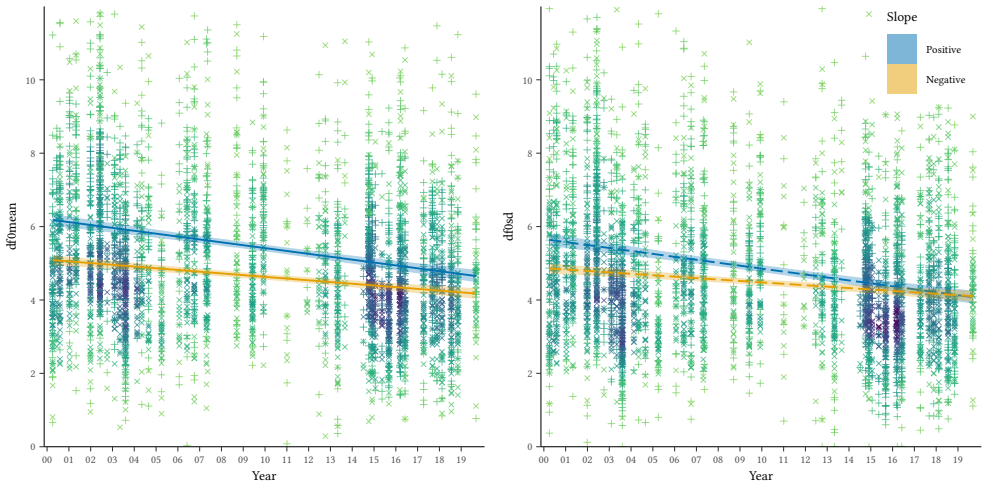


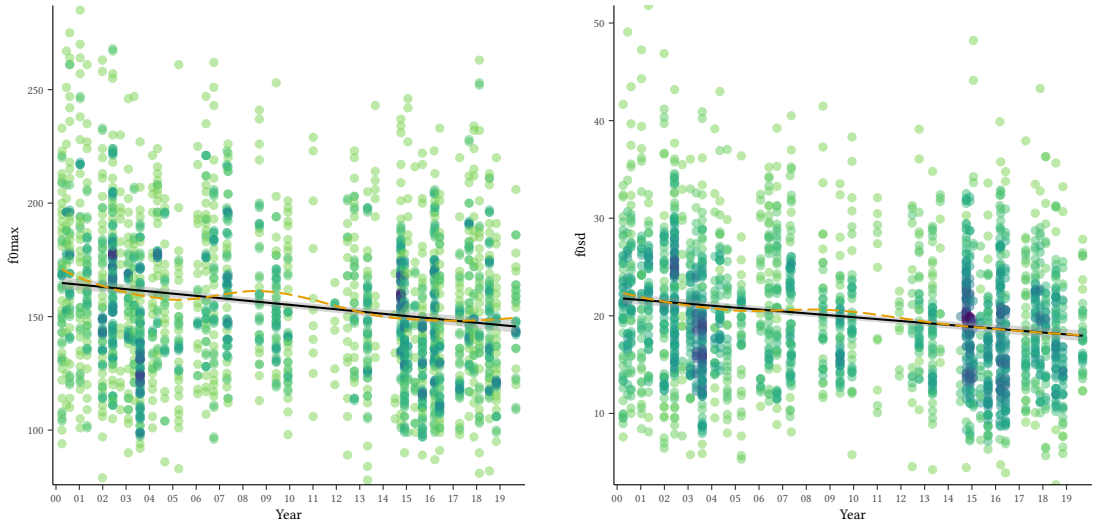
Figure 3.1: Means and standard deviations of f_0 derivatives by year

MAXIMUM f_0

A similar downward trend can be observed in the maxima of f_0. In Figure 3.2a, the regression line is plotted in black. Coupled with it it, the orange dashed curve represents a fit using a method of moving regression: the Locally Estimated Scatterplot Smoothing – LOESS (Guthrie, Filliben & Heckert, 2003). This resource allows for a better visualisation of the distribution, as the fitted curve is more sensitive to local dispersion in the data.

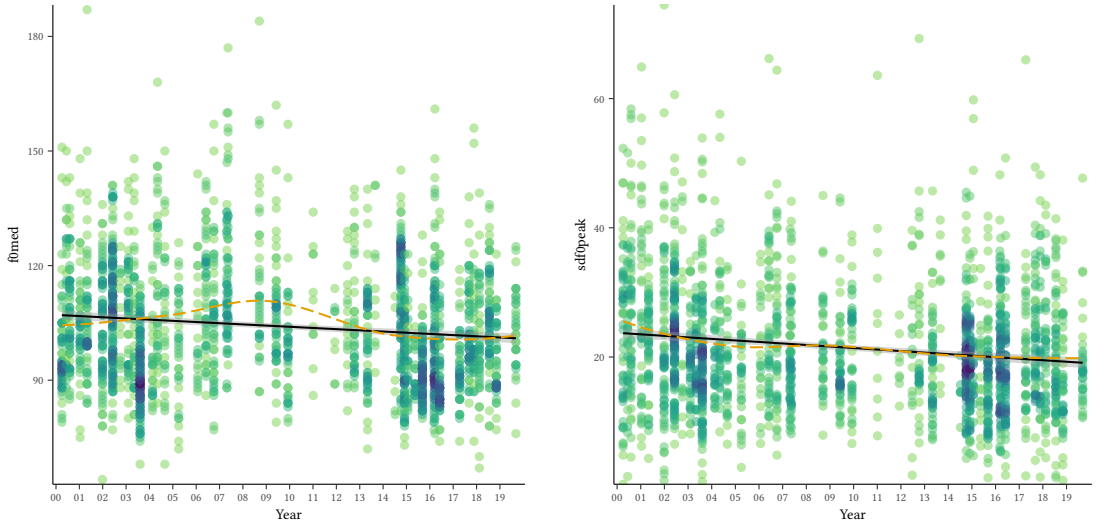
Each observation in the plot is represented by a colored dot. The colouring is applied in a similar fashion to that on Figure 3.1, that is, a bluer or darker dot represents an observation with a higher number of neighbouring dots than the observations represented by lighter greenish dots. This plotting "formula" is followed for the subsequent linear regression plots in this section.

In this case, we see a falling pattern, meaning that the maximum values observed in each chunk of audio get smaller over time, going from about 166 Hz in the year 2000 to 146 Hz in 2019. The difference is about 2 semitones (st) with reference value (re) of 1 Hz.



(a) Maximum values of f_0 (Hz) by year

(b) Standard deviation of f_0 (Hz) by year



(c) Median values of f_0 (Hz) by year

(d) Standard deviation of f_0 peaks (Hz) by year

Figure 3.2: Linear regression plots of intonational parameters

STANDARD DEVIATION OF f_0

Figure 3.2b shows the standard deviation of the fundamental frequency (in Hertz). The downward trend shows a decrease on the dispersion of the f_0 value observed through time, going from a mean value of 22 Hz

in 2000 to 18 Hz in 2019. Here, smaller values mean an overall lesser degree of f_0 variability, that could be generally perceived as lesser intonational variability, perhaps a more monotonous speech. The interquartile semi-amplitude (f_0 SAQ) was also measured but since the extreme values observed here are real observations and not outliers, the standard deviation seems to be a more reliable measure, since it doesn't discard what is outside of the interquartile range $\times$ 1.5 the standard deviation.

MEDIAN OF f_0

Figure 3.2c presents the medians of f_0 observed. Although there was a large amount of variation in the observed years, a slight change can be observed. This variable presented a slight change from 109 Hz in 2000 to 101 Hz in 2019, a difference of about 1 semitone. It's worth noting the high amount of variation seen on the data per edition. The mean overall standard deviation of the median f_0 is about 17 Hz. Considering that the speaker's average f_0 median is 105 Hz, his f_0 range while presenting the news would be about 5 semitones, typically between 88 Hz and 122 Hz, based on the standard deviation data.

STANDARD DEVIATION OF f_0 PEAKS

This measure corresponds to the standard deviation of the values of f_0 peaks in each observed chunk. Figure 3.2d shows a slight decrease of the standard deviation of the peaks from 24 Hz in 2000 to 19 Hz in 2019. Considering the global values of f_0 maxima seen on Section 3.1.1, the average maximum range went from between 142–190 Hz (diff = 48 Hz, 5 st) to 127–165 Hz (diff = 38 Hz, 4.5 st). That is, in the sampled time there is a small decrease in overall variability of the peaks, which might also be perceived a slightly less dynamic speech.

HARMONICS-TO-NOISE RATIO

This parameter corresponds to the ratio between harmonic signals and noise. It is used as a voice quality measure, as it is correlated with vocal roughness, hoarseness, and breathiness (Linville, 2001; Martin, Fitch & Wolfe, 1995). The higher the amount of noise, the higher these characteristics are perceived.

Figure 3.3 shows an upwards trend, with an increase in the ratio over time. This result corresponds to a reduction of noise in relation to harmonic sounds. In more recent news, thus, the newscaster's voice could be perceived as less rough and breathy, perhaps more clear.

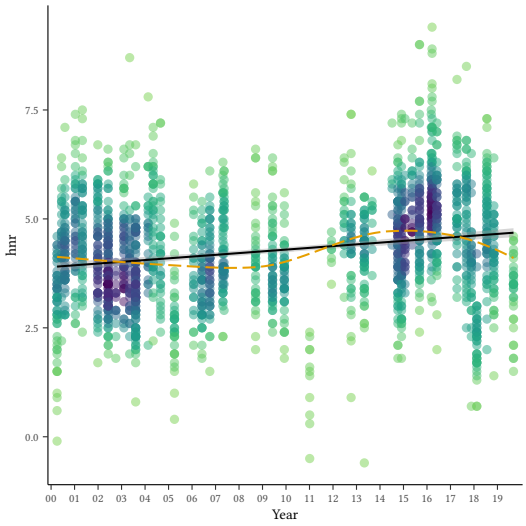
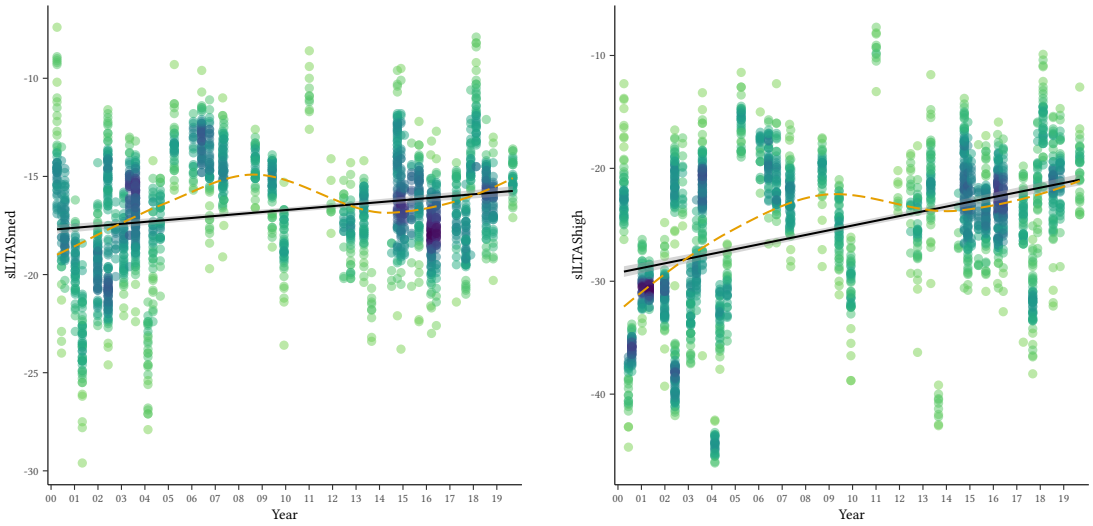


Figure 3.3: Linear regression of harmonics-to-noise ratio

LONG TIME AVERAGE SPECTRUM

The long term average spectrum is a measure used to describe the resonant frequencies of a speaker vocal tract, which allows for an assessment of speakers' invariant vocal characteristics (Figueiredo, 1993; Pittam, 1987). Thus, it is an interesting tool to speaker comparison and identification (Hollien & Majewski, 1977). It is calculated by averaging



(a) LTAS slope (medium frequencies)

(b) LTAS slope (high frequencies)

Figure 3.4: Linear regression of long term average spectra

spectra of continuous speech across a given frequency range (Pittam, 1987).

In this case, the slopes between the reference 0kHz–1kHz and medium (1kHz–4kHz) and high frequencies (4kHz–8kHz) are calculated. Lower absolute values correspond to higher perceived effort. Figure 3.4 shows that both parameters had great dispersion across the different news editions, trending towards lower absolute values, in this case, values closer to zero.

3.1.2 Continuity

In this section I present the results for parameters that, though variable, remained stable in the observed range of time. Some of these results have shown some but a negligible amount of change, being somewhat stable, proceeding to variables for which absolutely no change was observed. I first present the results regarding intonational parameters, followed by parameters associated with vocal effort, and then those associated with voice quality. The displayed figures follow the same order of the textual presentation of these results.

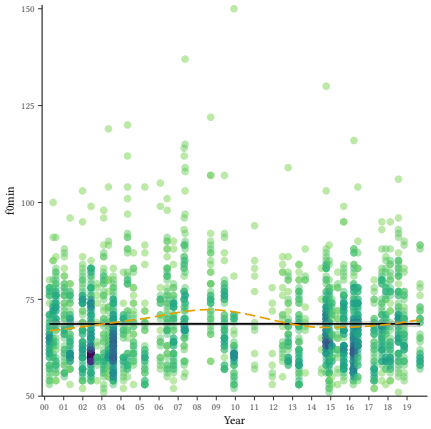
f_0 minimum

Figure 3.5a shows the minimum f_0 values for each analysed chunk. This parameter remained stable at around 67 Hz, approximately 72 st re 1 Hz.

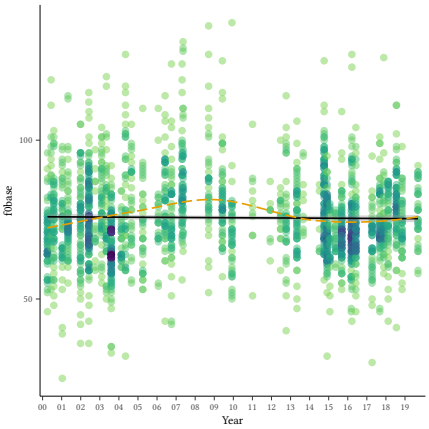
baseline of f_0

This acoustic parameter was proposed as an alternative f_0 descriptor, assumed to be more robust than the mean and median f_0 regarding potential sources of variation in the speaker's voice, such as differences in vocal effort and in speaking styles (Lindh & Eriksson, 2007; Traunmüller & Eriksson, 1995). Thus, it is of particular relevance for forensic analysis, as it deals with a potentially invariant vocal feature. It is calculated by the subtraction of $1.43 \times sd$ of the median of f_0.

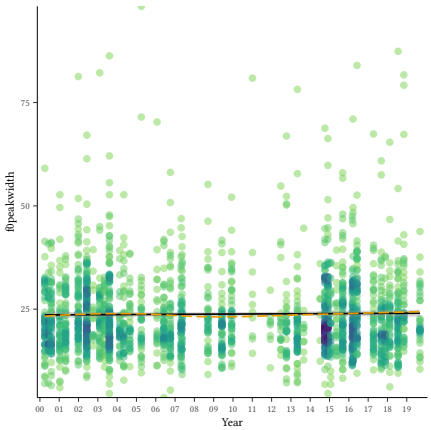
In the period analysed, the baseline of f_0 remained stable. This result corroborates the robustness of this parameter, as it remained stable in the sampled time while the median f_0 changed with time.



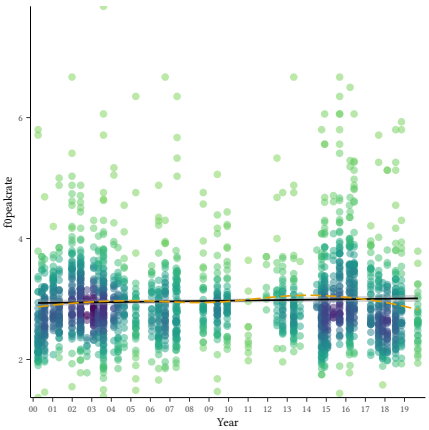
(a) f_0 min percentage by year



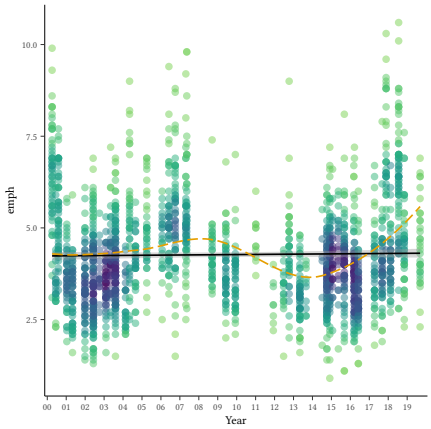
(b) f_0 baseline (Hz) by year



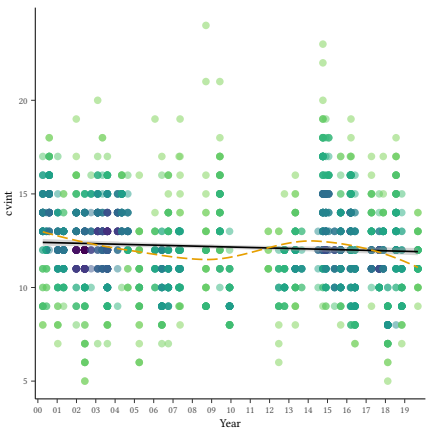
(c) Width of f_0 peaks (Hz) by year



(d) Rate of f_0 peaks (peaks/s) by year



(e) Spectral emphasis (Hz) by year



(f) Intensity coefficient of variation by year

Figure 3.5: Stability in intonational and vocal effort parameters

WIDTH OF f_0 PEAKS

The width of the f_0 peaks, also known as peakness, is given by the subtraction of the maximum f_0 frequency from the means of frequencies of two equidistant points before and after the peak. In other words, it is the difference between the peak and the average frequency of its neighbours given a time window. The bigger the difference between the two values, the flatter/wider the overall peak shape. On the opposite side, the smaller the value, the sharper/narrower the peak. A speaker with sharper peaks may be perceived as more charismatic and persuasive than one with flatter peaks (Niebuhr, Voße & Brem, 2016). The result in Figure 3.5(c) shows no difference across the observed interval, with the width stable at around 24 Hz.

RATE OF f_0 PEAKS

The rate of the f_0 peaks is given by the ratio of the number of peaks in a given interval by the duration of said interval. In this case the intervals are equivalent to the divided chunks. Higher values are equivalent to a higher number of peaks per second. Figure 3.5(d) shows that the average number of peaks remained the same: about 3 peaks per second on average.

SPECTRAL EMPHASIS

Spectral emphasis corresponds to the difference between the total intensity of the spectrum and the frequencies up to a given threshold, in this case: 400 Hz. It "may be described as the relative intensity in the higher frequency bands" (Heldner, 2003, p. 40). Perceptually, this parameter correlates with vocal effort, with higher values corresponding to higher perceived effort. As seen in Figure 3.5(e), the overall emphasis was overall stable at about 4 Hz.

INTENSITY COEFFICIENT OF VARIATION

The coefficient of variation is a measure of dispersion defined by the ratio of the standard deviation to its mean. In this matter, the statistic is applied to the global intensity values observed in the chunks, so both the standard deviation and the means are calculated from the intensity values. Resulting values are presented as a percentage. Figure 3.5(f) shows a slight tendency of reduction of the coefficient in about 0.6% of the total, from 12.4% to 11.8%.

SHIMMER

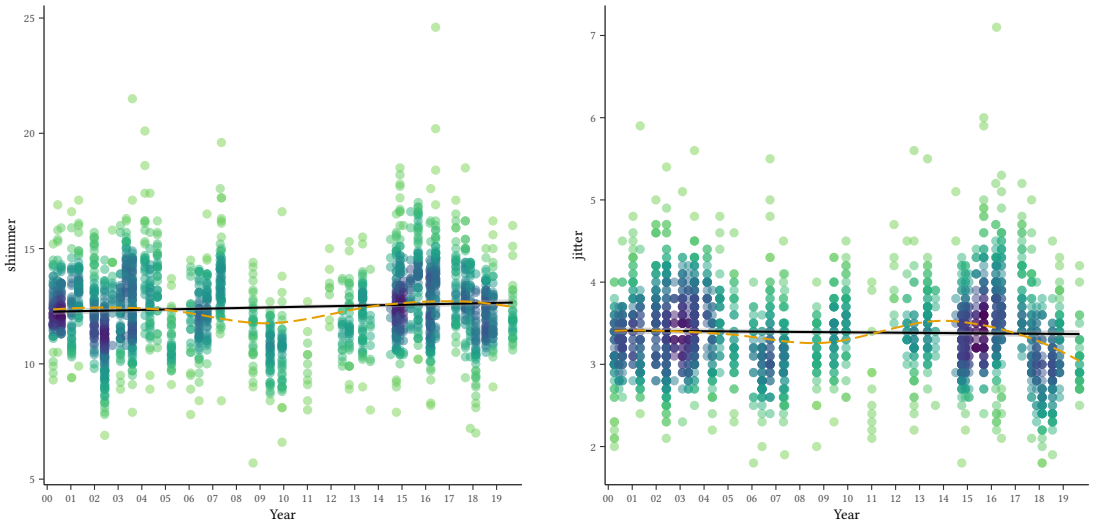
Shimmer is a measure of the amplitude perturbations in the sound wave. It can be seen as a measure of the voice stability, with lower values corresponding to more stability (i.e.: less perturbations). Higher values are correlated with the perception of vocal hoarseness and roughness. The results are presented in Figure 3.6(a) by means of the local shimmer percentage. The graph shows the slight increase of about 0.3% in shimmer.

JITTER

Like shimmer, jitter is also considered a measure of perturbation of the signal. It consists of small cycle-to-cycle variations of the fundamental frequency. The high incidence of jitter correlates with the perception of creakyness and whisperyness in the voice. In Figure 3.6(b) the local jitter percentage remained stable throughout the sampled time. The average observed value was 3.4%.

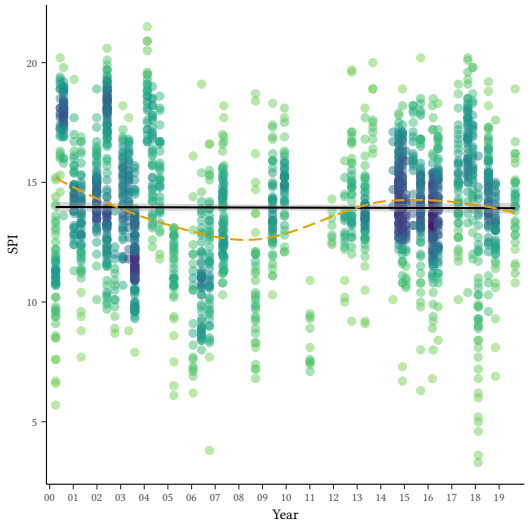
SOFT PHONATION INDEX

The soft phonation index corresponds to the average difference of harmonic energy between low (70–1600 Hz) and high (1600–4500 Hz) frequencies (Mathew & Bhat, 2009; Xue & Fucci, 2000). This index correlates with the amount of breathiness in the voice since rigidity of vocal



(a) Shimmer percentage by year

(b) Local jitter percentage by year



(c) Soft phonation index (dB) by year

Figure 3.6: Stability across voice quality measures

"may be an indication of incomplete or loosely adducted vocal folds during phonation" (Xue & Fucci, 2000). Figure 3.6(c) shows a variable but ultimately stable index of about 14 dB.

3.2 News themes

Although many parameters were stable over almost 20 years, the data from the previous section show a lot of variation of all of the parameters. Even in the cases that longitudinal change was observed, the correlation with time did not fully explain the variation in the acoustic data. These results indicate that there is another motivation for the wide variation observed. Furthermore, the data from the news themes was analysed in an attempt to better comprehend this phenomenon.

The results of the Dunn's multiple comparison between the theme of the news are presented in this section. The tests were applied only to the variables with significant results in the Kruskal-Wallis tests with $\alpha = 0.01$. The results of the these tests are shown in Table 3.2. For conciseness sake, only the variables that presented significant difference between at least two levels in the Dunn's tests ($\alpha = 0.01$) are displayed here. Results are presented in order of parameter type. First come the intonational parameters, followed by voice quality and then intensity parameters.

Six plots regarding the Dunn's multiple comparison tests are presented in Figure 3.7; each of the subfigures correspond to a different variable and its variance across the different levels of the news theme variable: science news (*sci*), disaster news (*dis*), diverse news (*div*), economy news (*econ*), election news (*ele*), sports news (*spo*), politics news (*pol*), violence news (*viol*) and emotional news (*sad*). See Table 2.1 for a description. Each set of same-coloured points represent the distribution of observations of a given theme. In each of these sets, there is also a minimalist boxplot, where the thicker box represents the interquartile range and its middle gap represents the median point. Horizontal bars on the upper part of each individual plot represent significant differences between the two themes directly below both edges of the bar ($\alpha = 0.01$).

In the maximum values of f_0 (Figure 3.7a), sports (*spo*) and election news (*ele*) showed the highest values, followed by science (*sci*). Sports news had an overall higher maxima than diverse news (*div*), which by its turn also had lesser maxima than election related news and than science

Variable	H-statistic	p
slLTAShigh	171.9	<.001
SPI	142.8	<.001
f0med	135.5	<.001
slLTASmed	90.6	<.001
f0base	89.1	<.001
hnr	83.9	<.001
df0posmean	83.5	<.001
df0sdpos	81.4	<.001
f0max	78.9	<.001
df0negmean	62.7	<.001
jitter	61.2	<.001
df0sdneg	56.4	<.001
f0sd	48.8	<.001
cvint	46.8	<.001
emph	45.4	<.001
f0SAQ	37.3	<.001
f0peakrate	32.5	<.001
f0min	23.7	.003
shimmer	16.7	.034
sdf0peak	13.8	.086
f0peakwidth	12.2	.144

Table 3.2: Kruskal-Wallis test results (n = 1969, df = 8).

news (*sci*).

Political news (*pol*), the category with the most observations, had lower f_0 maxima than the speech on the sports and election categories. Violence news (*viol*) presented a distribution similar to the political news, but was only different from the emotional news (*sad*). In fact, most of the observed differences are related to the emotionally affected news category (*sad*). These news, as can be seen in the other subfigures, had a remarkably lower overall f_0. In the case of f_0 maxima, it was significantly lower than all but the disaster (*dis*) and the diverse news (*div*).

On the other hand, fundamental frequency minima (Figure 3.7b) had mostly similar values across all categories. The only significant differences were the lower value in the sad (*sad*) than in the political, scientific and election related news.

In the f_0 median, there were large differences across the theme categories, all of them being different than at least one other theme. The highest values for can be seen again in the elections (*ele*), science (*sci*) and sports (*spo*) categories, while the lowest medians can be seen mainly in the emotional news (*sad*) and in the diverse news (*div*). Figure 3.7c shows that the political news have lower medians than all the three highest median categories (*sci, spo, ele*) as well as with the overall lowest (*sad*).

Fundamental frequency baseline was also smaller in the emotional news in relation to all the other categories except the diverse news (*div*). The other significant differences observed for this variable were the higher baseline values for election, sports and science (*ele, spo, sci*) in relation to the second lower baseline values from the diverse news.

As expected standard deviation and interquartile semi-amplitude showed coherent results with one another. Both showed that f_0 was more disperse in sports and election news (*spo, ele*) than in political news (*pol*). In both variables sports news also had higher values than diverse news, and election news had higher values than diverse news (*div*). Standard deviation values were also significantly lower for the

emotional news (*sad*) in comparison to the more disperse f_0 values in sports and election news.

Figure 3.8 shows the multiple comparisons for the measures of the first derivatives of f_0. Comparison of the positive first derivatives of f_0 mainly showed an overall higher value in sports news (*spo*) and an overall lower value for emotional news (*sad*) in relation to the other levels. That is, the rate of f_0 rises was higher in sports news and lower in the *sad* news.

Positive first derivative values from sports news (*spo*) were significantly greater than those on disaster (*dis*), diverse (*div*), political (*pol*) and emotional news (*sad*). Economy news (*econ*) showed a greater rate of rises than politics news. Emotional news also had significantly smaller values than economy, science (*sci*), election (*ele*) and violence news (*viol*).

The negative first derivatives of f_0 showed a somewhat similar pattern regarding sports news, but emotional news were not different from news about violence, science and economy. Diverse news showed overall lower absolute values than science news and political news had lower absolute values than election news. Thus, diverse news and political news had less steep f_0 falls than science and election news, respectively.

Values of the standard deviation of the rate of positive f_0 change (rises) showed great variability, again, mainly in relation to sports (more variable) and emotional news (less variable). News related to economy also had more variable rates of positive change than those about politics. The political theme (*pol*) was also less variable in the positive f_0 rate of change than sports (*spo*) and more variable than the emotional news (*sad*).

Regarding the negative first derivatives of f_0, their standard deviation was more homogeneous than the standard deviation of the positive ones. Nevertheless, the standard deviation of the rate of negative f_0 change was also overall greater in sports (*spo*) and lower in emotional news (*sad*), meaning that the rate of f_0 falls were generally more variable in sports and less variable in emotional news.

Jitter results on Figure 3.9 mainly showed an overall high jitter in disaster and emotional news (*dis, sad*) in relation to the other themes. News about violence (*viol*) also had significantly higher jitter than sport news (*spo*).

The rate of f_0 peaks had only a few significant differences in relation to the news themes. Science news (*sci*) had significantly slower rate than disaster, diverse and political news (*dis, div, pol*). Disaster news also had a higher f_0 peak rate than election news (*ele*).

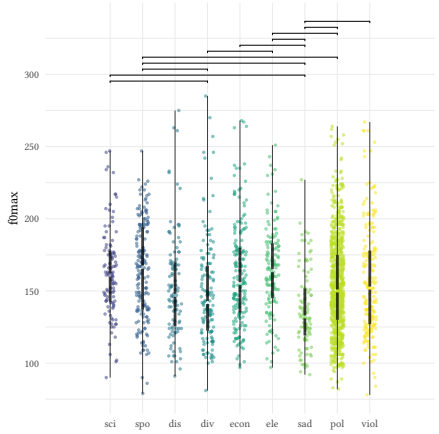
The long term average spectrum measures both showed a high amount of variability between groups, specially on the higher frequencies, presented on Figure 3.9d. Both cases showed higher absolute values for violence news (*viol*) in relation to most of the other themes. Election news (*ele*), on the other hand, displayed the opposite behaviour, with overall low absolute values in both frequency bands.

Voice quality and intensity measures are presented in Figure 3.10. Harmonics-to-noise ratio (Figure 3.10a) was higher in political news (*pol*) than on diverse news (*div*), which also had a lower value than sports news (*spo*). Emotional news (*sad*) had a ratio lower than all but diverse news.

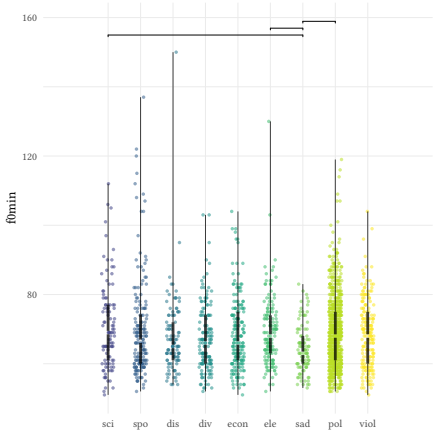
Regarding the coefficient of variation of intensity, on Figure 3.10d, the differences observed involved only election news (*ele*), which presented significantly higher coefficients than the ones from all themes but science (*sci*), sports (*spo*) and economy (*econ*).

Spectral emphasis also had a distinct behaviour on a single theme. In this case, the emphasis on violence news (*viol*) was significantly smaller than on news about politics, economy, sports, science and diverse news (*pol, econ, spo, sci, div*).

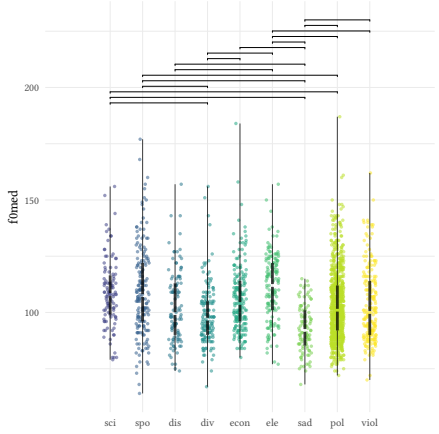
Soft phonation index had smaller values for the emotional news in relation to all other levels. News about violence (*viol*) also had a significantly higher index than political (*pol*), election (*ele*) and diverse news (*div*).



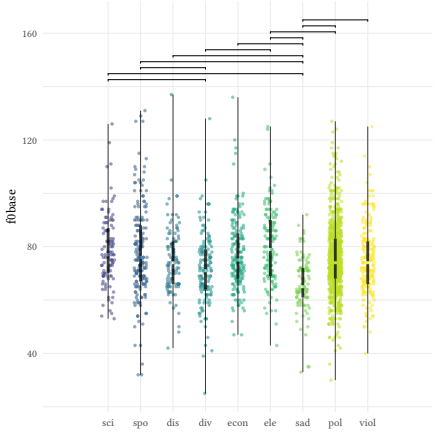
(a) Dunn's multiple comparison maximum f_0



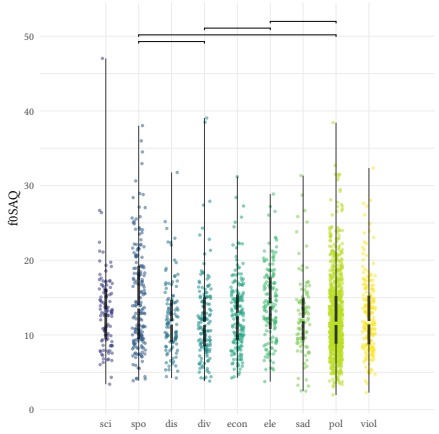
(b) Dunn's multiple comparison minimum f_0



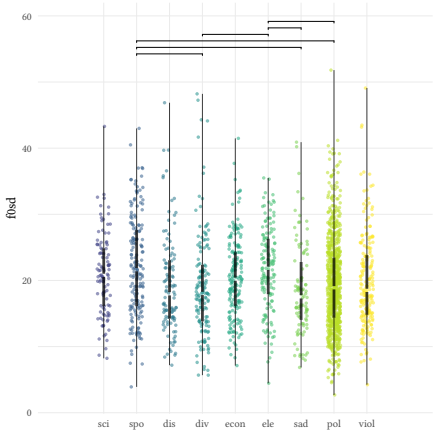
(c) Dunn's multiple comparison median f_0



(d) Dunn's multiple comparison f_0 baseline

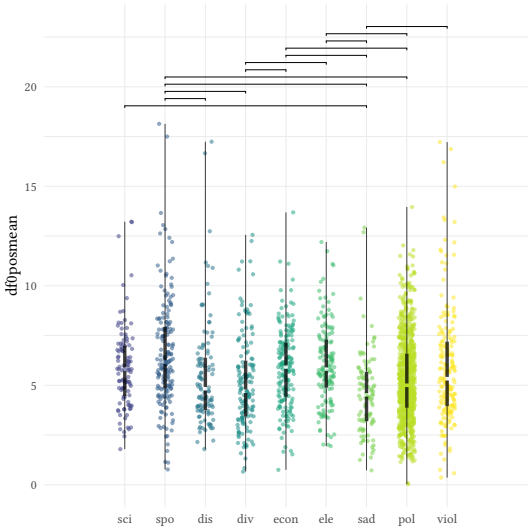


(e) Dunn's multiple comparison f_0 interquartile semi-amplitude

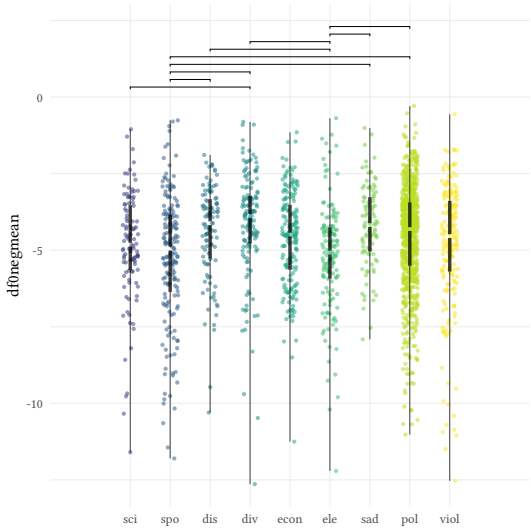


(f) Dunn's multiple comparison f_0 standard deviation

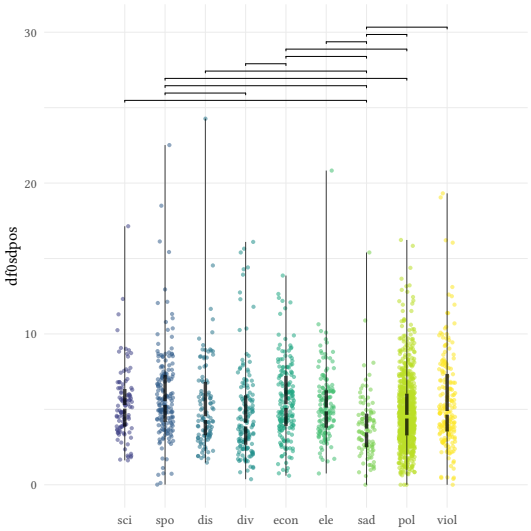
Figure 3.7: Dunn's multiple comparison of f_0 measures



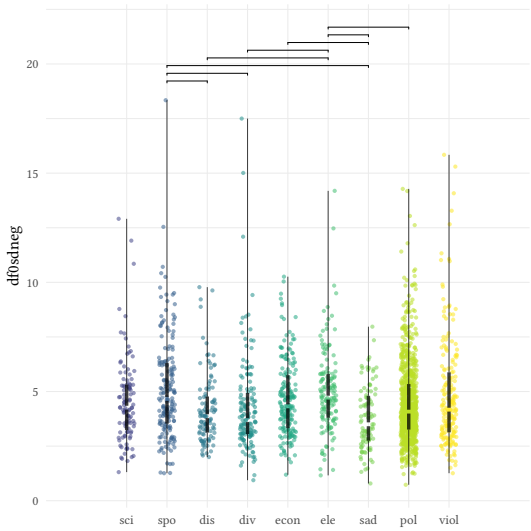
(a) Dunn's multiple comparison of positive 1st derivatives of f_0



(b) Dunn's multiple comparison of negative 1st derivatives of f_0

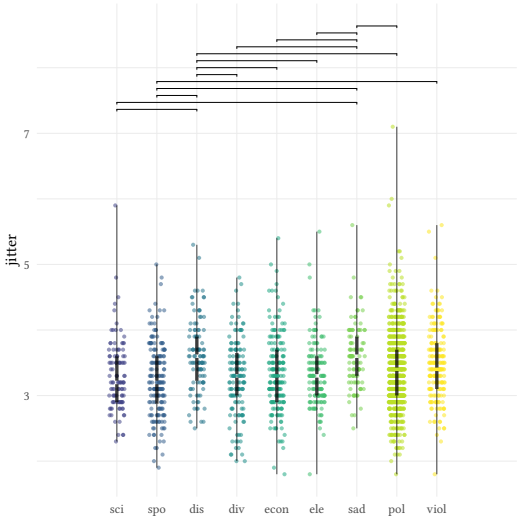


(c) Dunn's multiple comparison of the standard deviation of positive 1st derivatives of f_0

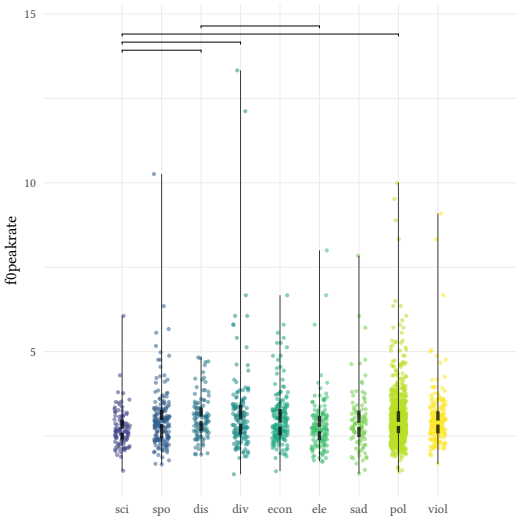


(d) Dunn's multiple comparison of the standard deviation of negative 1st derivatives of f_0

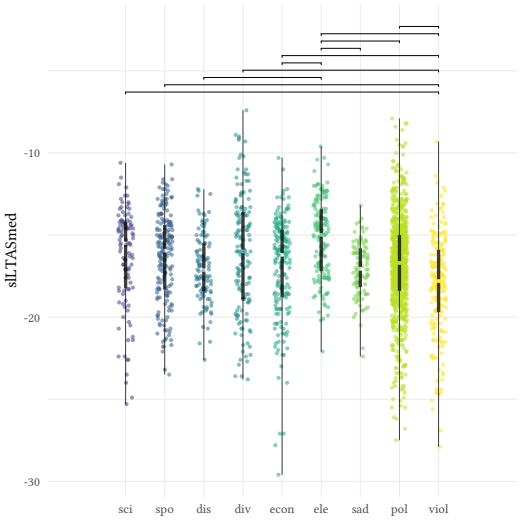
Figure 3.8: Dunn's multiple comparison of f_0 1st derivative measures



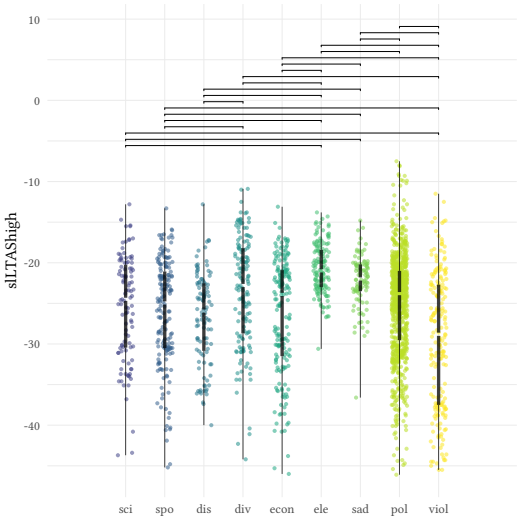
(a) Dunn's multiple comparison of local percent jitter



(b) Dunn's multiple comparison of f_0 peak rate

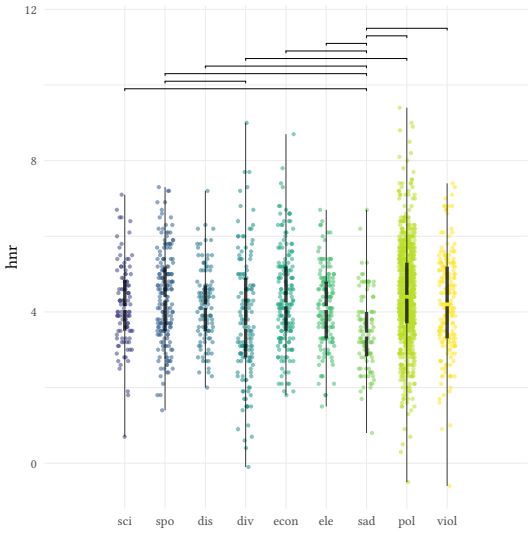


(c) Dunn's multiple comparison of slLTASmed

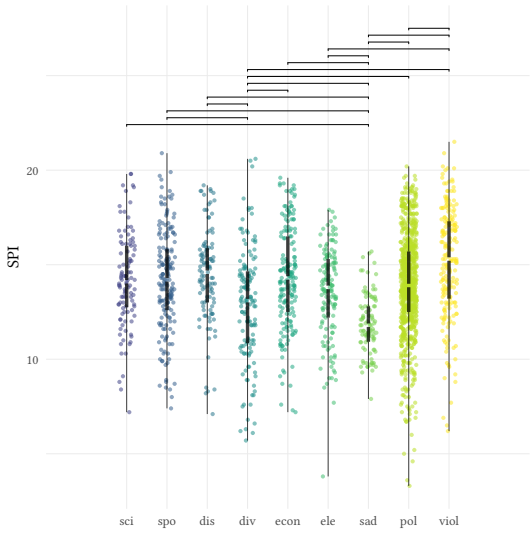


(d) Dunn's multiple comparison of slLTAShigh

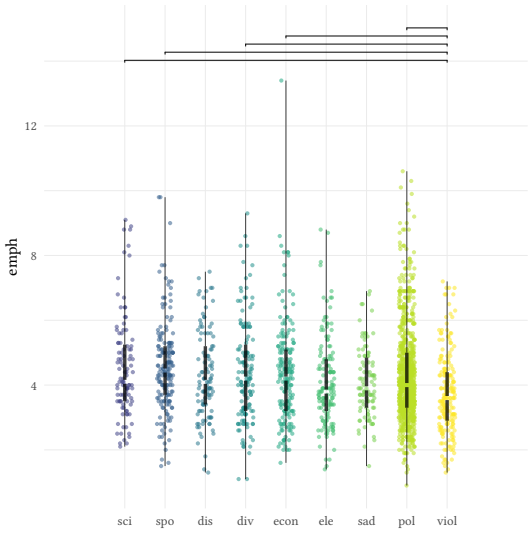
Figure 3.9: Dunn's multiple comparison of diverse measures



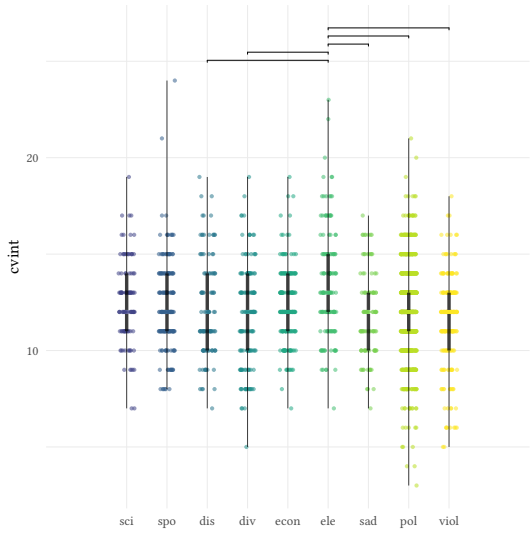
(a) Dunn's multiple comparison of harmonics-to-noise ratio



(b) Dunn's multiple comparison of soft phonation index



(c) Dunn's multiple comparison of spectral emphasis



(d) Dunn's multiple comparison of intensity variation coefficient

Figure 3.10: Dunn's multiple comparison of voice quality and intensity measures

3.3 Valence analyses

Given the results of the news themes analysis, it was hypothesised that the themes could be acting as a proxy variable for what a potentially more explanatory factor: the valence of the utterances. This valence was codified both manually and automatically, using Sentiment Analysis, in order to investigate its relationship with the acoustic parameters analysed in this study.

When comparing the data obtained with the manual classification with data of Sentiment Analysis, there was a stark difference in distribution. While the manual codification showed a more balanced distribution, with an overall higher number of negative news, the Sentiment Analysis shows opposite results with a clear bias towards positive classification.

Figure 3.11 shows a plot of the linear regression of manual valence. The vertical axis corresponds to the percentage of negative news in relation to the positive ones and the horizontal axis represents the broadcast date of the show in years. The black line represents a linear regression, with its shadow showing the correspondent 95% confidence interval, and the orange curve fits a LOESS regression, more sensitive to the local data behaviour.

Each of the blue dots represents a singular news edition, with larger dots and more opaque dots representing editions with a higher number of chunks. The vertical blue lines stemming from each dot represent the corresponding confidence interval based on the amount of observations. This figure shows that the proportion of negative-to-positive news was highly variable throughout the sampled programs.

Linear regressions for each of acoustic variable by each set of classification showed no correlation between the acoustic parameters and the automatic classification. Therefore, the Sentiment Analysis results were deemed unfit for the classification of the newscaster's utterances. The manual classification, on the other hand, showed some significant, yet not very strong, correlation with some acoustic parameters. The Table 3.3 show the F statistic, adjusted R^2 and significance of the linear

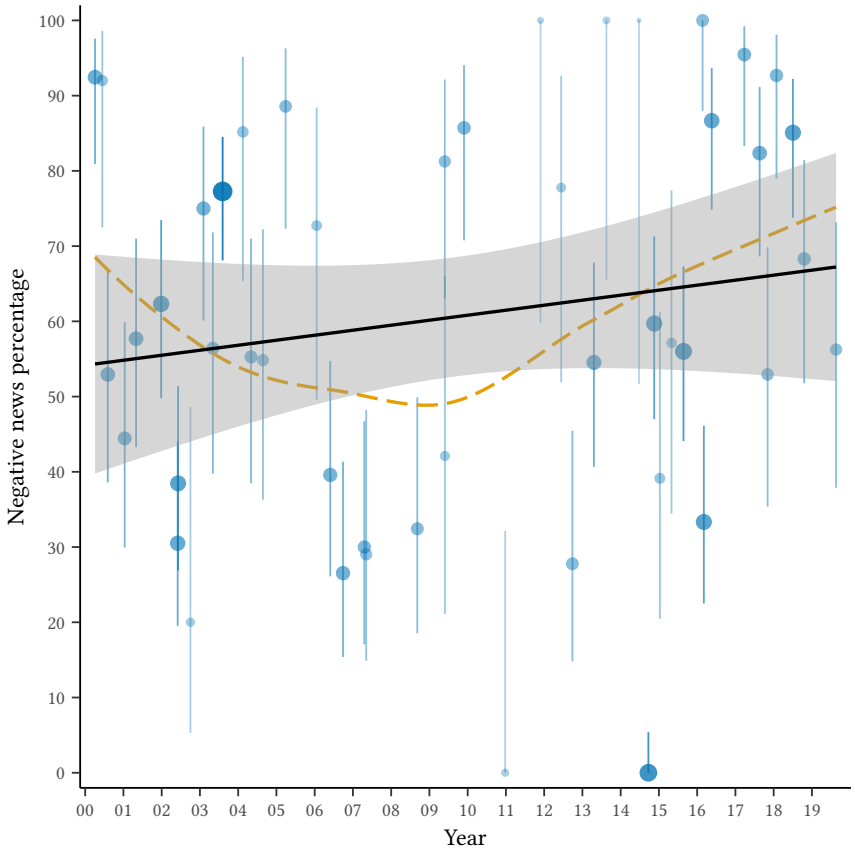


Figure 3.11: Regression of manually categorised valence by year

models corresponding to the manual valence analyses.

Checking the results of the linear regression, we can see that the highest coefficient of determination obtained was .03, that is, the manually-coded valence explain only 3% of the observed variation. These results suggest that the classification of the contents' valence are not strongly correlated to the observed variation in the analysed acoustic parameters.

Variable	F	$\bar{R}^2$	p
f0med	60.6	.03	< .01
f0base	38.0	.02	< .01
df0posmean	22.2	.01	< .01
cvint	19.6	.01	< .01
f0max	18.6	.01	< .01
f0SAQ	17.5	.01	< .01
df0negmean	17.0	.01	< .01
df0sdneg	16.4	.01	< .01
f0sd	15.2	.01	< .01
jitter	12.7	.01	< .01
f0min	8.34	< .01	< .01
sdf0peak	2.68	< .01	.1
f0peakwidth	0.95	< .01	.33
f0peakrate	0.03	< .01	.87
df0sdpos	4.33	< .01	.04
emph	0.06	< .01	.81
slLTASmed	10.7	< .01	< .01
slLTAShigh	8.39	< .01	< .01
hnr	3.53	< .01	.06
SPI	0.17	< .01	.68
shimmer	5.68	< .01	.02

Table 3.3: Linear regressions for valence codification

CHAPTER 4

DISCUSSION

In Chapter 3, I presented data resulting from the statistical analyses of a newscaster's speech. In this chapter I seek to provide interpretations of the main findings previously described.

Many of the analysed parameters have shown significant results regarding longitudinal change. Most of them were on parameters associated with intonation. In this regard, the median of f_0 is an important measure for describing one's speech, as it can be more robust than the mean when dealing with the characteristic positive skewness of f_0 samples (Arantes, 2014; Jassem, 1975).

A. D. Silva (2022) extracted reference f_0 values from studio recordings of 100 male BP speakers from São Paulo, between 18 and 62 years old. He saw that the mean f_0 value for the samples population was around 133 Hz. Comparing with the results from the present study, we can consider that the average median f_0 for the newscaster, about 105 Hz, is relatively low compared to most other BP speakers. The same applies for the f_0 baseline, with an average of 76 Hz in the newscaster's speech versus 95 Hz on the reference data. Standard deviation was also relatively smaller for the newscaster at around 20 Hz while the reference data was about 27 Hz.

Experimenting with the perception of vocal traits using synthesised voices, Brown, Strong and Rencher (1974) saw that an increased mean f_0 were correlated with a decrease in the ratings of "benevolence" and "competence" based on the opinion of 38 English-speaking judges. It is important to note, however, that some associations between prosody and evaluation may be language-related. Analysing cross-linguistic perception of prosody, Biadsy et al. (2008) saw that American, Palestinian and Swedish speakers associated a greater f_0 range with charisma in American English, but Palestinian judges associated the same feature

with lower charisma while rating stimuli in Arabic.

Moreover, linguistic and cultural differences may impact the association between prosodic behaviour and meaning on newscasting. While intonational patterns may be language dependent, its social conditioning is not restricted by the language/dialect domain. Distinctive intonational features can be linked to different communities of practice within the dominion of what can be recognised as a single language (Dolson, 1994).

Overall, there have been mixed results regarding the impact of newscasting on speech production and evaluation. In regard to Brazilian Portuguese, Panico and Fukushima (2003, apud Portinha, 2011) argue that it is of the interest of newscasters to avoid excessive excursions in frequency and intensity. This matches the findings reported by Portinha (2011), who suggests that this happens in order not to compromise the intended emphasis on specific parts of the news. Strangert (1991) reports a similar concern about the effective communication of news by Swedish newsreaders.

These findings seem to be in order with Gasser et al. (2019), who comment on these professionals' desire to avoid affectation. In fact, a growing aspiration of sounding more colloquial is an explicit part of the recent changes of news programs in Brazil (Catani, 2022; Portinha, 2011).

Gama (2003, apud Portinha, 2011) argues that Brazilian newscasters usually speak using lower fundamental frequencies while presenting news in comparison to their everyday speech. Analysing two Brazilian talk show hosts, Antunes (2017) observed that the hosts adopt a more contained melody when trying to convey credibility, speaking with a lower mean f_0 and with lesser f_0 maxima in relation to other moments in the show, where they use higher frequencies to portray humorous and polemic topics, circumstances that are related to a higher affective engagement. In the same direction, Campos (2012) saw that Brazilian radio announcers with older target audiences tend to speak with lower and less variable f_0 than those focusing on younger audiences (Campos,

2012).

Although some of the changes on f_0 measures seen on the present study may seem subtle, such as the change in median f_0, they shouldn't be underestimated. D'Errico et al. (2013) found relevant effect of changes as small as one semitone in listeners' ratings of the voices of French and Italian politicians. Furthermore, we may assume that the one semitone drop in the voice of the newscaster in the present study can potentially alter how the anchors voice is socially perceived. A similar assumption can be made about the shortening of fundamental frequency range and the smaller dispersion values of f_0 seen on the anchor's speech.

Although some studies have linked lower f_0 to lesser charisma, others such as Puts et al. (2007) and Fraccaro et al. (2013) link it to the perception of more dominance, which, in turn, may be related with the perception of more authority. Portinha (2011) observes that the first Brazilian newscasters came from radio newscasting where a lower pitch was common. In her study about radio announcers, Campos (2012) found that these speakers had an overall lower f_0 when talking in a more professional narration register as opposed to the use of a higher fundamental frequency when announcing in a more colloquial manner.

Regardless of recent changes towards more informal speech in Brazilian telejournalism, this low pitch tendency may still have some influence on more traditional TV newscasting. Based solely on the comparison with the general Brazilian normative data by A. D. Silva (2022), we could be compelled to think that the newscaster fundamental frequency measures from the present study could be seen as generally atypical. Nevertheless, not only can the speaker's age be responsible for his voice, but the specific setting in which his voice was recorded may also play a decisive role in his observed vocal behaviour.

Looking at the data from the news themes analysis we see that the topic of the presented news exhibits significant correlation with the analysed intonational parameters, such as median, maximum, baseline and standard deviation of fundamental frequency. Themes that are associated with higher excitement levels, such as sports and election, usually

had higher values of f_0 in these parameters.

Comparing BP football-news announcing with a more traditional narration, Campos (2012) also observed overall higher mean and standard deviation of f_0 on the sports news. These results are in agreement with the notion that emotions with higher arousal levels are typically associated with overall higher f_0 (Ayadi, Kamel & Karray, 2011; Murray & Arnott, 1993; Nwe, Foo & Silva, 2003; Palo, Mohanty & Chandra, 2017; Scherer, Johnstone & Klasmeyer, 2003).

In general, the f_0 in these news themes in the present study are also more disperse than in the more frequent political news and the less enthusiastic disaster and emotional news. These results also match previous findings associated with emotions and the relationship of their arousal level with intonation, such as Fonagy (1978), Murray and Arnott (1993), Nwe, Foo and Silva (2003) and Scherer, Johnstone and Klasmeyer (2003).

The minimum f_0 values found on the newscaster voice, however, did not change overtime and only showed to be slightly decreased on the emotional news. In this case the minima may have remained stable due to physiological limitations of basal phonation, given that this lower f_0 limit is considerably closer to the speaker's modal phonation f_0 than it is from the f_0 upper limit. In this sense, the reduction on the f_0 maxima can be related to the avoidance of higher frequencies due to a potential decrease in reliability of the finer motor control of the tract occasioned by age.

The decrease in range, mainly in the top frequencies, can be associated with loss of muscular elasticity and calcification or cartilages in the vocal tract (Kaplan, 1960). Xue and Hao (2003) reported an increase, with age, in the length of the oral cavity and in the volume of both the oral cavity and the vocal tract. Changes in breathing related with increasing age may also alter the range of speakers' phonatory frequency and their performance associated with maximum and minimum f_0 (Reich et al., 1990; Sataloff, Kost & Linville, 2017).

Another possible interpretation for the apparent stable f_0 minima is

that his basal f_0 could be changing slowly enough not to yield significant results, but can be gradually rising, consequently shortening the f_0 range. This trend of change with lower maximum f_0 and higher minimum f_0 has been previously observed among speakers younger than 40 years old and older than 65 (Ptacek et al., 1966).

Previous results regarding standard deviation of f_0 and age are also somewhat mixed. For example, while analysing 80 voices distributed in three age groups – 20–39, 40–65, and 66–75 years old –, Lortie et al. (2015) found that the f_0 standard deviation got progressively higher with age. On the other hand, a longitudinal study of the speech of 10 male subjects by Fouquet et al. (2016) observed no significant change in f_0 mean, standard deviation and f_0 minima on recordings produced by subjects of a similar age range to the newscaster in the present study. In this range, only f_0 maxima at the ages 49 and 56 were found to be overall higher than the corresponding 35-year-old voices.

In the derivative measures, that correspond to the rate of change of f_0, the overall reduction of the measures point to a decrease in the overall perceived vivacity. The pitch contour, therefore, can sound less dynamic, or more monotonous. This perception can be potentially heightened by the decrease in the standard deviation of both the positive and negative first derivatives – measures associated to the dispersion of the rate of change in melody. Such increase in uniformity of the rate of f_0 rises and falls could lead to a lower perceived emotional affectation, that is, the perception of a less emotive speech. It could also lead to the sensation of a more predictable melodic contour, given that the values are more uniform.

Both of these results are in order with a general social association of ageing with maturing. In line with the social expectations directed at them, an older person may want to sound more mature (Mirowsky & Ross, 2010). More mature subjects, on their turn, can be expected to behave in a more reserved manner, showing restraint as a sign of wisdom acquired with time.

These values are seemingly consonant with social expectations re-

garding news anchors: a professional, restrained and objective character. From this bearer of information and public influence, a high level of maturity is expected.

It is valid to note that conveying his experience through speech is certainly desirable for the newscaster under the analysed circumstances. These social expectations related with maturity, thus, are potentially convergent with the subjective evaluation that the anchor makes about how he should speak, aligning with a pursuit of sounding credible and respectable. It is not clear, however, to what extent these alterations in vocal behaviour are intentional, consciously perceived on some degree, or generally imperceptible to the speaker in question.

From a physiological standpoint, all of these changes and tendencies of change could be attributed to the stiffening of the vocal folds. Considering the age of the subject, however, it is unlikely that this is the case. It is usually not until the 60's that this phenomenon starts to impact voice substantially (Leeuw & Mahieu, 2004). Although some of these changes can be related to vocal degradation, in the case of professional speakers, this process can be lessened thanks to vocal training and more conscious usage of vocal resources.

The notable exceptions to the intonational changes in the longitudinal axis were the results regarding change in the long term average spectrum (LTAS) slopes on high and medium frequencies and in the harmonic-to-noise ratio. The slope of LTAS in high frequencies was by far the most correlated with time. It showed an $\bar{R}^2$ of 16%, while the "runner-up" variables had an $\bar{R}^2$ of about 6%. On medium frequencies, the slope of LTAS had $\bar{R}^2$ of 5%. Both of these parameters showed an increase, which are associated with higher perceived vocal effort.

Although differences in the audio quality could be associated with the observed long-term spectral characteristics, the most clear direction that it could impact the results would be observing a lower value in the older data, as we could expect greater audio quality, that is, less noise, in more recent recordings. Accordingly, the data on harmonics-to-noise ratio confirm this increase over time. This suggests that the observed

LTAS results are not due to differences in audio quality.

In the multiple comparison tests, both of the LTAS parameters had an overall high absolute value in news about violence in comparison with most other news, corresponding to the perception of lower vocal effort in the violence topic. This result may be related with a specific stance when reporting this type of news, that can be seen as the most delicate subject in the program. Deliberate or not, this behaviour can be the result of attempts to avoid shocking the viewer, in contrast with other over-emphatic news programs, which revolve around the exploitation of violent crimes and generally appeal to smaller audiences.

The smaller values for spectral emphasis are also found on violence news. A low spectral emphasis is also perceptually associated with lower perceived vocal effort, which can further add to the more self-contained posture in this particular news theme. In this parameter, violence was the only theme that showed significant differences in relation to others. Over the sampled time, spectral emphasis remained stable and didn't seem to relate with the vocal ageing during the middle age.

The coefficient of variation of intensity in the news about violence was only smaller than the one in election news – that had a generally higher coefficient than the others. These faint differences between the themes, as well as the subtle change over time, can be a particular characteristic of the newscasting genre and the involved demonstration of seriousness and self-control. In the analysed news program, the newscaster usually does not make radical alterations in intensity. The incursion in such behaviour could even be problematic from a logistic standpoint, as recording sounds with highly variable intensity might cause the less intense segments to be de-emphasised and less clear for the viewers.

Besides serving as a general indicator of audio quality, the calculation of the harmonics-to-noise ratio (HNR) applied to speech can be used to evaluate voice quality. As different configurations of the vocal tract imply in different harmonic amplitudes, the ratio between harmonics and noise in speech varies dynamically (Fernandes et al., 2018).

Moreover, conditions that disturb the regularity of the vocal folds' movement can affect the HNR on a given voice, as they may relate to increased turbulence and generate noise (Fernandes et al., 2018; Ferrand, 2002). These irregular vibrations can be generated by incomplete glottal closure, which can be associated with laryngeal pathologies and degeneration (Ramig, 1983; Yanagihara, 1967).

The HNR is perceptually correlated with hoarseness, roughness, and breathiness (Ferrand, 2002; Linville, 2001; Martin, Fitch & Wolfe, 1995). Linville (2001) suggests that the average perception of elder voice as breathy and hoarse indicates a connection between spectral noise and age. In this direction, Martin, Fitch and Wolfe (1995) show correlation of HNR with the classification of severe breathiness, roughness, and hoarseness in voice.

These notions are reinforced by Ramig's (1986) findings, in which spectral noise in speech was correlated to the perception of older age. A similar assumption regarding hoarseness is made by Mueller (1997), who further argues that as lower HNR is common in elder voices, it may be assumed as a typical characteristic of these voices. Decoster and Debruyne (1995, apud Ferrand, 2002) observed decreased harmonics-to-noise ratio in subjects of 60 years of age and beyond, suggesting that the increased noise was not necessarily due to abnormal ageing.

Ramig (1983) compared levels of spectral noise in the speech of subjects with different health conditions, assessed via heart rate, blood pressure, percent fat and vital capacity. Her results show correlation between spectral noise and physiological condition but not with chronological age, although healthy younger speakers appear to have lower spectral noise than healthy older speakers. The author suggests that as HNR is related to diverse alterations in the physiology of the larynx, affected by general changes in the body, chronological age alone may not suffice to explain a change in this parameter.

Nevertheless, while comparing multiple acoustic parameters in the voice of differently aged speakers, Xue and Deliyski (2001) observed significantly higher noise levels for elderly speakers in comparison with

younger ones. Somewhat similar results were found by Müller (2005), who observed significantly less noise in the voice of children (10–12 years old) and teenagers (13–19 years old) than in younger (20–64 years old) and older adults (more than 65 years old). The particular age grouping in this result, however, might be due to the age range of the used categories, as the young adults category ("Erwachsene") contained data ranging from the beginning of adulthood until the beginning of senescence. Regardless of that, the findings presented by the author still point in the direction of an increased presence of noise with age.

Schötz (2006) obtained different results while investigating acoustic correlates of speakers' age. She observed increasingly higher levels of noise in voices of 20 to 50-year-old males, but a decrease in noise of similar magnitude among those between 50 and 90 years old. In face of this finding, Schötz (2006, p. 123) recognises that this pattern is hard to explain. However, Olmedo's (2022) comparison of the speech of two voice professionals from middle to older age showed a trend coherent with Schötz's (2006) results: a decrease in noise after 50 years old. Regardless of the decrease in noise levels, different judges rated one of the speakers as significantly more breathy in his last recording, despite the increased HNR.

There are also studies showing stability in noise from young adulthood to older age. While analysing data obtained from BP speakers between 20 and 60 years old, Spazzapan et al. (2018) did not find any age-related difference in the noise levels for males.

The HNR data from the present research appears to follow the second presented trend where recordings close to the present have higher HNR and thus lower proportional noise in relation to harmonic sounds. This may be associated with a decrease in the noise in his speech but also with an increase in recording quality. It is not uncommon to witness changes in the audio recordings throughout the years while working with longitudinal data. The course of the development of audio technologies allows us to be optimistic enough to assume most of these changes tend to be for the better, with increasingly capable microphones, recording devices,

audio codecs and such. Nevertheless, the comparison of data spaced too far apart may be subject to the impact of these substantial acoustic differences.

All things considered, it is possible that both aspects had an effect on these results: there may have been a decrease in the noise on the subject's voice and also an increase in recording quality – resulting in overall less noise in the audio. It could also be the case that a strong increase in quality could mask a decrease in HNR on the voice of the subject, but given that the newscaster is not chronologically elder in the sampled period and that he is seemingly healthy, with a trained voice, it is unlikely that he suffers from premature larynx degradation that could impact his HNR.

Considering the HNR results from the multiple comparison tests, the main observed difference was between the emotional news and the others. In this case, while delivering emotional news, the speaker had a generally lower HNR. Nevertheless, the content valence data didn't present correlation with HNR. Since the emotional news were comprised of a specific negative type of emotion (sadness), it is probable that not all negative emotions associated with observations grouped under the negative valence have the same impact on HNR as sadness does.

Nunes, Coimbra and Teixeira (2010) analysed the voice of an actor performing the same sentence with different emotions in European Portuguese. In the study sadness had a slight but not significantly lower HNR than the neutral speech. The only significant difference regarding sadness was in relation to joyful speech, that had a higher HNR than sadness. This finding is somewhat similar to the present results, given that sports news, that can be considered overall joyous, had one of the highest median HNR.

The categories with lower HNR in Nunes, Coimbra and Teixeira (2010) were "anger" and "despair". In what concerns the present research, these are not usual emotions displayed on the news. In fact, most of the data in the newscasting corpus don't have such strong and polarized emotional engagement, which may explain the seemingly low

relevance of the valence in the results, differently from the themes. The usage of polar negative/positive classification can end up diluting the infrequent cases of speech with stronger emotional activation, mixing it with more frequent occurrences of news with dimmed emotional states such as what could be classified as "mild discontentment". Moreover, accounting for types of emotions and their activation level may provide more insightful results. Further research is needed, therefore, to better understand how HNR is affected by different emotions and also how different speech settings may affect this emotional display.

Although shimmer and coefficient of variation of intensity showed significant results, the $\bar{R}^2$ of longitudinal changes indicates that time explains less than 1% of the variation in these parameters. Due to the marginal significance value and this weak correlation with time, both of those parameters were considered to be overall stable. Regardless of the variation observed in each news program, shimmer was not correlated with either the news themes or the valence of the utterances.

According to Orlikoff (1990) and Ramig and Ringel (1983), shimmer increase is related with both age and health state of the speakers. Following Lieberman (1963), who indicated the connection between laryngeal diseases and fundamental frequency perturbations, Koike (1969) compared the amplitude signals of healthy speakers with those who had larynx pathologies. The author observed alterations in the fundamental frequency amplitude that were seemingly related to laryngeal neoplasms. Subsequently, Kitajima and Gould (1976) also found that shimmer was generally higher in the voices of subjects with vocal folds polyps. More recent work have similarly observed higher shimmer in dysphonic adults between 21 and 50 years old in relation to their normophonic counterparts (Hippargekar et al., 2021).

Regarding age-related changes, Spazzapan et al. (2018) analysed shimmer in male speakers from 20 to 60 years old. They found differences between the 30-to-40-year-old group and both the 40-to-50-years-old and 50-to-60-years-old groups, with higher values for older speakers. Xue and Deliyski (2001) found that percentage of shimmer

in males at around 75 was twice the amount present in the voice of younger speakers. Dehqan et al. (2012) similarly observed higher shimmer in the speech of elders (70–90 years old) in comparison to younger adults (20–49 years old). Goy et al. (2013) found that men around 72 years old had higher shimmer percentage than those around 19. Gorris et al. (2020) also saw increased shimmer in the older subjects (61–70 years old) while comparing males from 18 to 70 years old.

As these results indicate, most of the observed changes in shimmer in adults' speech seem to occur at around 65 years of age. Although some change can be observed before this period, it seems to be more subtle and gradual than those that appear to be prevalent on elderly ages. Even though researchers try to screen older patients for illness that could potentially bias the speech performance of the subjects, it may be the case that underlying health conditions, associated with normal and abnormal ageing, may have an impact on the results beyond the sole degradation of the larynx.

Most of the previous considerations are also valid for jitter, that remained stable through the sample years. Like shimmer, jitter is an f_0 perturbation measure, but in the frequency axis. While shimmer is related to changes in glottal resistance, jitter relates with a diminished control of the vibration of the vocal chords (Teixeira, Oliveira & Lopes, 2013). Nevertheless, they are both similarly subject to changes in the tract associated with age and both normal and abnormal vocal tract degradation.

Wilcox and Horii (1980), for example, found that average jitter was significantly higher in speakers around 70 years old than in younger ones, of about 23 years of age. On the other hand, the authors found that the values for the older speaker were smaller than those previously reported for speakers with laryngeal pathologies. Benjamin (1981) saw a supporting trend between speakers of about the same age. Orlikoff (1990) and Ramig and Ringel (1983) also observed the impact of the health of the subjects on jitter measures.

There is also a reported relationship between these perturbations

with loudness/phonation level. Huang et al. (1995) and Brockmann et al. (2008) found that as sound pressure level decreases, that is, on speech with diminished phonation, jitter and shimmer tend to increase. Deliyski, Shaw and Evans (2005) and Huang et al. (1995) also found that environmental noise and recording quality can impact on these and other vocal quality measures.

Specifically in relation to pathologies, it appears that the effects of medical problems are more distinctive in more serious health conditions. A similar reasoning can be read in Lieberman (1963), who suggests that small nodules and inflammations on the tract have comparatively little effect on the voice waveform. Moreover, potential health fluctuations may be suppressed by longitudinal samples from the same subject.

Differently from shimmer, jitter showed a significant correlation with the news themes. In this case, the highest observed values were in the themes associated with lower arousal negative emotions: disaster and emotional news. These news had an overall higher jitter than most other themes. In the same direction, news about violence weren't as jittery, yet the theme had significantly more jitter than the more excited sports news.

These results can be related with the previous discussion about the newscaster's behaviour in the violence news. These are more sensitive subjects than the much more common political theme. A higher jitter in this case might be related both with the expression of negative emotion, with a certain decrease in vocal effort, and with laryngealised speech. This behaviour seems in order with the expectancy of seriousness in these themes in contrast with more joyous ones, like sports. Regardless of that, unlike in violence news, there was no significant difference in spectral emphasis related to disaster and sad news. This indicates that there are fine and somewhat specific vocal behaviours associated with the presenting of different news themes, beyond the simple positive/negative axis.

Soft phonation index, that corresponds to the perception of breathiness, also did not show difference over time. The result can indicate that

the longitudinal differences reported in the literature may begin only at a more advanced age, given that the obtained index was in a similar range to the threshold for young and middle-age subjects (about 14 dB) (Deliyski and Gress, 1998, apud Xue and Fucci, 2000).

Nevertheless, this measure presented differences related to the news themes. The greatest disparity was between emotional news (lowest mean index) and violent news (highest mean index). This result is consonant with the above interpretation regarding specific vocal behaviours for specific themes and with the notion that the arousal level of emotion can be relevant in such cases.

In conjunction with the changes in the newscaster's speech based on the type of news, some of the parameters that did not change over time can also be related with the newscasting genre. This interpretation is applicable to both the intensity and the melodic peak parameters. f_0 peak width was also not significantly affected by news themes or valence. Results for f_0 peak rate regarding the themes showed only a slightly lower peak rate on science news (2.7 peaks/s) followed by election news (2.8 peaks/s), while the global average was about 3 peaks/s.

In their longitudinal study of prosody, Valente et al. (2021) observed no age-related difference in both of these parameters. In general, these peak parameters are probably not as related to age in the same sense that other f_0 measures are, since these parameters are more closely associated with durational aspects of speech than with changes on fundamental frequency and the corresponding perception of pitch.

The only f_0 peak parameter that changed over time was the standard deviation of f_0 peaks. This parameter shows the dispersion of the f_0 peak heights. While news theme did not show influence over this measure, there has been a decrease in variability of the peaks with time. This can be related with the lowering of f_0 maxima and overall f_0 range, that could reduce the disparity between the peak heights. This result goes in the opposite direction of the changes observed by Valente et al. (2021), who observed a prominent increase in dispersion between ages 51 and 74. One possible explanation is that the physiological processes that

generate the increase in this parameter could occur only in later ages and may not have started on the newscaster's voice. Another possibility is that the themes explored by the EP speakers required more vivacity across emphasised lexical items.

Given that much of the variability on the analysed parameters were not explained by either time or the news themes, the valence codification was done considering it as a proxy variable to the emotional content of the news. However, the results obtained from the linear regressions for this variable showed an overall weaker correlation than with the two other predictors.

The news segments were initially manually classified into three discrete categories: positive, neutral and negative. Seeking a replicable method of classification, Sentiment Analysis seemed like a valid tool, potentially scalable for working with larger datasets, where there would be a larger cost for manual codification.

On the manual codification, the number of observations for each category indicated a strong bias against positive news, with less than 15% of the data in this category and more than 58% on the negative category. For sentiment analysis, three different lexicons were used, but none of the results were accurate, showing a somewhat opposite distribution compared to the manual data.

Proceeding only with the manual codification data, both the neutral and positive-rated news were merged into a new and more balanced positive category (with about 41% of the data). This was done in order to allow for stronger generalisations, considering that news that were not explicitly negative in the program could be reckoned as positive, or at least not-negative.

The stronger correlation with valence was seen in f_0 median, followed by the f_0 baseline. Overall, higher f_0 was associated with more positive news. The median f_0 result can be related with the aforementioned level of activation of the emotions involved in the program.

A more interesting result is the one for the baseline of f_0. The idea behind this parameter is to be a robust descriptor of speakers' voices,

regardless of changes in their vocal effort, audio quality, speaking style, and emotional expression (Lindh & Eriksson, 2007; Traunmüller & Eriksson, 1995). Accordingly, this parameter did not show longitudinal change in the sampled time, regardless of changes in the other common descriptor, median f_0.

Nevertheless, the results associated with the news themes showed that the baseline was significantly lower in the sad news. Coupled with the results for valence, this result can indicate a sensitivity of this parameter to negative emotions. In this sense, f_0 minima appeared to be an overall more stable measure, since it did not change with age, and was only marginally affected by news themes and valence. More investigation is needed in order to verify if these descriptors would still show similar results on the vernacular speech of the newscaster or under different types of emotions with different arousal levels.

Considering that emotional expressiveness is responsible for a reasonable amount of variation in the corpus, the codified valence might not have been fully successful in capturing this type of information. Banse and Scherer (1996) suggest that the perception of emotions in speech may be related not only to valence but also to their activation level.

One apparent source of the weak correlation observed was the method used for classification, based on the content and not on the perception of the contextualized voice of the speaker. While abstracting from the vocal dimension could facilitate codification by isolating the content from the other meaning-bearing dimensions of speech, a classification based solely on content can be incomplete without some potentially crucial information from other dimensions.

For instance, a relevant axis for television newscasting is the visual information given by body and face expressions. The sole fact that there is an investment by the broadcaster to create a visual scenario for the show, with journalists in specific types of outfits and scenic behaviour, confirms the relevance of this resource. In fact, even the camera position in relation to the eye-level has been shown to alter the evaluation of the speaker by the public (Kepplinger & Donsbach, 1990). Consequently, the

perception of the valence may be severely altered due to a high number of factors associated not only with voice itself, but with the evaluation of the speaker, the situation and so on.

On the other hand, considering information from the prosodic domain to check for correlation with the prosodic variables themselves could be seen as a flawed approach, specially with a single non-naive judge – that is, the researcher. Given that meaning association is socially dependent, the classifier is also a source of complexity that needs to be accounted for, specially on manual classification. Thus, even though the manual coding showed some correlation with the variation in the data, a more robust approach, as said earlier, would require multiple judges to rate each observation.

In the case of automatic coding of content, the usage of more sophisticated Sentiment Analysis could certainly generate more meaningful results or results closer to the ones that would be obtained via manual coding. Syntactic information handling and contextualization might be optimal solutions, albeit costly ones. A simpler step might be the development of specialized lexicons accounting for genre particularities.

Regarding the overall classification results, the program bias towards negative news might be associated with ideals of what would constitute news in the target audience community. Perhaps the seriousness associated with traditional television news can be related to the seriousness of the news given. There could be also an implicit association between negative news and importance. Although there are news shows with clearly more positive or uplifting content in Brazilian television, they are mostly transmitted during daytime TV and not on prime time.

While it would be hard to question the pertinence of news about nation-wide political scandals in prime time programming – typical enough to generate unbalance in the categories from this study – the presentation of less dramatic topics might be seen as a waste of valuable time and a lost of seriousness of the show. In this scenario, less seriousness equals to less credibility, which in turn might lead to lower audience ratings and lower advertisement earnings.

CHAPTER 5

CONCLUSION

This research focuses on the speech of a single middle-aged newscaster in the course of 19 years. Prosodic parameters were extracted from 50 audio recordings from this period and were qualitatively and quantitatively analysed regarding their overall stability over time and their relation to the topic and valence of the news. The extracted parameters relate to different aspects of speech prosody: intonation, voice quality and intensity. Data were statistically analysed, mainly with the usage of Linear Regressions, Kruskal-Wallis and Dunn's multiple comparison tests.

Even in a highly controlled environment with a highly trained speaker, it was possible to observe language variation and change within a speaker's voice. The results show that gradual intonational changes can be observed throughout middle age. Although not drastic, some of these changes can be observed before senescence and can be potentially more pronounced in the voice of speakers without vocal training.

In general, the newscaster's speech at an older age was associated with less dynamic intonational patterns even in the midst of a very expressive genre with much variability, as seen in measures of central tendency, measures of dispersion, and measures associated with the rate of change of intonation. Other prosodic factors associated with increased age, such as perturbations in amplitude and frequency of f_0 may not be present on the voice of normophonic speakers of this age or, at least, may be unobservable in chained speech.

The following answers were found for the questions initially posed for this research:

- *How variable is the prosody of a newscasting anchor while presenting*

the news?

The prosody of the newscaster is quite variable. Both through the years and inside each edition, all the parameters show a high dispersion, which correspond to different vocal behaviour in different parts of the show. This finding further confirms that there is intraspeaker variation in the prosody even on a highly restricted context and in a professional and trained speaker who is regarded as a reference of standard language.

- *Did the newscaster's prosody change throughout the course of 19 years? If so, which acoustic parameters changed? How did they change? Which parameters remained stable?*

  Many changes occurred in the voice of the newscaster in the course of the sampled 19 years. Although not all were substantial, parameters related to intonation and voice quality showed significant change over time. The parameters that changed were: the median, maximum, standard deviation and interquartile semi-amplitude of f_0; first derivatives of f_0 and their standard deviation; standard deviation of f_0 peaks; harmonics-to-noise ratio; slope of LTAS on high and medium frequencies. Shimmer and coefficient of variation of intensity were only marginally significant and thus considered stable, but further investigation regarding the typicality of these results is needed. The parameters that did not have significant change were: jitter; spectral emphasis; Soft Phonation Index; f_0 peak rate and width; f_0 minima and f_0 baseline.

  Overall there was a general tendency of decrease of several f_0 measures, such as: f_0 range, f_0 dispersion, f_0 rate of change, variability of f_0 rate of change, variability of f_0 peak height, median f_0 and maximum f_0. Perceptually, these parameters may correspond to the sensation of a lower pitch with less vivacity due to smaller and more uniform excursions. Given the increase in the slopes of LTAS, a potentially higher vocal effort could be perceived, in con-

junction with an overall less noisy voice, thanks to the increase in the proportion of harmonics to noise.

- *Do the theme of the news and the positive–negative valence of the utterances affect the speech prosody of the newscaster? If so, which parameters are affected? How they are affected?*

  The theme of the news and the valence of the utterances did affect the prosody of the newscaster. News themes showed correlation with most of the analysed variables. From all of the analysed parameters (see Table 2.2, on 57), only shimmer, standard deviation of f_0 peaks and f_0 peak width were not affected by these variables.

  On the other hand, valence had little impact on most of the analysed parameters. It was mainly correlated with f_0 median and baseline, but it also had an effect on f_0 minima, maxima, standard deviation and interquartile semi-amplitude; f_0 first derivatives; and the coefficient of variation of intensity.

  Positive utterances were associated with higher f_0, more intensity and overall more vivacity on the speech. Similarly, themes more associated with positive news and with higher positive emotional arousal such as sports and election showed a similar trend, while more negative themes such as disaster news had an opposite behaviour.

Television data from the current century has an overall satisfactory quality for acoustic prosodic analysis. Furthermore, this research attests the viability of the application of sophisticated acoustic phonetics methods for analysing speech. With the remarkable volume of speech recordings that are done nowadays and the development of associated technologies, material previously recorded without acoustic analysis purposes can be a fruitful source of linguistic data. This type of data is in a middle ground between recorded sociolinguistic interviews and laboratory recordings, but with the good addition of this speech being usually non-elicited.

However, it is important to consider the potential limitation of the recordings given the parameters herein analysed. There is always some degree of difference in longitudinally collected data, from local disparities associated with the speakers' physiology, to greater changes in the recording equipment and environment. Larger corpora may be able to deal with more disparity in these factors, but may also be subject to less specialised results.

In this sense, further research is needed to fully confirm some of the results regarding the voice quality and intensity parameters. Despite the manual screening of the data performed in this research, avoiding the inclusion of low audio quality segments in the final corpus, there is likely some impact of the mixed recording quality in these sensitive parameters. Nevertheless, most of the findings were generally in order with previous results concerning the behaviour of the analysed prosodic features.

It is also important to note that most data on this subject come from voice samples of sustained vowels. In this sense, the present findings may also help to establish some acoustic thresholds for these parameters on chained speech.

Moreover, the findings here reported can be used as a source of normative data for future studies as well as for forensic analysis. They may be specially useful for analysing data of public speech from speakers with professional media training such as media figures and politicians.

It can also be used as a general point of comparison for vocal therapists as it provides insight about what can be expected from a middle-aged professional voice. Therefore, it can be useful for comparing measures associated with age with speakers without vocal training, as well as serve a reference of the pace of the ageing process in the mid-30's to mid-50's age range.

Multiple factors can have an impact on speakers' voices, such as the speech setting and associated formality levels, the audience, the topic, the expression of emotions and of personae that may or may not be in conformity with social expectations (Bell, 1984; Eckert, 2003; Labov,

1972, inter alia). Therefore, although it is important to recognise the available frameworks associated with intraspeaker variation, it is also valuable to keep an open mind in relation to the conditioning factors of this phenomenon. Adopting a single narrow view of style can negatively impact the understanding of intraspeaker variation and even bias forensic analyses of voice.

For example, the results regarding news themes in this research show not only the importance of the topic of the enunciation, but also the relevance of more specific local aspects in linguistic performance. Although topic can be seen as a part of news themes, they are not equivalent. The themes, for instance, are always subject to prior consideration in the news show, which is certainly not always the case for general conversational topic. News themes are intrinsically related to the newscasting genre and, as such, are subject to the influence of other specific local factors related to the genre and its associated practices.

Suggestions for further research

Since this work comprises many different complex subjects it certainly does not exhaust all the pertaining literature nor all the possible intepretations. Future research may provide further relevant results by exploring linguistic domains that did not fall within the scope of this project.

In the field of prosodic research, the study of rhythmic aspects of newscasting speech may grant valuable insights, considering that the usage of pauses and of particular rhythmic patterns that could, for instance, be responsible for emphasis, can shed more light on the speech performance of the newscasters and thus on linguistic behaviour in general.

Another crucial dimension for the comprehension of the meaning of the obtained parameters is how they are perceived by others and how the listeners evaluate the changes in news speech, and, more generally, how they evaluate longitudinal change.

Television presenting is hardly a long and continuous monologue. In fact, much of traditional linguistic data is not. Thus, we may consider that speech analysed here may be subject to the effects of long pauses in the subjects speech due to various reasons. These reasons can and should be explored considering the particularities of the structures typically present in different social settings. This would allow for a better comprehension of the role that these structuring elements might carry in speech and how they may affect spoken behaviour and vocal production.

For example, in the journalistic setting, one commercial break may not be a simple equivalent of a couple of minutes without speaking. The broadcaster might drink water, receive criticism about the recent performance, talk to colleagues in a more casual manner, regain attention of the expected way of speaking when narrating news, adjust the lavalier microphone position and audio and only then restart the broadcast. In this mean time, the speaker can be asked to cover a relevant event that happened after the start of the broadcasting. They can also receive a very uplifting or very worrying personal news on the smartphone or even remember it and so on.

Although the classification of valence did not seem erroneous, as the information given on the news was trully majorly skewed towards negative topics, it may be of interest to verify its validity through a comparison with the ratings of different judges. Moreover, an analysis about emotions on newscasting speech accounting for activation and arousal level also seem in order to provide a better understanding of prosodic variation in public speech.

Beyond the reporting and analysis of the longitudinal newscasting voice data, this study sought to make more explicit the relevance of a broad sociophonetic approach for the forensic linguistics field. The growing employment of phonetics theoretical and methodological knowledge in forensics has even greater potential when associated with a sociolinguistic perspective of language variation and how societal factors may influence speech.

REFERENCES

Antunes, L. B. (2017). Escolhas prosódicas e a construção do ethos do apresentador de talk shows. *Anais do VI Colóquio Brasileiro de Prosódia da Fala*, *4*.

Arantes, P. (2014). Estimativas de longo termo da frequência fundamental: Implicações para a fonética forense. *ReVEL*, *12*, 217–236.

Ayadi, M. E., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, *44*(3), 572–587. https://doi.org/10.1016/j.patcog.2010.09.020

Bailly, A. (1935). ᾨΙδή. In *Dictionnaire grec–français* (p. 2179). Hachette. https://archive.org/details/BaillyDictionnaireGrecFrancais/page/2179/mode/1up

Banse, R., & Scherer, K. S. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636.

Barbosa, P. A. (2009). Measuring speech rhythm variation in a model-based framework. *Proc. Interspeech*, 1527–1530. https://doi.org/10.21437/Interspeech.2009-463

Barbosa, P. A. (2019). *Prosódia*. Parábola.

Barbosa, P. A. (2021). *Prosody descriptor extractor [v3.1]*. https://www.github.com/pabarbosa/prosody-scripts/tree/master/ProsodyDescriptorExtractor

Barbosa, P. A., & Boula de Mareüil, P. (2018, January). Imitating broadcast news style: Commonalities and differences between french

and brazilian professionals. In A. V. et al. (Ed.), *Lecture notes in computer science* (pp. 419–428). Springer International Publishing. https://doi.org/10.1007/978-3-319-99722-3_42

Bell, A. (1984). Language style as audience design. *Language in society, v.13, n*(2), 145–204.

Bell, A. (2001). Back in style: Reworking audience design. In A. Bell, P. Eckert & J. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 139–169). Cambridge University Press.

Benjamin, B. J. (1981). Frequency variability in the aged voice. *Journal of Gerontology, 36*, 722–726. https://doi.org/10.1093/geronj/36.6.722

Biadsy, F., Rosenberg, A., Carlson, R., Hirschberg, J. B., & Strangert, E. (2008). A cross-cultural comparison of american, palestinian, and swedish perception of charismatic speech. *Speech Prosody 2008, 37*, 579–582. https://doi.org/10.7916/D8M3342B

Bloomfield, L. (1927). Literate and illiterate speech. *American speech, 2*(10), 432–439.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*.

Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer [software].

Bolinger, D. (1985). *Intonation and its parts: Melody in spoken english*. Stanford University Press.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford University Press.

Brockmann, M., Storck, C., Carding, P. N., & Drinnan, M. J. (2008). Voice loudness and gender effects on jitter and shimmer in healthy adults. *Journal of Speech, Language, and Hearing Research, 51*(5), 1152–1160. https://doi.org/10.1044/1092-4388(2008/06-0208)

Brown, B. L., Strong, W. J., & Rencher, A. C. (1974). Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *The Journal of the Acoustical Society of America*, *55*(2), 313–318. https://doi.org/10.1121/1.1914504

Brugman, H., & Russel, A. (2004). Annotating multimedia: Multi-modal resources with elan. *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation.*

Campos, L. C. P. (2012). *Radialista: Análise acústica da variação entoacional na fala profissional e na fala coloquial* (Master's thesis). Unicamp. Campinas.

Carvalho, A. M. (2004). I speak like the guys on tv: Palatalization and the urbanization of uruguayan portuguese. *Language variation and change*, *16*(2), 127–151.

Castro, L., Freitas, M., Moraes, J. A. d., & Serridge, B. (2010). Listeners' ability to identify professional speaking styles based on prosodic cues. *Speech Prosody 2010-Fifth International Conference.*

Castro, L., Serridge, B., Moraes, J., & Freitas, M. (2010). The prosody of the tv news speaking style in brazilian portuguese. *Third ISCA workshop on experimental linguistics.*

Catani, G. (2021). *O jornal nacional do século xxi: Mudança, estilo e norma* (Bachelor's thesis). Unicamp.

Catani, G. (2022). *A fala telejornalística do século xxi: Mudança, estilo e norma.* Asa da Palavra.

Chomsky, N. (1965). *Aspects of the theory of syntax.* MIT Press.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic

speech sounds. *The Journal of the Acoustical Society of America*, *24*(6), 597–606. https://doi.org/10.1121/1.1906940

Cooper, F. S., Liberman, A. M., & Borst, J. M. (1951). The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Sciences*, *37*(5), 318–325. https://doi.org/10.1073/pnas.37.5.318

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge University Press.

Deffenbacher, K. A., Cross, J. F., Handkins, R. E., Chance, J. E., Goldstein, A. G., Hammersley, R., & Read, J. D. (1989). Relevance of voice identification research to criteria for evaluating reliability of an identification. *The Journal of Psychology*, *123*(2), 109–119. https://doi.org/10.1080/00223980.1989.10542967

Dehqan, A., Scherer, R. C., Dashti, G., Ansari-Moghaddam, A., & Fanaie, S. (2012). The effects of aging on acoustic parameters of voice. *Folia Phoniatrica et Logopaedica*, *64*(6), 265–270. https://doi.org/10.1159/000343998

Delattre, P., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color: Observations on one and two-formant vowels synthesized from spectrographic patterns. *Word*, *8*(3), 195–210. https://doi.org/10.1080/00437956.1952.11659431

Deliyski, D. D., Shaw, H. S., & Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice*, *19*(1), 15–28. https://doi.org/10.1016/j.jvoice.2004.07.003

D'Errico, F., Signorello, R., Demolin, D., & Poggi, I. (2013). The perception of charisma from voice: A cross-cultural study. *2013 Humaine*

Association Conference on Affective Computing and Intelligent Interaction. https://doi.org/10.1109/acii.2013.97

Dolgun, A., & Demirhan, H. (2017). Performance of nonparametric multiple comparison tests under heteroscedasticity, dependency, and skewed error distribution. *Communications in Statistics - Simulation and Computation*, *46*(7), 5166–5183. https://doi.org/10.1080/03610918.2016.1146761

Dolson, M. (1994). The pitch of speech as a function of linguistic community. *Music Perception*, *11*(3), 321–331. https://doi.org/10.2307/40285626

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, *6*(3), 241–252.

Eckert, P. (2003). The meaning of style. *Proceedings of the Eleventh Annual Symposium about Language and Society*, 41–53.

Eckert, P. (2008). Variation and the indexical field. *Journal of sociolinguistics*, *12*(4), 453–476.

Fant, G. (1958). Modern instruments and methods for acoustic studies of speech. *Proceedings of the VIIIth International Congress of Linguists*, 282–358.

Fernandes, J., Teixeira, F., Guedes, V., Junior, A., & Teixeira, J. P. (2018). Harmonic to noise ratio measurement - selection of window and length. *Procedia Computer Science*, *138*, 280–285. https://doi.org/10.1016/j.procs.2018.10.040

Ferrand, C. T. (2002). Harmonics-to-noise ratio: An index of vocal aging. *Journal of Voice*, *16*(4), 480–487. https://doi.org/10.1016/s0892-1997(02)00123-6

Figueiredo, R. M. (1993). A eficácia de medidas extraídas do espectro de longo termo para a identificação de falantes. *Cadernos de Estudos Linguísticos*, *25*, 113–127.

Fonagy, I. (1978). A new method of investigating the perception of prosodic features. *Language and Speech*, *21*(1), 34–49. https://doi.org/10.1177/002383097802100102

Foulkes, P. (2015). Sociophonetics. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*.

Foulkes, P., & French, P. (2012). Forensic speaker comparison: A linguistic-acoustic perspective. In *The oxford handbook of language and law* (pp. 558–572). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199572120.013.0041

Foulkes, P., Scobbie, J. M., & Watt, D. (2010). Sociophonetics. In W. J. Hardcastle, J. Laver & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (2nd ed., pp. 703–754). Wiley-Blackwell.

Fouquet, M., Pisanski, K., Mathevon, N., & Reby, D. (2016). Seven and up: Individual differences in male voice fundamental frequency emerge before puberty and remain stable throughout adulthood. *Royal Society Open Science*, *3*(10). https://doi.org/10.1098/rsos.160395

Fraccaro, P. J., O'Connor, J. J., Re, D. E., Jones, B. C., DeBruine, L. M., & Feinberg, D. R. (2013). Faking it: Deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour*, *85*(1), 127–136. https://doi.org/10.1016/j.anbehav.2012.10.016

Gasser, E., Ahn, B., Napoli, D., & Zhou, Z. (2019). Production, perception, and communicative goals of american newscaster speech. *Language in Society*, *48*, 233–259. https://doi.org/10.1017/S0047404518001392

Gold, E., & French, P. (2019). International practices in forensic speaker comparisons. *International Journal of Speech Language and the Law*, *26*(1), 1–20. https://doi.org/10.1558/ijsll.38028

Gorris, C., Maccarini, A. R., Vanoni, F., Poggioli, M., Vaschetto, R., Garzaro, M., & Valletti, P. A. (2020). Acoustic analysis of normal voice patterns in italian adults by using praat. *Journal of Voice*, *34*(6). https://doi.org/10.1016/j.jvoice.2019.04.016

Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., & van Lieshout, P. (2013). Normative voice data for younger and older adults. *Journal of Voice*, *27*(5), 545–555. https://doi.org/10.1016/j.jvoice.2013.03.002

Guthrie, W., Filliben, J., & Heckert, A. (2003). Process modeling. In *Nist sematech e-handbook of statistical methods*. https://doi.org/10.18434/M32189

Hay, J., & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology*, *36*(1), 89–103. https://doi.org/10.1146/annurev.anthro.34.081804.120633

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in swedish. *Journal of Phonetics*, *31*(1), 39–62.

Hellwig, B., van Uytvanck, D., Hulsbosch, M., Somasundaram, A., Tacchetti, M., Geerts, J., & Sloetjes, H. (2022). *Elan - linguistic annotator*. Full manual. Version 6.4. Nijmegen, The Netherlands, The Language Archive, MPI for Psycholinguistics. https://archive.mpi.nl/tla/elan

Hernández-Campoy, J. M., & Cutillas-Espinosa, J. A. (2012). Introduction: Style-shifting revisited. In *Style-shifting in public* (pp. 1–18). John Benjamins.

Hippargekar, P., Bhise, S., Kothule, S., & Shelke, S. (2021). Acoustic voice analysis of normal and pathological voices in indian population using praat software. *Indian Journal of Otolaryngology and Head Neck Surgery*, *74*(S3), 5069–5074. https://doi.org/10.1007/s12070-021-02757-9

Hirst, D., & di Cristo, A. (1998). *Intonation systems: A survey of twenty languages*. Cambridge University Press.

Hollien, H., & Majewski, W. (1977). Speaker identification by long-term spectra under normal and distorted speech conditions. *The Journal of the Acoustical Society of America*, *62*(4), 975–980.

Hollien, H., & Schwartz, R. (2001). Speaker identification utilizing non-contemporary speech. *Journal of Forensic Science*, *46*(1), 63–67.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Huang, D. Z., Minifie, F. D., Kasuya, H., & Lin, S. X. (1995). Measures of vocal function during changes in vocal effort level. *Journal of Voice*, *9*(4), 429–438. https://doi.org/10.1016/s0892-1997(05)80206-1

Irvine, J. T. (2001). "style" as distictiveness: The culture and ideology of linguistic differentiation. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation*. Cambridge.

Jassem, W. (1975). Normalisation of f0 curves. In G. Fant & M. Tatham (Eds.), *Auditory analysis and perception of speech*. Academic Press.

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, *2*(4), 671–711. https://doi.org/10.1111/j.1749-818x.2008.00066.x

Jockers, M. L. (2015). *Syuzhet: Extract sentiment and plot arcs from text*. https://github.com/mjockers/syuzhet

Johnstone, B., & Bean, J. M. (1997). Self-expression and linguistic variation. *Language in Society*, *26*(2), 221–246.

Kaplan, H. M. (1960). *Anatomy and physiology of speech*. McGraw-Hill Companies.

Kepplinger, H. M., & Donsbach, W. (1990). The impact of camera perspectives on the perception of a speaker. *Studies in Educational Evaluation*, *16*(1), 133–156. https://doi.org/10.1016/s0191-491x(05)80076-8

Kitajima, K., & Gould, W. J. (1976). Vocal shimmer in sustained phonation of normal and pathologic voice. *Ann Otol 85*.

Koike, Y. (1969). Vowel amplitude modulations in patients with laryngeal diseases. *The Journal of the Acoustical Society of America*, *45*(4), 839–844. https://doi.org/10.1121/1.1911554

Künzel, H. J. (2007). Non-contemporary speech samples: Auditory detectability of an 11 year delay and its effect on automatic speaker identification. *International Journal of Speech, Language & the Law*, *14*(1).

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*(3), 273–309. https://doi.org/10.1080/00437956.1963.11659799

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (2001a). The anatomy of style-shifting. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation*. Cambridge.

Labov, W. (2001b). *Principles of linguistic change, volume 2: Social factors*. Blackwell.

Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics*, *34*, 500–515. https://doi.org/10.1016/j.wocn.2006.05.002

Labov, W. (2006 [1966]). *The social stratification of english in new york city*. Cambridge: Cambridge University Press.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress* (Report on National Science Foundation). University of Pennsylvania.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press.

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics* (7th ed.). Cengage learning.

Laver, J. (1968). Voice quality and indexical information. *International Journal of Language & Communication Disorders*, *3*(1), 43–54. https://doi.org/10.3109/13682826809011440

Laver, J. (2000). Phonetic evaluation of voice quality. *Voice quality measurement*, 37–48.

Leeuw, I. M. V. D., & Mahieu, H. F. (2004). Vocal aging and the impact on daily life: A longitudinal study. *Journal of Voice*, *18*(2), 193–202. https://doi.org/10.1016/j.jvoice.2003.10.002

Liddell, H. G., & Scott, R. (1940). ᾨΙδή. In *A greek-english lexicon* (Revised and augmented by Sir Henry Stuart Jones). Clarendon Press. http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0057:entry=w)%7Cdh/

Lieberman, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *The Journal of the Acoustical Society of America*, *35*(3), 344–353. https://doi.org/10.1121/1.1918465

Lindh, J., & Eriksson, A. (2007). Robustness of long time measures of fundamental frequency. *Eighth Annual Conference of the International Speech Communication Association*.

Linville, S. E. (2001). *Vocal aging*. Singular.

Lortie, C. L., Thibeault, M., Guitton, M. J., & Tremblay, P. (2015). Effects of age on the amplitude, frequency and perceived quality of voice. *AGE*, *37*(6). https://doi.org/10.1007/s11357-015-9854-1

Madureira, S. (2016). Intonation and variation: The multiplicity of forms and senses. *Dialectologia*, (2016.2016), 57–74. https://doi.org/10.1344/dialectologia2016.2016.4

Martin, D., Fitch, J., & Wolfe, V. (1995). Pathologic voice type and the acoustic prediction of severity. *Journal of Speech, Language, and Hearing Research*, *38*(4), 765–771. https://doi.org/10.1044/jshr.3804.765

Mathew, M. M., & Bhat, J. S. (2009). Soft phonation index – a sensitive parameter? *Indian Journal of Otolaryngoligy and Head Neck Surgery*, *61*, 127–130.

McMenamin, G. R. (2002). Forensic linguistics. In G. R. McMenamin (Ed.), *Forensic linguistics: Advances in forensic stylistics*. CRC Press.

Mendes, R. B., & Oushiro, L. (2012). O paulistano no mapa sociolinguístico brasileiro. *Alfa: Revista de Linguística (São José do Rio Preto)*, *56*(3), 973–1001.

Mendes, R. B., & Oushiro, L. (2013). Documentação do projeto sp2010 – construção de uma amostra da fala paulistana.

Mirowsky, J., & Ross, C. E. (2010). Well-being across the life course. In T. L. Scheid & T. N. Brown (Eds.), *A handbook for the study of mental health* (2nd ed., pp. 361–383). Cambridge University Press Cambridge.

Mohammad, S. M., & Turney, P. D. (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. *Proceedings of the NAACL HLT 2010 Workshop on Compu-*

tational Approaches to Analysis and Generation of Emotion in Text, 26–34. https://aclanthology.org/W10-0204

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, *29*(3), 436–465.

Mueller, P. (1997). The aging voice. *Seminars in Speech and Language*, *18*(02), 159–169. https://doi.org/10.1055/s-2008-1064070

Müller, C. (2005). *Zweistufige kontextsensitive sprecherklassifikation am beispiel von alter und geschlecht* (Doctoral dissertation). Universität des Saarlandes. Saarbrücken.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, *93*(2), 1097–1108. https://doi.org/10.1121/1.405558

Niebuhr, O., Voße, J., & Brem, A. (2016). What makes a charismatic speaker? a computer-based acoustic-prosodic analysis of steve jobs tone of voice. *Computers in Human Behavior*, *64*, 366–382. https://doi.org/10.1016/j.chb.2016.06.059

Nunes, A., Coimbra, R. L., & Teixeira, A. (2010). Voice quality of european portuguese emotional speech. In *Lecture notes in computer science* (pp. 142–151). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12320-7_19

Nwe, T. L., Foo, S. W., & Silva, L. C. D. (2003). Speech emotion recognition using hidden markov models. *Speech Communication*, *41*(4), 603–623. https://doi.org/10.1016/s0167-6393(03)00099-2

Olmedo, F. (2022). *Envelhecimento vocal: Estudo das alterações segmentais e variações prosódicas no português brasileiro* (Undergraduate Research). Unicamp.

Orletti, F., & Mariottini, L. (Eds.). (2017). *Forensic communication in theory and practice: A study of discourse analysis and transcription*. Cambridge Scholars Publishing.

Orlikoff, R. F. (1990). The relationship of age and cardiovascular health to certain acoustic characteristics of male voices. *Journal of Speech, Language, and Hearing Research*, *33*(3), 450–457.

Ostendorf, M., Price, P. J., & Shattuck-Hufnagel, S. (1995). The boston university radio news corpus. *Linguistic Data Consortium*, 1–19.

Oushiro, L. (2014). Tratamento de dados com o R para análises sociolinguísticas. In R. M. K. Freitag (Ed.), *Metodologia de coleta e manipulação de dados em sociolinguística*. Blucher. https://doi.org/10.5151/blucheroa-mcmds-10cap

Özdil, K., Çatıker, A., & Bulucu Büyüksoy, G. D. (2022). Smartphone addiction and perceived pain among nursing students: A cross-sectional study. *Psychology, Health & Medicine*, *27*(10), 2246–2260.

Palo, H. K., Mohanty, M. N., & Chandra, M. (2017). Emotion analysis from speech of different age groups. *Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering*. https://doi.org/10.15439/2017r21

Passetti, R. R., & Barbosa, P. A. (2015). O efeito do telefone celular no sinal da fala: Uma análise fonético-acústica com implicações para a verificação de locutor em português brasileiro. *Anais do Congresso Brasileiro de Prosódia*, (3).

Passetti, R. R., Madureira, S., & Barbosa, P. A. (2022). Voice perception on a voice messaging app: Implications for forensic phonetics. *Proceedings of Speech Prosody 2022*, 490–494. https://doi.org/10.21437/speechprosody.2022-100

Paul, H. (1891 [1889]). *Principles of the history of language*. Longmans, Green; Co.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184. https://doi.org/10.1121/1.1906875

Pittam, J. (1987). The long-term spectral measurement of voice quality as a social and personality marker: A review. *Language and speech*, *30*(1), 1–12.

Portinha, S. (2011). A voz das emoções no jornalismo de televisão. In U. de Lisboa (Ed.), *Actas das jornadas a linguagem da voz*.

Ptacek, P. H., Sander, E. K., Maloney, W. H., & Jackson, C. C. R. (1966). Phonatory and related changes with advanced age. *Journal of Speech and Hearing Research*, *9*, 353–360.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344. https://doi.org/10.1016/j.evolhumbehav.2007.05.002

R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. http://www.R-project.org/

Ramig, L. A. (1983). Effects of physiological aging on vowel spectral noise. *Journal of Gerontology*, *38*(2), 223–225. https://doi.org/10.1093/geronj/38.2.223

Ramig, L. A. (1986). Aging speech: Physiological and sociological aspects. *Language & Communication*, *6*, 25–34.

Ramig, L. A., & Ringel, R. L. (1983). Effects of physiological aging on selected acoustic characteristics of voice. *Journal of Speech, Language, and Hearing Research*, *26*(1), 22–30.

Reich, A. R., Frederickson, R. R., Mason, J. A., & Schlauch, R. S. (1990). Methodological variables affecting phonational frequency range in adults. *Journal of Speech and Hearing Disorders*, *55*(4), 124–131. https://doi.org/10.1044/jshd.5504.804

Rose, P. (2002). *Forensic speaker identification*. Taylor & Francis.

Sataloff, R. T., Kost, K. M., & Linville, S. E. (2017). The effects of age on the voice. In R. T. Sataloff (Ed.), *Clinical assessment of voice* (2nd ed.). Plural Publishing.

Saussure, F. d. (1916). *Cours de linguistique générale* (2nd ed.).

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences*. Oxford University Press.

Schilling, N. (2013). Investigating stylilstic variation. In J. Chambers & N. Schilling (Eds.), *The handbook of language variation and change* (2nd ed., pp. 327–349). John Wiley & Sons.

Schilling-Estes, N. (2002). Investigating stylilstic variation. In J. Chambers & N. Schilling (Eds.), *The handbook of language variation and change* (1st ed., pp. 375–401). John Wiley & Sons.

Schötz, S. (2006). *Perception, analysis and synthesis of speaker age* (Vol. 47). Lund University.

Silva, A. D. (2022). Dados de referência de f0 em corpus de falantes do português brasileiro na variedade falada na capital paulista. *Revista Brasileira de Criminalística*, *11*(2), 92–105. https://doi.org/10.15260/rbc.v11i2.612

Silva, M. J., Carvalho, P., Costa, C., & Sarmento, L. (2010). Automatic expansion of a social judgment lexicon for sentiment analysis.

Silva, M. J., Carvalho, P., & Sarmento, L. (2012). Building a sentiment lexicon for social judgement mining. *International Conference on Computational Processing of the Portuguese Language*, 218–228.

Silveira, G. d. C. P. d. (2022). *The prosody of speech in a dialect contact situation: A sociophonetic study of the speech of alagoan migrants in são paulo.* (Master's thesis). Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

Silveira, G. d. C. P. d., & Oushiro, L. (2022). Sociofonética (verbete).

Souza, M., & Vieira, R. (2012). Sentiment analysis on twitter data for portuguese language. *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language*, 241–247. https://doi.org/10.1007/978-3-642-28885-2_28

Souza, M., Vieira, R., Busetti, D., Chishman, R., Alves, I. M., & Unisinos, F. D. L. (2011). Construction of a portuguese opinion lexicon from multiple resources. *In 8th Brazilian Symposium in Information and Human Language Technology - STIL, Mato Grosso.*

Spazzapan, E. A., Cardoso, V. M., Fabron, E. M. G., Berti, L. C., Brasolotto, A. G., & de Castro Marino, V. C. (2018). Características acústicas de vozes saudáveis de adultos: Da idade jovem à meia-idade. *CoDAS*, *30*(5). https://doi.org/10.1590/2317-1782/20182017225

Strangert, E. (1991). Phonetic characteristics of professional news reading. *PERILUS XII*, 39–42.

Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis – jitter, shimmer and HNR parameters. *Procedia Technology*, *9*, 1112–1122. https://doi.org/10.1016/j.protcy.2013.12.124

Thomas, E. R. (2010). *Sociophonetics: An introduction: An introduction*. Palgrave.

Thomas, E. R. (2019). Innovations in sociophonetics. In W. F. Katz & P. F. Assmann (Eds.), *The routledge handbook of phonetics* (pp. 448–472). Routledge.

Tielen, M. (1990). Perception of the voices of men and women in relation to their profession. *Speaker Characterization in Speech Technology*.

Traunmüller, H., & Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. *Unpublished manuscript*, *11*.

Tronick, E. C., Als, H., & Brazelton, T. B. (1980). The infant's communicative competencies and the achievement of intersubjectivity. *The relationship of verbal and nonverbal communication*, (25), 261–273. According to bolinger1985: In Key 1980.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban british english of norwich. *Language in Society*, *1*(2), 179–195. https://doi.org/10.1017/s0047404500000488

Valente, A. R., Oliveira, C., Albuquerque, L., Teixeira, A., & Barbosa, P. A. (2021). Prosodic changes with age: A longitudinal study on a famous european portuguese native speaker. *International Conference on Speech and Computer*, 726–736.

Weinreich, U., Labov, W., & Herzog, M. (1968). *Empirical foundations for a theory of language change* [Tradução: Marcos Bagno. São Paulo: Editora Parábola]. University of Texas Press.

Wilcox, K. A., & Horii, Y. (1980). Age and changes in vocal jitter. *Journal of Gerontology*, *35*, 194–198.

Wolfram, W. (n.d.). *Sociolinguistics* (L. S. of America, Ed.). https://www.linguisticsociety.org/resource/sociolinguistics

Xue, S. A., & Deliyski, D. (2001). Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. *Educational Gerontology*, *27*(2), 159–168. https://doi.org/10.1080/03601270151075561

Xue, S. A., & Fucci, D. (2000). Effects of race and sex on acoustic features of voice analysis. *Perceptual and motor skills*, *91*(3), 951–958.

Xue, S. A., & Hao, G. J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production. *Journal of Speech, Language, and Hearing Research*, *46*(3), 689–701. https://doi.org/10.1044/1092-4388(2003/054)

Yanagihara, N. (1967). Significance of harmonic changes and noise components in hoarseness. *Journal of Speech and Hearing Research*, *10*(3), 531–541. https://doi.org/10.1044/jshr.1003.531

Zimman, L. (2020). Sociophonetics. In J. Stanlaw (Ed.), *The international encyclopedia of linguistic anthorpology* (pp. 1–5). Wiley. https://doi.org/10.1002/9781118786093.iela0363