**FEDERAL UNIVERSITY OF TECHNOLOGY - PARANÁ**

**ACADEMIC DEPARTMENT OF LANGUAGE AND COMMUNICATION**

**ACADEMIC DEPARTMENT OF MODERN FOREIGN LANGUAGES**

**LANGUAGE ARTS ENGLISH/PORTUGUESE COLLEGE**

**ROBINSON LUIS KREMER**

# FUNDAMENTAL FREQUENCY: AN ANALYSIS OF AGE AND GENDER

**THESIS**

**CURITIBA**

**2015**

**ROBINSON LUIS KREMER**

# FUNDAMENTAL FREQUENCY: AN ANALYSIS OF AGE AND GENDER

Thesis presented to the Language Arts English/Portuguese College of the Academic Department of Modern Foreign Languages – DALEM – and the Academic Department of Language and Communication– DALIC – of the Federal University of Technology – Paraná, as a requirement to obtain the Teaching degree.

Advisor: Doctor Maria Lúcia de Castro Gomes

**CURITIBA**

**2015**

Ministério da Educação
**Universidade Tecnológica Federal do Paraná**
Campus Curitiba

Departamento Acadêmico de Linguagem e Comunicação
Departamento Acadêmico de Línguas Estrangeiras Modernas
Licenciatura em Letras Português-Inglês

# TERMO DE APROVAÇÃO

FUNDAMENTAL FREQUENCY: AN ANALYSIS OF AGE AND GENDER

por

ROBINSON LUIS KREMER

Este Trabalho de Conclusão de Curso (TCC) foi apresentado em 13 de fevereiro de 2015 como requisito parcial para a obtenção do título de Licenciado em Letras Português-Inglês. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

_____
Profa. Dra. Maria Lúcia de Castro Gomes
Profa. Orientadora

_____
Prof. Dr. Gustavo Nishida
Membro titular

_____
Profa. Jennifer I. A. de Albuquerque
Membro titular

**- O TERMO DE APROVAÇÃO ASSINADO ENCONTRA-SE NA COORDENAÇÃO DO CURSO -**

I dedicate this thesis to my beloved Advisor
Malu who challenged me from the very
beginning, making me believe that I could do
anything as long as I believed in myself.

## SPECIAL THANKS

First of all, I need to thank my Mother and my Father for making me pursue my dreams and never forcing me to study or do anything against my will.

My big sister, Patricia and my brother-in-law Nilton, without their support (emotional and financial) I would not be writing this Thesis today.

My sister Letícia, for being the best publicist I could hire if I ever wanted to.

My soul mate Alessandra Zaia, my great friend Joyce Corrêa and all my Ortiz family, they kept showing me my light. We are miles away but never too far. And without Alessandra I would never made it into this institution.

My friends: Renata, Bruna and Cristiano for making my days at University worth it. Especially Renata (my body and soul), for giving me the 'click' that I could write this work in English. We are the real deal.

What would I do without Will Pitkin and Andrew Lindsay? It was a pleasure working with you. And now, it is a pleasure having your friendship. Will, thank you for reviewing and improving my thesis' writing.

My trinity: Juçara, Eline and Raquel, for guiding me through my English life (in the very beginning).

Another very important trio, Joice, Denise and Andréa, I could not learn about F0 and voice quality without you. Also, thank you for helping the study group.

Renata Badi, she gave me the best opportunity of improvement in my life and also she was always supporting my personal/professional growth.

Last but not least, my advisor, academic mom, sister and friend Malu, for pushing me and always making me believe that everything was possible, she made me grow and helped me deciding about my future and I quote "Do you want to write about something that is already well explored or do you want to write about something challenging and brand new over here?". And here I am in Forensic Phonetics.

I still have so many names and so many things to say. You know who you are, I love you all and I thank you for being part of my life someway, somehow.

“Anything you say can and will be used

against you in a court of law”.

(Excerpt of The Miranda Warning based on
the Fifth Amendment, USA)

## ABSTRACT

KREMER, Robinson Luis. **Fundamental Frequency:** an analysis of age and gender. 2015. 55p. Thesis (Major in Language Arts English/Portuguese) – Federal University of Technology - Paraná. Curitiba, 2015.

The phenomenon of speech is very complex and every time a speaker produces it, s/he presents personal characteristics, making the speech an individualizing element. Acoustic analysis of speech helps the identification of these personal traits of the speaker, with the help of numerical results and graphics. Acoustics is one of the objects of study of Phonetics, which has two other areas: Physiologic and Perceptual.One of the fields that uses the three areas of Phonetics is Forensic Phonetics, offering knowledge and methods of linguistic analysis in contexts of criminal investigation involving speech. One of the activities of Forensics is speaker identification, and fundamental frequency (F0) is a promising phonetic parameter in this area due to the fact that most of the disguises directly affect it. Gender and age also have influence on it, because physiological differences for both sexes and ages can be seen when analyzing F0 separately, for example, high or low pitch. In this study, we analyzed F0 comparing normal and disguised voice of 50 speaker, men and women in two age groups,and investigated which disguises were most used in the different groups. We used perceptual and acoustic analyses, and provided numerical data to observe the changes in the voice, after applying the VPAS protocol (LAVER, 1980) adapted to Brazilian Portuguese (CAMARGO; MADUREIRA, 2008) to label the features/parameters of voice used in the disguises.The objective of this study was purely instructional. We wanted to learn how to deal with this new area of linguistics: forensic phonetics.Partial results have shown that most of the participants raised their F0 and that the choice of raising F0 was most frequent among men and lowering F0 among women. Concerning age related to the type of disguise, we could not find very significant difference rather than more young participants lowered their F0 than adults, and more adults had their F0 higher.



**Keywords:**Speech. Forensic phonetics. Fundamental frequency. Disguise. Vocal quality.

## RESUMO

KREMER, Robinson Luis. **Fundamental Frequency:** an analysis of age and gender. 2015. 55f. Trabalho de Conclusão de Curso (Licenciatura em Letras Português/Inglês)– Universidade Tecnológica Federal do Paraná. Curitiba, 2015.

O fenômeno da fala é bastante complexo e cada vez que um falante a produz, ele apresenta características pessoais, tornando a sua fala um fator individualizante. A análise acústica da voz ajuda a identificar estes traços pessoais do falante, pelo uso de resultados numéricos e gráficos. A fala é também o objeto de estudo da Fonética, que se divide em três áreas: Fisiológica, Acústica e Perceptual, tendo a Fonética Forense como um campo de estudo da fonética, oferecendo métodos e entendimento de análise linguística no contexto de investigação criminal em que envolva a fala. Uma das atividades da fonética forense é a identificação de falantes, e a frequência fundamental (F0) é um parâmetro promissor nessa área devido ao fato de que a maioria dos disfarces a afeta diretamente. A idade e o sexo também possuem influência, pois diferenças fisiológicas para ambos os sexos e entre idades podem ser encontradas quando analisada a F0 separadamente, como o uso de *pitch* alto ou baixo, por exemplo. Neste estudo, analisamos a F0 comparando a voz normal e disfarçada de 50 participantes, homens e mulheres em duas faixas etárias, bem como investigamos quais disfarces foram mais utilizados nos diferentes grupos. Usamos dados de percepção, acústicos e numéricos para observar as mudanças e aplicamos o protocolo VPAS (LAVER, 1980), adaptado ao português brasileiro (CAMARGO; MADUREIRA, 2008) para nos ajudar na classificação dos parâmetros de voz, ou qualidade vocal, utilizados no disfarce. O objetivo desse estudo foi instrucional. Buscamos aprender como lidar com esta nova área da lingüística: a fonética forense. Os resultados parciais indicam que a maioria dos participantes aumentou a F0 e que esta escolha por aumentar a F0 ocorreu com mais frequência entre os homens e a diminuição da F0 entre mulheres. Com relação à idade relacionada aos tipos de disfarce, nós não encontramos diferenças muito significativa a não ser que os participantes mais jovens optaram por baixar a F0 e os mais adultos optaram por aumentar a F0.



**Palavras-chave:**Fala. Fonética Forense. Frequência fundamental. Disfarce. Qualidade vocal.

**SUMMARY**

## 1 INTRODUCTION

A variety of definitions can be found for the phenomenon of speech, Kent and Read provide one of them in their book, when citing Raymond H. Stetson – a pioneer of the studies of speech – who states that speech is movement made audible (STETSON, 1928 apud KENT; READ, 2002).The authors also complement that "speech gains its unique importance as the primary means by which language is expressed in all human cultures except for people who are deaf. […] The end product of speech is an acoustic signal." (KENT; READ, 2002, p.01).

In order for the phenomenon of speech to occur, around 80 different muscles are used in a precise and fast way. Each speaker, when making these movements necessary to produce the speech, is going to present personal characteristics, turning his/her speech an individualizing element. Acoustic analysis of speech makes it possible to identify speaker's personal traits. By means of graphics, numerical results, it is possible to analyse events that happen, even when very fast, due to the dynamic of information during the speech.

Our area of research, Phonetics, beyond studying the production and perception of the sounds of speech, has the speech itself as the object of study. According to Kent and Read (2002, p.02): "[...] speech has three major arenas of study: the physiologic arena (or *physiologic phonetics*), the acoustic arena (or *acoustic phonetics*), and the perceptual arena (typically called *speech perception*)". As part of the group of phonetics, we have Forensic Phonetics (from now on being referred as FP), part of applied linguistics, and offers the knowledge and methods of linguistic analysis in context of criminal investigation involving speech. According to Braid (2003, p.05 e 06) "forensic phonetics is not limited to speaker identification, it focuses on all forensic trades that involve aspects of the speech or, extrapolating, sounds in general".[1] Consequently, speech analysis, mainly through Phonetics, serves as a multidisciplinary tool, aiding different areas, being Forensic as one example of them.

One of the activities of FP is speaker identification, in order to possibly determine if two utterances were performed by the same individual. Once recordings started being used as evidences, a professional (or an expert) in the area could be requested by a Deputy Chief, Attorney, Prosecutor or Judge, as in the context of Brazil (GOMES; CARNEIRO, 2014). Therefore, s/he could analyze samples and prepare an expert report, either in court or in the investigation stage. In order to determine the identity of the speaker, oftentimes, it is possible for experts to infer the speaker's profile, e.g. social class or origin, by means of the language

[1] "a fonética forense não se encerra na identificação de falantes, mas incide em todos os misteres criminalísticos que envolvam aspectos da fala ou, extrapolando, de sons em geral".(Author's translation)

they produce. In cases of poor-quality recordings, an expert – phonetician or speech scientist, for instance – can also help to determine what has been said or even if the recording has been edited. As stated by Nolan (1999, p.746): "The resultant field of activity has become widely known as forensic phonetics", i.e., works involving these activities as well as a range of others that include the speech or the sounds in general recorded in any type of media.

Human beings have an immense ability to recognize people by their own voices, even in untoward situations, establishing a standard capable of recreating/reproducing such ability stands as a major challenge to speech scientists. Though it still lacks more research, meeting this challenge has motivated numerous studies in the area of speaker identification, and with the development of digital signal processing systems, experts of FP studies have noticed an increase of research in the area.

I am going to present in this thesis the results of one of the studies that is being conducted at the Federal University of Technology- Paraná (UTFPR), by the FP study group created to develop more research in the area. We analyzed the difference produced with normal and disguised voices performed by female and male speakers, by means of perceptual and acoustic analysis. The first analysis, performed in two sessions by the author and other members of the study group, primarily tested the efficiency of the disguise[2] through the possible recognition or not of the voices. Later on, we analyzed disguise strategies that were chosen by each of the participants. In the latter (acoustic analysis), also performed in two sessions, we measured, firstly, the fundamental frequency (F0) from part of a speech and verified the difference of the values between normal and disguised voice. Secondly, we analyzed the curve of F0, also comparing both registers of voice: normal and disguise.

This thesis is divided in five sections: starting with some concepts and terminologies related to the area of forensics, we are then going to the field of voice quality in which we explain the VPAS protocol, also giving a better understanding about fundamental frequency, gender and age and disguise, all of them related to normal and disguised voice. Finally, we get to the research, where we will give details about it and the methodology applied to it. After this, we will present the results and discussion andwill finish it with the conclusion of this thesis.

As mentioned above, we will start by presenting some concepts and terminologies related to Forensic Phonetics, more specifically, identification, verification and comparison of

[2] The efficiency was attributed only when studying female voices.On further sections, I am going to explain the reason.

speaker, and explaining the differences among them, as well as the difference between technical and non-technical identification.

## 2 FORENSIC PHONETICS – CONCEPTS AND TERMINOLOGY

After having a better comprehension about the phenomenon of speech, complex in its way, it is necessary to point the existence of a science that has speech as the object of study: Phonetics. It is actually a group composed by phonetics sciences – mentioned in the previous section of this work -, as presented by Braid (2003):

> Articulatory Phonetics, which studies the physiology of speech and the sounds produced by the vocal tract; Acoustic Phonetics, which studies the physics ofthe sound waves produced by the speaker, their measure and instrumental analysis; or Auditory Phonetics, which studies the perception of the sounds of speech. Phonetics intercepts these areas in order to comprehend such a complex phenomenon as the speech.[3] (BRAID, 2003, p.5)

As stated previously, each speaker presents personal characteristics when producing speech. This way, Phonetics can be one of the tools that provide the knowledge of these individual traits/characteristics of a person. In other words, the information provided by Phonetics can be a way of identifying somebody, i.e., voice analysis can be a multidisciplinary tool, helping different areas, especially forensic.(BRAID, 2003).

Forensic Phonetics consists of different examinations in order to get to the voice identification of a person. It is a comparison of two recorded voices: one of the suspect and the other of the likely author of the suspected voice. Examinations as physical, linguistic, and graphic digitalization by computer, along with other modalities, make it possible to get to significant results. (ESPINDULA, 2005).

When committing a felony, the criminal can leave some remains behind, and that includes the voice. Sometimes, this voice is recorded in a media which preserves it, e.g., tape recording and others. Most of the times, the identity of the speaker can be hard to get due to some characteristics of the recording, as psychological and physiological conditions of the speaker. In order to minimize such effects, "different interdisciplinary techniques are applied, in the areas of linguistics, engineering, computing and forensic, having as a result a set of tests […]" (ESPINDULA, 2005, p.339). That is why Forensic Phonetics is part of a

[3] "Fonética Articulatória, estudando a fisiologia da fala e os sons produzidos pelo aparelho fonador; a Fonética Acústica, com o estudo físico das ondas sonoras produzidas pelo falante, suas medições e análises instrumentais; ou a Fonética Auditiva, com o estudo da percepção dos sons da fala. A Fonética intercepta todas essas áreas para que se possa compreender o fenômeno tão complexo do ato de falar." (Author's translation).

multidisciplinary area, and by means of a team of professionals – each one from a specific area – it is possible to get a satisfactory result in those examinations.

Speaker recognition is a generic term that encompasses Verification and Identification of speaker, being them two ways of discern people by their voices. (ESPINDULA, 2005). Though very similar, these two classifications present a slight difference, and that is what the next section will be about.

## 2.1 UNDERSTANDING IDENTIFICATION, VERIFICATION AND COMPARISON

When writing and/or reading about Forensic Phonetics, it is very common to come across some terms that at firstseem very similar – or even synonyms –, but we must be careful with our understanding about their significance in these studies, in order to avoid having a miscomprehension of a whole section. Different authors have been writing about the differences in these terminologies, stating the importance of having a clear understanding of which one is going to be used to give the right idea of the topic discussed. We are talking about identification, verification and comparison and, in the next paragraphs, we will present some definitions that may help avoid misunderstanding in this field of study.

Authors as Hollien (2002) and Nolan (1999) have been doing a distinction between two terms very commonly used in speaker recognition studies: verification and identification. Starting with the easiest definition, speaker verification states for the situation in which the person who is talking needs to be determined rather than what is being said (HOLLIEN, 2002). The choice of using this type of recognition is limited. As an example of this use, we have the access of a restricted area by voice command and, in situations like these, the speaker not only wants but also needs to be recognized from a database (HOLLIEN, 2002).

Regarding speaker identification, in this case, it is necessary to identify the speaker – obviously an unknown person – using voice analysis in contexts that, generally, reside some problems: background noise or poor signal quality – in a conversation on the telephone – and also some distortions in the voice, for example, in the case of the speech being uttered under a stressful situation. Such distortion can also be a result of fear or excitement (when committing a felony) or in the attempt of disguising the voice. In this case, the speaker does not want to be identified (HOLLIEN, 2002).

However, here in Brazil, the use of speaker verification[4]has been largely used in forensic context (BRAID, 2003). It is not possible to individualize a speaker within a huge database due to the nature of human speech and existing mechanisms of analysis. For this reason, there is a preference for using "verification" rather than "identification". The latter implies the idea of a precise recognition of voice, similar to a fingerprint, when the truth is that we have a variety of voice data, which would not makeit possible to assert that certain characteristics of voice could be foundonly in an appointed person. (CARNEIRO; GOMES, 2014).

Regardless of the term used, speaker identification can be performed by a professional/specialist or by somebody without technical knowledge, in other words, technical and naïve speaker identification.

## 2.2 TECHNICAL AND NON-TECHNICAL IDENTIFICATION

Non-technical identification (or naïve speaker recognition) is basically performed by non-trained people. "Naive speaker recognition in the forensic arena often goes under the heading 'Earwitness evidence'" (NOLAN, 1999,p.755). In situations like this, the criminal can be identified during the felony; in this case s/he must be somebody familiar to the witness. In other cases, the witness can be called to testify and identify which voice is from the author of the crime, among the presented voices.

In technical identification, the recording is recorded during the felony and it can be, for instance, a bomb threatening or hostage negotiation. The suspect is called to give his/her statement and, during the interview, the specialist tries to control the language context in order that the recording can be compared to the previous one – that of the crime. Regarding this type of identification, both auditory perception and acoustic analysis are used (NOLAN, 1999).

Specialists such as phoneticians or speech scientists can perform auditory perception. They analyze the speech sample, so as to verify the vowels and consonant quality, pronunciation and speaker's characteristics. In the environment of acoustic analysis, it is possible to obtain much more than only speech analysis it "[...] also reveals information which our auditory system obscures." (NOLAN, 1999,p.762).

[4]The Portuguese term is *verificação de locutor*.

In order to follow with this thesis, it is also necessary to clarify what the verifications cited above consist of, in order words, to understand voice quality.

## 3 WHAT IS VOICE QUALITY?

Every time we speak, our voice carries information about us as individuals, so that "speaker may sound young, or tired, or elated, or distracted". (KREIMAN; SIDTIS, 2011, p.2). In addition, when speaking, adults can reveal their biological sex (male or female), their age, and even where they come from. These last examples are part of the individual characteristics that somebody can have, as discussed previously in this work. Yet the question still remains: what is voice quality?

First of all, it is necessary to understand what voice is, that is why, according to scientific usage, the term *voice* "has a physical and physiological base that refers to the acoustic signal" (KREIMAN; SIDTIS, 2011, p.5). Therefore, voice quality refers to the perception of this voice by the listener, as if an impression that voice production can cause to us, listeners, as it is better explained by Kreiman and Sidtis when they state that "The term voice quality belongs properly to the realm of perception, and refers to how the voice sounds to a listener." (KREIMAN; SIDTIS, 2011, p.6).

It is important to bear in mind that pitch and loudness cannot be separated or excluded from the concept of vocal quality, according to the authors, they are highly important characteristics of voice, which listeners depend on to have their judgments on voice. (KREIMAN; SIDTIS, 2011). According to them:

> In the absence of empirical evidence for the validity of particular descriptors or dimensions, it is unclear why some should be included, and others excluded, in a descriptive framework for vocal quality. […] This makes it difficult to understand precisely how qualities differ from one another, or how seemingly similar qualities are related. […] The phonetic/articulatory features for voice quality proposed by Laver (1980, 2000; Ball, Esling, and Dickson, 2000) were designed in response to these limitations. (KREIMAN; SIDTIS, 2011, p.18/19)

One problem noted by the authors, is the fact that some studies refer to pitch and loudness as analogous, due to the fact that "these factor can be scaled unidimensionally, from low to high or faint to strong" and "in fact, some authors even treat pitch and fundamental frequency, or loudness and intensity, as synonymous in informal writing", (KREIMAN; SIDTIS, 2011, p.9). The authors continue:

> This creates the expectation that the acoustic correlates of quality should be fairly consistent from listener to listener, and that the same cues should operate across all voices, as fundamental frequency is the major cue to pitch, but not the only cue (see,

> for example, Thomas 1969), and intensity is the primary cue to loudness (Fletcher and Munson, 1933). However, quality is multidimensional. It cannot be successfully scaled unidimensionally; and because more than one possible cue to quality exists, the possibility of listener differences is always present, so that quality can never have fixed acoustic determinants (Kreiman, Gerratt, and Berke, 1994). (KREIMAN; SIDTIS, 2011, p.9).

Compared to the concepts for linguistics descriptions, voice quality started recently to have a descriptive theory. "With the current expansion of research in the area, this may be appropriate time to try to suggest the broad outline of an overall descriptive model of voice quality." (LAVER, 1968, p.43). For this reason, Laver suggested a Vocal Profile Analysis Scheme, also known as the VPAS protocol.

## 3.1 VPAS PROTOCOL

The VPAS protocol (Attachment A) was created in order to help having a perceptual evaluation of the voice's characteristics, i.e., "linguistic, paralinguistic and extralinguistic uses of the voice settings" (CAMARGO; MADUREIRA, 2008). This protocol was based on a phonetically grounded description of voice quality and it kept the tendency of being user-friendly and having a clear presentation.

Gillier (2011) points out that vocal quality can also provide clues to identify the speaker, because according to Laver (1980), each speaker has a tendency to use specific configurations of his/her vocal tract when producing his/her normal speech, (i.e. the natural form of speaking), because the anatomy of each speaker determines his/her way of speaking (vocal quality).

In his protocol, the author pointed two components of voice quality: the organic component and the articulatory setting component. The first one cannot be controlled by the speaker while the second can be totally controlled by the speaker. In his work, Laver described vocal quality by distinguishing two types of settings: supralaryngeal and phonatory, and they are described as deviations of the neutral configuration of the vocal tract. According to him, any change in supralaryngeal tract of the neutral configuration, causes changes in the longitudinal, latitudinal and velopharyngeal axis. As to longitudinalaxis changes, we can have the vertical movement of the larynx, that is, raised and lowered larynx, lip protrusion and the retraction and raising of the lower lip, in other words, labiodentalization. Regarding latitudinal axis settings, we have labial, lingual, pharyngeal, mandibular and velopharyngeal settings, the latter being responsible for the development of nasal and non-nasal voices.These settings

result in an effect of constriction or expansion along the vocal tract. The last is the phonation settings, they are related to the way the larynx can vibrate, originating falsetto, creak(y), whisper(y) and harsh voice, those being considered different from the 'modal' voice, in which occurs a periodical, efficient and without audible friction vibration of the true vocal folds. The term 'modal' is used to specify the neutral mode of vibration and it includes a range of fundamental frequencies that are used normally in speaking and singing. (LAVER apud GILLIER, 2011).

A Brazilian version of the protocol, BP-VPAS (Attachment B), was created by Camargo and Madureira (2008) from the protocol developed by Laver (1980). Applying the protocol in a Brazilian context came from the idea of improving and systematizing the material that was aimed at instructional purposes. When adapting the English protocol to Brazilian Portuguese, the authors made some modifications "considering more recent advances in speech science research related to studies on the physiology and speech signal research" (CAMARGO; MADUREIRA, 2008). It is also important to keep in mind that:

> As the VPAS poses that the same setting, depending on the degree, can be thought as normal or deviant, the recordings of voice qualities to constitute the database for the training were intended to illustrate the variety of voice quality settings and not the difference between normal and deviant voice qualities. (CAMARGO; MADUREIRA, 2008, p.58).

In other words, the protocol was not designed to point if the voices studied are "normal" or "disguised", the protocol helps to describe the characteristics noticed in both modes of voice.

Regarding the adaptations made by the Brazilian authors, we can mention that in the English version it is possible to see the setting "tremor" and "harsh" in "Laryngeal irregularity" – part of the "Phonation features" –, whereas in the Brazilian version of the protocol we only have "harsh". The same happens with "temporal organization" which has the term "dyplophonia" only in the English version.

Most of the vocal settings can cause an interference in the voice signal, resulting in change of fundamental frequency and that is easily perceived by our ears, "F0 displays correlations with an individual's physical characteristics and internal state, but is also finely controllable and highly salient to listeners" (KREIMAN; SIDTIS, 2011, p.139).

## 3.2 FUNDAMENTAL FREQUENCY

As stated previously in this work, human beings have an enormous capacity for voice recognition, independent of the situation. Frances McGehee was the first ear witness researcher, as in 1937 she performed a study that is considered the first significant experiment to the area of acoustic analysis. According to Tosi (1979), her research was based on long-term memory process. McGehee used a total of 31 males and 18 females, all university graduate students, who read a paragraph of 56 words. After that, a total of 740 undergraduate students participated as listeners in two sessions that varied from one day to five months between each session. In this experiment, she investigated – among other aspects – the effects of the disguise by changing the pitch and observed that alterations in F0 reduced considerably the recognition of the speakers. Künzel (2000) states that F0 is a widely popular parameter in forensic speaker recognition, once it is easy to be extracted and measured whether compared to other parameters, for instance, the formants[5].

Regarding the variation, at least in non tonal languages, fundamental frequency is not affected by the lexical content of the speech. Nevertheless, a lot of non-linguistic factors that affects F0 are already recognized as health and emotional states, discourse genre, environmental noises and whether the person is speaking on the telephone or not. One single person can show a great variation in F0depending on the situation and even during one only recording session, as stated in the researches of French (1994) and Braun (1995) (apud Kinoshita et al, 2009).

According to Kinoshita (2009):

> Since the inherent strength of forensic speaker recognition parameters relies primarily on the ratio of within- to between-speaker variance, F0 is considered not very effective as a FSR parameter, and although some, for instance (Nolan 1983, Sambur 1975, Jiang 1996), have suggested that F0 as a potential speaker identification parameter. (KINOSHITA, 2009, p.92)

When citing his own work of 2005, Kinoshita (2009) states that due to its poor variance ratio, the duration of LTF0 (Long-term formant) shows very poor strength as evidence in identification. Forensic analysis of F0 has yet concentrated in its long-term mean and standard deviation.

[5] According to Kent and Read (2002:302), a formant is a resonance of the vocal tract. A formant is specified by its center frequency (commonly called formant frequency) and bandwidth. Formants are denoted by integers that increase with the relative frequency location of the formants. F1 is the lowest-frequency formant, F2 is the next highest, and so on.

In this work we are referring to F0 (fundamental frequency) constantly, but what is F0 exactly? It is the smallest periodic component resulting from the vocal folds vibration. The opening between the vocal folds is called the glottis, and F0 is the first frequency produced there. Regarding the voice, F0 "determines the perception of speaker's *vocal pitch*" (KENT; READ, 2002, p.06), in other words, high or low pitch, and it also determines the intensity of the sound – weak or strong.

People who present a voice perceived as 'high voice' (layman term), are considered as presenting a high pitch – or pitched voice –, on the other hand, people with deep voice, as low-pitched voice. In a more technical statement: "A speaker with a high-pitched voice has a relatively high frequency of vocal fold vibration, and a speaker with a low-pitched voice has a relatively low frequency of vocal fold vibration" (KENT; READ, 2002, p.06). Yet independently of personal characteristics of the high- or low-pitched voice, people usually change the pitch during speech, as it (the pitch) is responsible for intonation. According to Roach (2009), there is not a completely satisfactory definition to this term of intonation, which is often mistaken with the term prosody that is not easily defined, as explained by Kent and Read (2002, p.229):

> One major disagreement is with the pair of terms, prosody and intonation. Some writers regard them as synonyms, while others mark an important distinction between them. […] Intonation is similar to prosody in that its parameters are vocal frequency, intensity, and duration, but intonation refers to a narrower range of phenomena, generally the patterns of pitch rises and falls and the patterns of stress in a given language. Prosody includes these effects but also embraces *tempo*(pause and lengthening), vocal effort, loudness, and other phenomena.

The speaker, when uttering speech, can produce a variation of pitch, once s/he used a diversity of tones to express her/himself. Acoustic analysis of F0 provides the drawing of graphics that is named as pitch curve analysis, in which the instantaneous values of fundamental frequency versus time can be found (Braid, 2003).

Fundamental frequency is, definitely, a promising phonetic parameter in speaker identification. As confirmed by Künzel (2000) it is the most known parameter used in voice recognition/speaker identification research. In his study, the author acknowledges the fact that the most frequent forms of disguise directly affect the F0. Other facts also had some influence or power over fundamental frequency, but in these cases, one cannot control it, as in disguise. We are talking about gender and age.

## 3.3 GENDER AND AGE

From the moment we are born, our vocal tract and respiratory system grow and mature. Though these patterns of development are quite the same for male and female during puberty, acoustic evidences points to the sexual dimorphism in the vocal tract. (KREIMAN; SIDTIS, 2011)[6]. Adding to that, studies made by Harrington et al (2007) state that, when analyzing the Christmas broadcast of Queen Elizabeth II, there is evidence of age-related changes due to the maturation of the vocal tract with increasing age "the fundamental frequency decreased from an average of 253Hz to 189Hz." (HARRINGTON et al, 2007). In addition, Kent and Read (2002) in their book cite a study of Peterson and Barney (1952) which includes acoustic data of vowels for women, men and children, "making it clear that acoustic values vary markedly with age and gender characteristics of speakers". (KENT; READ, 2002, p.189).

Regarding gender, boys usually have lower formant frequencies than girls, in adulthood, females average fundamental frequency drops to about 220Hz, while male voices drop to 130Hz by age 18. (TITZE apud KREIMAN; SIDTIS, 2011, P.114). In general, according to Gillier (2011), when citing studies from Stevens (2000),the in the value of F0 value in women is higher (average of 200Hz) if compared to men (average of 120Hz), but it could not be the only parameter to be analyzed if trying to identify the gender of the speaker. As stated by Kent and Read (2002, p.223) "women's voice may differ from men's in many ways".They also say that woman's voice is not a man's voice produced with higher fundamental frequency and formant frequencies and that fundamental frequency is not the only dimension which differs male from female voices. The authors provide a list of some attributes of women's voice: breathy, weak, less abrupt flow termination, higher fundamental frequency, and so on.

Regarding gender identification/differences, other clues can be used to help and they are not attached to the glottis specifically. For example, Gillier (2011) also points to a study from Schwartz (1968) in which it is shown that the production of voiceless fricatives has a higher energy concentration point in women than in men, due to the fact that female vocal tract is shorter.Gillier (2011) also cites the studies of Khuen and Moll (1976) stating that the difference in the female vocal tract also causes a difference in the rhythm of the speech, that is, women speak faster than men do. (GILLIER, 2011).

[6]Even though it is cited here, sexual dimorphism and deeper details about biological functions/changes of the body are not the focus in this work. It is possible to have a furtherreading in Kreiman and Sidtis (2011)

Still focusing on gender differences, due to some biological changes in the vocal folds of men and women (regarding the airflow), vocal folds of most women do not close completely during each phonatory cycle, resulting in what is called a "glottal gap". This gap causes to women's voice an aspiration noise making their voices being described as breathier or weaker if compared to men's. (KREIMAN; SIDTIS, 2011). The authors also add to that:

> Men and women also differ in patterns of frequency variation. Women's intonation is often described as "more dynamic" than men's: A number of authors have reported that women use a wider pitch range and change pitch more sharply, while men's voice are typically monotone [...]. Received wisdom states that women (particularly younger women) also use rising, question-like intonation on statements ("uptalk") more often than men do, and rising pitch contours have been associated with more effeminate speech in males [...]. (KREIMAN; SIDTIS, 2011, p.129)

The authors Kreiman and Sidtis (2011) also mention the fact that women have been considered to have a more precise pronunciation than men. As an example of the English language, they pronounce more accurately the individual difference sound for 'thin' and 'tin' and make less omission, for instance, producing the 'ing' of the last syllable in place of 'in'. Men are considered to omit or reduce vowels and simplify consonant clusters. (KREIMAN; SIDTIS, 2011). They continue citing studies that have been carried out with other languages regarding thesedifferences; they use speakers of Canadian French as an example.

Other differences related to both gender and age are mentioned by Kreiman and Sidtis (2011) when saying that, for instance, male speakers over 70 years old expend a greater amount of their lung volume on each syllable spoken, and produce fewer syllables per breath when compared to speakers under 25. The same does not happen with older women and younger women. When analyzing the variability of the fundamental frequency, older speakers differ significantly from young and middle-aged speakers. Due to physiological differences, male and female voices present a notable difference in fundamental frequency, so as in the perceived pitch. (KREIMAN; SIDTIS, 2011).

As mentioned before, F0 can be easily changed by speakers when producing a disguised speech and this parameter is also the most notable distinction between man and women (male and female voices). Disguises vary from speaker to speaker, for example, in an experiment held by Künzel (2000), he noticed that raised pitch was used only by males and lowered pitch only by females, but some characteristics are common choice to both genders.

Aside from those characteristics that provide us the differences that can occur in our vocal tract,helping us to identify the speaker, a person can change his/her own particular characteristic when trying to disguise the voice. Studies about voice disguise, according to

Gillier (2011), are not very large, though some works already points that the presence of disguise can cause interference in an acoustic parameter.

Before moving on, it is necessary to see what exactly is understood by disguise and how important it is in forensic phonetics.

## 3.4 DISGUISE

In varied research in forensic phonetics and speaker identification, disguise hasdrawn the attention of some researchers, even though they are still not common (Eriksson, 2010). According to this author, disguise can cause serious problems to speaker verification, especially in situations where the speaker uses electronic methods of voice manipulation. However, most of forensic cases involving disguises use simple methods, i.e., speakers rely only on their ability to voice alteration.

By disguise (in voice), it can be inferred that it isan action used consciously by the speaker to manipulate and change the voice in order to hide his/her identity and cannot be mistaken with pathological state of the person, e.g., flu, cold or a psychological state. (GILLIER, 2011).In his work, Künzel presents a scheme (Attachment C) with frequent types of disguise and complements that:

> The scheme […], which is far from exhaustive, relies on the definition of voice disguise as a voluntary change of features of voice, speech and language, produced by a speaker in order to conceal his identity. Consequently, involuntary changes of features such as hoarseness due to laryngitis, or hyponasality due to flu, are excluded. (KÜNZEL, 2000, p.150)

According to theaforementioned classification of disguise, we can have two comparisons to make: intentional vs. unintentional, and electronic vs. non-electronic. Basically, with them, we have some combinations: intentional-electronic and unintentional-electronic. The first one is simply by using some electronic device to disguise/alter the voice while the second is caused by some limitations of the transmission channel, for example, a low signal (or bad quality) of a telephone call.Unintentional-non-electronic and intentional-non-electronic, being the first combination the disguise already mentioned here, is caused by some emotional state, disease and so on and the second combination is caused when the person alters his/her vocal quality or during phonation, for example, using falsetto(RODMAN apud GILLIER, 2011). In this thesis, we are analyzing the combination intentional-non-electronic, as our subjects were oriented to use any disguise they wanted to during the recording session.

Besides the fact that most of disguises affect F0 directly (Künzel, 2000), fundamental frequency is also a feature "commonly used in forensic cases of speaker identification" (HUDSON et al, 2007).Though it is variable within a single speaker and when using a single type of disguise, F0 is "relatively undisturbed by background noise, […] its measurement is unaffected by telephone transmission." (NOLAN et al, 2007, p.1809). However, F0 measurements' interpretation cannot be isolated from population data for inter- and intra-speaker variation (HUDSON et al, 2007).

In order to recognize/detect a disguise it is necessary to have knowledge of the undisguised voice behavior of a particular speaker, without this it would not be possible to determine if a specific feature of the voice is disguised or natural. (KÜNZEL, 2000).

When talking about disguise in forensic context, another word emerges – efficiency – , in other words, the more efficient is the disguise the more difficult it is to identify the speaker. Nonetheless, some disguises are more efficient than others, i.e., they make more changes in the acoustic signal of the voice. This efficiency is also related to the individual that produces it, once the ability of disguising one's voice is directly related to and/or dependent on the capacity of manipulating the vocal tract, that is why one disguise can be considered efficient with one speaker and not efficient with another. (GILLIER, 2011).

Hollien (2002) provides us with a general idea of disguise, but not studying its behavior, the author studies how effective a disguise can be in speaker identification context. For example, a whisper eliminates or reduces information about the fundamental frequency or heard pitch, making it very difficult the process of identification.

Gillier (2001) presents a study on voice disguise in forensic phonetics. She studied not only fundamental frequency, but also the formants, the vowel triangle. The author begins by explaining forensic phonetics and then defining acoustics, characteristics and factors to attribute the "identity" of one's voice, speaker identification – in which gender differences are taken into account – and the auditory and acoustic analysis that provide parameters such as the vocal quality and the fundamental frequency. She then goes on explaining about disguise and how itcan affect voice identification. Focusing on the results of F0, once it is the subject of my work, Gillier states that fundamental frequency is an efficient parameter to distinguish speakers of the same sex and regarding the vowels, F0 values are very relevant to distinguish different sex, once female voices have higher average of F0 than male voices.

Künzel (2000) presents a study about the effects of voice disguise on speaking fundamental frequency. In his work, he presents the gender preferences for disguises. Among the results, he found that male chose to disguise their voices by lowering the F0 more often

than by raising it and the contrary was observed for women. This study follows the same idea of Masthoff's (1996) in which the goal was to determine preferred forms of disguise, not exactly focusing on the comparison of the participants' gender, but inevitably providing some comparison whatsoever. He found that most of the participants decided for using a single disguise rather than multiple disguises and that no disguise involved more than two parameters at a time. Raising F0 was seen in male and lowering F0 in females.

Bearing the importance of having more studies in the area, not only the study group was created, but also a series of research and works are being surrounded by the importance of studying the effects of a disguise in voice recognition.

## 4 THE RESEARCH

In 2012, a group of Professors and experts in forensic investigation created a study group at UTFPR with the purpose of carrying out research in forensic phonetics. After a yearofreadingsand discussions, the group recordedtheir own voicesto start some acoustic analysis. Subsequently some undergraduate students joined the group and we started carrying out a series of experiments, expanding the corpus, now having speech samples from 50 speakers.

So far, the experiments we are conducting have instructional purposes; in other words, the group is studying and developing an understanding in Forensic Phonetics and areas related to it. We started dividing the big group into small groups or pairs (a Professor and an undergraduate student or a Speech-language Pathologist and an undergraduate student). Each group or pair was responsible for studying one acoustic parameter of voice, and in my case, the parameter chosen was fundamental frequency.

The members of the group havepresented papersat seminars and conferencesand havea goodscientific productionfordisseminationof research conducted. I myself presented at CELLIP 2013 and had a paper published at the proceedings (KREMER, 2013)[7], another paper at the conference Language and The Law: Bridging Gaps (KREMER, 2013)[8],I submitted an article to a magazine having as co-author my Advisor (GOMES; KREMER, 2014)[9]. More recently, finishing my scientific initiation at the University, I submitted a report to a seminar and I received an Honorable Mention from the scientific committee of the event (KREMER,

[7]Articles published in: http://cellip.files.wordpress.com/2013/08/anais-do-xxi-cellip.pdf
[8] Poster and Talk presented at Language and The Law: Bridging Gaps Conference http://pattie.fe.up.pt/ocs/index.php/LL/LL2013/
[9] Article published in: http://www.revel.inf.br/files/1cf60caccb7480accf465bc241e04e76.pdf

2014)[10]. In this thesis, I present a continuation of my study in F0. I started by just comparing two groups (GG and CG)[11], all women, and the focus was attributing the efficiency of disguise when perceptually analyzing the voices and comparing with the graphic. Later on, I presented a talk and a poster, doing the same comparison, but now comparing three groups (GG, CG and WG), all women, and also taking into account disguise strategies. Following with the same analysis (women, disguise strategies, graphics and difference between F0 values), I wrote the article and the report mentioned above (GOMES; KREMER, 2014, KREMER, 2014).

## 4.1 METHODOLOGY

As stated before, the experiment has 50 participants. All of them recorded in our laboratory a text with 75 words created to simulate a hostage situation, they were asked to read it four times: twice in normal voice and twice with a freely chosen kind of disguise. The reading was recorded twice for both normal and disguise in case some of them presented problems with the recording, allowing us to have another sample available to perform our analysis. The participants were divided into 5 groups with 10 people each, separated by gender and age and labelled: Control Group (25-55 years old)[12] [Table 1], Women Group (30-55 years old) [Table 2], Girls Group (19-25 years old) [Table 3], Men Group (30-55 years old) [Table 4] and Boys Group (19-25 years old) [Table 5]. The Control Group – being called hitherto as CG – received its name based on our inferences that it can have most of the efficiency with disguises due to the studies these participants have been performing.

Data collection was performed in an acoustic laboratory and they were sampled at a rate of 44100 Hz. We used a Pentium Dual Core 5.300 2.60 GHz computer, 1.99GB of RAM, XP 2002 Service Pack 3, M-AudioFastTrack Pro 4x4 external sound card, microphone AKG 3000 B. The software Audacity was used to make all the recordings.

In order to have a good analysis of vowels, the text read by the participants was created having two-syllable-words stressed on first syllable with the seven vowels of Brazilian Portuguese between the plosive consonants /p/ /t/ /k/. The context of the text (Appendix A) was created to simulate a hostage situation. Different experiments are being

[10] Talk and Report presented at http://sicite.com.br/2014/
[11] All the analyses and abbreviations are going to be explained in details in further sections of this work.
[12] This group was the first recorded and was composed of linguistics professors and forensic experts.

conducted to analyze, for example, formants, intensity, vowel length, vocal quality and fundamental frequency, always comparing normal and disguised voice.

For this study, for the acoustic analysis, we selected just one sentence and we considered only the first recording of normal and disguised voice, since the participants could not hold the same rhythm of disguise twice, as Kent and Read (2002, p.14) state "[…] that the speech signal is lost quickly as its acoustic energy dissipates into the atmosphere. We may repeat what was said but we can never retrieve the original production", in other words, as Nolan (1995) also states: "No utterances are identical, even if they are by the same speaker".

For the aforementioned studies, only female voices, three members of the group – the author of this Thesis and two others – performed the perceptual analysis. As we knew the participants, our inferences were according to the recognition or not of the author of the disguise. Firstly, the evaluation was made individually and then we reached an agreement as to the assigned level: efficient and not efficient. For this Thesis, as male voices were included, we decided not toattribute efficiency of disguised voices,as we did not know all the participants involved.

In order to analyse the voice quality in the disguise, we used VPAS – Voice Profile Analysis Scheme, created by Laver (1968) and adapted to Brazilian Portuguese – BP-VPAS by Camargo and Madureira (2008). This scheme allows performing a perceptual analysis of voice quality configurations, e.g., vocal tract features: open jaw, lip rounding/protrusion, backed tongue body, etc.; phonation features: falsetto, harsh voice, whisper, etc. It is very important to highlight that we only considered the items of the protocol and not the scalar degrees, which go from 1 to 6, 1 to 3 as moderate and 4 to 6 as extreme (CAMARGO and MADUREIRA, 2008). The voice analysis was made by six members of the group and the result was checked by a Speech-language Pathologist (also a group member). The protocol was applied in the whole passage, i.e., the whole speech of each participant, different from the acoustic analysis, in which part of the speech was selected.

Regarding acoustic analysis, we used Praat, software developed by Paul Boersma and David Weenink– University of Amsterdam – in which was possible to have a spectrogram of the voice, extract the values of F0 as well as the graphics.

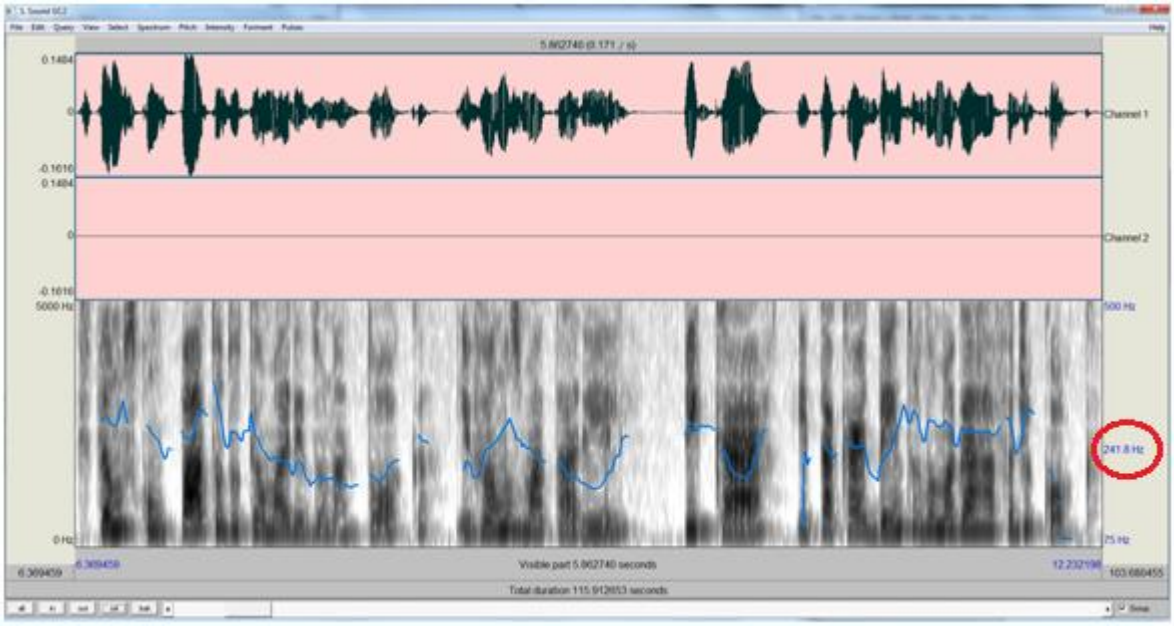



**Figure 1–Spectrogram of voice for acoustic analysis**

In figure 1 we have the spectrogram of the voice. In the lower part,we have a blue line showing the pitch curve of the selected part in pink (upper part). Circled in red, we have the value of F0 that we used to compare the voices: disguised and normal. In the example, we have a normal voice with F0 measuring 241Hz. We did the same with the disguised voice[13] and obtained the value of F0, then getting the difference between them, as we are going to show in further sections in the tables. After that, for each normal and disguise separately, we selected the option of drawing the pitch contour, which gave us the graphics to compare normal and disguised voices, as we are going to show in the next section.

After considering these three aspects – perceptual analysis of voice quality, F0 measurement and F0 curve analysis (graphics),we have some data to discuss and start understanding voice disguise.

## 4.2 RESULTS

Wedivided the 50 participants in five different groups, as stated previously, separating them by gender and age. In the following tables we present the vocal quality obtained with the perceptual analysis, the value of F0 of normal voice, the value of F0 of disguised voice and the difference between those values (F0 disguised – F0 normal).

[13]It is important to keep in mind that we selected part of the whole speech for studying F0.

**Table 1 – Control Group (CG): study group women from 25 to 55 years old**

| Participant | Vocal quality (perceptual analysis) | F0 – normal voice (Hz) | F0 – disguised voice (Hz) | F0 difference: disguised – normal (Hz) |
|---|---|---|---|---|
| CG1 | Backed tongue body<br>Tense vocal tract<br>Foreign accent | 230 | 288 | +58 |
| CG2 | Harsh voice | 250 | 213 | -37 |
| CG3 | Falsetto | 222 | 330 | +108 |
| CG4 | Backed and lowered tongue body | 260 | 307 | +47 |
| CG5 | High pitch[14]<br>Child's voice | 213 | 304 | +91 |
| CG6 | Open jaw | 226 | 238 | +12 |
| CG7 | Lip protrusion<br>Low pitch | 242 | 215 | -27 |
| CG8 | Tense and raised larynx<br>Nasal voice | 238 | 364 | +126 |
| CG9 | High pitch<br>Nasal voice | 262 | 371 | +109 |
| CG10 | Backed tongue body<br>High pitch | 231 | 275 | +44 |

Table 1 presents ten participants from the Control Group, all women between 25 to 55 years old and participants of the study group – linguistics professors and forensic experts. They were the first to record their voices. We have 8 participants who raised their F0 and 2 that lowered it, having 3 participants with a raise of more than+100Hz (CG8 presented the highest value of +126Hz and CG2and WG5 with -37Hz the lowest). Regarding vocal quality, we havemore variety if compared to other groups and also more resourceful than just using high or low pitch, for instance, we have the presence of a falsetto to achieve a raise in F0.

**Table 2 – Women Group (WG): Women from 30 to 55 years old**

| Participant | Vocal quality (perceptual analysis) | F0 – normal voice (Hz) | F0 – disguised voice (Hz) | F0 difference: disguised – normal (Hz) |
|---|---|---|---|---|
| WG1 | High pitch | 194 | 228 | +34 |
| WG2 | Nasal voice | 223 | 234 | +11 |
| WG3 | Tense larynx<br>Open jaw | 220 | 222 | +2 |
| WG4 | Lip protrusion<br>Low pitch | 217 | 201 | -16 |
| WG5 | Raised and tense larynx | 220 | 183 | -37 |
| WG6 | Low pitch | 234 | 198 | -36 |
| WG7 | Pinch noise | 265 | 278 | +13 |
| WG8 | Low pitch | 240 | 223 | -17 |
| WG9 | High pitch<br>Regional accent | 242 | 299 | +57 |
| WG10 | Backed tongue body<br>Foreign accent | 213 | 234 | +21 |

[14] As our study was based on the analysis of voice variation (disguise), when stating that a voice presented a high or low pitch, we are considering the variability setting that is given in BP-VPAS Protocol.

Table 2 is composed by women between 30 to 55 years old, they are students, professors, and forensic workers. We have 6 participants that raised their F0 and 4 that lowered it. The difference is very slight, having the highest raise in WG9 with +59Hz and the lowest in WG5 with -37Hz. Different from CG, some disguised were repeated, for instance, the use of low or high pitch.

**Table 3 – Girls Group (GG): girls from 19 to 25 years old**

| Participant | Vocal quality (perceptual analysis) | F0 – normal voice (Hz) | F0 – disguised voice (Hz) | F0 difference: disguised – normal (Hz) |
|---|---|---|---|---|
| GG1 | High pitch<br>Lip protrusion | 280 | 295 | +15 |
| GG2 | Low pitch<br>Covered mouth with a piece of cloth | 251 | 218 | -33 |
| GG3 | High pitch<br>Denasal | 282 | 302 | +20 |
| GG4 | Low pitch | 254 | 231 | -23 |
| GG5 | Lip protrusion | 232 | 236 | +4 |
| GG6 | Lip protrusion<br>Low pitch | 271 | 254 | -17 |
| GG7 | Fronted tongue body<br>High pitch | 269 | 334 | +65 |
| GG8 | Tense larynx | 249 | 334 | +76 |
| GG9 | Harsh voice | 214 | 187 | -27 |
| GG10 | Lip protrusion<br>Low pitch | 220 | 202 | -18 |

Table 3 gives the information of Girls Group, composed by students from 19 to 25 years old. Here we have half of the group raising their F0 and the other half lowering it. The differences were also very slight, with the highest raise in GG7 with +65Hz and the larger decrease in GG2 with -32Hz. Here we have also a repetition of disguises: lip protrusion and low pitch were widely used.

**Table 4 – Men Group (MG): men from 30 to 55 years old**

| Participant | Vocal quality (perceptual analysis) | F0 – normal voice (Hz) | F0 – disguised voice (Hz) | F0 difference: disguised – normal (Hz) |
|---|---|---|---|---|
| MG1 | Fronted tongue body<br>Tense larynx<br>Harsh voice | 144 | 187 | +67 |
| MG2 | Open jaw<br>High pitch<br>Pitched voice | 148 | 342 | +194 |
| MG3 | High pitch<br>Soothing voice | 135 | 303 | +168 |
| MG4 | Open jaw<br>Accent | 174 | 195 | +21 |
| MG5 | High pitch<br>Nasal | 177 | 216 | +39 |
| MG6 | Open jaw<br>High pitch<br>Nasal | 133 | 180 | +47 |
| MG7 | High pitch<br>Nasal<br>Accent | 111 | 212 | +101 |
| MG8 | Harsh voice | 165 | 183 | +18 |
| MG9 | Open jaw<br>High pitch<br>Nasal | 143 | 170 | +27 |
| MG10 | Tense larynx<br>Harsh voice | 139 | 328 | +189 |

Table 4 demonstrates the data ofthe Men Group, whose ages were from 30 to 55 years old. All of them raised their F0, some significantly, for instance, MG2 with +194Hz. The slightest raise is seen in MG8 with +18Hz. Again, we have a repetition of disguises, with 6 participants using the high pitch setting when disguising their voice and, as we can see, they were combined with other strategy.

**Table 5 – Boys Group (BG): boys from 19 to 25 years old**

| Participant | Vocal quality (perceptual analysis) | F0 – normal voice (Hz) | F0 – disguised voice (Hz) | F0 difference: disguised – normal (Hz) |
|---|---|---|---|---|
| BG1 | Without alteration | 148 | 156 | +8 |
| BG2 | High pitch<br>Creaky voice | 117 | 126 | +9 |
| BG3 | Lip protrusion<br>Fronted tongue body | 149 | 133 | -16 |
| BG4 | Raised tongue body<br>Nasal | 126 | 144 | +18 |
| BG5 | Low pitch | 138 | 132 | -6 |
| BG6 | Tense larynx | 120 | 106 | -14 |
| BG7 | Open jaw<br>Regional accent | 140 | 170 | +30 |
| BG8 | Open jaw<br>Harsh voice<br>Regional accent | 121 | 164 | +43 |
| BG9 | Open jaw<br>Nasal | 110 | 109 | -1 |
| BG10 | High pitch<br>Harsh voice | 102 | 171 | +69 |

Table 5 showsinformation about the Boys Group, with agesbetween 19 to 25 years old. We have 6 participants raising their F0 and 4 lowering it. In this group the differences were the smallest, the highest raise in BG10 is +69Hz and the lowest,in BG3 with -16Hz. Here the strategies did not repeat very often as in other groups, though some disguises appeared more than once: harsh voice was used by 2 participants and open jaw, by 3.

As demonstrated above, there are two groups of young participants– GG and BG – and two groups of adult participants– WG and MG – and the CG, composed by members of the study group, making up two groups of young voices (one male and one female) and three groups of adults voice (one male and two female). As we stated before, due to the fact of the participants from the CG being members of the study group and, therefore, aware of vocal quality studies, we decided to separate them from the others. It also helped us with some comparisons regarding inferences made in previous studies, for example, CG would have more efficient disguises when compared to the GG and WG, as the study showed, they had. (GOMES; KREMER, 2014).

After analyzing data, we came up with the conclusion that most of the participants raised theirvalue of F0 when disguising their voices – 35 participants out of 50 – having the groups CG and MG with most of them, 8 and 10 participants respectively. The participant MG2 raised his F0 in +194Hz (see Table 4), having a total of 342Hz in his disguised voice. On the other hand, the lowest F0 was found in a female voice, WG5 lowered her voice in -37Hz (see Table 2), having a total of 183Hz in her disguised voice.

When analyzing vocal quality, we can listthe disguises that were most used: harsh voice, change inaccent, nasalvoice, high pitch, low pitch and lip protrusion. Of course those vocal features were not used separately, having most of them being used in a combination with another vocal feature.

Participants who decided to use harsh voice were 7 in total, being 5in male groups (MG1, MG8, MG10, BG8 and BG10) and 2 in female (CG2 and GG9). A similar result appeared with the use of nasal, having 8 participants in total: 6 in male (MG5, MG6, MG7, MG9, BG4 and BG9)and 2 in female groups (CG8 WG2). Regarding the accent, we have 4 male participants (MG4, MG7, BG7 and BG8)and4 female participants (CG1, WG9, WG10 and GG8)changing the accent, only two (CG1 and WG10) faking a foreign accent and the others faking a regional Brazilian accent. High pitch was the vocal feature used by

most of the participants, a total of 17, having 9 in male (MG2, MG3, MG5, MG6, MG7, MG9, BG2, BG5 and BG10)and 8 in female groups (CG5, CG9, CG10, WG1, WG9, GG1, GG3 and GG7). On the other hand, participants who decided to lower their pitch were most in female groups, a total of 8 participants (CG7, WG4, WG6, WG8, GG2, GG4, GG6 and GG10) some of them combining this feature with lip protrusion and 1 in male group (BG5), totalizing 9 participants using this feature. Still regarding the use of high pitch, the concentration of its use was in the MG (see Table 4) with 6 participants, having BG (see Table 5), GG (see Table 3) and CG (see Table 1) with 3 participants in each and WG (see Table 2) with 2 participants.

In relation to harsh voice, it can be understood as an unpleasant and rough sound, providing the aperiodicity of the F0. (LAVER apud GILLIER, 2011).Most of the participants in this study used this feature alone, having only 3 participants combining it withanother feature: MG1 – fronted tongue body, tense larynx, harsh voice, MG10 – harsh voice and tense larynx, BG10 – harsh voice and high pitch, BG8 – harsh voice, regional accent and open jaw.

Concerning the features lip protrusion and low pitch, in most of disguises they appeared combined, having 5 participants using them, all women (CG7, WG4, GG2, GG6 and GG10). Using only low pitch we have 4 participants:3women (WG6, WG8 and GG4) and 1 man (BG5). Finally, we have 2 participants using only lip protrusion, one woman and one man (GG5 and BG3).In total, we had 7 participants using lip protrusion (when counting this feature isolated).

Regarding the age differences, it seems that most of the young group of female voices decided to use lip protrusion and low pitch as disguise, maybe to sound more threatening due to the context of the text that was read. The same did not happen with adult groups of female voices, in which we had less lip protrusion and low pitch. It also seems to reflect in the changes of F0, most of the raised F0 appeared among the adults, 24 against 11 young and although the difference is smaller, the lowering in F0 was more presented in young voices, more precisely in 9 participants against 6 adults.

As mentioned in the previous section, we performed the F0 curve analysis, that is, graphics obtained with Praat, providing a visual analysis of voices (see more in Appendix B). The aim of this analysis was to ‘see’ how different the normal and disguised voices would be, considering the change in F0. We could infer that big changes in F0 would produce a deep difference in the curves, and the disguise would sound efficient,on the

other hand, a small change in F0 would produce a slight change in the disguise, and it would not be efficient. Here are two examples, one with deep change and the other with a slight one:

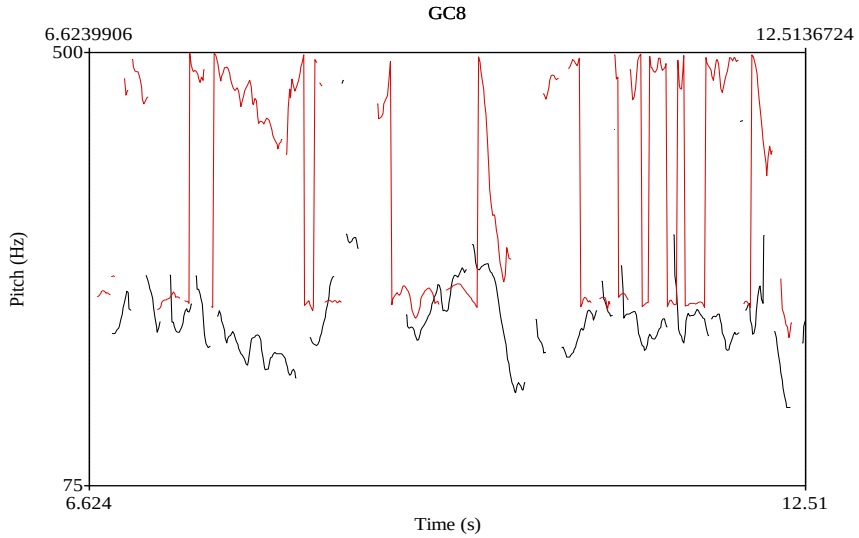



**Figure 2 - Disguise considered efficient**

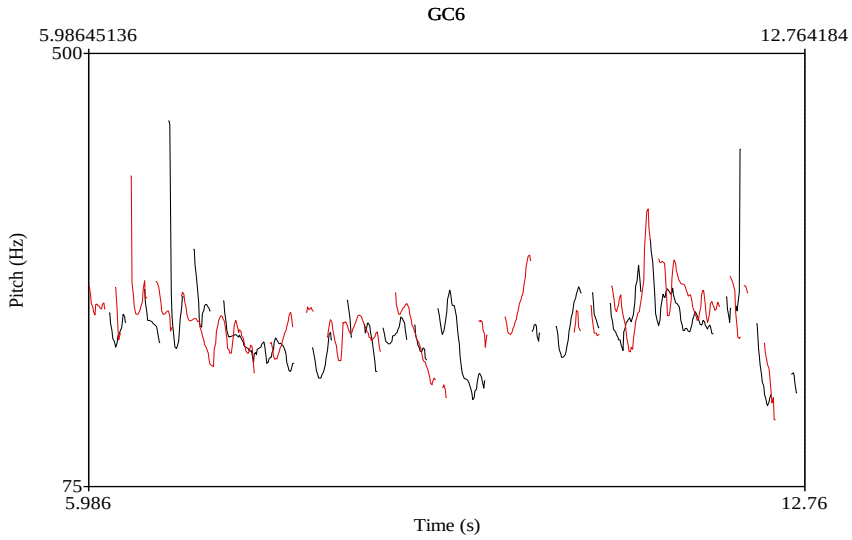



**Figure 3 - Disguise considered not efficient**

Both images show us visually the production of the sentence we had selected for this experiment, the black line represents the normal voice (the one without disguise) and the red line, the disguised voice. It is important to clarify that we are not seeking for efficiency in disguises in order to avoid having a subjective analysis, as the character of

this thesis is instructional, we are going to work now hypothetically just to understand how to interpret data, in this case, the graphs.

After collecting the data, we extracted the F0 values of both disguise and normal voices and the perceived voice quality, as we had shown in the five tables before. In our previous studies, we just analyzed female voices. Due to the fact that these samples were from the participants known by the members of the study group who performed the analysis, we decided to attribute the efficiency, that is, if we could say to whom each voice belonged to. So, we recurred to the graphs to corroborate our judgment of "efficient" or "not efficient".

The images show us two distinct voice samples, both belonging to CG[15] (Control Group). As we can see in figure 2, the red line is not following the same pattern as the black line at all. Neither the peaks nor valleys (highs or lows) are equal, seeming as if two distinct people are uttering the same sentence. In addition, the red line is very distant from the black line, showing that in this sample we had a raise of F0 when producing the disguise. In the case of figure 3, both red and black lines are following the same pattern, also the change of F0 was minimum, so if we have same F0 value for both voices (or very approximate) we could attribute easily the identity of the person. Of course, as stated before, we knew the participants and altogether with other analysis – for instance, this participant just adjust her voice by opening her jaw and nothing else, it caused only a slight change of F0 (see Table 1) – we could attribute the failure in the efficiency of this disguise.

For this work, the use of the graphics is to have a visual perspective of the changes in F0 perceived by us in auditory analysis. We are now going to observe how the "harsh voice" setting can cause a change, if any, in F0 in female and male voices.

[15] As the graphs were extracted from samples labeled in Brazilian Portuguese, the abbreviation that appears is "GC", standing for "GrupoControle" (Control Group in English).

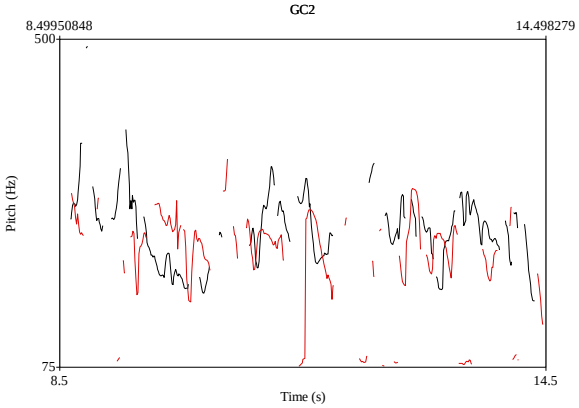



**Figure 4–Harsh voice (CG2)**

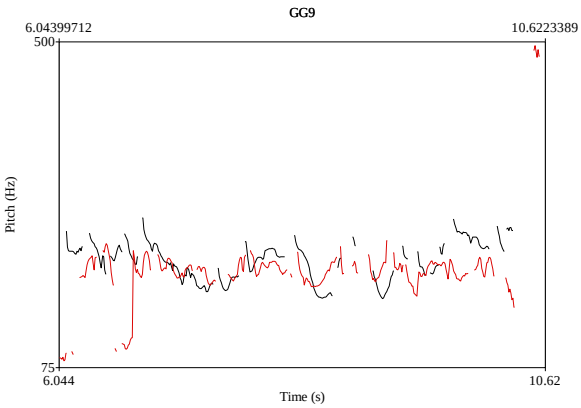



**Figure 5–Harsh voice (GG9)**

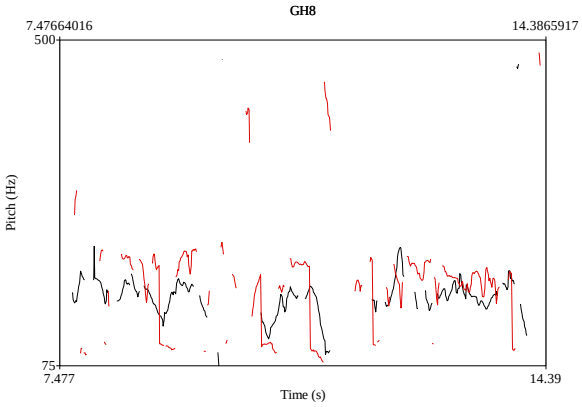



**Figure 6–Harsh voice (MG8)**

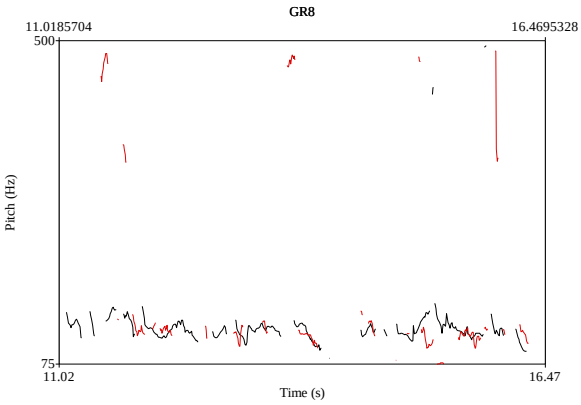



**Figure 7–Harsh voice (BG8)**

In the four examples, the participants chose to use only harsh voice, except for BG8 that with this phonation feature also used a vocal tract feature (open jaw). We can have a visual of what a harsh voice can cause to F0, or course, one could not tell just by looking the graphics which strategy of disguise was applied, it would be necessary to have a perceptual analysis (auditory) in order to do so. Although perceptually effective, visually the use of harsh voice do not cause a huge difference in F0, as we can see that in all graphics the red line (disguise) is very close to the black line (normal). The differences reside in the peaks and valleys of the curves as they practically do not follow the same pattern. We could not say, precisely, with this analysis alone that both voices, in each figure, belong to the same person, again, this type of inference can only be made after analysis combining acoustic, perceptual and numerical results.

## 4.3 DISCUSSION

Notably, all the participants altered the value of F0 when disguising their voices, some presented a big change and others, not so big. As Künzel (2000) had already pointed in his study, disguise affects/causes changes of spectrographic parameters such as, among others, fundamental frequency. It also happened because F0 is easy to be controlled, as stated by Kreiman and Sidtis (2011, p.138) that "we are also able to control F0 very finely".

Curiously, different from other participants using lip protrusion – a feature that caused a decrease in F0 – the participant GG5 presented otherwise, a small raise in F0 value. It can be explained by the fact that her choice of disguise was almost imperceptible; she used the lip protrusion feature very slightly.

Another fact that is worth mentioning is that harsh voice raised the F0 in male voices and lowered the F0 in female voices. We can see the highest raise in MG10 (see Table 4), with a raise in F0 of +189Hz. This characteristic also appeared in the study of Pontes et al (2002) in which they compared the F0 of normal, hoarse and rough[16] voices, the results showed that in male voices F0 had a more than expected rising in hoarse voices. We can also see the preference for using this parameter due to the context given to the participants – hostage situation –, making them trying to sound as threatening as they could, as harsh voice, according to Laver (1968, p.50) "is correlated with more aggressive, dominant, authoritative characteristics".

Consistent with the study of Masthoff (1996) in which he states that lowering of F0 was observed only for the females, in our study, when compared to men, a large number of women lowered their F0: 11 women against 3 men.

Though F0 is easy to be controlled (as it is responsible for tone and intensity), most of the participants decided to use other vocal settings when disguising their voices. For example, most speakers who decided to use high pitch or low pitch, combined it with another feature, lip protrusion with low pitch (as mentioned before) or high pitch with different accent.

Regarding the spectrograms (from Praat), their role were only to provide us a value of F0 and to extract the graphics, we did not use them to analyze voices in order to

[16] In this study, the authors used the term "hoarse" when referring to "harsh" voice. This second term was adopted in this Thesis to follow up with the protocol developed by Laver (1980).

identify them or to show efficiency in disguises. As defended by Braun (1995, p.148) “voice identifications based on spectrograms were found to be much less reliable than those based on aural perceptual judgments”. In addition, the focus of our study was to analyze fundamental frequency in contexts of normal and disguised voices and not to attribute efficiency on them.

## 5 CONCLUSION

The objective of this study was to compare normal and disguised voice through the analysis of fundamental frequency. In addition, our idea was to have an instructional research of forensic phonetics, to have some learning and understandingabout this field of study.

Results showed that most of the participants raised their F0, ones more significantly than others, being the highest raise F0 in a male participant and the lowest F0 in two female participants. The choice of raising F0 was more frequentamong men and the lowering F0 was more frequent among women, the same result as in Masthoff’s (1996) study. The use of harsh voice featuring caused a raise F0 in men’s voice and lowered the F0 in women’s voice, this characteristic also appeared in Pontes et al (2002). Concerning the age related to the type of disguise chosen, we could not find significant difference other than the use of lip protrusion with low pitch that was used the most by young females, making their F0 decrease. It also caused young voices to have lower F0 than adults and, certainly, more adults having higher F0 than the young.

This thesis was developed after previous studies of fundamental frequency made by our study group, but the main goal was to learn how to deal with this new area of linguistics: forensic phonetics. Also, the objective was purely instructional and our learning development generated a diversity of scientific studies as we mentioned before. While studying, we noticed the lack of research in Brazilian context as well as studies in Brazilian Portuguese (one of the reasons for submitting a study in this language to the scientific magazine ReVEL). Further ideas are to improve our range of studies in Brazil as well as presenting internationally these studies can stimulate more research in the area.

## REFERENCES


BRAID, A. C. M. **Fonética Forense:** Tratado de perícias criminalísticas. Campinas, SP: Millenium, 2003.

BRAUN, A. **Procedures and perspectives in forensic phonetics.** *Phonetic Sciences: Proceeedings of the XIIIth International Congress of …*, ICPhS ’95, Stockholm, Sweden, v. 3, pp. 146-1153, 1995

CAMARGO, Z., MADUREIRA, S. **Voice quality analysis from a phonetic perspective**: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). Campinas, SP, 2008.

CARNEIRO, D. O., GOMES, M. L.C. **A fonética forense no Brasil**: cenários e atores, 2014.

ESPINDULA, A. **Perícia criminal e cível:** uma visão geral para peritos e usuários da perícia. 2ª edição. Campinas, SP: MilleniumEditora, 2005. p.338-341.

ERIKSSON, A. **The Disguised Voice**: Imitating Accents or Speech Styles and Impersonating Individuals. Edinburg: Kapitel, p.86 – 96, 2010

GILLIER, R. **O disfarce da voz em fonética forense.** 113 f. Thesis (Master’s) – Romanic and General Linguistics Department, University of Lisbon, 2011.

GOMES, M. L. C., KREMER, R. L.**A eficiência do disfarce em vozes femininas: uma análise da frequência fundamental.**ReVEL – Revista Virtual de Estudos da Linguagem, vol.12 (23), Rio Grande do Sul, 2014.

HARRINGTON, J., PALETHORPE, S., WATSON, C. I. **Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers.**in Proc. INTERSPEECH, 2007, p.2753-2756.

HOLLIEN, H. **Forensic Voice Identification.** London, UK: Academic Press, 2002.

HUDSON, T., JONG, G., McDOUGALL, K., HARRISON, P., NOLAN, F. **F0 statistics of 100 young male speakers of standard Southern British English.**ICPhS XVI, Germany, p.1809 – 1812, 2007.

KENT, R. D., READ, C. **The Acoustic Analysis of Speech**. Connecticut, USA: Thomson Learning, 2002.

KREIMAN, J., SIDTIS, D. **Foundations of voice studies:** an interdisciplinary approach to voice production and perception. West Sussex, UK: Wiley-Blackwell, 2011.

KINOSHITA, Y., ISHIHARA, S., ROSE, P. **Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition**. Equinox Publishing Online, p. 92, 2009.

KÜNZEL, H. J. **Effects on voice disguise on speaking fundamental frequency**. Birmingham, UK: University of Birmingham Press, 2000.

LAVER, J. D. M. **Voice quality and indexical information**. In: *British journal of disorders of communication.* Edinburgh, UK: University of Edinburgh, 1968.

MASTHOFF, H. **A report on voice disguise experiment.** Forensic Linguistics, vol.3 (1), Equinox Publishing Online, UK, pp. 160-167, 1996

NOLAN, F. **Can the definition of each speaker be expected to come from the laboratory in the next decades?***Phonetic Sciences: Proceeedings of the XIIIth International Congress of ...*, ICPhS '95, Stockholm, Sweden, v. 2, pp. 130-137, 13-19 August, 1995

_____. **Speaker Recognition and Forensic Phonetics**. In: HARDCASTLE, William J.; LAVER, John (eds.).*The Handbook of Phonetic Sciences.* The United States: BlackwellPublishing, 1999.

PONTES, P. A. L., VIEIRA, V. P., GONÇALVES, M. I. R., PONTES, A.A.L. **Características das vozes roucas, ásperas e normais:** análise acústica espectrográfica comparativa. *RevistaBrasileira de Otorrinolaringologia,vol.68,* pp. 182-188, março/abril, 2002.

ROACH, P. **English Phonetics and Phonology:** A practical course. Cambridge: CUP, 2009.

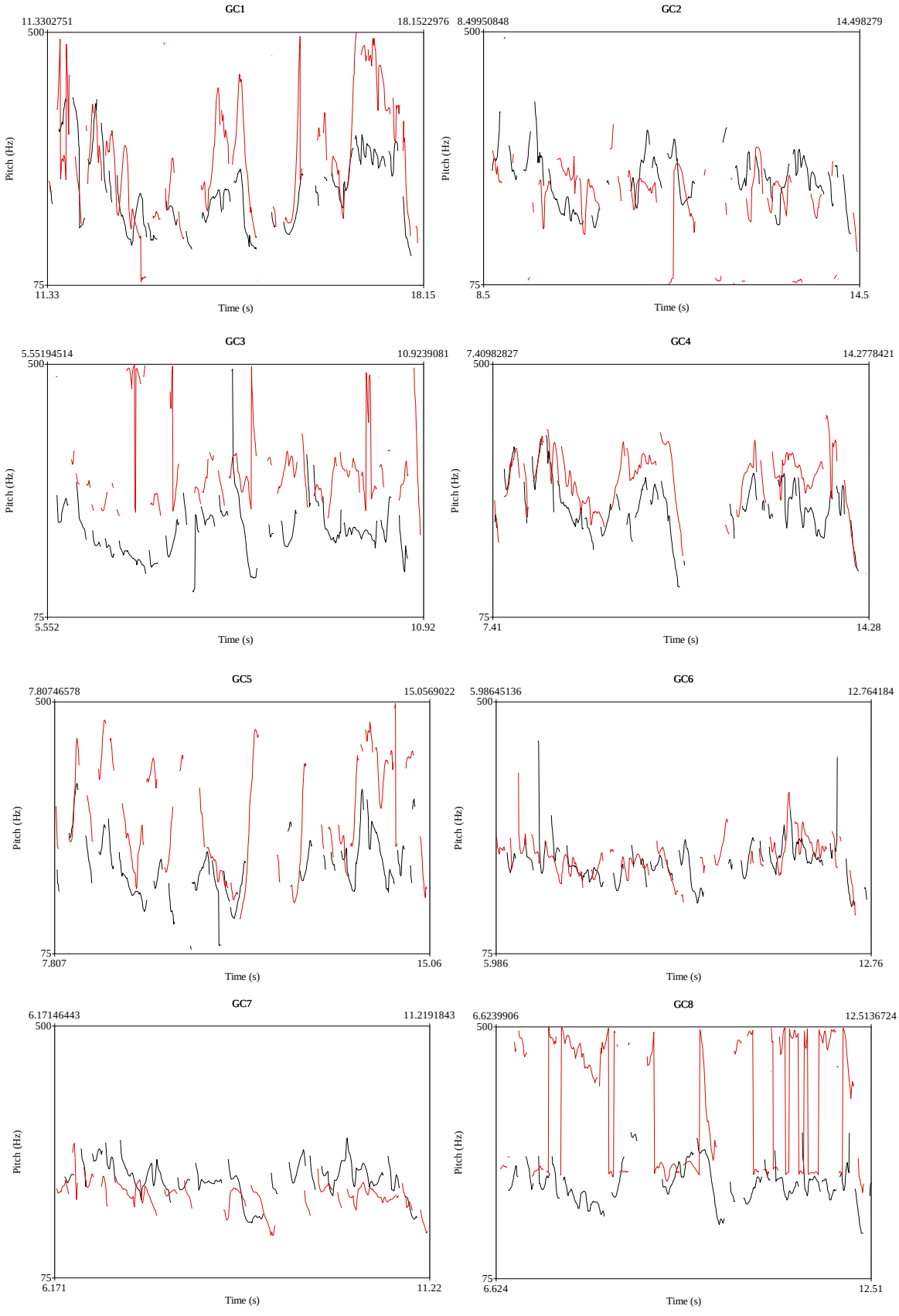
TOSI, O. **Voice identification**: theory and legal applications. Baltimore, MD: University Park Press, 1979

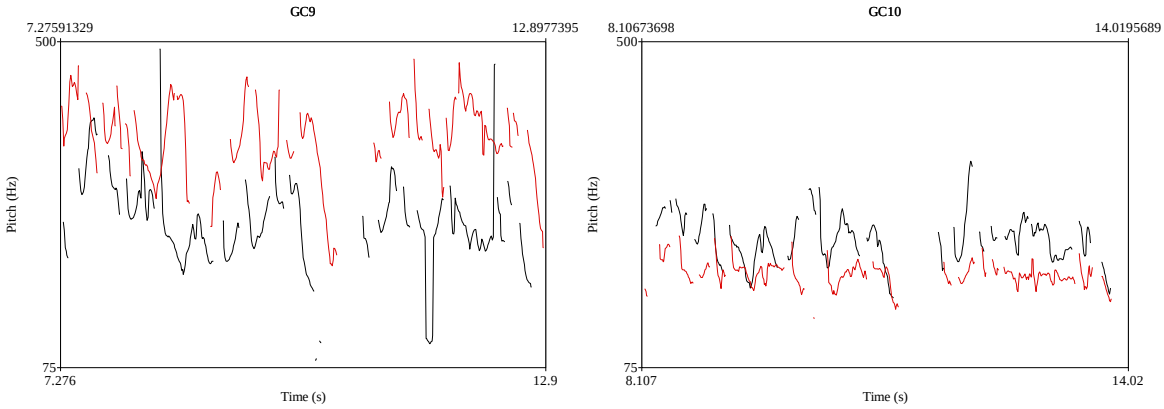
**APPENDIX A –** Simulating ahostage negotiation

“Alô?! Quero falar com a Dona Teca. Dona Teca, aqui fala o Capeta, estamos com o pato do teu marido na toca e vamos matar ele, picar e tacar dentro de uma garrafa de coca. Quer salvar ele, então não me deixa puto. Quero um milhão, coloca tudo num pacote perto da pipa do cateto e se pica. Não faz caca, senão vai levar toco. Não chama a polícia, senão te cutuco e espeto o teu coco”.

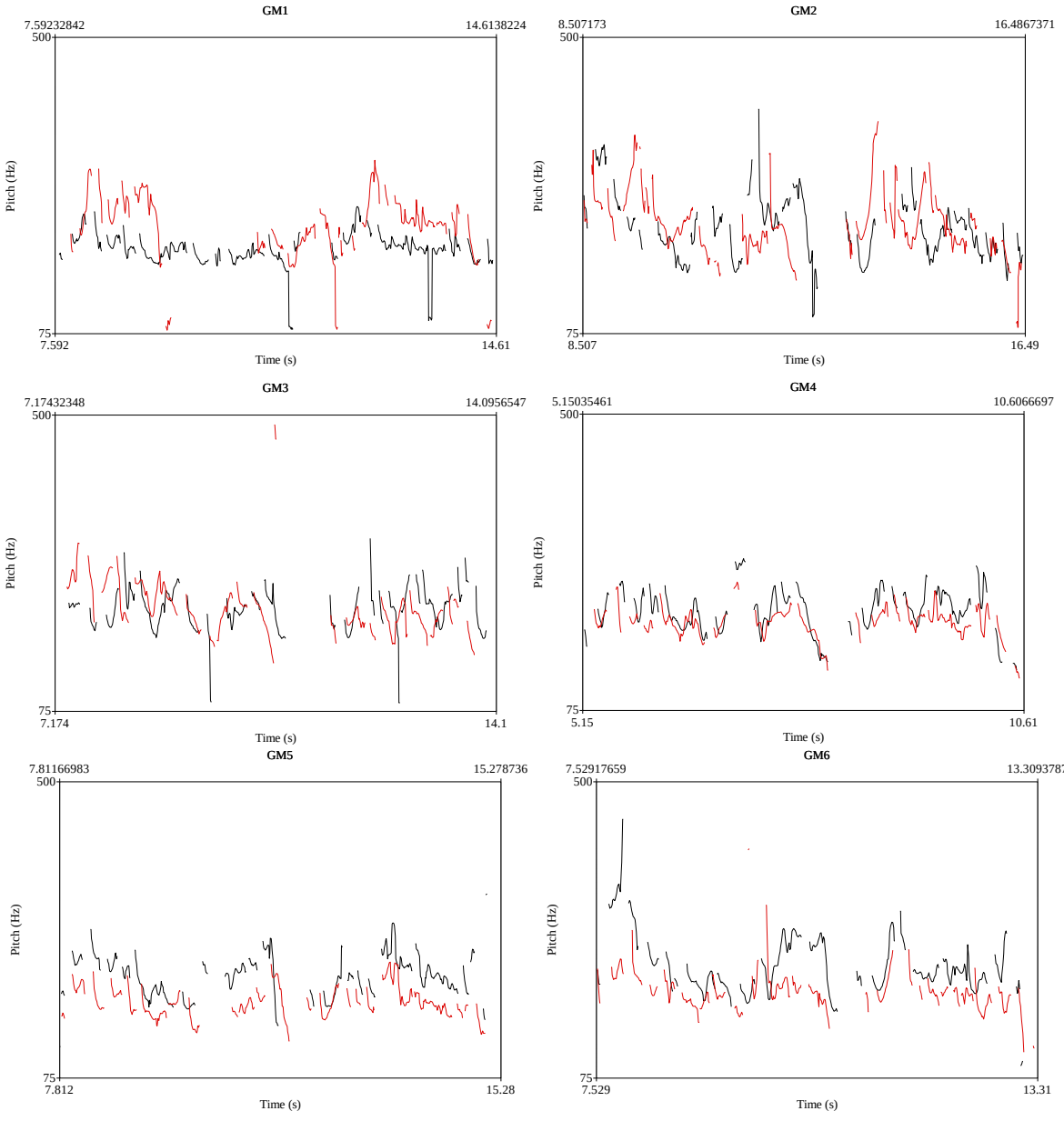
**APPENDIX B –** F0 curves (graphs)

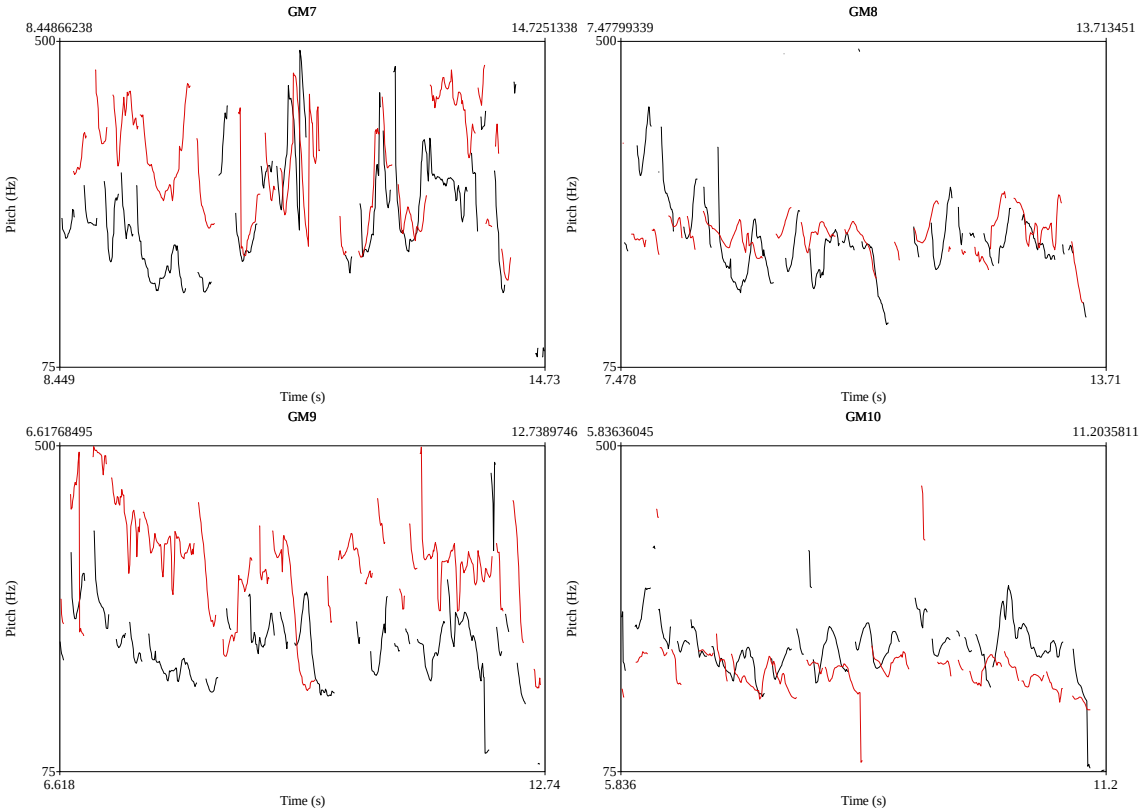
# CONTROL GROUP



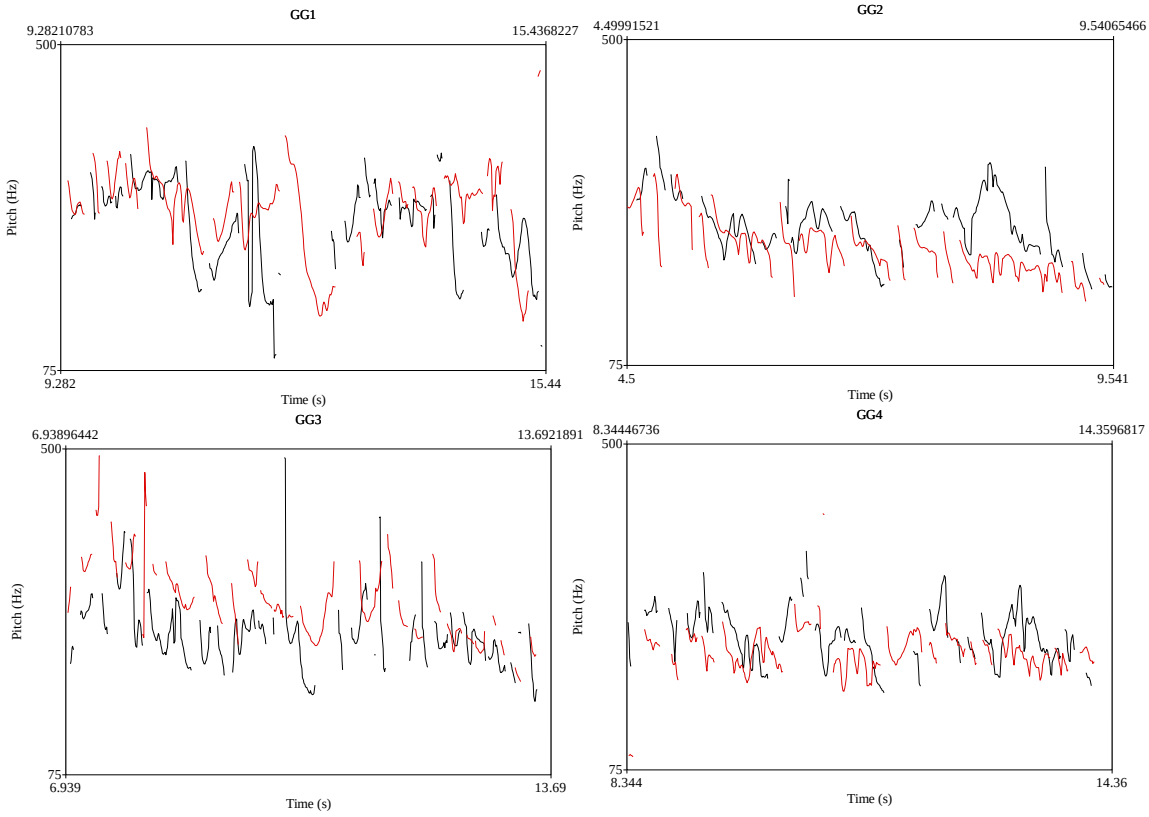

GC9
7.27591329
12.8977395
500
75
Pitch (Hz)
7.276
12.9
Time (s)
GC10
8.10673698
14.0195689
500
75
Pitch (Hz)
8.107
14.02
Time (s)

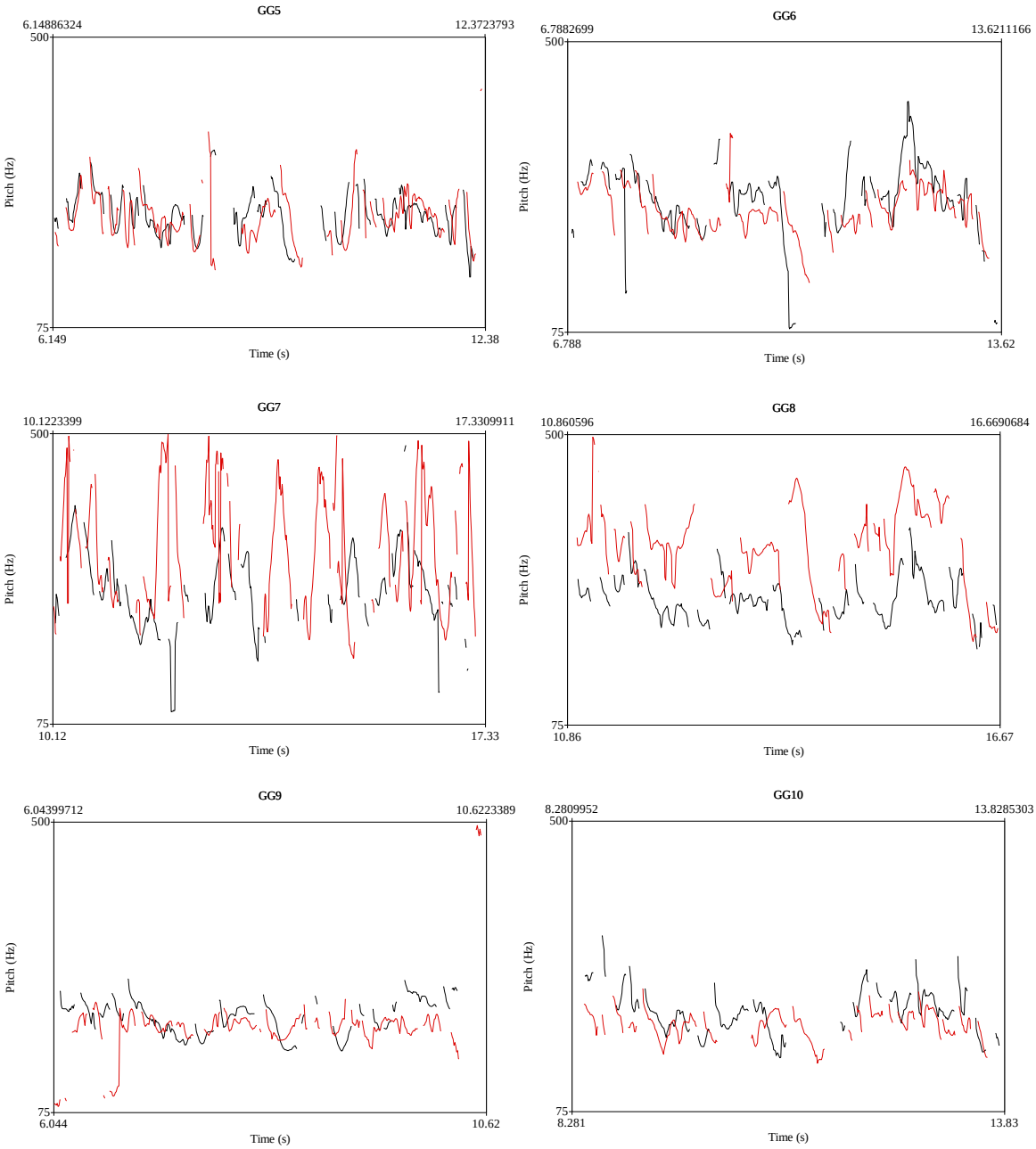
# WOMEN GROUP


GM1
7.59232842
14.6138224
500
75
Pitch (Hz)
7.592
14.61
Time (s)
GM2
8.507173
16.4867371
500
75
Pitch (Hz)
8.507
16.49
Time (s)
GM3
7.17432348
14.0956547
500
75
Pitch (Hz)
7.174
14.1
Time (s)
GM4
5.15035461
10.6066697
500
75
Pitch (Hz)
5.15
10.61
Time (s)
GM5
7.81166983
15.278736
500
75
Pitch (Hz)
7.812
15.28
Time (s)
GM6
7.52917659
13.3093787
500
75
Pitch (Hz)
7.529
13.31
Time (s)

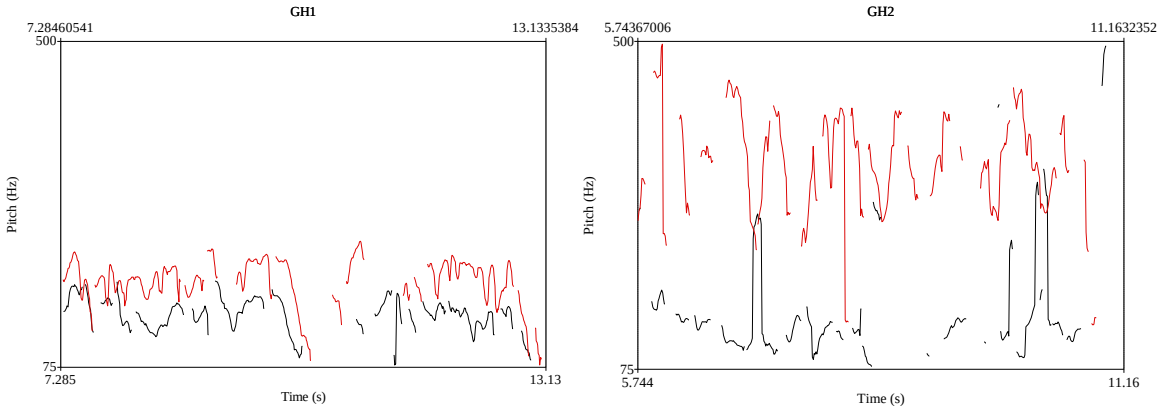

GM7
8.44866238
14.7251338
500
75
Pitch (Hz)
8.449
14.73
Time (s)
GM8
7.47799339
13.713451
500
75
Pitch (Hz)
7.478
13.71
Time (s)
GM9
6.61768495
12.7389746
500
75
Pitch (Hz)
6.618
12.74
Time (s)
GM10
5.83636045
11.2035811
500
75
Pitch (Hz)
5.836
11.2
Time (s)

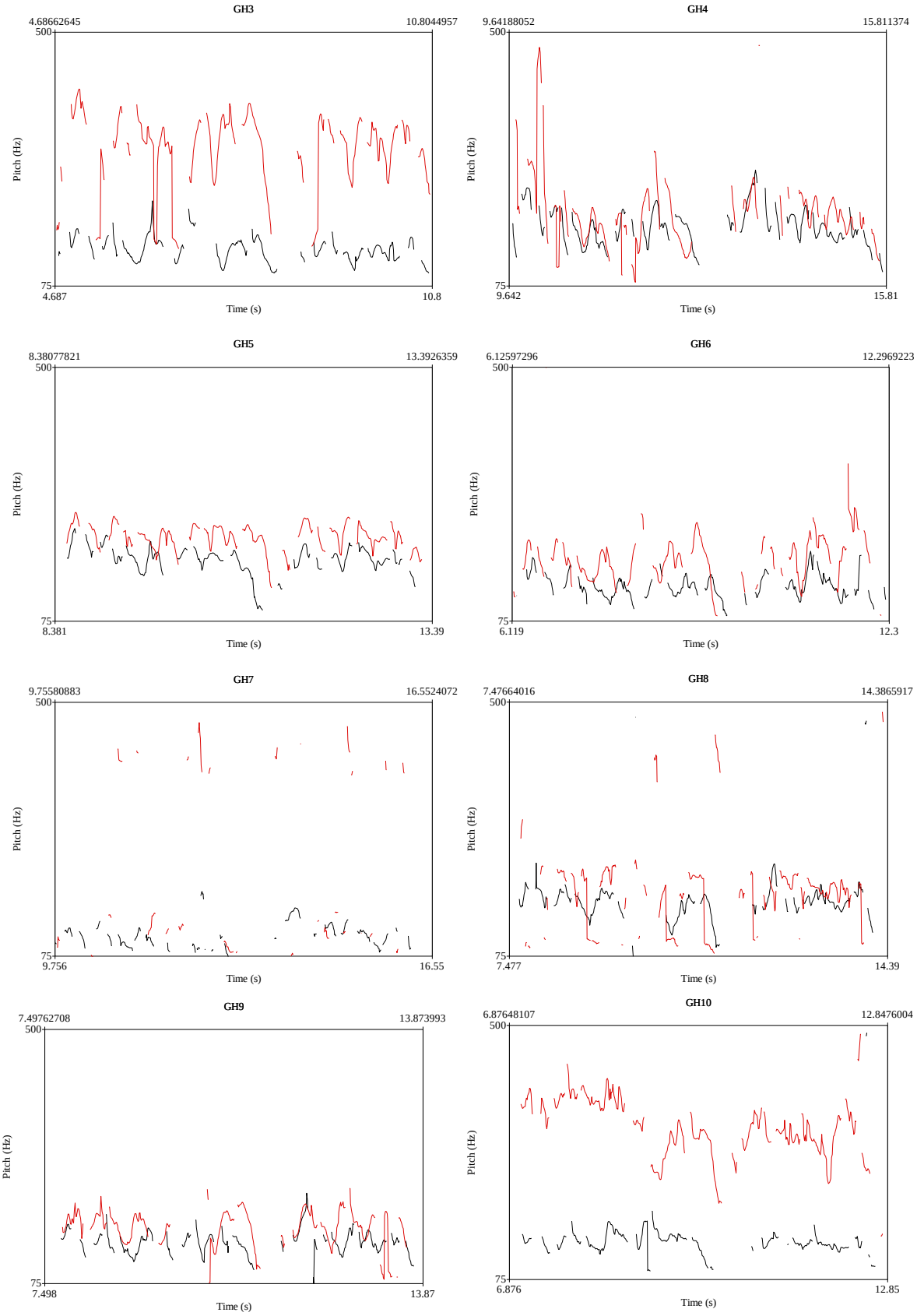
# GIRLS GROUP


GG1
9.28210783
15.4368227
500
75
Pitch (Hz)
9.282
15.44
Time (s)
GG2
4.49991521
9.54065466
500
75
Pitch (Hz)
4.5
9.541
Time (s)
GG3
6.93896442
13.6921891
500
75
Pitch (Hz)
6.939
13.69
Time (s)
GG4
8.34446736
14.3596817
500
75
Pitch (Hz)
8.344
14.36
Time (s)

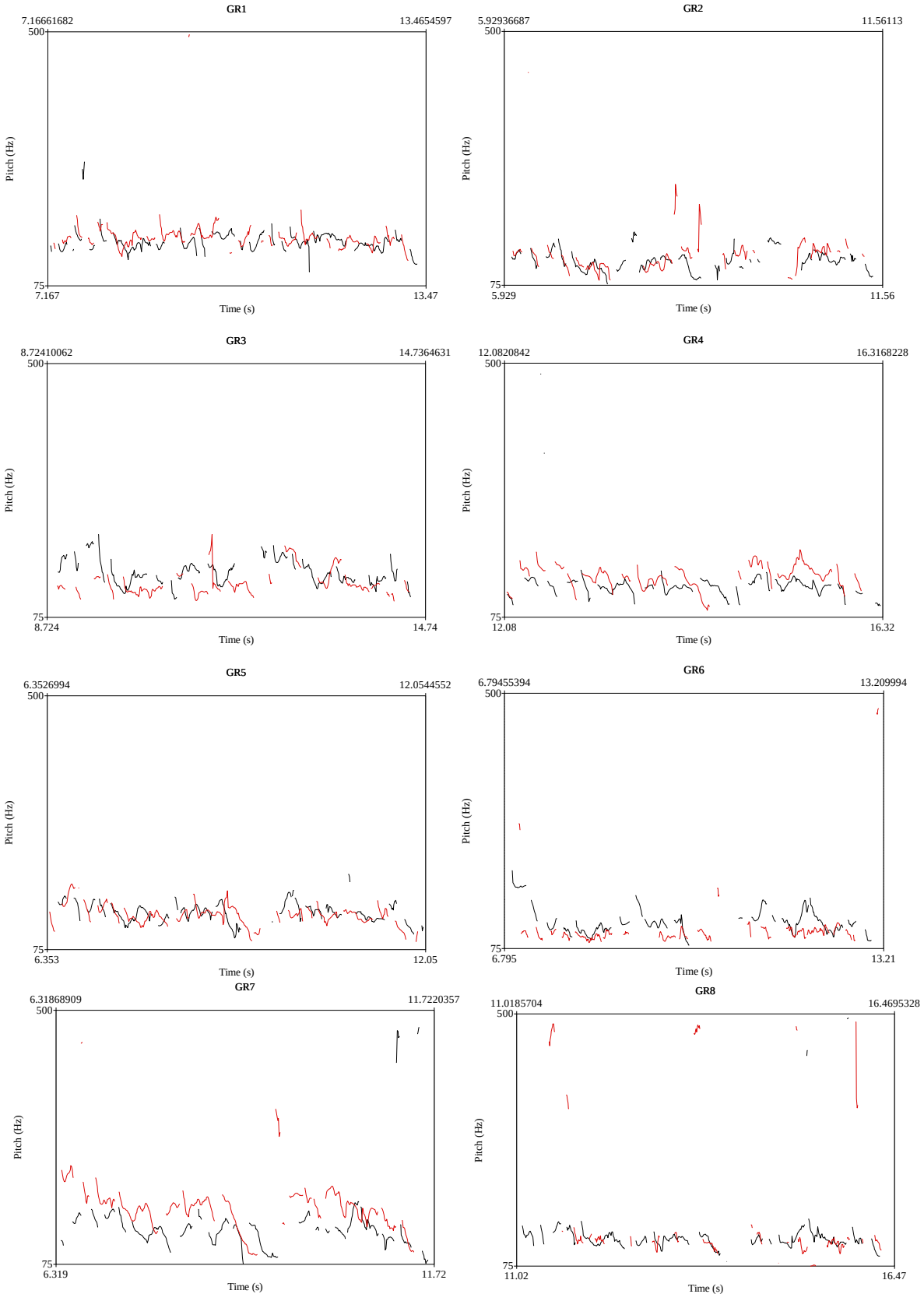

GG5
6.14886324
12.3723793
500
75
6.149
12.38
Pitch (Hz)
Time (s)
GG6
6.7882699
13.6211166
500
75
6.788
13.62
Pitch (Hz)
Time (s)
GG7
10.1223399
17.3309911
500
75
10.12
17.33
Pitch (Hz)
Time (s)
GG8
10.860596
16.6690684
500
75
10.86
16.67
Pitch (Hz)
Time (s)
GG9
6.04399712
10.6223389
500
75
6.044
10.62
Pitch (Hz)
Time (s)
GG10
8.2809952
13.8285303
500
75
8.281
13.83
Pitch (Hz)
Time (s)

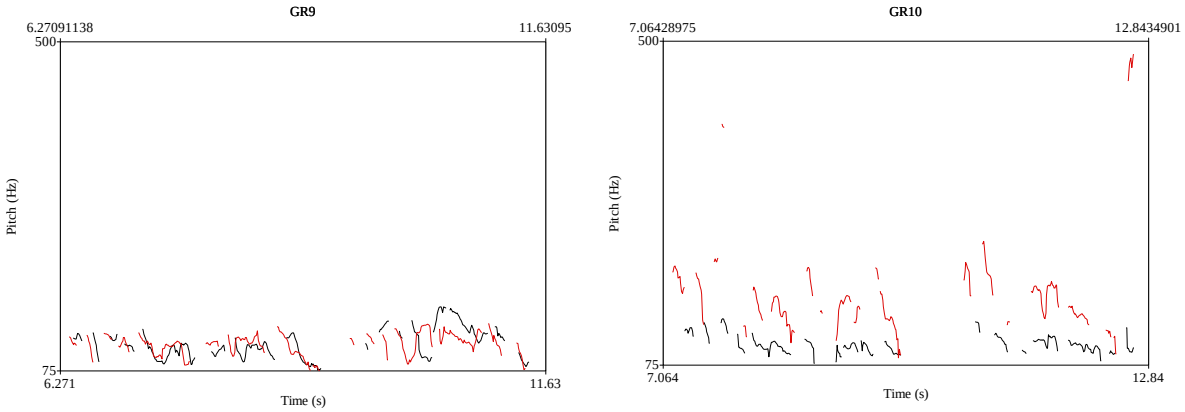
# MEN GROUP


GH1
7.28460541
13.1335384
500
75
7.285
13.13
Pitch (Hz)
Time (s)
GH2
5.74367006
11.1632352
500
75
5.744
11.16
Pitch (Hz)
Time (s)


GH3
4.68662645
10.8044957
500
75
Pitch (Hz)
4.687
10.8
Time (s)
GH4
9.64188052
15.811374
500
75
Pitch (Hz)
9.642
15.81
Time (s)
GH5
8.38077821
13.3926359
500
75
Pitch (Hz)
8.381
13.39
Time (s)
GH6
6.12597296
12.2969223
500
75
Pitch (Hz)
6.119
12.3
Time (s)
GH7
9.75580883
16.5524072
500
75
Pitch (Hz)
9.756
16.55
Time (s)
GH8
7.47664016
14.3865917
500
75
Pitch (Hz)
7.477
14.39
Time (s)
GH9
7.49762708
13.873993
500
75
Pitch (Hz)
7.498
13.87
Time (s)
GH10
6.87648107
12.8476004
500
75
Pitch (Hz)
6.876
12.85
Time (s)

# BOYSGROUP


GR1
7.16661682
13.4654597
500
75
Pitch (Hz)
7.167
13.47
Time (s)
GR2
5.92936687
11.56113
500
75
Pitch (Hz)
5.929
11.56
Time (s)
GR3
8.72410062
14.7364631
500
75
Pitch (Hz)
8.724
14.74
Time (s)
GR4
12.0820842
16.3168228
500
75
Pitch (Hz)
12.08
16.32
Time (s)
GR5
6.3526994
12.0544552
500
75
Pitch (Hz)
6.353
12.05
Time (s)
GR6
6.79455394
13.209994
500
75
Pitch (Hz)
6.795
13.21
Time (s)
GR7
6.31868909
11.7220357
500
75
Pitch (Hz)
6.319
11.72
Time (s)
GR8
11.0185704
16.4695328
500
75
Pitch (Hz)
11.02
16.47
Time (s)


GR9
6.27091138
11.63095
500
75
Pitch (Hz)
6.271
11.63
Time (s)
GR10
7.06428975
12.8434901
500
75
Pitch (Hz)
7.064
12.84
Time (s)

**ATTACHMENT A –** The VPAS Protocol – Voice Profile Analysis Scheme

| | FIRST PASS | | SECOND PASS | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Neutral | Non-neutral | SETTING | moderate | | | extreme | | |
| | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| **A. VOCAL TRACT FEATURES** | | | | | | | | | |
| **1.Labial** | | | Lip rounding/protrusion | | | | | | |
| | | | Lip spreading | | | | | | |
| | | | Labiodentalization | | | | | | |
| | | | Minimised range | | | | | | |
| | | | Extensive range | | | | | | |
| **2. Mandibular** | | | Close jaw | | | | | | |
| | | | Open jaw | | | | | | |
| | | | Protruded jaw | | | | | | |
| | | | Extensive range | | | | | | |
| | | | Minimised range | | | | | | |
| **3.Lingual tip/blade** | | | Advanced tip/blade | | | | | | |
| | | | Retracted tip/blade | | | | | | |
| **4. Lingual body** | | | Fronted tongue body | | | | | | |
| | | | Backed tongue body | | | | | | |
| | | | Raised tongue body | | | | | | |
| | | | Lowered tongue body | | | | | | |
| | | | Extensive range | | | | | | |
| | | | Minimised range | | | | | | |
| **5.Pharyngeal** | | | Pharyngeal constriction | | | | | | |
| | | | Pharyngeal expansion | | | | | | |
| **6.Velopharyngeal** | | | Audible nasal escape | | | | | | |
| | | | Nasal | | | | | | |
| | | | Denasal | | | | | | |
| **7. Larynx height** | | | Raised Larynx | | | | | | |
| | | | Lowered Larynx | | | | | | |
| **B. OVERALL MUSCULAR TENSION** | | | | | | | | | |
| **8. Vocal tract tension** | | | Tense vocal tract | | | | | | |
| **9. Laryngeal tension** | | | Tense larynx | | | | | | |
| | | | Lax larynx | | | | | | |
| **C. PHONATION FEATURES** | | | | | | | | | |
| | **SETTING** | | **Present** | | | **Scalar Degree** | | | |
| | | | **Neutral** | **Non-neutral** | | **Moderate** | | **Extreme** | |
| | | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| **10. Voicing type** | Voice | | | | | | | | | |
| | Falsetto | | | | | | | | | |
| | Creak | | | | | | | | | |
| | Creaky | | | | | | | | | |
| **11. Laryngeal frication** | Whisper | | | | | | | | | |
| | Whispery | | | | | | | | | |
| **12.Laryngeal irregularity** | Harsh | | | | | | | | | |
| | Tremor | | | | | | | | | |

| | | Neutral | SETTING | moderate | | | extreme | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| **D. PROSODIC FEATURES** | | | | | | | | | |
| **13.Pitch** | **Mean** | | High | | | | | | |
| | | | Low | | | | | | |
| | **Range** | | Minimised range | | | | | | |
| | | | Extensive range | | | | | | |
| | **Variability** | | High | | | | | | |
| | | | Low | | | | | | |
| **14. Loudness** | **Mean** | | High | | | | | | |
| | | | Low | | | | | | |
| | **Range** | | Extensive range | | | | | | |
| | | | Minimised range | | | | | | |
| | **Variability** | | High | | | | | | |
| | | | Low | | | | | | |
| **E. TEMPORAL ORGANIZATION** | | | | | | | | | |
| **15. Continuity** | | | Interrupted | | | | | | |
| **16. Rate** | | | Fast | | | | | | |
| | | | Slow | | | | | | |
| **F. OTHER FEATURES** | | | | | | | | | |
| **17. Respiratory support** | | | Adequate | | | | | | |
| | | | Inadequate | | | | | | |
| **18. Dyplophonia** | | | Absent | | | | | | |
| | | | Present | | | | | | |

**ATTACHMENTB –** The BP-VPAS Protocol– Voice Profile Analysis Scheme for Brazilian Portuguese

| QUALIDADE VOCAL | PRIMEIRA PASSADA | | SEGUNDA PASSADA | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Neutro | Não neutro | AJUSTE | Moderado | | | Extremo | | |
| | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| A. ELEMENTOS DO TRATO VOCAL | | | | | | | | | |
| 1.Labial | | | Arredondamento/protrusão | | | | | | |
| | | | Estiramento | | | | | | |
| | | | Labiodentalização | | | | | | |
| | | | Extensão diminuída | | | | | | |
| | | | Extensão aumentada | | | | | | |
| 2. Mandibular | | | Mandíbula fechada | | | | | | |
| | | | Mandíbula aberta | | | | | | |
| | | | Mandíbula protraída | | | | | | |
| | | | Extensão diminuída | | | | | | |
| | | | Extensão aumentada | | | | | | |
| 3.Lingual ponta/lâmina | | | Avançada | | | | | | |
| | | | Recuada | | | | | | |
| 4. Corpo de língua | | | Avançada | | | | | | |
| | | | Recuada | | | | | | |
| | | | Elevada | | | | | | |
| | | | Abaixada | | | | | | |
| | | | Extensão diminuída | | | | | | |
| | | | Extensão aumentada | | | | | | |
| 5.Faringe | | | Constrição | | | | | | |
| | | | Expansão | | | | | | |
| 6.Velofaringe | | | Escape nasal audível | | | | | | |
| | | | Nasal | | | | | | |
| | | | Denasal | | | | | | |
| 7. Altura de laringe | | | Elevada | | | | | | |
| | | | Abaixada | | | | | | |
| B. TENSÃO MUSCULAR GERAL | | | | | | | | | |
| 8. Tensão do trato vocal | | | Hiperfunção | | | | | | |
| 9. Tensão laríngea | | | Hiperfunção | | | | | | |
| | | | Hipofunção | | | | | | |

| C. ELEMENTOS FONATÓRIOS | AJUSTE | Presente | | Graus de escala | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Neutro | Não Neutro | Moderado | | | Extremo | | |
| | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| 10. Modo de fonação | Voz Modal | | | | | | | | |
| | Falsete | | | | | | | | |
| | Crepitância/ *(creak)* | | | | | | | | |
| | Voz crepitante *(creaky voice)* | | | | | | | | |
| 11. Fricção laríngea | Escape de ar | | | | | | | | |
| | Voz soprosa | | | | | | | | |
| 12.Irregularidade laríngea | Voz áspera | | | | | | | | |

| DINÂMICA VOCAL | | Neutro | AJUSTE | Moderado | | | Extremo | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| D. ELEMENTOS PROSÓDICOS | | | | | | | | | |
| 13.*Pitch* | Médio | | Elevado | | | | | | |
| | | | Abaixado | | | | | | |
| | Extensão | | Extensão diminuída | | | | | | |
| | | | Extensão aumentada | | | | | | |
| | Variabilidade | | Alta | | | | | | |
| | | | Abaixada | | | | | | |
| 14. *Loudness* | Médio | | Aumentado | | | | | | |
| | | | Diminuído | | | | | | |
| | Extensão | | Extensão diminuída | | | | | | |
| | | | Extensão aumentada | | | | | | |
| | Variabilidade | | Alta | | | | | | |
| | | | Baixa | | | | | | |
| 15. Tempo | | | | | | | | | |
| Continuidade | | | Interrompida | | | | | | |
| Taxa de elocução | | | Rápida | | | | | | |
| | | | Lenta | | | | | | |
| 16.OUTROS ELEMENTOS | | | | | | | | | |
| Suporte respiratório | | | Adequado | | | | | | |
| | | | Inadequado | | | | | | |
| | | | Presente | | | | | | |

**ATTACHMENTC –** Frequent types of disguise in forensic speech analysis

| | |
|---|---|
| Voice source | -raised pitch (with or without change of register)<br>- lowered pitch<br>- creaky voice, vocal fry<br>- artificial hoarseness<br>- use of the false vocal folds<br>- whispered speech |
| Resonance features | - foreign object in vocal tract (e.g. ball pen)<br>- additional resonator (e.g. beer can held close to mouth)<br>- hypo- / hypernasality (mostly hypo-, by nose pinching)<br>- tissue held to mouth (e.g. handkerchief) |
| Language | - change of degree of dialectal colouring<br>- use of a different dialect (of the same language)<br>- faking a foreign accent |
| Manner of speaking | - reduction of normal pitch variation (artificial monotony)<br>- exaggeration of normal pitch variation<br>- change of speaking tempo (mostly slow-down)<br>- change of stress patterns (‘robotic voice’) |