Courses on Speech Prosody

Courses on Speech Prosody

Edited by

Alexsandro Rodrigues Meireles

Cambridge
Scholars
Publishing

Courses on Speech Prosody

Edited by Alexsandro Rodrigues Meireles

This book first published 2015

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-7600-3
ISBN (13): 978-1-4438-7600-1

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

We would like to thank Roseli de Fátima Dias A. Barbosa for her permission to publish her painting on the book cover, and Nathália Reis for the design of the book cover. Also, we acknowledge the Brazilian research institutions that financed the II School of Prosody: CAPES, FAPES, UFES, and LBASS.

INTRODUCTION

In recent years, speech prosody research in Brazil has grown significantly, mainly due to a series of events organized in this country and through the support of Brazilian government.

The 1st Brazilian Colloquium on Speech Prosody, organized by Dr. João Antônio de Moraes, took place at the Federal University of Rio de Janeiro in 2007, and gathered researchers from all over the country. The main objectives of the Brazilian Colloquium on Speech Prosody series are to promote discussion, to disseminate work throughout the Brazilian territory on the various fields of prosody research, as well as to meet researchers from related fields of interest such as linguistics, speech therapy, phonetics, neuroscience, engineering, psychology, and language teachers, among others.

Due to the success of this first enterprise, the Luso-Brazilian Association of Speech Sciences (LBASS) was created in 2007 to support the organization of the Speech Prosody 2008 Conference, which was held in Campinas in May 2008. This international conference, organized by Doctors Plínio Barbosa, César Reis, and Sandra Madureira, followed the success of the previous avenues of study, and gathered prosodic researchers from all over the world.

The 2nd Brazilian Colloquium on Speech Prosody, organized by Dr. Plínio Barbosa, took place at the State University of Campinas in 2009. This was the first Brazilian event supported by LBASS. Also, as a means of complementing the efforts to develop Brazilian prosody research, at this event, there came a proposal from the I School of Prosody, which had the mission of providing a ground for the development of prosody studies with the interplay of young and senior prosody researchers.

In order to accomplish the objective of the dissemination of Brazilian prosody studies, these two events take place regularly, on a biannual basis and in alternation. Therefore, the following events have been realized in previous years: a) 3rd Colloquium on Speech Prosody, organized by Dr. César Reis at the Federal University of Minas Gerais in 2011; b) I School of Speech Prosody, organized by Sandra Madureira and Zuleica Camargo at PUC-SP in 2010; c) II School of Speech Prosody, organized by Dr. Alexsandro Meireles at the Federal University of Espirito Santo in 2012; d) 4th Colloquium on Speech Prosody, organized by Dr. Miguel Oliveira at the Federal University of Alagoas in 2013.

Due to increasing support by Brazilian grant agencies in recent years, the II School of Speech Prosody was able to bring international researchers to Brazil for the first time, which greatly contributed to improving as well as to giving visibility to speech prosody research in Brazil. The internationalization of speech prosody events continued at the 4th Colloquium on Speech Prosody, and will continue at the upcoming III School of Prosody in 2014 in Campinas, next August.

It is clear from the above that there is an increasing number of quality works on speech prosody research in Brazil, which undoubtedly is the direct result of the initiatives proposed by LBASS.

This book came from the fact that there are very few textbooks dealing directly with speech prosody methodology and data. Because of this, we propose a book from a selection of the most attended prosody courses given during the II School of Prosody, and expect that this enterprise will help not only to contribute to prosody education in Brazil but also to the education in other countries. Also, we expect this book to be the first of a series on speech prosody research in Brazil.

Alexsandro Rodrigues Meireles (Editor)

CHAPTER ONE

ASSESSMENT OF SPEECH PRODUCTION IN SPEECH THERAPY DATA

ALINE N. PESSOA[1] AND LÍLIAN K. PEREIRA[2]

Abstract

Based on phonological therapy, this course addressed the interface of the complex relationships between perception and speech production from the integration of information from physiological, perceptive and acoustics spheres. Such foundations allow speech therapists to explore and make inferences about the manifestations of speech in cases without alterations as well as in cases with diverse disturbances. The content aimed to cover practical issues from the composition of the corpus to be assessed to the possible uses of instruments for phonetic analysis. From lectures and examples with data from speech therapy, this study aimed to exemplify the different analyses from instruments and their correlations with clinical data, which allowed the detailing of short- and long-term speech instances. In the case of short-term data, the focus was speech segments (consonant and vowel sounds). Regarding long-term data, recurring aspects of speech emission were highlighted, such as the prosodic properties and, among them, vocal quality and dynamics. Thus, the goal of this course was to perform phonetic analysis using different instruments as clinical auxiliary tools for understanding speech characteristics.

Keywords: clinical phonetics, speech therapy, acoustic analysis, auditory perception, speech production.

[1] Department of Speech Therapy/Audiology - Federal University of Espirito Santo (UFES), Vitória-ES-Brazil. Laboratory of Cognition and Acoustic Analysis (LIAAC) – Pontifical Catholic University of Sao Paulo (PUCSP).

[2] Laboratory of Cognition and Acoustic Analysis (LIAAC) – Pontifical Catholic University of Sao Paulo (PUCSP).

1. Clinical context

From reflection on the interface between two fields of study – speech therapy and phonological therapy – we identified the relationships between perception and speech production in an indissoluble and dynamic perspective. From theoretical perspectives involved in the relationships between such spheres, we identified the different areas of knowledge in speech therapy: the knowledge field related to neurovegetative functions and human communication (orofacial motricity, swallowing, breathing, oral language, hearing and balance, voice, reading and writing).

The challenge in speech therapy, in favor of the need for consistent theoretical-practical training, is notorious and debatable in the area, i.e., the choice of adequate instruments for the evaluation and monitoring of perception skills and speech production. This results from the need to obtain data and for analytical interpretation, in order to understand evidences, repercussions and implications of the phenomenon, especially from a clinical perspective.

Starting from such a discussion, we proposed this short course in order to promote an introductory action to reflect on clinical data from patients with hearing loss and users of hearing aid devices (personal amplification device [PAD] and/or cochlear implant [CI])[3], interpreted in consonance with the phonetic-acoustic theory, a science responsible for the study of speech sounds in order to characterize the mechanisms involved in the production and perception of languages sounds, in this case Brazilian Portuguese (BP).

Thus, we pointed out the instances involved in the interpretation of data from the contributions of the principles of phonetics, which include articulation aspects (transmitter, speaker, production of speech sounds or phonemes), acoustic aspects (related to the message, transmission, sound wave considering parameters of frequency (Hz), intensity (dB), duration

[3] PAD is an electronic hearing device comprising a system of individual sound amplification from electro-acoustic algorithms. It consists of the following components: microphone; amplifier (with analog, digital or hybrid processing of electro-acoustic signals); and receiver. The CI is a type of electronic hearing device that provides the sensation of hearing to users via electrical stimulation in the auditory system. It consists of an external part (microphone, microprocessor and transmitter) and an inner part (receiver and stimulator, a reference electrode and a set of electrodes that are surgically inserted into the cochlea (inner ear) and/or the auditory neural pathways). The goal is to take the acoustic signals, electrically decoded, to the brain, where they will be decoded and interpreted as sounds.

(ms) and their possible relationships in short and long term), and perceptive aspects (involving receiver/listener, understanding and the complexity of central auditory processing).

2. Theoretical Perspectives

Based on the acoustic theory of speech production (Fant, 1960) and considering the acoustic characteristics of each speech sound determined by the constrictions and bifurcations of the vocal tract and the fundamental frequency (f_0), we proposed to address methodological aspects involved in the development of the understanding of speech data from these theoretical assumptions involved and the delimited aspects to be considered in the short and long term steps of acoustic analysis, such as possible approaches from other spheres: physiological and perceptive-hearing; and basic principles for the adoption of a method of systematization and interpretation of evidence from correlations with statistics.

2.1 Acoustic analysis

The possibility of applying acoustic phonetics for the analysis of voice and speech production in the clinical, educational, artistic, and technological or expert fields is indisputable. However, we face the methodological challenges involved in the application of experimental phonetics that could be detailed and punctuated according to the following procedures: 1) formulation of hypothesis; 2) selection/constitution of data – corpus/subjects; 3) recording of the corpus, environmental conditions (noise, reverberation, control of intensity (dB) and microphone); 4) data analysis (choice of instruments, spectral analysis); 5) data processing; and 6) interpretation of results.

We highlight the necessary cautiousness and sensitivity when addressing the corpus (delimitation of the corpus, spontaneous/semi-mediated/mediated speech or text reading; context, phonetically balanced or unbalanced; random productions or speech excerpts) since these factors define and are directly related to the interpretation of the data.

With respect to data collection in cases of studio recording, it is recommended to use a soundboard for digitization of sound files, a headset microphone and sound editing software. In the specific situation of recordings with children in speech therapy, namely, *in loco*, it is recommended to use a portable digital recorder and a headset microphone. It is relevant to take the acoustics of the environment into consideration.

As a tool for acoustic analysis, we used Praat free software (www.fon.hum.uva.nl/praat/), which allowed the presentation of speech data obtained from speech therapy, giving a representation of the sound wave in spectrum and spectrogram.

Spectra are diagrams that represent the amplitude and frequency of simple waves at a given point in time, and spectrograms are diagrams that allow the visualization of sound spectral evolution through time, by means of broadband and narrowband resolution. The representations described allowed introductory concepts to be addressed, for adoption of the theoretical and methodological assumptions involved in the linkage between speech therapy and clinical phonetics.

Acoustic signal segmentation is a basic step and, at the same time, a complex acoustic analysis procedure. There must be a rigorous adoption and maintenance of the criteria used for segmentation, as well as consideration of the articulatory characteristics of segments, their acoustic correlates and the notion of coarticulation.

3. Brazilian Portuguese researches

3.1 The relevance of segmental aspects

With regard to segments of BP, the articulatory and acoustic characteristics were specified, and the characterization of the following points were discussed: a) vowels: periodic complex sound with acoustic characteristics determined by the sum of sine waves with no obstructions in the air passage, but in resonance regions in which there are bigger or smaller concentrations of energy in the tube; and b) consonants: sounds produced with some kind of obstruction (partial/total) in the vocal tract, which causes interruption of the air passage.

The criteria for classification of consonantal segments are: 1) manner of articulation; 2) place of articulation; 3) voicing; and 4) nasality x orality. We highlight the current path of egressive air from the chamber, the role of the vocal folds, the soft palate, and the place and manner of articulation. Especially for speech therapy, we emphasized the peculiarities of rothics (Gregio et al., 2012); although they are grouped under this nomenclature, the group is not defined by common articulatory characteristics.

The examples and citations of studies on the specific case of vowels mention the exploration of the formant patterns that reveal indications of movement limitation in vowel articulation. In this sense, findings of decreased frequency of the first (F1) and second (F2) formant,

respectively, related to the degree of openness of the vocal tract (vertical positioning of tongue and jaw) and anterior-posterior tongue movements, and even findings of anteriorized and lowered tongue tendency have been reported (Benedicte et al., 1989; Mendes, 2003; Barzaghi-Ficker, 2003; Cukier & Camargo, 2005; Campisi et al., 2005; Pereira, 2007; Serkhane et al., 2006; Seifert et al., 2002; Peng et al., 2008). Such evidence has also been described by means of another methodological approach such as in supra-segmental areas, according to Pessoa's thesis (2012).

Regarding the segmental level analysis of speech data, we identified studies (Barzaghi-Ficker, 2013; Pereira, 2008, 2012) that, from precepts of phonetic-acoustic analysis and based on the acoustic theory of speech production (Fant, 1960) and articulatory phonology (Browman, Goldstein, 1986, 1990, 1992), detailed the production of alveolar plosive consonants of BP in subjects with hearing impairments. In the study conducted by Pereira and Madureira (2012), using the Praat software, they extracted measures of duration (ms), f_0 values and formants (Hz) of a corpus made up of words consisting of plosive consonants ('tata', 'data', 'cata' and 'cada') inserted in the carrier phrase.

The acoustic analysis of the study (Pereira and Madureira, 2012) shows that the most changed parameter was the percentage of voicing during the total duration of consonants. In this sense, the assessment of the production characteristics of alveolar plosive consonants of BP, according to different positions (in this case, the unstressed position of two-syllable words stressed on the second syllable) in the speech of a subject with hearing impairment, especially as regards the voicing contrast, pointed to clinical developments. In view of this research, in which some hypotheses related to the importance of considering factors such as focus, degree of accentuation and coarticulation as interference were listed, the results will be analyzed from the assumptions of articulatory phonology and acoustic phonetics. Thus, the decrease in synchronicity between articulatory gestures, which may be present in the speech of subjects with hearing impairment, would cause retardation in voicing interruption, and, consequently, would change the perception of voicing parameters of these subjects' speech. This type of description produces notorious evidence for therapy, which without the acoustic tool would not be possible to detail.

At this point, it is important to highlight the importance of the methodological adequacy and consistency of the statistical analysis adopted: significance of data found; correlations between production and perception of speech; intra- and inter-subjects comparison; analysis procedures; and predetermined significance level. This is due to the fact that some situations require greater care and consideration; for example,

pre-judgments, auditory memory and linguistic knowledge in cases of individuals with hearing impairment (limited auditory perception).

3.2 Approaching to supra-segmental aspects

Inextricably linked to segmental aspects, we dealt in a prosodic approach with the set of speech phenomena that include frequency variability, intensity and duration in the long term. The variation with respect to the field should be noted, which, according to Barbosa (2010), comprises the analysis of the phonic units and their relationships, from the syllable to the oral text. Prosody makes it possible to: define the mode of utterance (declarative, interrogative or exclamatory); organize speech structurally through chaining and prominence, interacting with the syntax; offer pragmatic function (integration of the message with its context); define attitudes through speech and, therefore, establish the relationship between the speakers; express emotions and even characterize the speaker (social and physically, for example) (Barbosa, 2010).

It is known that prosodic elements, from the earliest vocalizations and gurgles of babies and young children, have a fundamental role in the perception and production of speech sounds, being connected to the symbolic and cognitive development that pervades the relationship between sound and sense.

3.3 Sounds and prosody – complementarity

In this context, permeated by segmental and supra-segmental instances, we identify the methodological challenge of undertaking phonetics-based studies that consider the many variables involved in order to delimit the research corpus, exploring contexts for recording speech samples in therapeutic frames and without comparisons with adults' patterns. It is possible not to depend on a standardized corpus, offering conditions for analyzing spontaneous speech excerpts (Pessoa et al., 2010, 2011, 2012).

Complementarity between segmental and prosodic elements in speech therapy has been addressed in studies on speech production (Guedes, 2003; Magri, 2003; Andrade, 2004; Camargo et al., 2004; Peralta, 2005; Cukier et al., 2006; Lima et al., 2007; Magri et al., 2007; Blaj et al., 2007, Camargo & Madureira, 2010; Madureira & Camargo, 2010; Rusilo et al., 2011), especially due to the importance of reflecting on the impact of speech development on the role of communication and interaction in the paralinguistic field (short term adjustments of vocal quality used to signal excitement, communicative intentions, etc.) and in the extra-linguistic

field (long-term vocal quality) which is not always consciously controlled (Mackenzie-Beck, 2005).

Considering the variability of patterns contained in the speech and the complex interactions between perception and production mechanisms related to the dynamic model of speech (Boothroyd, 1986; Fujimura & Hirano, 1995; Lindblom, 1990; Barbosa, 2006, 2007, 2009; Xu, 2011; Hirst, 2011; Fourcin & Abberton, 2008), and the variety of possible and predictable adjustments, allows us to understand combinations resulting from detailed compensations. This fact is justified and correlated by the acoustic sphere, which is based on interdependent functioning of the structures of the laryngeal and super-laryngeal vocal tract (Cukier, 2006; Gregio et al., 2006). Thus, we highlight the refinement of the action of the vocal tract to phonation and the various compensations that may be caused due to their plasticity.

4. Perception analysis – VPAS

From the vocal quality approach and vocal dynamics, the voice profile analysis scheme (VPAS) (Camargo & Madureira, 2008) has allowed the exploration of clinical data about the long-term tendencies of speech production that characterize a particular speaker (product of respiratory, laryngeal/phonatory and supra-laryngeal/articulatory activities). Mackenzie-Beck (2005) stated that they are: “those features that are present more or less all the time in which a person is speaking”, i.e., “An almost-permanent quality traversing all the sounds that emanate from the speaker’s mouth” (Abercrombie, 1967). From the phonetic point of view, there is the notion of setting (adjustment): “Recurrent feature translated as a tendency of the vocal apparatus to be subjected to a particular muscle long-term adjustment” (Laver, 1980).

From the acoustic point of view, vocal quality and dynamics have been explored in our group by combining a group of acoustic measures (Hammaberg & Gauffin, 1995; Barbosa, 2009; Camargo & Madureira, 2009; Rusilo et al., 2011). Long-term acoustic measures extraction was performed using the Expression Evaluator script (Barbosa, 2009). Such measures are extracted from speech excerpts and excerpt labeling is not required. In this way, this method of analysis does not require a standardized speech sample and it is applied to assess acoustic correlates of quality adjustments and vocal dynamics.

Perceptive-hearing (through the VPAS-PB script) and acoustic correlations (through the Expression Evaluator script), based on dynamic models and methodological procedures of experimental phonetics, allow

speech production to be approached in contexts of speakers with and without speech alterations. Such instruments addressing spontaneous speech allow discussion of production without characterizing a dichotomy between normality and pathology. Methodologically, controlled experimental situations are indisputably necessary and such systematic and experimental control can be indispensable in our quest to understand the mechanisms underlying speech. However, with the advent of prosodic studies, especially those of expressiveness, such procedures can compromise and hinder the recognition of factors that contribute to the description and understanding of particular elements of human speech (Xu, 2010).

A combination of a group of acoustic measures (Barbosa, 2006, 2007, 2009) relating to the f_0, the first derivative of f_0, intensity, spectral decline and the long-term spectrum was used for the acoustic approach (Camargo & Madureira, 2010; Pessoa et al., 2010, 2012; Rusilo et al., 2011; Lima-Bonfim, 2012; Camargo et al., 2012; Queiroz, 2012). Methodologically, the use of acoustic measures taken through long-term techniques (from the processing of speech excerpts and not from isolated units) is highlighted. This procedure will not require labeling of vowel and consonantal segments, which may not be well delimited in certain productions, both in earlier stages of language development and in particular characteristics of the speakers, as those productions considered altered for the age bracket.

From the earliest babblings – especially in children with hearing impairment – the perception of acoustic signals in the articulatory movement at the moment of babbling may show evidence that points to the importance of discovering the vocal tract skills and learning the relationships between movements and perception from the sequence of motor gestures and adjustments from auditory feedback (Meier et al., 1997; Bailly, 1997; Boysson-Bardies et al., 1999; Serkhane et al., 2006; Iverson et al., 2007).

In children, the phonatory system is oscillating and, due to the non-linear relationships between the elements, it can feature patterns of great variability. A description method able to include all the mobilizations can indicate, over time, changes in adopted patterns that define the maturation of the mechanism, learning phonatory control categories and, possibly, the use of different vocalizations in social contexts (Buder et al., 2007).

This way, the relevance of speech therapies for the therapeutic process is confirmed, because perceptual and acoustic data enable the speech therapist to explore and make inferences about speech manifestations, both in cases without alterations and cases of disturbances from diverse origins. Such evidence can outline the definition of clinical outcome, i.e.,

indicators of hearing care services. In addition, this evidence can outline clinical indicators and therapeutic management for therapies, as demonstrated in this short course for the case of children with hearing impairment regarding strategies related to articulatory control and precision on the acoustic target.

There are two challenges: a) appropriate and articulated decision to approach the data from consistent methodological assumptions to be adopted; and b) reflection on the relationship between segment and supra-segment that unfolds: influence of temporal processes, variability and combinations between the parameters of frequency, intensity and duration offered by devices. Still, we emphasize the need for performing phonetic-acoustic analysis that allows the inference of modes of speech and voice production, and correlates them with the plan of speech perception.

5. Final considerations

Delimiting evidence and clinical outcome indicators has been the great challenge in the studies that relate to the spheres involved in speech production and perception. Through the innovative technologies offered by the CI, it has been possible to achieve a good performance in hearing acuity, revealing speakers with excellent quality of detection, discrimination, and recognition of speech sounds of a wide range of frequencies (reaching higher frequencies) and even the weakest intensities, which can be observed through consistent responses to sound stimuli. Such speakers show good results on specific tasks of speech sounds perception (vowel and segmental), thus confirming the efficiency of the electrical stimulation technology provided by the CI.

The correlation between speech therapy and clinical phonetics is a fertile field and it allows understanding clinical evolution. The data of this study highlighted the importance of creating a database with a population character, aimed at service indicators, decision making regarding population and technological procedures and, above all, based on a speech corpus whose collection occurs during therapeutic procedures, with discussion of data collected in a longitudinal character. In view of the contributions of the analysis of prosodic elements, we suggest the possibility of incorporating tools for speech assessment in routine clinical monitoring of this population and in research that consider speech as the central object.

References

Barbosa, P. A. (2009) Detecting changes in speech expressiveness in participants of a radio program. In: *Proceedings of Interspeech*. v. 1, 2155-2158. Brighton, United Kingdom.

Camargo, Z. A.; Madureira, S. (2009) Dimensões perceptivas das alterações de qualidade vocal e suas correlações aos planos da acústica e da fisiologia. Delta. Documentação de Estudos em Linguística Teórica e Aplicada (PUCSP. Impresso), v. 25, p. 285-317.

Camargo, Z., Navas, A. L. (2008) Fonética e fonologia aplicadas à aprendizagem. In: Zorzi, J.; Capellini, S. *Dislexia e outros distúrbios de leitura-escrita*. São José dos Campos: Pulso. p. 127-157.

Fant, G. (2000) *Half a century in phonetics and speech research*. Fonetik 2000, Swedish phonetics meeting in Skövde, May 24-26.

Gregio, F. N.; Gama-Rossi, A.; Madureira, S.; Camargo, Z. N. (2006) *Modelos teóricos de produção e percepção da fala como um modelo dinâmico*. Rev CEFAC, São Paulo, v. 8, n. 2, 244-247.

Johnson, K. (2003) Acoustic & Auditory Phonetics. Malden: Blackwell.

Kent, R. D. & Read, C. (2002) *The Acoustic Analysis of Speech*. 2nd ed. Albany, NY: Singular Thomson Learning.

Pereira, L. K.; Madureira, S. (2012) *A produção das plosivas alveolares /T/ e /D/ por um sujeito com deficiência auditiva: Um estudo fonético-acústico*. Intercâmbio (PUCSP), v. XXIII, p. 128-151.

Pessoa, A. N.; Novaes, B. C. A.; Pereira, L. K.; Camargo, Z. A. (2011) *Dados de dinâmica e qualidade vocal: correlatos acústicos e perceptivo-auditivos da fala em criança usuária de implante coclear*. Journal of Speech Sciences 1(2): 17-33.

Rusilo, L. C., Madureira S., Camargo Z. (2011) *The validity of some acoustic measures to predict voice quality settings: trends between acoustic and perceptual correlates of voice quality*. Proceedings of the Fourth ISCA Tutorial and Research Workshop on Experimental Linguistics. Paris: ISCA, p. 115-118.

CHAPTER TWO

SETFON: THE PROBLEM OF THE ANALYSIS OF PROSODIC, TEXTUAL AND ACOUSTIC DATA

ANA CRISTINA FRICKE MATTE,[1]
RUBENS TAKIGUTI RIBEIRO,[2]
ALEXSANDRO MEIRELES[3]
AND ADELMA L.O S. ARAÚJO[4]

Abstract

Setfon is an open source web information system for data collecting in the field of speech sciences. The system is component based and emerged from the need to process an increasing amount of acoustic-phonological data, in order to solve the problem of statistical significance in emotional and stylistic speech. Moreover, this chapter presents the software and its components under different approaches of data collecting: Acoustic Phonetics, Phonology, Semiotics, Information Technology and Computation.

Keywords: technology, acoustic phonetics, phonology, phonostylistics.

[1] Universidade Federal de Minas Gerais, UFMG, Faculdade de Letras, POSLIN, Belo Horizonte, MG, Brasil, anacrisfm@ufmg.br.
[2] Universidade Federal de Lavras, UFLA, Faculdade de Ciência da Computação, TecnoLivre, Lavras, MG, Brasil, rubens@tecnolivre.com.br.
[3] Universidade Federal do Espírito Santo, UFES, Departamento de Línguas e Letras, PPGEL, Vitória, ES, Brasil, meirelesalex@gmail.com.
[4] Universidade Federal de Minas Gerais, UFMG, Faculdade de Letras, POSLIN, Belo Horizonte, MG, Brasil, adelmaa.ufmg@gmail.com.

1. What is Setfon?

Research on speech is generally divided into acoustic, acoustic-phonetic, phonological and expressive (emotion, attitude, etc.) approaches. The last ones especially focus on the content of what has been said, trying to relate a communicative intention to a particular sound expression. Therefore, physical, linguistic and semiotic information is important to determine the existence of this relationship and, if it exists, the degree and type of the relationship. Data collection, hence, cannot focus on only one aspect of speech production, which significantly increases the number of parameters to be collected and analyzed.

Setfon is an open source web information system for data collecting in the field of speech sciences. The system addresses the problem of collecting and managing data through software components. It was created from the need to process an increasing amount of acoustic-phonological data, in order to attend to the demands of statistical significance on expressive speech studies. In addition, since it is available online, Setfon permits the creation of a national database that can be shared among speech scientists.

This work would not have been possible without an interdisciplinary team, with very different concerns, but common goals. This chapter aims to explore the different facets of the work. To accomplish this goal, we will present the program's context and technology.

2. Phonetic-Acoustic Database

According to many scholars, working with phonetics is to be in between linguistics and non-linguistics studies. The situation is further complicated when it comes to articulatory and acoustic phonetics, given the amount of knowledge of physics and human anatomy involved in those fields. However, phonetic studies may not be disconnected from phonological studies. In this case, the rejection of phonetics from linguistic studies by phonologists should not be applied.

From our point of view, the idea of phonetics not being considered part of linguistics is largely the result of a reversal of priorities in terms of the time the phonetician dedicates to linguistic research, which can be divided into raising hypotheses and preparation of the experiment, data collection and analysis. The elaboration of the problem and analysis of data are epistemologically and linguistically well founded, but occupies a third or less of the time of the phonetician's research, who spends most of the time collecting acoustic or articulatory data of speech sounds. The work of

segmentation and labeling of sound samples is the main issue which causes misinterpretation of the linguist phonetician's work.

Unfortunately, it is not for language concerns that many researchers, for instance electrical engineers, elaborate automatic speech segmentation software. The automatic speech segmentation programs, which have been developed by engineers, usually aim at speech synthesis or recognition, and therefore restrict phonetic transcription as a single element linked to a phonic segment. On the other hand, programs such as Praat – open source speech analyzer software for the speech science community – allow the connection of several levels of information with a sound file. Nevertheless, the whole process of labeling is necessarily manual (conceiving labeling as the process of linking information – e.g., phonetic transcription, speech rate applied to the speaker, emotion reported by the speaker – to a pre-determined sound unit). These automatic speech segmentation programs undoubtedly represent major advances for the linguist, but they are insufficient for the reversal of priorities in the schedule of research in phonetics. This reversal is vital so that phonetic studies applied to speech technology can reach international competitiveness.

3. Setfon: Proposal optimization

To solve this problem, a semi-automatic labeling system was elaborated that not only meets the needs of the linguist phonetician but also provides future interaction with speech recognition as well as speech synthesis. This device is Setfon: an algorithm for the production and organization of phonological semiolabelers.

Phonological semiolabelers are products of a tool for annotating and organizing data from textual, syntactic and semantic analyses, information about the recording, and any other information relevant to acoustic phonetics data (e.g. phonological and phonetic labels). This tool aims to speed up the process of preparing data for phonetic analysis, and represents an important advance for phonetics research in Brazil, given the originality of the proposal.

Setfon's algorithm combines a sound file (.wav or other format) with a text file (.txt) in order to obtain a labeled segmentation with access to information such as duration, intonation curve, and labeling – which is able to receive new information according to the researcher's demands. Finally, it returns in tables, information about each segment. Given the nature of this process, the tool manager, from a computational viewpoint, is a system of various programs, each responsible for one of the tasks necessary for the work of labeling sound samples and obtaining tables.

Some of these programs pre-existed, such as Praat[1], Ortofon (Albano and Moreira, 1996) and SilWeb (Matte, Meireles and Fraguas, 2006).

Setfon is dedicated to the linguistic analysis of speech, it being optional to support research on speech synthesis. Thus, it is not a segmentation program, but an algorithm that automates and manages the relationship of linguistic and extralinguistic information to segments, whose size is determined by the needs of each piece of research.

Importantly, due to the versatility of Setfon, this proposed labeling is more than a simple program: it is designed to be an application server that can be easily adapted for many different purposes in experimental phonetics research, and whose components' maintenance is very straightforward. Since the goal of the project is the design of this general algorithm, the implementation of the analysis, initially restricted to a sentence in terms of duration and intonation curve, enables the immediate realization of tests utilizing casual speech. Using casual and controlled corpora yielded results that were immediately applicable to research on Brazilian Portuguese phonostylistics (Mendes, 2009). Since this is a web resource, its application in speech technologies is wide in terms of man/machine interaction: telephone, web, home appliances, and speech disorders.

4. Software Development

The design of the semiolabeler followed the phonological process of software development based on components proposed by Brito et al. (2005), which is divided into the following steps: (i) domain analysis, (ii) modeling the components, (iii) implementation of the components, (iv) testing the components, (v) implementation of the web interface, and (iv) integration testing.

During the development of each Setfon component, interfaces were created for individual use (shell scripts), making it possible to perform independent tests. The Oriented Programming Components is a technical approach to solve computational problems through atomic logical structures and well defined interfaces. Components encapsulate black-box processes, or processes that do not require detailed knowledge of the implementation strategy, since they do not have coupling at the modular level. The phonological semiolabeler forms a complex data set. In order to get this, each attribute is treated by a differentiated component.

The component-based process has great proximity to semi-automatic and manual activities performed by researchers to obtain acoustic data. Most operations are atomic, and they have well-defined inputs and

outputs. In this sense, four essential components have been identified: (i) audio segmentation, (ii) fonotranscriber of text to phonological transcription, (iii) TextGrid1 handler, and (iv) data noise extraction. These components are handled by a web tier that functions both as the controller of the steps involved in the extraction process and as the direct interface with the user researcher.

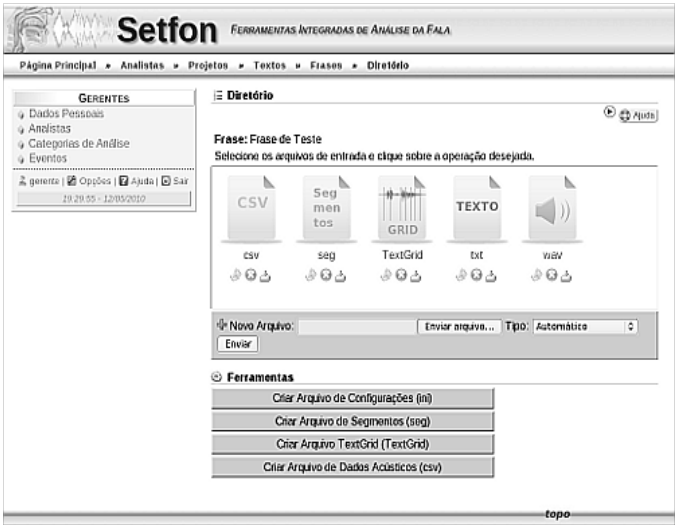

Figure 1: Main page of Setfon

The main tool of Setfon is represented in Figure 1. This is a web interface that starts the process from a sound file and a text file (with corresponding semantic values). It is necessary to evaluate the sound file with its respective TextGrid, which is filled with phonological segments and other relevant data, to obtain the acoustic data. On the other hand, three sub-steps are necessary to obtain the TextGrid: (i) phonologically transcribe the text and segment using VV units, (ii) generate a TextGrid only with the data from the sound file, but still without phonological segments, and (iii) insert the phonological segments into the TextGrid.

The main strategy to address the solution to this problem is to define the inputs and outputs of each component as files of different types. Each component hence receives one or more input files and produces a resulting file. The web tier presents the files (central region of Figure 1) and the possible operations on these files (bottom of Figure 1). In regard to performing an operation, one must select the input files (by clicking on

them) and then trigger the desired operation. Each component in this process uses the most suitable technologies and techniques for the purpose.

5. A disassembly line: speech analysis

Setfon works as a disassembly line: the product 'speech' is decoupled in time and its qualities are analyzed and arranged in order to allow the visualization of its parts before displaying the whole set. Two types of segmenting are needed to obtain phonostylistics data: a macro-segmentation based on stress groups as well as a micro-segmentation based on VV units. Although the automation of the process has been mandatory on the choice of the segment types and has been responsible for the use of phrase and syllable, the method of the inclusion of data was created in order to enhance the semiotic approach that considers the text as a whole.

The starting point of Setfon was the SilWeb software, which was originally designed to return to every word and syllable its accentual classification. Working with UML, despite not been brought to completion, allowed a neat algorithm compatible with other applications, some of which were eventually incorporated into the program.

The programming was started in Matlab and was completed in PHP. It was based on the phonological studies by Mattoso Câmara (1970), and predicts, with 99% accuracy, any Brazilian Portuguese word (henceforth BP) or logatome that follow the rules of BP phonotactics. The program also returns syllables and consonant-vowel(-glide) masks.

The program was tested on a large corpus, CETEN-Folha, during the time that three researchers who were involved in the project were working on a thematic project called "Integrating Parameters in Continuous and Discrete Models of Knowledge and Lexical Phonic", coordinated by Eleonora Albano, held at UNICAMP, and funded by FAPESP until January 2005.

The behavior of units larger than a phoneme such as speech rate and inter-group perceptual-center (VV unit) – consisting of syllable-sized units from the first unit following the stressed segments up to the next stressed segment (Marcus, 1981) – has been significantly related to the tensile potential behavior of the text, according to the results of Matte (2005) on emotional speech.

The application of phonological semiolabelers, with the implementation of Setfon, attends researchers' demands to test which independent

variables imply variation at the expression level, instead of sticking to a single working hypothesis.

For instance, we mention the hypothesis of the tensile curve of temporality M (Matte 2004a), proposed in 2001. M is a combination of three elements directly derived from a semiotic analysis of five levels of temporality in text content (Matte, 2004b), two of which are discarded by the formula M due to a theoretical obsolescence. During this phase of the research, the component prosodic speech rate was significantly correlated with the variation of M.

Setfon speeds the process of gathering and organizing data to allow the testing of different hypotheses; for example, a test with each temporal component alone and in different combinations, including those dropped by the hypothesis of the original formula of M. In addition, it also enables a leap towards a predictable and desirable step of the research regarding a semiotic analysis of the lexicon, by testing the relationship between semio-temporal keyword analysis and prosodic results. It is acceptable to predict that the analysis of a possible tensile content of the lexicon, linked to a syntactic analysis, may enable an automation of the semiotic analysis, taking into account speech synthesis, based on the assumption of vocal caricature (Matte, 2004a). Besides enabling the testing of a greater number and variety of cases in less time, the agility guaranteed by Setfon allows changes in strategy whenever the results point to it, without causing significant delays to the research.

The project can be divided into three blocks: phonological study, computational study of the management, and programming of subsidiary tools in the interface between computer science and linguistics.

6. Components

The linguistic study, using the methods of experimental phonetics and phonology, the behavior of speech rate, silent pauses and intonation curve (f0), enables a segmentation grounded in the linguistic sense of the sentence prosody. The segmentation of the sentence follows the concept of the VV unit, starting in the first vowel and ending at the beginning of the last vowel of the sentence, given the greater accuracy of the perception of the transition between a consonant and a subsequent vowel than of the transition between a vowel and a consonant, as shown by Barbosa (1996), Cummins (2002), and Pompino-Marschall (1989).

The length of the sentence, segmented with this method, can only generate information about the speech rate if the units are also VV units, so it was necessary to remodel the SilWeb program (Matte, Meireles and

Fráguas, 2006) to perform accentual analysis and phonic decomposition of words with phonological transcription, and the currently capability of splitting CV syllables, for use in the database of the thematic project mentioned above. This program, SilWebVV, also allows data to be obtained to calculate the z-score of the sentence, a relative measure of duration that takes into account the intrinsic duration of phonic segments, which is also done automatically.

The overall design of the component-oriented program allows you to add other existing programs to the process, improving, therefore, the final result. It works like a grid of text that links different layers of information to each piece of sound-sentence. In a strict sense, it is a network of informational classes of different natures, connected to continuous media; in this case, sound through digital identities.

Implemented in this way, the phonological semiolabeler Setfon is receptive to updates, some of which are predictable and desired by the phonetics community, such as the replacement of the sentence segmenter with a phone segmenter, as well as the implementation of an updated f0 analyzer. The three blocks that organize this project were developed as needed, and often simultaneously.

7. Phonological Labels

The concept of a phonological semiolabeler, here proposed, is an approach to speech analysis that deals with speech sounds as objects. A phonological semiolabeler is a class of objects whose attributes are intrinsic or acquired data.

The objects are segments of speech sound that may have different sizes. At this point, we have adopted the VV unit (vowel-to-vowel) (Marcus, 1981) and the stress group to support the segmentation (Barbosa, 2006). These objects are obtained by automatic analysis of stretches of speech accompanied by an orthographic transcription (Barbosa, 1996). The stress group is a sequence of VV units obtained by quantitative and qualitative analysis of VV durations. It is, therefore, an analysis dependent on the original attributes.

On the one hand, the intrinsic data are essentially qualitative independent variables, which can be obtained by automatic acoustic analysis. On the other hand, acquired data are parameters whose automation is still an unexplored possibility, given its reliance on a qualitative analysis. Currently, it is possible to have syntactic and semantic parsers to help the process, although the semiotic analysis is completely manual.

While different in nature, both the stress group and the VV unit can receive intrinsic and acquired attributes. The first attribute of the acquired VV unit is a phonological label obtained by transcription and phonological segmentation of the text corresponding to a speech sound. It is only possible to calculate its intrinsic attributes (duration, intensity, frequency, formant configuration) and create objects of the class Accentual Group by getting the phonological label. Regarding the objects of this class, they have intrinsic prosodic attributes such as speech rate, melodic curve, intensity variation, duration variation, stress position, and number of VV units. Acquired attributes are totally dependent on the type of the desired result, which may arise from syntactic parsers, semantic parsers and/or specific content, such as tensile or narrative semiotic analysis, just to name a few.

8. Ortosil

Matte, Meireles and Fraguas (2006) developed a phonological syllabic-accentual parser for linguistic applications: SilWeb. Briefly, the program returns the following lexical information from a phonological input: 1) the word accentual class, 2) the number and type of syllable (stressed, pre-stressed and post-stressed), and 3) syllabic masks (with or without the presence of glides). An example of this analysis is shown in Figure 2 below.

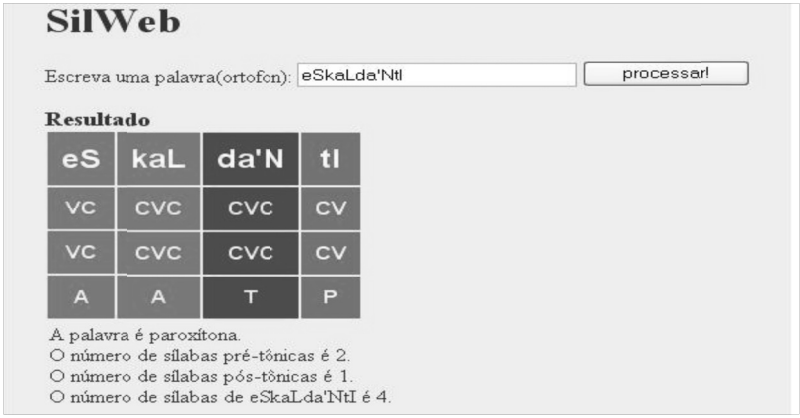

Figure 2: Phonological syllabic-accentual analysis of a word written in “Ortofon” transcription (Matte, Meireles, and Fraguas, 2006, p. 47)

In figure 2, it is possible to notice that the researcher has entered the word transcribed for "Ortofon" (eSkaLda'NtI), and the program returns linguistic information that is pre-programmed in the source code. This type of transcription (conversion letter–phone) was proposed by Albano & Moreira (1996) for speech synthesis, and was performed by the program Ortofon (restricted use).

In this way, even though SilWeb generates linguistic information, if linguists and others interested in corpus linguistics want to take advantage of the benefits of this program, they should know how to transcribe in "Ortofon", which makes the practical implementation of the program more difficult. Thus, in order to facilitate the use of the program for the scientific community, we developed a suitable program for converting data using phonological spelling: Ortosil. This computational tool follows a similar theory, but independent of Ortofon principles.

Ortosil emerged from our experience with Ortofon (Albano and Moreira, 1996); however, as our aim was to develop a program that reflects the phonological knowledge of Portuguese, we based our transcription, more precisely, on the phonological analysis proposed by Mattoso Câmara, but with some fundamental changes. According to Mattoso Jr. (1970), the phonological system of the Portuguese language is composed of the following phonemes: 19 (nineteen) consonants, 7 (seven) vowels and two (2) archiphonemes. Within this phonemic framework, one can transcribe any BP word. A comparative analysis of our phonological analysis contrasted with Mattoso Câmara Jr.'s follows below.

Analysis of consonants: Mattoso Câmara Jr. proposes 19 BP consonant phonemes: /ʃ ʒ p b t d k g f v s z m n ɲ ʎ ɾ l R/. All of these phonemes occur at the beginning of the syllable, and therefore have little articulatory variability (see, among others, Taurosa, 1992; Keating et al. 1999). Our transcription of these phonemes is identical to this analysis, except for symbolic changes related to ease of computational implementation, namely: /sh zh p b t d k g f v s z m n nh l lh R/. Furthermore, we chose to represent the tap [ɾ] with the same archiphoneme symbol /R/, since it is one of its possible pronunciations. As can be seen, the phonemic consonantal chart for the beginning of the syllable and/or word (except the phoneme /r/), due to the issue of stability articulation, is uncontroversial. However, there is wide variation in the dialectal pronunciation of consonants in the final syllable and/or word in Brazilian Portuguese. To explain this variation, Mattoso Câmara Jr. used the classical notion of archiphoneme from Russian structuralism.

Consonantal archiphonemes: According to Trubetzkoy (1939), archiphonemes are symbols that represent the loss of phonemic contrast in