Language and Law

Linguagem e Direito

VOLUME 1.2
ISSN 2183-3745

Language and Law
Linguagem e Direito

ISSN: 2183-3745 (online)
Volume 1, Issue 2, 2014

Publisher / Editora
Faculdade de Letras da Universidade do Porto

U.PORTO Faculdade de Letras Universidade do Porto
Centro de Linguística da Universidade do Porto
Linguística Forense
ALIDI ASSOCIAÇÃO DE LINGUAGEM & DIREITO

Language and Law / Linguagem e Direito

Language and Law / Linguagem e Direito is a free, exclusively online peer-reviewed journal published twice a year. It is available on the website of the Faculty of Arts of the University of Porto http://ler.letras.up.pt.

All articles should be submitted by email to the journal email address (llldjournal@gmail.com). See the guidelines for submission at the end of this issue.

Requests for book reviews should be sent to llldjournal@gmail.com.

Abstracts of PhD theses should be sent to the PhD Abstracts Editor, Dayane de Almeida (daycelestino@gmail.com).

Language and Law / Linguagem e Direito é uma revista gratuita publicada exclusivamente online, sujeita a revisão por pares, publicada semestralmente e disponível no website da Faculdade de Letras da Universidade do Porto http://ler.letras.up.pt.

Os materiais para publicação deverão ser enviados por email para o endereço da revista (llldjournal@gmail.com), e devem seguir as instruções disponíveis no final deste volume.

As propostas de recensão de livros devem ser enviadas para llldjournal@gmail.com.

Os resumos de teses de doutoramento devem ser enviados para a Editora de Resenhas de Teses, Dayane de Almeida (daycelestino@gmail.com).

PUBLISHED BIANNUALLY ONLINE / PUBLICAÇÃO SEMESTRAL ONLINE

ISSN: 2183-3745

THE ARTICLES ARE THE SOLE RESPONSIBILITY OF THEIR AUTHORS.
THE ARTICLES WERE PEER REVIEWED.

OS ARTIGOS SÃO DA EXCLUSIVA RESPONSABILIDADE DOS SEUS AUTORES.
OS ARTIGOS FORAM SUBMETIDOS A ARBITRAGEM CIENTÍFICA.

Contents / Índice

ARTICLES / ARTIGOS

Guest Editors' Introduction

Maria Lúcia de Castro Gomes,
Andrea Alves Guimarães Dresch &
Denise de Oliveira Carneiro

Universidade Tecnológica Federal do Paraná – UTFPR &
Instituto de Criminalística do Paraná, Polícia Científica

It was a great honor for our Forensic Phonetics Study Group, which is part of the Speech Sounds Research Group at the Technical Federal University of Parana in Curitiba, to accept the invitation of Malcolm Coulthard and Rui Sousa-Silva, to organize this special issue on Forensic Phonetics. These last six months spent interacting with authors, reviewers and editors was a delightful learning experience for us all. We hope that the outcome, this issue of the journal, will contribute to the development of forensic speaker recognition, especially in Brazil, where there is a notable lack of consistent and coordinated research.

The primary goal of our group has been to develop and disseminate research in the areas of production and perception of speech, focusing particularly on forensic applications, and we have advocated the importance, for the development of the discipline, of multidisciplinary studies and of a close relationship between academics and forensic practitioners. Therefore, we are very happy with the result because we have managed to put together articles produced by scholars and experts from the fields of linguistics, engineering, speech therapy and computer science, professionals drawn from universities, forensic institutes and technology centers. The articles discuss topics that are highly relevant for both speech and forensic sciences. As befits a bilingual publication, the volume includes three articles in English and three in Portuguese.

Helen Fraser's article deals with the transcription problems of indistinct recordings, and the use of these transcriptions for investigative and evidentiary use. The author draws attention to collaborative research between the phonetic sciences and law. Eugenia San Segundo and Pedro Gomez-Vilda analyze and compare the glottal parameters of vowel fillers produced by pairs of monozygotic and dizygotic twins, as well as pairs of non-twin siblings and pairs of unrelated people. The objective was to assess the genetic influence on several glottal parameters. These same authors, San Segundo and Gomez-Vilda, joined by Luis M. Mazaira, Agustin Alvarez and Victoria Rodellar,

report a study of phonation distortion and present a methodology that can be used both in Speaker Verification and in Dysphonic Voice Grading.

The articles in Portuguese approach forensic speaker comparison from three different perspectives. First, Cintia Gonçalves Schivinscki and Claudia Regina Brescancini defend the sociophonetic approach to forensic speech analysis. Paloma Alves Miquilussi, Marilisa Exter Kalovski and Denise de Oliveira Carneiro evaluate the contributions of the speech sciences to forensic studies, discussing the National Curriculum Guidelines for the course of Speech Sciences in Brazil. Finally, Aline de Paula Machado and Plinio Almeida Barbosa investigate the effectiveness of a set of acoustic measures for speaker verification analysis. We hope that this publication stimulates further research, and that this journal becomes a standard reference for forensic linguistics, not only in Brazil, but also in all countries where there are studies on the interface between the sciences of language and law.

Maria Lúcia de Castro Gomes
Universidade Tecnológica Federal do Paraná – UTFPR

Andrea Alves Guimarães Dresch
Instituto de Criminalística do Paraná, Polícia Científica, Brasil

Denise de Oliveira Carneiro
Instituto de Criminalística do Paraná, Polícia Científica, Brasil

Nota Introdutória

Maria Lúcia de Castro Gomes,
Andrea Alves Guimarães Dresch &
Denise de Oliveira Carneiro

Universidade Tecnológica Federal do Paraná – UTFPR &
Instituto de Criminalística do Paraná, Polícia Científica

Foi uma grande honra para o nosso Grupo de Estudos em Fonética Forense, ligado ao Grupo de Pesquisa 'Estudos dos Sons da Fala' da UTFPR Curitiba, aceitar o convite dos Professores Malcolm Coulthard e Rui Sousa-Silva, Editores da Revista *Language and Law / Linguagem e Direito*, para organizar este volume especial sobre fonética forense. Estes últimos seis meses de contato com os autores, os revisores e os editores foram um período de grande aprendizado e alegria para nós. Esperamos que o resultado final desta edição venha a contribuir para o desenvolvimento desta área, tão carente de trabalhos de pesquisa, especialmente no Brasil – a fonética forense.

O objetivo primordial do nosso grupo tem sido desenvolver e difundir pesquisas na área de produção e percepção da fala, com foco na área forense, e tem defendido a multidisciplinaridade e a estreita relação da universidade com os praticantes da perícia estimulando o desenvolvimento da pesquisa. Assim sendo, estamos muito satisfeitos com o resultado final deste volume, pois conseguimos reunir aqui trabalhos realizados por estudiosos das áreas da linguística, das engenharias, da fonoaudiologia e da computação, sendo eles profissionais de universidades, institutos de perícia e centros de tecnologia. Os textos versam sobre temas de alta relevância, tanto para a fonética e as ciências da fala, quanto para as ciências forenses. Sendo uma revista bilíngue, o volume se apresenta com seis artigos, os três primeiros em inglês e os outros três em português.

O artigo de Helen Fraser trata dos problemas de transcrição em gravações com pouca nitidez, e do uso dessas transcrições para investigação ou para constituição de prova. A autora atenta para a importância da pesquisa colaborativa entre as ciências fonética e jurídica. Eugenia San Segundo e Pedro Gomez-Vilda analisam parâmetros glotais da vogal de preenchimento produzidas por pares de irmãos gêmeos monozigóticos e pares dizigóticos, assim como por pares de irmãos não gêmeos e pares de pessoas sem parentesco. O objetivo do trabalho foi verificar a influência genética de diversos parâmetros glotais. Esses mesmos autores, San Segundo e Gomez-Vilda, em conjunto

com Luis M. Mazaira, Agustín Álvarez e Victoria Rodellar, descrevem um estudo sobre distorção de fonação e apresentam uma metodologia que pode ser utilizada tanto no ambiente forense de comparação de locutor, como no monitoramento patológico de voz.

A seguir, na sequência de artigos em português, a comparação de locutor vai ser tratada em três perspectivas. Primeiro, Cintia Schivinscki Gonçalves e Claudia Regina Brescancini defendem a abordagem sociofonética para a perícia. Paloma Alves Miquilussi, Marilisa Exter Kalovski e Denise de Oliveira Carneiro, por sua vez, apresentam as contribuições da Fonoaudiologia para os estudos forenses, discutindo sobre as Diretrizes Curriculares para o curso de fonoaudiologia no Brasil. E, finalmente, Aline de Paula Machado e Plinio Almeida Barbosa investigam sobre a eficácia de um conjunto de medidas acústicas para exames de verificação de locutor.

Esperamos que este trabalho seja um incentivo para novas pesquisas e que esta revista se torne uma referência para a linguística forense não só no Brasil, mas em todos os países onde haja estudos sobre a interface entre as ciências da linguagem e do direito.

Maria Lúcia de Castro Gomes
Universidade Tecnológica Federal do Paraná – UTFPR

Andrea Alves Guimarães Dresch
Instituto de Criminalística do Paraná, Polícia Científica, Brasil

Denise de Oliveira Carneiro
Instituto de Criminalística do Paraná, Polícia Científica, Brasil

Transcription of indistinct forensic recordings: Problems and solutions from the perspective of phonetic science

Helen Fraser

Forensic Phonetics Australia

Abstract. *Covert recordings (speech captured on audio or video without the knowledge of the speakers) can provide powerful forensic evidence. Unfortunately, however, since it is difficult to control their recording conditions, covert recordings are often indistinct, to the extent the speech is unintelligible to listeners who do not already have knowledge or expectations about their content. This raises the question of how to present the speech evidence in court so that the trier of fact (jury, magistrate, judge, etc.) can make use of the information contained in the recording. In many jurisdictions, the answer is to have a transcript made by those who know the context (typically police working on the case), and provided to the trier of fact as an aid to perception of the speech. The present article outlines several problems with this approach, then suggests some solutions that allow maximal value of the intelligence contained in covert recordings, while reducing the risk of injustice through biased perception of indistinct audio. A key part of the suggested solution is to make a clear distinction between investigative and evidentiary uses of indistinct covert recordings, and to ensure that transcripts of evidentiary recordings be produced by professional transcribers independent of the case.*

Keywords: *Forensic phonetics, forensic transcription, covert recordings, digital forensics, forensic surveillance.*

Resumo. *Gravações clandestinas (fala capturada em áudio ou vídeo sem o conhecimento dos locutores) podem fornecer poderosa evidência forense. Infelizmente, no entanto, uma vez que é difícil controlar as condições de captura, as gravações por interceptação são muitas vezes indefinidas, na medida em que a fala pode ser ininteligível para os ouvintes que ainda não têm conhecimento ou expectativas sobre o seu conteúdo. Isso levanta a questão de como apresentar essa evidência no tribunal para que o juiz de fato (júri, magistrado, juiz, etc.) possa fazer uso das informações contidas na gravação. Em muitas jurisdições, a solução é ter uma transcrição feita por alguém que conheça o contexto (alguém da polícia que trabalhe no caso), e fornecida ao juiz de fato como auxílio à compreensão da fala. O presente artigo descreve vários problemas com essa abordagem, e depois sugere algumas soluções que permitam aproveitar ao máximo o conteúdo dessas*

gravações clandestinas, reduzindo o risco de injustiça por percepção enviesada de áudio com problemas para compreensão. Uma parte fundamental da solução é fazer uma clara distinção entre as utilizações dessas gravações clandestinas pouco claras, para investigação ou para prova, e para garantir que as transcrições de gravações de prova sejam produzidas por transcritores ou profissionais não envolvidos com o caso.

Palavras-chave: *Fonética forense, transcrições fonéticas, gravações por interceptação.*

Introduction

Technological advances are making it ever easier to collect covert recordings (conversations recorded electronically without the knowledge of the speakers). Legally obtained[1] covert recordings can potentially yield powerful evidence in criminal trials, allowing the court to hear speakers making admissions or giving information they would not have been willing to provide in person, or in an overt recording (one made with all speakers' knowledge, for example in a police interview).

A major limitation of covert recordings[2], however, is that it can be hard to control their recording conditions, with the result that the audio is often of very poor quality, to the extent it is difficult to hear what is said. This is the problem that is the topic of the present paper.

In certain circumstances, it is possible to overcome this limitation by providing a transcript to assist listeners, notably the trier of fact (the judge, magistrate or jury deciding the verdict of the trial), to hear what words are spoken. Of course, the reliability of a transcript used for this purpose is crucial. Otherwise there is a danger it might 'assist' the listeners to hear words other than those spoken in the conversation originally captured by the covert recording, and thus to reach an inappropriate evaluation of the evidence.

Unfortunately, ensuring reliable transcription, especially of poor quality audio, can be highly problematic. This is well known across multiple branches of linguistic science, where transcripts are frequently used for research purposes (Bucholtz, 2007; Heselwood, 2013), and specifically in forensic phonetics, which has a small but well-established branch dealing specifically with transcription of indistinct audio (French and Stevens, 2013; Coulthard and Johnson, 2007; Shuy, 1993).

However, these issues in transcription seem to be less well known in legal circles, where practices for the handling of audio evidence have developed with little reference to the relevant science. For example, it is common, internationally, for indistinct covert recordings played in court to be accompanied by transcripts produced by police investigating the crime. This may seem sensible, on the grounds that investigators' knowledge of the context of the recording helps them to make out words that are unclear to others — and indeed it is important to recognize that contextual knowledge can aid perception. However, as discussed below, use of transcripts by those involved in the case raises serious problems, which can affect the fairness of trials.

The present paper starts by setting out some general background about transcription. Though each element of this background is well known in the sciences of language and speech, they are rarely brought together to be viewed as a whole in the legal context. This background is then used to highlight several problems associated with use of

police transcripts of indistinct covert recordings. The paper finishes by suggesting, as a starting point for collaborative research between phonetic science and the law, some directions in which solutions to the problems might be sought, while still retaining the advantages of access to police contextual knowledge.

Definitions and distinctions

It will be useful to start by clarifying some concepts and terminology regarding speech recordings and their transcription. This will enable important distinctions to be maintained throughout the discussion, and allow forensic transcription (i.e. transcription of indistinct audio evidence) to be set in a broader framework. One aim of this section is to outline some features that make transcription in general a more difficult task than is sometimes recognised. Another is to demonstrate several ways in which the handling of forensic transcription differs from standard practice in general transcription.

What is a transcript?

The term 'transcript' was first used in the middle ages, before the development of the printing press, to denote a copy of a hand-lettered text, 'transcribed' or 'written across' from an original (Oxford Dictionary). Later, the word was used for a 'fair copy', written up from notes made during a meeting or event. This usage continues to the present day in relation to work such as that of court or parliamentary stenographers, who take shorthand notes and transcribe them into a written text, which becomes (after appropriate checking) the official version of the proceedings.

Alongside this has developed a new usage, made possible by the introduction and rapid spread of audio-recording technology. Nowadays, 'transcribe' most commonly means 'to write out (i.e. to represent in written form) the speech captured in an audio recording', and a whole new industry of audio-transcription has emerged, servicing legal, medical, research and general markets (Mills, 2010).

An unfortunate consequence of this semantic evolution is that connotations that had, reasonably, accrued to earlier uses of the word have been inappropriately transferred to the modern usage. In particular, it is often assumed by those who have never tried it that transcribing from an audio recording is a simple matter of 'writing down what you hear', a little like schoolroom dictation, making the product essentially a copy of the audio, albeit in a different medium.

Despite its ubiquity, however, this idea is inaccurate. As discussed below, it is well known in the linguistic sciences that a transcript is never anything like a copy of the audio. Nor is transcription ever simple — not even when the recorded speech seems easy to hear, and certainly not when it is indistinct. Ensuring reliability of a transcript, therefore, requires careful attention to a range of factors, some of which are canvassed briefly in the following sections.

Speech recordings: Purpose

Speech by nature is fleeting, disappearing the moment it is uttered. With audio technology it can be captured, allowing it to be heard again, in a different context or by different listeners. This can be done for a range of different reasons. One familiar purpose is to create a record of an official event, for example, a police interview or court proceedings.

An important aspect of overt recordings like these is that everyone knows their speech will be transcribed. For this reason, the speech is monitored for clarity, by the speakers themselves, or by others admonishing them to 'speak up for the tape'.

Speech recordings are also used by language scientists, to capture speech for various kinds of analysis and research. This naturally requires high-quality recordings. In the early days, this restricted research to short, clear utterances, sometimes called 'laboratory speech'. Later it became possible to capture high quality recordings of spontaneous conversational speech, though this remains notoriously difficult (Wray and Boomer, 2013).

Spontaneous conversational speech also proved extremely difficult to transcribe, even in a good quality recording. Though the overall meaning may be quite clear to the listener, detailed identification of every word is surprisingly difficult, and transcription is extremely time-consuming. However, the efforts of researchers across several decades have provided many insights into the vast differences between monitored and spontaneous speech (Shockey, 2003; Cauldwell, 2013).

These insights have had profound implications for our understanding of how speech is processed in the mind — which unfortunately remain little known outside academic circles (Cutler, 2012; Fraser, 2003). Most importantly, they indicate that speech is not perceived 'bottom up', i.e. purely from information in the acoustic signal. To a far greater extent than is evident from everyday experience, perception relies on the hearer's knowledge of the context, both internal context (available from the speech itself) and external context (from the background situation within which the recording was obtained).

This is seen, for a topical example, in the effects of recording courtroom proceedings and outsourcing the audio to casual transcribers. The 'bottom up' view would predict that having the audio should increase accuracy, as the transcribers can listen again to the actual words. In fact the opposite is true. Transcribers who have witnessed the proceedings produce better transcripts, even when relying only on shorthand notes, without reference to the audio (Wilson, 2013).

All the examples mentioned so far have been overt recordings, where the purpose is to retain a record for reference or analysis by those who heard, or could have heard, the original speech in its context. It is often possible, or even required, that those present be consulted before finalisation of the transcript. This is quite different from the purpose of a covert recording, which is to provide evidence of unmonitored conversation that would not have occurred in the presence of those listening to the recording. Checking the transcript with those present is problematic.

Within the overall category of covert recordings, an important distinction is whether the purpose is investigative or evidentiary/evidential (Haworth, 2010; French and Harrison, 2006).

Investigative uses are those related to the investigation of a matter, before it goes to trial, when detectives or other investigators attempt to uncover the facts surrounding an alleged crime. For example, if a covert recording reveals suspects making plans to meet at a particular address, this may prompt police to raid the premises at the relevant time. If the raid is successful, its results become evidence to be used in the trial. The recording itself may never be played again.

Evidentiary uses of covert recordings are those where the audio is played in court as evidence of the crime itself. Importantly, it is not just the fact of particular words having been uttered that is relevant. The manner in which the words are uttered is essential to the trier of fact in evaluating the intentions of the speakers, and the significance of the

audio evidence in relation to all the other evidence being weighed to reach a verdict. This 'manner of speech' can never be represented objectively in a transcript, no matter how detailed, but must be interpreted by the human ear. This is the reason most courts insist that it is the audio, not the transcript, that is the evidence, with the transcript provided only as an aid to listeners' perception of indistinct audio (more will be said about this below).

Speech recordings: quality

Another factor to be considered in ensuring reliable transcription is the technical quality of the recording itself. In principle, the technical quality of a recording is a separate issue from the clarity of the speech being recorded. However, in practice these interact substantially. For example, careful, pre-planned speech may be understood even when recorded with cheap equipment, while casual conversation can be surprisingly hard to understand even when the best equipment is used (as discussed above). For present purposes, then, it is useful to define quality in terms of clarity of speech, and to distinguish three levels.

Clear recordings, are those where most of the speech can be understood readily, in one sitting, by anyone with competence in the language, even if they know little about the context of the recording. This does not require studio quality, such as that appropriate in listening for pleasure. Many overt recordings are of only fair quality, and may have significant background noise. All that is required for them to fit into the *clear* category is that the speech in general is easily intelligible. Thus, while there may be issues of analysis, or of interpretation of speakers' intentions, overall there is little room for disagreement about what words are actually spoken.

Poor or *unclear* recordings are those where the speech itself is hard to understand, and listening is significantly unpleasant. On first hearing, listeners, especially those lacking background knowledge of the context, may pick up only an impression of the conversation, and not its detail. To hear it properly requires headphones, software to enable isolation of particular sections of audio, and repeated, patient listening, with note-taking. However, with these requirements in place, it is possible to hear most of what is said, to the extent that several independent transcribers are liable to reach overall agreement on most of the content — though of course there may still be differences of opinion regarding the analysis or interpretation of the recorded conversation.

Some poor quality recordings can be further classified as *indistinct*. These are recordings so poor that casual listeners are liable to understand little or nothing of the content. Even with specialised equipment, repeated listening, and contextual knowledge, it is still difficult to hear the conversation, and multiple transcribers are liable to produce significantly different versions of the content.

For most purposes, indistinct recordings are discarded, and for many, even poor quality audio is not used. Another unusual feature of covert recordings used as evidence in trials, then, is that they may be, and often are, used even if substantial portions are indistinct.

Transcripts: Purpose

In general, the most common purpose for a transcript is convenience. Written text is far easier to navigate, index, annotate and refer to than audio, which must be listened to in real time. Of course, this convenience depends on the transcript being reliable.

With court proceedings, for example, or even the minutes of meetings, elaborate checking processes have been developed over centuries to ensure the written record is not compromised by any kind of fraud or personal bias on the part of the transcriber.

Convenience also depends on the written text containing the relevant information in accessible form. The criteria for this vary considerably according to specific aspects of the intended purpose. For general use, an orthographic transcript (i.e. one using ordinary spelling) in a standardised format is usually appropriate, as it can be easily read and referenced by a wide range of users. It is worth noting, however, that this accessibility comes at the cost of considerable loss of detail. Thus even so-called 'verbatim' transcripts do not literally represent each and every word spoken (Eades, 1996).

Transcripts intended for research such as discourse analysis, for example, require far more information to be represented than is possible with normal orthographic conventions, and make extensive use of special symbols and conventions relevant to the specific topic being studied (Sidnell and Stivers, 2012; Edwards, 2008). The advantage is, again, that once the transcript has been accepted as reliable, analysis can be conducted on the written symbols, reducing (but not eliminating) the need to refer repeatedly to the actual audio. Of course, this level of detail comes at the cost not just of increased time required for preparation of the transcript, but also of reduced legibility to a general audience.

Most importantly, it is very widely recognised that even the most detailed transcription involves selection and interpretation, whether according to explicit or implicit criteria, of the features to be represented (Jefferson, 2004; Green *et al.*, 1997). This goes not only for discourse-level transcripts but also for those used in phonetic research (Heselwood, 2013; Cox, 2012). Speech is an extremely complex signal (Kreiman and Sidtis, 2011; Laver, 1994), and it is literally impossible to include every aspect in a transcript (Kerswill and Wright, 2008). This is why it is misleading to treat any transcript, no matter how detailed, as a copy, or even an objective representation, of the audio it represents.

Turning now to indistinct covert recordings used as evidence in criminal trials, we find a totally different purpose for the transcript. It does not simply allow convenient reference to words that have previously been heard clearly by multiple people. It is emphatically not intended as any kind of substitute, however limited, for the audio itself. Rather, it is intended to assist the listener in making out words that they would find difficult or impossible to hear without the transcript, leaving them free to concentrate on interpreting the speakers' intentions without undue influence from the transcriber's opinions (see Fraser, 2015 for further discussion).

Transcribers: skill level

Recognition that there is much more to producing a transcript than simply 'writing down what you hear', makes clear that a major factor in the reliability of a transcript, and its usefulness for its intended purpose, is the skill level of the transcriber. For the present context, we can distinguish several categories.

Untrained transcribers are those who undertake transcription with little or no training or systematic, reflective experience. (Of course this refers to training and experience specifically in transcription; they may have high levels of expertise in other professional skills.) Untrained transcribers may work very hard on a transcript, listening many times to the audio, and updating frequently to reflect changes in their hearing. However, due

to their lack of experience with the variability of speech perception, they may be inclined simply to accept their most recent hearing as accurate. They may also have little understanding of the effect of layout on transcript usability.

Professional transcribers have considerable experience in the kind of analytic listening needed to choose among multiple potential perceptions, as well as explicit training in representing speech and laying out transcripts in the manners suitable for particular purposes. They may even have been tested to ensure the accuracy and usability of their output.

Experts in transcription are those with a theoretical understanding of the nature of transcription, and its many complexities — some of which are summarised in this paper.

Experts in forensic transcription are those with high-level qualifications in relevant branches of phonetics, enabling them to evaluate the acoustic evidence supporting (or not) a particular transcript. It is important to acknowledge that such experts may be employed by police-based forensic analysis units. These officers have little direct involvement in investigation of cases, creating a very different situation from the one discussed below, where indistinct audio is transcribed by police working on the case, who lack any expertise in transcription.

Expertise in forensic transcription is often sought when there is a 'disputed utterance' (French, 1990). Without it, resolving the dispute becomes a mere 'battle of the ears'. It is worth emphasising though, that forensic transcription is a broader topic than resolution of disputed utterances. Many covert recordings have indistinct portions lasting hours or even days. Unfortunately, however, as discussed below, these seem less likely to be sent for expert analysis than shorter segments with explicit alternative transcriptions.

It is notable that in virtually all contexts in which transcripts are used, it is a standard requirement that they will be prepared by professional transcribers, and that the less clear the audio the more skill the transcriber needs. Even with covert audio, transcripts of clear or poor recordings are normally produced by transcribers with at least some training and experience. Again indistinct covert recordings are an exception, being often transcribed by police with no training whatsoever in transcription.

Transcribers: relationship to material

Another factor that is usually given considerable attention in ensuring the reliability of transcripts is the relationship of the transcriber to the material being transcribed. We have seen already that, while general background knowledge of the context is useful or essential in preparing a transcript that is reliable and useful for its purpose, steps must be taken to ensure effects of personal bias are avoided.

However there is another kind of bias which is more difficult to avoid. Cognitive bias (Kahneman, 2011) is the tendency for people to perceive something they expect, assume or want to be present, even if it is not objectively there. Importantly, cognitive bias is unconscious. It can exist even in those who feel themselves to be neutral, and it cannot be controlled by a mere effort of will (Thompson, 2011). This is well known as the reason that medical and other sciences insist on 'double-blind' analysis of experimental results. A very similar, though less widely publicised, effect exists in speech perception, where it is often called 'priming'. It is discussed further below.

In view of these observations, we can classify transcribers according to their relationship to the audio being transcribed. An *independent* or *impartial* transcriber has

little or no information about the specific context of the recording, beyond general background knowledge essential for producing the desired type of output, and no prior opinion about, or particular interest in, the ultimate use that will be made of the transcript.

By contrast, an *involved* transcriber does have such an opinion or interest. In linguistic research, considerable effort is made to ensure transcripts are prepared by independent, impartial transcribers. Even where this is difficult for practical reasons, academic standards require at least a proportion to be transcribed in a way that 'blinds' the transcriber to the purpose, and the level of agreement between this and other portions to be reported.

Here again, transcription of indistinct audio diverges from standard practice, with transcripts by involved transcribers, notably police, often used.

How does forensic transcription relate to transcription in general?

With this brief background, we are in a position to locate forensic transcription within a broader framework of understanding regarding what a transcript is, and the standard processes that must be followed to ensure it is reliable.

Most importantly, the discussion makes clear that transcription of lengthy indistinct covert recordings is very different from most other forms of transcription. First, the quality of such audio is far worse than that used for most other purposes. Second, the intended purpose of the transcript is very different, going beyond mere convenience of reference, to influence on perception. Third, despite this, there is heavy reliance on transcripts by untrained, involved transcribers. Finally, it is worth noting that the negative consequences of errors in a forensic transcript are considerably higher than those in transcripts used for most other purposes. When the comparison is set out explicitly like this, it becomes unsurprising that it results in a range of problems.

Problems in current practice regarding transcription of indistinct forensic recordings

The majority of covert recordings are obtained on behalf of police, so it is natural that the audio would go first to them to determine if any of the material is pertinent to their investigation. To save time, they may obtain a summary transcript (notes on the content of the audio) prepared by an independent transcription agency, to help guide them to relevant parts. Police then note any information of investigative value, which they act on appropriately. At this investigative stage, to the extent police transcribers' contextual knowledge of the case helps them to hear indistinct audio accurately, the intelligence provided by the recording will be valuable, while inaccuracy in their transcription is likely to result at worst in some waste of time or effort.

It is only later, when the investigation is complete, and a case is being prepared for trial, that it is necessary to decide which parts of the covert recording are relevant as evidence in their own right, and require detailed transcription for use in court. Clear and poor recordings are sent to independent professional transcribers. It is indistinct parts, too hard for the professionals, that may be transcribed by detectives from the case.

The audio, along with the transcript, is then disclosed to the defendant or representatives for checking, admission of the evidence to the trial is sought, and the audio is ultimately played to the trier of fact with the transcript provided as an aid to perception of indistinct sections. As indicated earlier, this process has numerous problems.

Inaccuracy of content

In terms of the classifications above, police transcribers are typically *untrained* and *involved*, and it is these characteristics that create the problems discussed here, not the fact that the transcribers are police. Thus the present remarks apply equally to transcripts by any untrained, involved transcriber, regardless of whether their transcripts are used by prosecution or defence.

The most notable problem with transcripts by untrained, involved transcribers is that they are liable to be inaccurate. The advantage such transcribers have over a professional, independent transcriber — and the reason their transcripts are used — is their knowledge about the specific background and context of the recording. And it is important to acknowledge, as discussed above, that such knowledge can sometimes enable them to hear words that are unintelligible to others. However, due to the effects of cognitive bias, this contextual knowledge is a double-edged sword, creating more problems than it solves.

First, involved transcribers tend to focus on sections they consider to have most relevance to their investigation. This may lead them to pay less attention to parts they consider unimportant, with heavy use of fillers such as 'indistinct' or 'indecipherable' or simply '[...]'. Of course, from a less involved, or differently involved, perspective, these sections may contain crucial information.

Second, they may mis-hear (i.e. hear words or phrases even when these are contradicted by the acoustics) or over-interpret the audio (i.e. hear words or phrases even when they are not well supported by the acoustics). Unfortunately, to untrained listeners, the experience is the same whether valid contextual knowledge is helping them to hear accurately, or assumptions or expectations are creating perceptual error. In both cases, they feel confident they are simply 'hearing what is there to be heard', and accept their perception uncritically.

Since this phenomenon is rather little known outside phonetic science, it can seem hard to accept when first encountered. However, it is a very well established feature of human speech perception, with strong experimental and experiential support going back at least to the 1950s (Miller, 1951; Cutler, 2010). A particularly clear example is given by one of the experiments that first brought this interesting feature of human speech perception to light. Bruce (1958), investigating the effect of listeners' mental 'set', or context-based anticipations, created a number of sentences with the following form:

Sentence 1: I tell you that our team will win the cup next year

Sentence 2: You said it would rain but the sun has come out now

Participants heard these sentences 'masked' with a hissing noise, which he calibrated so as to make the sentences around 25% intelligible. Next, he presented the same sentences in the same level of masking noise — but this time he preceded each with a keyword that gave it a context. For example, the keyword for Sentence 1 was SPORT, for Sentence 2, WEATHER, and so on.

As predicted, sentences preceded by their keyword were more intelligible. However, an unexpected discovery was what happened when he played the masked sentences with the wrong keyword. This did not hinder perception, as had been predicted. Rather it created a different perception. For example, playing Sentence 1 with the keyword

FOOD (instead of SPORT) led participants to hear a range of sentences quite different from the one that had actually been spoken, such as:

Sentence 1 (FOOD): I tell you that I feel more hungry than you are

Playing the same Sentence 1 with the keyword TRAVEL led participants to hear yet another range of sentences, again different from the one that had actually been spoken, such as:

Sentence 1 (TRAVEL) I tell you that I too will leave next year

Crucially, these erroneous perceptions were heard with no diminution of confidence. Listeners felt they were simply hearing what was there to be heard, in just the same way they did when their perception was accurate — an early indication that listeners' personal confidence in their perception of indistinct audio is a poor guide to the accuracy of a transcript.

Though these findings were surprising at the time, and remain less widely known by the general public than similar effects in other areas of forensic evidence (Ridley *et al.*, 2013), the role of this kind of contextual priming in speech perception is now well understood in the phonetic sciences, as discussed above. Therefore, the argument that police transcribers might be similarly affected carries no suggestion of any personal or moral bias. However it is essential to recognize the role of cognitive bias, in audio as in other forms of evidence. It would be more surprising to find that police transcripts were not affected by cognitive bias than that they are.

Inadequacy of checking procedures

A common legal response, upon mention of these problems with police transcripts, is to point out that the transcripts are not just accepted uncritically. Most jurisdictions have processes of checking that must be undergone before the transcript is admitted. Unfortunately, however, these processes are frequently inadequate. They commonly involve the defendant and/or legal representatives checking the transcript against the audio. This may be acceptable, if not ideal, when the audio is fairly clear, and the transcript is professionally laid out. Such checking may lead to identification of one or more 'disputed utterances' which can be sent for expert analysis.

However, well known findings of phonetic science, discussed above and taken up again shortly, suggest that with indistinct audio it is not effective to evaluate transcripts in this way. Inexpert listeners may hear (accurately or not) a few of the phrases they read in the transcript, and assume the rest of it must be reliable — and of course, at this stage they do not know which utterances are going to be picked out as relevant in the trial, and thus require special attention.

A further feature of police transcripts can make this kind of checking even less likely to be effective than it might otherwise be, namely: their poor layout. We have emphasised above that an important part of training in transcription is learning how to format a transcript in a manner appropriate to its purpose. Police have no such training, and their transcripts are typically quite unsuited to the purpose of careful checking against the audio by another listener. Here, as an example, is an exact replica of a section of a police transcript from a real case (please note this represents less than a minute from within a 136-page transcript of a covert recording featuring several hours of barely audible conversation).

Male voice (F) huh?'

Male voice (P) '(indec — possibly fix the counter) (indec).' Male voice (F?) 'fix the what?'

Voices (indec).

Motor attempting to start.

Male voice (R?) 'no petrol.'

Motor attempting to start.

Voices (indec) swearing (fuck).

Male voice "What, What you want me to do?" (F?)

"No it's all good" (P)

Female voice (idec)

The question of exactly how forensic transcripts should be laid out for maximal usability by listeners in court is a complex one (the subject of research in progress by the present author), but it is clear layouts like this one are far from ideal (regardless of accuracy — which in this case was low). The mixing-up of speaker attributions, words heard, unconfident suggestions, reference to background noises, etc., makes it extremely hard to read this against rapidly-passing indistinct voices with a great deal of loud background noise.

Unrealistic expectations of trier of fact

Ultimately, the content of the audio must be evaluated by the trier of fact, as one piece of evidence to be weighed in with all the other evidence in a case, with the aim of reaching a verdict as to the guilt of the person undergoing trial. As discussed earlier, it is the audio that is the evidence. The transcript is intended only as an aid to the perception of listeners who may find an indistinct recording difficult to hear.

Again, while this use 'only as an aid' may sound reasonable based on everyday understanding of speech perception, phonetic science vigorously opposes it. Two recent experimental studies have demonstrated the reasons for this opposition, using indistinct audio and police transcripts from real cases, and closely simulating the experience of juries in interpreting a poor quality recording with the aid of an inaccurate police transcript.

Fraser *et al.* (2011) used a disputed utterance from the famous the case of David Bain (Innes, 2011), convicted in 1995 of murdering his entire family, then acquitted on all charges in 2009, after 13 years in prison. Subjects, divided into two experimental groups, listened repeatedly to the same audio, while evidence about the case, including the inaccurate police transcript, was gradually revealed to them across seven 'evidence points'. At each evidence point, they were asked what they heard in the audio, along with various other questions.

At first, virtually no one in either group heard anything remotely like the police transcript. However, at evidence point 4, when a transcript was explicitly suggested, over 30% of Group A, who were given the inaccurate police transcript, confidently heard the exact words of the transcript, with many others displaying perception clearly influenced by the transcript.

Subsequent evidence points attempted to convince participants that the police transcript was inaccurate, but brought about little change in perception. The final evidence

point explained the purpose of the experiment, with information the materials had been deliberately chosen to show how easily a demonstrably inaccurate transcript could mislead listeners' perception of indistinct audio. However, even after being told that experts on both sides of the case had agreed the police transcript was inaccurate, 17% of Group A claimed to hear its exact words, with, again, many more influenced by it in a variety of ways.

Most interesting of all was the response of Group B at this last evidence point. Group B had been primed with a different transcript, and had never heard anything like the police version. However, mention of the police transcript at the final evidence point prompted 12% of Group B to hear the audio exactly in line with the transcript, and many others to be influenced by it.

These results suggest that it may be unrealistic to expect a jury to use a transcript 'only as an aid', and to reach their own independent conclusion as to what is said in an indistinct recording. A useful and important follow-up study (Bonifaz, 2014) demonstrated that explicitly informing participants at the outset that they might potentially be primed by the suggested transcript did not significantly reduce their tendency to be affected by the prime.

The second experiment (Fraser and Stevenson, 2014) carried this work forward by looking at how knowledge, or assumptions, about the context can affect perception of words in an indistinct recording. Again, the experiment used audio from a real case, this time a short excerpt from a 38-minute recording of extremely poor quality — along with the police transcript used in the trial and later demonstrated to be inaccurate. It was conducted in two parts.

The first part demonstrated that, in the absence of contextual knowledge, the police transcript was actually quite implausible. Its priming effect was less than usual when first presented to participants, and even the few who initially accepted it, quickly abandoned it when presented with a more plausible alternative. This raises the question of how the transcript could ever have been accepted in court. An answer was suggested by the second part of the experiment.

The second part began by giving participants contextual knowledge about the trial, similar to the background available to lawyers in the case, and, later, to the jury. These participants were far more likely than those in the first part to hear the audio in line with the inaccurate police transcript, and far less likely to be swayed by the alternative, more plausible, transcript. As with the 2011 experiment, many subtle effects were found on the perception even of participants who rejected the police transcript.

Again, these results were interpreted as evidence it is unrealistic to instruct a trier of fact that they should use the transcript only as an aid. They also went further to suggest that leaving evaluation of forensic transcripts to defendants and their legal representatives gives insufficient protection from inaccuracy — and to demonstrate other problems with the law regarding police transcripts.

Finally, in both experiments, before-and-after questions revealed a strong effect of seeing the inaccurate transcript on participants' ultimate opinions about the guilt of the defendant, even in those who rejected its exact words. Participants believed their understanding of the case was influenced by the audio, not realising how much their perception of the audio was influenced by their understanding of the case.

Towards some solutions

It is hoped that this brief discussion has demonstrated some dangers of using transcripts by untrained, involved transcribers as an aid to perception of indistinct covert recordings presented as evidence in criminal trials. In summary, it is highly likely that such a transcript will be inaccurate in its representation of what is said in the recording, highly unlikely that errors will be picked up through the currently standard checking processes, and highly likely that the perception of the trier of fact will be unconsciously influenced by the transcript, with consequences for their evaluation of the significance of the audio in relation to the overall verdict.

The effect is that transcribers provide a 'view' of the audio evidence that may unwittingly influence those who believe themselves to be reaching an independent interpretation — much in the way that eye witnesses can be unwittingly influenced by others' descriptions of what they have seen (Loftus and Palmer, 1974).

This section turns to consideration of how to resolve these problems. To start, it may be worth noting one potential solution that is unlikely to be effective: letting the trier of fact hear indistinct covert recordings with no transcript. While this clearly reduces the direct influence of the transcript on a trier of fact's perception, it certainly does not ensure they will hear the audio accurately, due to the strong effect of the context itself on perception — demonstrated by Fraser and Stevenson (2014). Listeners need a reliable transcript to hear indistinct audio properly.

The question that remains is: how to ensure the transcript provided is reliable. In particular, are there practical ways to limit the disadvantages of using police transcripts for this purpose, while still retaining their potential to provide intelligence essential to the solving of crimes?

In fact, it may be relatively easy to achieve this desirable outcome through close collaboration between phonetic science and the legal system, keeping in mind the distinctions outlined above. This section outlines some suggestions that emerge from the earlier discussion, and may provide a starting point for such collaborative research.

Distinguish clearly between investigative and evidentiary uses of covert recordings

Perhaps the most important recommendation is to recognise the major difference in transcripts used for investigative purposes, as opposed to those used for evidentiary purposes. As long as indistinct covert recordings are used only for investigative purposes, little harm and much advantage can accrue from reliance on interpretations of involved transcribers, even if they are untrained in transcription.

The dangers described earlier arise when police transcripts are used for evidentiary purposes. Here, then, it can be recommended that the barriers to moving indistinct covert audio with police transcripts from investigative use to trial evidence should be far higher than is currently common. It may be worth emphasising that this is true even if perception of some words in the transcript may have been shown to be accurate by their value during investigative stages of the case. While this may lend credence to the transcription of those particular words, it is no guarantee of overall reliability of the transcript.

Decrease reliance on indistinct audio as evidence

The next recommendation is an overall reduction in use of indistinct audio as evidence in trials. In some cases, especially with shorter utterances, detailed analysis can allow experts in forensic transcription to provide a reliable transcript. However, in many other cases, even expert analysis has inconclusive results. This indicates the audio is simply untranscribable. It has a status similar to that of a smudged fingerprint, or inconclusive DNA results. The appropriate response, with audio as with other kinds of forensic evidence, is to exclude the evidence from the trial — not to admit it, with a police interpretation, 'for the jury to decide'.

While exclusion of indistinct audio may be frustrating at first, it might have the advantage of greater attention being given to ensuring that covert recordings are obtained in such a way as to ensure the speech is clear.

For evidentiary uses, insist that covert recordings be (re-)transcribed by an independent, professional transcriber

Where indistinct covert recordings are admitted as evidence, it is essential that they should be accompanied by a reliable transcript. A first step is to ensure the transcript was produced by a transcriber who is independent of the case, with no stake in its outcome, and minimal contextual knowledge, and no influence from seeing a police version. However, the value of gradually providing specific contextual knowledge to the transcriber through a process of 'sequential unmasking' (Thompson, 2011) is a topic of current research by the present author.

Any disputes regarding words heard in the transcript should be resolved before the audio is admitted as evidence, through evaluation of the audio by a genuine expert in forensic transcription. Naturally this expert should also be independent of the case, in line with requirements increasingly being enforced in other branches of forensic science (Edmond and San Roque, 2012).

Present indistinct audio to the trier of fact in a way that enables reliable evaluation

Indistinct audio evidence should be prepared so as to make realistic demands of listeners. That means, for example, restricting the overall amount of audio (in one trial known to the author, the court had to listen to covert recordings for more than six days straight); dividing it into short sections containing coherent parts of conversations; providing headphones and software allowing replay at will; and allowing time for it to be listened to carefully — i.e. many times longer than the duration of the recording (cf. French and Stevens, 2013.

It also, of course, means presenting the audio with a reliable transcript, laid out in a way that assists listeners to follow the speech and find their way around the recording, enabling them to pay attention to intonation, tone of voice and other aspects affecting interpretation of the speakers' intentions. Finally, reading or quoting from the transcript by barristers should be discouraged (Haworth, 2010), and emphasis placed on the need for interpretation of the evidence to be based on listening to the audio.

Conclusion

This paper has demonstrated a paradoxical situation in the legal process, whereby audio of worse than usual clarity is subjected to transcription practices of less than usual rigour.

It is notable, for example, that relatively clear, overt recordings, such as police interviews, are transcribed with more accountability than indistinct covert recordings. A range of resulting problems has been discussed, and directions for solutions suggested on the basis of well established research in the linguistic sciences.

A first step in solving these problems is increased publicity for general issues such as those raised in this paper. Unfortunately, it may not be sufficient for experts in forensic transcription to wait to be asked for assistance with specific disputes in individual recordings. In many cases, arguably those with the worst audio and the most unreliable transcripts, the legal process relies on its own checking procedures, not realising their inadequacy.

Ultimately, there is a need for collaborative research between the phonetic sciences and the law in developing evidence-based practices to ensure indistinct covert recordings used as evidence in criminal cases are accompanied by reliable and usable transcripts. It is hoped the present paper may spark interest in this kind of initiative.

Acknowledgments

This is a modified and expanded version of a paper presented at IAFPA, Zurich, 1-3 September 2014. It has benefitted from useful discussion with colleagues at that meeting, and later at *Theories, Practices and Instruments of Forensic Linguistics*, Rome, 1 December 2014, as well as from comments by Jose Antonio Mompeán González, and two reviewers. Of course, the views expressed are those of the author and all responsibility for errors or omissions rests with her.

Notes

[1]This paper is based mainly on experience with the Australian context, where covert surveillance is governed by a variety of legal instruments, including the *Commonwealth Surveillance Devices Act 2007* and related State-based legislation (see Australian Law Reform Commission 2008). It is hoped that parts of the discussion may also be relevant in other jurisdictions.

[2]Covert recordings can be divided into two broad classes: telephone intercepts and environmental recordings (made with a microphone placed in the environment of the speaker). This paper deals mainly with environmental recordings, though some of the remarks may be relevant to telephone intercept material, where this is of poor quality.

References

Australian Law Reform Commission, (2008). *For Your Information: Australian privacy law and practice.* Commonwealth of Australia: ALRC Report 108.

Bonifaz, S. (2014). The effect of priming awareness when listening to disputed utterances. Master's thesis, University of York.

Bruce, D. (1958). The effect of listeners' anticipations on the intelligibility of heard speech. *Language and Speech*, 1, 79–97.

Bucholtz, M. (2007). Variation in transcription. *Discourse Studies*, 9(6), 784–808.

Cauldwell, R. (2013). *Phonology for Listening.* Birmingham: Speech in Action.

Coulthard, R. and Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence.* London and New York: Routledge.

Cox, F. (2012). *Australian English Pronunciation and Transcription.* Cambridge: Cambridge University Press.

Cutler, A. (2010). Abstraction-based efficiency in the lexicon. *Laboratory Phonology*, 1(2), 301–318.

Cutler, A. (2012). *Native Listening*. Cambridge, MA: MIT Press.

Eades, D. (1996). Verbatim courtroom transcripts and discourse analysis. In H. Kniffke, Ed., *Recent Developments in Forensic Linguistics*. Frankfurt: Peter Lang.

Edmond, G. and San Roque, M. (2012). The cool crucible: Forensic science and the frailty of the criminal trial. *Current Issues in Criminal Justice*, 24(1), 51–64.

Edwards, J. (2008). The transcription of discourse. In D. Schiffrin, D. Tannen and H. E. Hamilton, Eds., *The Handbook of Discourse Analysis*. Oxford: Blackwell Publishing Ltd.

Fraser, H. (2003). Issues in transcription: Factors affecting the reliability of transcripts as evidence in legal cases. *International Journal of Speech Language and the Law*, 10(2), 203–226.

Fraser, H. (2015). Transcription of indistinct covert recordings used as evidence in criminal trials. In H. Selby and I. Freckelton, Eds., *Expert Evidence*. Thomson Reuters.

Fraser, H. and Stevenson, B. (2014). The power and persistence of contextual priming: More risks in using police transcripts to aid jurors' perception of poor quality covert recordings. *International Journal of Evidence and Proof*, 18, 205–229.

Fraser, H., Stevenson, B. and Marks, T. (2011). Interpretation of a crisis call: Persistence of a primed perception of a disputed utterance. *International Journal of Speech Language and the Law*, 18(2), 261–292.

French, P. (1990). Analytic procedures for the determination of disputed utterances. In H. Kniffke, Ed., *Texte zu Theorie und Praxis Forensischer Linguistik*. Tübingen: Niemeyer.

French, P. and Harrison, P. (2006). Investigative and evidential applications of forensic speech science. In A. Heaton-Armstrong and E. Shepherd, Eds., *Witness Testimony: Psychological, investigative and evidential perspectives*. Oxford: Oxford University Press.

French, P. and Stevens, L. (2013). Forensic speech science. In R. Knight and M. Jones, Eds., *Bloomsbury Companion to Phonetics*. London: Continuum.

Green, J., Franquiz, M. and Dixon, C. (1997). The myth of the objective transcript: Transcribing as a situated act. *TESOL Quarterly*, 31(1), 172–176.

Haworth, K. (2010). Police interviews as evidence. In M. Coulthard and A. Johnson, Eds., *Routledge Handbook of Forensic Linguistics*. Abingdon and New York: Routledge.

Heselwood, B. (2013). *Phonetic Transcription in Theory and Practice*. Edinburgh: Edinburgh University Press.

Innes, B. (2011). R v David Bain: A unique case in New Zealand legal and linguistic history. *International Journal of Speech, Language and the Law*, 18(1), 145–155.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. Lerner, Ed., *Conversation Analysis: Studies from the first generation*. Amsterdam: John Benjamins.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar Straus Giroux.

Kerswill, P. and Wright, S. (2008). The validity of phonetic transcription: Limitations of a sociolinguistic research tool. *Language Variation and Change*, 2(3), 255–275.

Kreiman, J. and Sidtis, D. (2011). *Foundations of Voice Studies: An interdisciplinary approach to voice production and perception*. Oxford: Wiley-Blackwell.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Loftus, E. and Palmer, J. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5), 585–589.

Miller, G. (1951). *Language and Communication*. New York: McGraw-Hill.

Mills, L. (2010). *Jump-Start Your Work at Home General Transcription Career: The Fast and Easy Way to Get Started!* Amazon.

Ridley, A., Gabbert, F. and La Rooy, D. (2013). *Suggestibility in Legal Contexts*. Oxford: Wiley-Blackwell.

Shockey, L. (2003). *Sound Patterns of Spoken English*. Oxford: Blackwell.

Shuy, R. (1993). *Language Crimes: The use and abuse of language evidence in the courtroom*. Oxford: Blackwell.

Sidnell, J. and Stivers, T. (2012). *The Handbook of Conversation Analysis*. John Wiley & Sons.

Thompson, W. (2011). What role should investigative facts play in the evaluation of scientific evidence? *Australian Journal of Forensic Sciences*, 43(2–3), 123–134.

Wilson, R. (2013). Warning in Patel case highlights court transcript weakness. *Sunshine Coast Daily*, Available from: http://www.sunshinecoastdaily.com.au/news/warning-patel-case-highlights-court-transcript-wea/2050544/.

Wray, A. and Boomer, A. (2013). *Projects in Linguistics and Language Studies*. London: Routledge, 3 ed.

Evaluating the forensic importance of glottal source features through the voice analysis of twins and non-twin siblings

Eugenia San Segundo & Pedro Gómez-Vilda

Consejo Superior de Investigaciones Científicas,
Universidad Internacional Menéndez Pelayo
&
Center for Biomedical Technology,
Universidad Politécnica de Madrid

Abstract. *In this study we have analyzed 853 tokens of the vowel filler [eː], extracted from spontaneous speech fragments of 54 male Spanish speakers (North-Central Peninsular variety), each one recorded on two separate sessions. The speakers — to be compared in a pairwise fashion – were divided in four groups: 24 monozygotic (MZ) twins, 10 dizygotic (DZ) twins, 8 non-twin brothers and 12 unrelated speakers. From the extracted vowel fillers, considered long enough for a glottal analysis (around 160 milliseconds), a vector of 68 glottal parameters was created. Our hypothesis that higher similarity values would be found in the intra-pair comparison of MZ twins than in DZ twins, brothers or unrelated speakers was confirmed, which suggests that the glottal parameters under investigation are genetically influenced. This finding seems of great forensic importance, as a phonetic parameter is considered forensically robust provided that it exhibits large between-speaker variation while it remains as consistent as possible for each speaker (i.e. small within-speaker variation).*

Keywords: *Forensic phonetics, glottal source, twins, siblings, biometry, phonation.*

Resumo. *Neste trabalho foram analisadas 853 amostras de preenchimento da vogal [eː], extraídas a partir de fragmentos de fala espontânea de 54 falantes espanhóis do sexo masculino (variedade de fala Norte-Central Peninsular), cada um gravado em duas sessões separadas. Os falantes — comparados dois a dois — foram divididos em quatro grupos: 24 gêmeos monozigóticos (MZ), 10 gêmeos dizigóticos (DZ), 8 irmãos não gêmeos e 12 falantes sem parentesco. A partir das vogais de preenchimento extraídas, consideradas suficientemente longas para uma análise glotal (cerca de 160 milissegundos), um vector de 68 parâmetros glotais foi criado. Nossa hipótese de que seriam encontrados valores de similaridade mais elevados na comparação intra-par dos gêmeos monozigóticos do que na dos*

gêmeos DZ, dos irmãos ou dos falantes sem parentesco foi confirmada, o que sugere que os parâmetros glotais sob investigação são geneticamente influenciados. Essa descoberta parece ser de grande importância forense, na medida em que um parâmetro fonético é considerado robusto para a área forense desde que contenha uma grande variação entre-falantes, enquanto permanece tão consistente quanto possível para cada falante (ou seja, pequena variação intra-falante).

Palavras-chave: *Fonética forense, fonte glotal, gêmeos, irmãos, biometria, fonação.*

Introduction

In this investigation we have explored the voice characteristics of four speaker groups: monozygotic (MZ) twins, dizygotic (DZ) twins, male non-twin siblings (i.e. brothers, B) and unrelated speakers (US). Among other possible phonetic parameters that could be analyzed in these speakers with forensic purposes (see San Segundo, 2014), on this occasion we have focused on a group of glottal features reported to show good identification results in previous studies (Gómez-Vilda *et al.*, 2010, 2012).

In this introduction we will first describe — in a succinct way — the scientific field to which this study mostly contributes: Forensic Phonetics, and more specifically Forensic Speaker Comparison (FSC)[1] Secondly, we will explain the relevance of the *twin methodology* for this discipline. In a third stage, we will specifically detail how glottal source features have proved useful to discriminate speakers in several studies. This will serve as a state-of-the-art background against which the research hypothesis can be set, together with the methodology, in the next section.

Forensic Phonetics is the application of Phonetics aimed at solving any type of legal issue, or, in the words of Jessen (2008: 671), "the application of the knowledge, theories and methods of general phonetics to practical tasks that arise out of a context of police work or the presentation of evidence in court". There are many tasks which a phonetician may be requested to perform for forensic purposes. French (1994), Rose (2002) and French and Stevens (2013) are only some references where all these forensic tasks are explained in some detail. A brief overview of task classifications by the above-mentioned authors can be read in San Segundo (2014), where five task subgroups are mentioned: (1) determination of unclear or contested utterances — closely related to phonetic transcription; (2) examination of the authenticity of audio recordings; (3) design and validation of voice line-ups; (4) speaker profiling and LADO (Language Analysis for the Determination of Origin of Asylum Seekers)[2]; and (5) Forensic Speaker Comparison (FSC from now on). Out of all these tasks, the one for which speech experts are more frequently required, according to French and Stevens (2013), is the last one. In such cases, the experts have to compare the voice of an *offender* (i.e. the speech samples of an unknown speaker) with the voice of a suspect or several suspects (i.e. the speech samples of known origin). The kind of criminal offenses which are typically involved in FSC usually take place over the telephone, whether they are cases of drug dealers arranging illegal transactions, fraudulent bank deals, bomb hoaxes, kidnappers' ransom demands, or stalking offenses.

Twin studies are not especially widespread in Forensic Phonetics, despite the fact that other forensic disciplines have evinced a clear interest in twin discriminability, particularly in recent times. In San Segundo (2013), some examples were mentioned belonging to DNA testing, fingerprint identification and handwriting discrimination of

twins. A more extensive review of voice-related studies focusing on twins is provided in San Segundo (2014), where thirty-nine works were described, encompassing the year span 1948-2014. What all of them have in common is that they tackle the issue of voice similarity in twins and non-twin siblings, either from an articulatory, acoustical, perceptual or automatic point of view. Of course, not all of the reviewed studies stem from a forensic-phonetic perspective (e.g. trying to answer research questions of interest for this field) but most of them draw on the twin methodology. In other words, they involve the comparison of monozygotic (MZ) and dizygotic (DZ) twins with the aim of finding the relative contribution of genetic and environmental factors on the differences found between them. The twin research methodology offers several design variations of the classic twin method, which compares the resemblance within MZ twin pairs to the resemblance within DZ twin pairs, assuming equal environment influences for both types of twins:

> Differences within MZ twin pairs are explained by environmental effects because all genetic inheritance is commonly shared. In contrast, differences within DZ twin pairs are associated with both genetic and environmental influences because these twins share half their genes, on average, by descent. (Segal, 1990: 613)

In other words, what the twin methodology suggests is that any 'excess' of similarity in MZs over DZs refers to "the proportion of phenotypic variation that can be attributed to genetic variance" (Tomblin and Buckwalter, 1998: 189). In San Segundo (2014) we described the genetic endowment of MZ and DZ twins (100% shared genes in the former, and 50% shared genes in the latter) as well as the environmental influences possibly affecting their voice and speech. In relation to this last aspect, we tried to link "environmental influences" not only to the prenatal-perinatal-postnatal division provided for instance by Stromswold (2006), but also to sociolinguistic perspectives which provide insightful observations about the effects exerted by the family on the linguistic output of individuals (Hazen, 2002). Equally important in this respect are the existing investigations evolving around the idea of 'intratwin mimetism' (Debruyne *et al.*, 2002), which would be more commonly found in MZ than in DZ twins.

All in all, the forensic importance of investigating twins' voices lies in the fact that these speakers are the most extreme cases of physical similarity in human beings. The fact that they are genetically identical — in the case of MZ twins — or very similar — in the case of DZ twins — and most frequently raised in the same circumstances, make their voices highly confusable. Distinguishing them is therefore a challenge in a forensic context, as acknowledged by authors such as Künzel (2010). Some real cases involving the forensic comparison of speech samples in twins and non-twin siblings can be found in Rose (2002) or Rose (2006). Furthermore, Mora (2013) described in a recent piece of news how the perpetrator of six rapes in France could not be clearly identified on the basis of DNA, resulting in the arrest of two MZ twins. Having acknowledged the existence of real offences involving twins — which suggests that the study of these speakers is not so exotic as one could a priori think — it should be pointed out that there is an interest in this kind of investigations *per se*. As explained in San Segundo (2014: 1), "the study of genetically identical speakers (MZ twins) and their comparison with non-identical siblings [...] allows gaining insight into the contribution of nurture and nature in the speech patterns of speakers in general". See next section for a more in-depth explana-

tion of our main hypothesis: the more genetically influenced a phonetic parameter is, the more robust it will be for general speaker comparison.

A final aspect that we would like to highlight in this introduction is related to the use of glottal features for speaker comparison. Current methodologies in FSC are varied and they imply the analysis of multiple features. Indeed it is not uncommon to characterize this forensic-phonetic subdiscipline by its lack of consensus over the analysis and comparison techniques used, but also over issues like the expression of conclusions. Cambier-Langeveld (2007) and Gold and French (2011) provide good summaries of the most common international practices in FSC, with some detailed information about most frequent acoustic measures, relative weighting attached to those parameters, as well as an attempt to classify the different methods. Yet it is interesting to note that glottal source features do not specifically appear in either work. It could be inferred that this kind of features are subsumed within the broader category 'voice quality', which only appears in Gold and French (2011). However, what the authors mean by voice quality is not actually explained in the article[3]. As a matter of fact, the definition of this parameter is not absent of complexity and ambiguity, as Gil and San Segundo (2014) tried to show. This concept is most frequently associated with perceptual analyses, mainly following the phonetic description of voice quality and the perceptual protocol described in Laver *et al.* (1981), known as Vocal Profile Analysis (VPA). The investigation of Stevens and French (2012) represents an example of the application of the VPA scheme to the characterization of voices for forensic purposes. While this protocol tries to objectify voice quality and it actually includes analysis categories related to the voice source (i.e. laryngeal tension, larynx position and phonation types), it remains a perceptual evaluation. The search for acoustical correlates of those perceptual measures is yet open to further investigation. The importance of undertaking this kind of research was already mentioned by Nolan (1983).

> It will be worth developing and improving this work [the work of specifying the acoustic correlates of an auditory phonetic framework for classifying voice qualities] since, from the point of view of speaker identification it provides an approach to the problem of classifying voices alternative, and complementary, to the more usual one of picking readily measurable acoustic features and investigating, in a relatively unguided way, how these features vary among a population of speakers. (Nolan, 1983: 108)

Taking into account the distinction (e.g. Jessen, 1997) between supralaryngeal voice quality and laryngeal voice quality, if we focus on the latter (i.e. voice aspects related to the glottal source), some forensic studies have aimed to investigate the speaker discriminatory potential of this type of features, from classical distortion parameters like jitter and shimmer (Künzel and Köster, 1992) to other laryngeal parameters related to the ratio between harmonics (Jessen, 1997), or later approaches suggesting the use of vocal source information to improve speaker recognition systems (Zheng, 2005). In this line, studies like Gómez-Vilda *et al.* (2008, 2009) or Gómez-Vilda *et al.* (2012) have proved that their voice analysis methodology — based on previous voice pathology investigations such as Gómez-Vilda *et al.* (2007) — is also useful for forensic speaker comparison. San Segundo and Gómez-Vilda (2013) or San Segundo and Gómez-Vilda (2014) represent some preliminary studies that have specifically tested in twins this methodology, which presents the advantage of splitting vocal from glottal information –by means of inverse filtering — thus opening the possibility of independently studying vocal and glottal components.

Research hypothesis and methodology

We start from the premise that a parameter that is genetically influenced will be a robust parameter for FSC. In other words, it will be highly speaker-discriminant for the comparison of the unknown and known speech samples. It is widely known in this discipline that some criteria exist for selecting a useful or robust forensic-phonetic parameter. Wolf (1972) set out these criteria and since then, other authors such as Nolan (1983) have spread and also redefined them. The first criterion (*high between-speaker variability*) and the second one (*low within-speaker variability*) are probably the most important, or at least they have been the most repeated criteria in many publications thereafter. These two criteria could be reformulated as: "the parameter needs to exhibit a high degree of variation from one speaker to another" (Nolan, 1983: 1) and "it should be as consistent as possible for each speaker" (Wolf, 1972: 2044). It seems logical to think that a parameter which is very dependent on the genetic endowment of the speaker will fulfill these two criteria.

For the purpose of evaluating whether a voice parameter is more or less 'genetic', San Segundo (2014) suggested the hypothesis that higher similarity values would be found in the comparison of MZ twin pairs than in DZ twin pairs, in pairs of non-twin siblings (in this case, male siblings, i.e. brothers) or in a population of unrelated speakers. This hypothesis applied to the three different analyses carried out in that study (aimed at investigating not only glottal source features but also formant trajectories of vocalic sequences and cepstral features). Since the current study focuses on glottal source features, our hypothesis would be as follows: *Glottal parameters will be genetically related: higher similarity values will be found in MZ twins than in DZ twins. These, in turn, will obtain higher similarity values than brothers (B), who will obtain higher similarity values than unrelated speakers (US).* The expected decreasing scale of similarity values in these speakers would then be: MZ $>$ DZ $>$ B $>$ US[4] According to this, we can establish the five following hypotheses:

H1. Intra-speaker comparisons should yield large likelihood ratios (LRs).
H2. MZ intra-pair comparisons should yield also large LRs.
H3. DZ intra-pair comparisons should yield large LRs although not as large as H1 or H2.
H4. B intra-pair comparisons should yield LRs at least over the background baseline.
H5. US intra-pair comparisons should yield LRs aligned with the background baseline.

Taking into account that the results of the speaker comparisons will be shown in the form of log-likelihood ratios (LLRs), the decision thresholds (λ) for the hypotheses described above could be represented as:

H1.$\lambda >$ -1
H2.$\lambda >$ -1
H3.$\lambda >$ -10
H4.$\lambda >$ -10
H5.$\lambda <$ -10

For the execution of this study, we have recruited 54 male speakers, distributed in four different groups:

- Monozygotic twins (MZ), also called identical twins: 24 speakers.
- Dizygotic twins (DZ), also called non-identical or fraternal twins: 10 speakers.

- Full brothers (B), i.e. of the same mother and same father: 8 speakers.
- Unrelated speakers (US), who for the most part were pairs of friends or work colleagues: 12 speakers.

Friends or work colleagues — the fourth speaker group — served to create a reference population, whose relevance for Likelihood-Ratio-based forensic comparison has been acknowledged elsewhere (e.g. Morrison, 2010). In short, a reference population is aimed at considering *typicality* in addition to *similarity*[5]. The ages of all the speakers recruited for this study ranged between 18 and 52 years old (median age: 28.96). The age difference between the siblings in each pair varied between four and eleven years. The language variety spoken by all the subjects was North-Central Peninsular Spanish. Speakers were recorded on two different recording sessions (separated by 2-4 weeks) in order to account for intra-speaker variability.

Although the participating speakers were recorded carrying out five different speaking tasks — for a full description of the *ad hoc* collected corpus, see San Segundo (2013) and San Segundo (2014) — in the current study we have specifically extracted the speech material from the fifth speaking task: informal interview with the researcher. This task was carried out on the telephone in the following way: the researcher is at one end of the telephone and one member of each speaker pair at a time is at the other end of the telephone[6]. In this task, which lasts around 5-10 minutes, the researcher asks the speaker about any of the topics[7] that they had been discussing with their conversational partner — either his sibling or friend — in the first task (semi-structured spontaneous conversation). Since there is a considerably long time gap between the execution of the first and the fifth task, the speakers do not remember clearly the whole conversation and they exhibit hesitating responses, resulting in *pause fillers* (Cicres, 2007).

The complete speech material consisted of 853 tokens of the [eː] vowel (average tokens per speaker and session: 7.89) naturally sustained in pause fillers. For the selection of the sustained [eː] vowels we made an auditory and spectrographic examination in *Praat* (Boersma and Weenink, 2012) for every speaker and session's audio files recorded in the fifth task. We did not select those vowels where we perceived a marked creak realization, a high degree of nasalization, overlap with extraneous noise, laughter, etc. In average, the duration of the vowels was around 200 milliseconds. These phonetic units were manually located with *Praat* and the most stable part of them was marked and extracted, avoiding the beginning and the end of the vowel. These pause fillers or hesitation marks, which most people use — as the name suggests — when they hesitate in a conversation, while they are thinking of what they are going to say next, or when they are trying to remember something, were found very useful for our study, as they are longer than vowels in connected speech. Obtaining a relatively long vowel is highly important in order to estimate glottal parameters, which in the clinical tradition have been normally elicited upon asking the subject to sustain a long vowel for as long as possible. This technique — which is foreign to the forensic realm — could be replaced by the use of naturally sustained pause fillers.

Using the software BioMet®Soft (2010), a vector of 68 parameters was created from each vocalic segment. These parameters were estimated from the glottal source by inverse filtering (Gómez-Vilda *et al.*, 2009) and they can be distributed in the following seven subgroups: 1) f0 and distortion parameters; 2) cepstral coefficients of the glottal source power spectral density (PSD); 3) singularities of the glottal source PSD; 4) biome-

chanical estimates of vocal fold mass, tension and losses; 5) time-based glottal source coefficients; 6) glottal gap (closure) coefficients; and 7) tremor (cyclic) coefficients. For a detailed description of these parameters, see San Segundo (2014). BioMet®Soft (2010) was also used to carry out the speaker comparisons in the form of pairwise parameter matching experiments, yielding the results in LRs, as in Ariyaeeinia *et al.* (2008). The specific methodology is described in Gómez-Vilda *et al.* (2012).

The vector of glottal features will be referred as x_{sij}, where *s* refers to the speaker, *i* is for the session, and *j* for the vowel filler. The two voice samples under test — in each comparison — will be denoted by Z_a=$\{x_{aij}\}$ and Z_b=$\{x_{bij}\}$ for subjects *a* and *b*. Thus, we will be evaluating our two-hypotheses contrasts in terms of a logarithmic likelihood value:

$$\lambda_{ab} = log\left(\frac{p(Z_b|\Gamma_a)}{\sqrt{p(Z_a|\Gamma_R)p(Z_b|\Gamma_R)}}\right)$$

where we evaluate the conditional probability of each speaker relative to a Reference Speaker's Model Γ_R and calculate if the conditional probability between the two voice samples *a* and *b* is larger than the conditional probabilities relative to the reference model. The conditional probabilities have been evaluated using Gaussian Mixture Models (Γ_a, Γ_b, Γ_R) as:

$$p(Z_b|\Gamma_a) = \Gamma_a(Z_b)$$

$$p(Z_a|\Gamma_R) = \Gamma_R(Z_a)$$

$$p(Z_b|\Gamma_R) = \Gamma_R(Z_b)$$

The forensic-comparison evaluation framework used is a two-step process, which could be described as follows:

- Step 1. Model Generation: A model representative of the reference population (male subjects between 18-52 years old) was created using recordings $\mathbf{Z}_R$=$\{x_{Rjk}\}$ as a Gaussian Mixture Model Γ_R=$\{w_R, \mu_R, C_R\}$ where w_R, μ_R and C_R are the set of weights, averages and covariance matrices, respectively, associated to each Gaussian Probability Distribution in the set.
- Step 2. Score Evaluation: The material under evaluation is composed of different parameterized voice samples grouped in a matrix $\mathbf{Z}_a$=$\{\mathbf{x}_{aj}\}$ where $1 \leq j \leq J_a$ is the sample index, each sample being a vector $\mathbf{x}_{aj}$=$\{\mathbf{x}_{aj1} \ldots \mathbf{x}_{ajM}\}$ from vowel segments conveniently parameterized. Similarly, the set of the corresponding speaker to be matched will be given as $\mathbf{Z}_b$=$\{\mathbf{x}_{bj}\}$ where $1 \leq j \leq J_b$ will be the sample index, each sample being a vector $\mathbf{x}_{bj}$=$\{\mathbf{x}_{bj1} \ldots x_{bjM}\}$. The conditioned probability of a sample from speaker *a* x_{aj} matching speaker *b* will be estimated as

$$P(x_{bj}|\Gamma_a) = \frac{1}{(2\pi)^{M/2}|C_a|^Q} \cdot e^{-1/2(x_{bj} - \mu_a)^T C_s^{-1}(x_{bj} - \mu_a)}$$

Similarly the conditioned probability of a sample from speaker *a* matching the Reference Model will be:

$$P(x_{aj}|\Gamma_R) = \frac{1}{(2\pi)^{M/2}|C_R|^Q} \cdot e^{-1/2(x_{aj} - \mu_R)^T C_s^{-1}(x_{aj} - \mu_R)}$$

Finally, the conditioned probability of a sample from speaker *b* matching the Reference Model will be:

$$P(x_{bj}|\Gamma_R) = \frac{1}{(2\pi)^{M/2}|C_R|^Q} \cdot e^{-1/2(x_{bj} - \mu_R)^T C_s^{-1}(x_{bj} - \mu_R)}$$

Results

The results of the different comparison tests are shown in tables 1 to 4, where we have marked whether the LLR values of each comparison entail the confirmation or the refutation of the hypotheses described in the previous section.

MZ speakers			
Speakers compared	**Type of comparison**	**LLR**	**Hypothesis confirmation**
01v01	Intra-speaker	2.4	√
02v02	Intra-speaker	-0.5	√
01v02	Intra-pair	-0.0	√
03v03	Intra-speaker	-1.1	×
04v04	Intra-speaker	-8.3	×
03v04	Intra-pair	-1.0	√
05v05	Intra-speaker	12.5	√
06v06	Intra-speaker	6.1	√
05v06	Intra-pair	5.8	√
07v07	Intra-speaker	12.0	√
08v08	Intra-speaker	6.6	√
07v08	Intra-pair	12.1	√
09v09	Intra-speaker	-7.0	×
10v10	Intra-speaker	23.0	√
09v10	Intra-pair	12.6	√
11v11	Intra-speaker	4.3	√
12v12	Intra-speaker	14.1	√
11v12	Intra-pair	-14.6	×
33v33	Intra-speaker	-5.0	×
34v34	Intra-speaker	0.2	√
33v34	Intra-pair	0.6	√
35v35	Intra-speaker	-1.6	×
36v36	Intra-speaker	-0.2	√
35v36	Intra-pair	-1.5	×
37v37	Intra-speaker	-7.0	×
38v38	Intra-speaker	15.7	√
37v38	Intra-pair	9.9	√
39v39	Intra-speaker	3.1	√
40v40	Intra-speaker	4.9	√
39v40	Intra-pai	2.9	√
41v41	Intra-speaker	6.9	√
42v42	Intra-speaker	-4.1	×
41v42	Intra-pair	0.2	√
43v43	Intra-speaker	0-0	√
44v44	Intra-speaker	3.0	√
43v44	Intra-pair	-0.1	√

Table 1. Results for the MZ speakers
LLR means log-likelihood ratio.

DZ speakers			
Speakers compared	**Type of comparison**	**LLR**	**Hypothesis confirmation**
13v13	Intra-speaker	6.4	√
14v14	Intra-speaker	-0.7	√
13v14	Intra-pair	1.7	√
15v15	Intra-speaker	-8.7	×
16v16	Intra-speaker	5.2	×
15v16	Intra-pair	-3.2	√
17v17	Intra-speaker	1.6	√
18v18	Intra-speaker	4.3	√
17v18	Intra-pair	-10.1	√
19v19	Intra-speaker	0.6	√
20v20	Intra-speaker	-7.7	√
19v20	Intra-pair	-0.4	√
45v45	Intra-speaker	-1.0	×
46v46	Intra-speaker	0.0	√
45v46	Intra-pair	3.4	√

Table 2. Results for the DZ speakers
LLR means log-likelihood ratio.

Non-twin brothers (B)			
Speakers compared	**Type of comparison**	**LLR**	**Hypothesis confirmation**
21v21	Intra-speaker	6.4	√
22v22	Intra-speaker	-0.7	√
21v22	Intra-pair	1.7	√
23v23	Intra-speaker	-8.7	√
24v24	Intra-speaker	5.2	√
23v24	Intra-pair	-3.2	√
47v47	Intra-speaker	1.6	√
48v48	Intra-speaker	4.3	×
47v48	Intra-pair	-10.1	√
49v49	Intra-speaker	0.6	×
50v50	Intra-speaker	-7.7	×
49v50	Intra-pair	-0.4	√

Table 3. Results for the B speakers
LLR means log-likelihood ratio.

Unrelated Speakers (US)			
Speakers compared	**Type of comparison**	**LLR**	**Hypothesis confirmation**
25v25	Intra-speaker	-42.2	×
26v26	Intra-speaker	-0.7	√
25v26	Intra-pair	-11.2	√
27v27	Intra-speaker	10.2	√
28v28	Intra-speaker	11.9	√
27v28	Intra-pair	-9.7	×
29v29	Intra-speaker	-0.2	√
30v30	Intra-speaker	7.5	√
29v30	Intra-pair	-13.2	√
31v31	Intra-speaker	6.1	√
32v32	Intra-speaker	5.2	√
31v32	Intra-pair	-12.7	√
51v51	Intra-speaker	-4.9	×
52v52	Intra-speaker	4.9	√
51v52	Intra-pair	-10.4	√
53v53	Intra-speaker	8.1	√
54v54	Intra-speaker	5.7	√
53v54	Intra-pair	-12.1	√

Table 4. Results for the US speakers
LLR means log-likelihood ratio.

In relation to H1, we have computed all the cases of intra-speaker dissimilarity in the four tables, and we have found that five out of the total 54 participating speakers seem to be in the limit of the established threshold (subjects 03, 35, 48, 49 and 50) while eight speakers show strong intra-speaker dissimilarity (subjects 04, 09, 15, 20, 33, 37, 42 and 51), and only one shows very strong dissimilarity (subject 25). Therefore, 14 out of 54 do not fulfil H1. However, since five speakers out of 54 obtain values very close to the established threshold, we could speak of 9 out of 54 speakers not fulfilling the hypothesis of intra-speaker similarity.

Regarding H2, we find two out of 12 pairs not fulfilling it (MZ pairs 11-12 and 35-36). The third hypothesis is not fulfilled in one out of five pairs (DZ pair 17-18), while H4 — which refers to non-twin siblings — is fulfilled in all four cases. Finally, only one pair of unrelated speakers is slightly over the baseline (speakers 27-28) out of 5 cases fulfilling H5. Therefore, in view of the results, the degree of hypothesis corroboration could be summarized as:

H1: 40/54; a relaxed threshold would be 45/54 = 83.3%
H2: 10/12 = 83.3%
H3: 4/5 = 80%
H4: 4/4 = 100%
H5: 5/6 = 83%

We will present our comparison results by means of a Tippett plot, since this is a standard graphical method for representing the LR results of a forensic comparison

system as well as a method for the evaluation of a system performance. As recalled in San Segundo (2014: 106), "this type of representation was proposed by Evett and Buckleton (1996) in the field of DNA analysis and it owes its name to the work of Tippett *et al.* (1968), who first referred to the concepts of 'within-source comparison' and 'between-source comparison' (cf Drygajlo *et al.*, 2003)". In this type of graph, two types of curves are displayed, each one representing the probability for one of the competing hypothesis: *Hp* or *Hd*. Typically the hypothesis of the prosecution (*Hp*) is that the offender and the suspect samples come from the same speaker, while the hypothesis of the defense (*Hd*) is that they belong to different speakers. However, for the speaker types that we are testing (MZ, DZ, B or US), our Tippett plot needs to be based on a more specific *Hd*. In other words, the hypothesis of the defense is not simply that the voice samples belong to different speakers but — depending on the type of speakers compared at each time — that the voice samples belong to either (a) MZ twins, (b) DZ twins, (c) non-twin siblings, or (d) unrelated speakers. For that reason, figure 1 shows only one line rising to the right (the black line), representing the cumulative distribution of LLRs for all the intra-speaker comparisons — *targets* in Automatic Speaker Recognition (ASR) terminology — while there are four different lines rising to the left (red, magenta, cyan and blue), each one representing a different type of intra-pair comparison (a-d), depending on the type of kinship relationship between the speakers being compared. These cases of intra-pair comparisons are also inter-speaker comparisons *sensu stricto* and they would be named *non-targets* in ASR terminology.

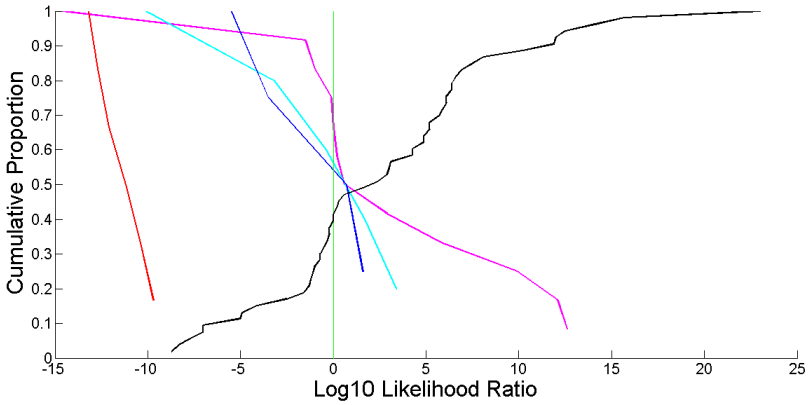

Figure 1. Tippett plot.
The black line represents the intra-speaker comparisons (for all the speaker types) and the following colours represent the intra-pair comparisons: red for US, magenta for MZ, cyan for DZ and blue for B.

As it can be seen in figure 1, the black line (intra-speaker comparisons) extends largely on the right, which implies a good performance of the system, but there are still some LLRs which support the contrary-to-fact hypothesis, represented in the black line from 0 to the left[8]. If we look at the intra-pair comparisons, different results are found:

- For the US (red line), the system performance is optimal, as there are only LLRs supporting the consistent-with-fact hypothesis. Note that all cases fall within the field to the left of 0. More specifically, the LLR values seem to be grouped

around -10, as could be also observed in table (cf. intra-pair comparisons). This indicates a *very strong support*[9] to the different-speaker hypothesis.

- In the case of MZ, DZ and B comparisons, the following trends are observed: The strongest support for the contrary-to-fact hypothesis occurs in MZ twins. Note that the magenta line stretching from 0 to the right is the longest. However, for the DZ (cyan) and B (blue) comparisons, the system shows a similar performance, with most cases falling within the consistent-with-fact hypothesis and only some cases supporting the contrary-to-fact hypothesis.

Discussion of the results

Our main hypothesis was that the glottal parameters analyzed would be genetically influenced, i.e. higher similarity values would be found in MZ twins than in DZ twins, non-twin brothers or in the reference population. Therefore, we predicted a decreasing scale of similarity values, expected to follow this order: MZ > DZ > B > US. According to this, we suggested five specific hypotheses, depending on whether the comparisons were intra-speaker comparisons (H1), or intra-pair comparisons of some of these types: MZ intra-pair comparison (H2), DZ intra-pair comparison (H3), B intra-pair comparison (H4) and US intra-pair comparison (H5). We further established some decision thresholds for each of these hypotheses in order to assess whether the LLR values obtained in the comparisons could be deemed large or small — and could consequently allow the rejection or the confirmation of the hypotheses.

In view of the results, the degree of hypotheses corroboration was very high: three of our hypotheses were corroborated in 83.3% of the cases (H1, H2 and H5), another one was corroborated in 80% of the cases (H3) and a further one was corroborated in 100% of the cases under study (H4). In the rest of this section we aim to discuss these results, distinguishing between the intra-speaker comparisons (which relate to H1) and the inter-speaker comparisons, referring to H2, H3, H4 and H5.

Intra-speaker results

There are no clear reasons why 14 out of 54 intra-speaker comparisons (or 9 out of 54 if we relax the threshold, as explained above) yield very low LLRs, indicating a strong dissimilarity of those speakers towards themselves. For the intra-speaker comparisons, the vowel fillers extracted from the first recording session are tested against the vowel fillers obtained in the second recording session. Some possible explanations for the cases of hypothesis rejection could then be associated with changes in phonation due to emotional stress or with the existence of temporary pathological conditions. Despite the fact that the speakers were only recorded when they exhibited a healthy condition — and the health troubles potentially affecting their voice had to be indicated by the speakers in a questionnaire —, it is still possible that they could have experienced temporary and minor voice maladies at one recording session but not at the other, this being behind the dissimilarity results of certain speakers.

Another possible explanation could be related with the speaker classification first proposed by Doddington *et al.* (1998). It is a truism in speaker recognition that not all speakers affect the performance of a forensic-comparison system in the same way, or in the words of the above-mentioned authors, there are "striking performance inhomogeneities among speakers within a population" (Doddington *et al.*, 1998: 1). The existence of these inhomogeneities allowed the authors to classify speakers in *sheep*,

lambs, *wolves* and *goats*; basically depending on whether they are more or less difficult to recognize by the system. In this sense, the percentage of speakers in our study who obtained low LLRs could probably be considered 'goats' in Doddington's Zoo, as these type of speakers "tend to adversely affect the performance of systems by accounting for a disproportionate share of the missed detections" (Doddington *et al.*, 1998: 1). Yet, independently of the fact that the existence of 'goats' is acknowledged since long in ASR, the question of what makes a speaker so different from himself is a key issue in Forensic Phonetics and it remains largely unexplored.

Finally, a distinction should be made between the average LLR values of the discordant cases of intra-speaker comparisons (around -8, -7 or lower) and a single case with a striking LLR value of -42.2 (speaker 25). As the study by San Segundo (2014) explained, this was a clear exception which deserved detailed analysis. Indeed, upon examination of the anamnesis of this speaker, it became apparent that he suffered from hypothyroidism. We suggested in the above-mentioned study that this hormonal problem could be the cause of the strikingly large intra-speaker variation found for this speaker. Often called underactive thyroid hormone, one of its symptoms is hoarse voice, according to Longo and Fauci (2011). This type of phonation, especially if it appears intermittently in the speaker's vocal output, could explain the strong dissimilarity in the comparison of the first and the second recording sessions of speaker 25. Nevertheless, more research would be necessary to investigate how this disease specifically affects voice.

Intra-pair (inter-speaker) results

Having focused on H1 in the previous subsection, we have to consider now separately H2, H3, H4 and H5. As far as H2 is concerned, only 2 out of 12 MZ pairs did not obtain LLRs above -1, as our hypothesis established. While one case is that of MZ pair 11-12 (LLR = -14.6), with a strong deviation from the established threshold, the other case is that of MZ pair 35-36 (LLR = -1.5), i.e. certainly close to the threshold. It seems evident that their cases are not comparable and that the most interesting pair to examine in detail is the first one. The most plausible reason for their striking differences — despite being identical twins — is twofold. On the one hand, the existence of smoking habits in one of them made his f0 much lower than that of his cotwin, and this could affect the rest of the glottal parameters analyzed in this study. On the other hand, the questionnaire that the speakers had to fill at the time of the recordings included some questions about their attitude towards being twins. In view of the answers given by this specific pair, it was made clear that they were not especially close to each other, which could have made them separate in personality and possibly also phonetically [10]. In other words, "the learned speech habits aimed at attaining divergence patterns may have outweighed their anatomical similarities" (San Segundo, 2014: 188).

As far as H3 is concerned, only in one DZ pair out of five the hypothesis was not corroborated. It is the case of DZ pair 17-18, who obtained a LLR = -10.1. In our hypothesis formulation, we considered that DZ twins should show large LLRs but not as large as MZ twins, being the decision threshold λ = -10. The only exception found is therefore almost irrelevant. In all the other cases, the LLR values were as expected: relatively large but not that large as those found for MZ twins, on average. For that reason, the third hypothesis is well corroborated.

If we consider now H4, all the non-twin brothers corroborate our hypothesis: LLR values above -10 are obtained in 100% of the pairs analyzed. Since full siblings (i.e. broth-

ers) and DZ twins both share the same genetic load, H3 and H4 were established at the same level: λ = -10. Finally, H5 established that US would obtain LLRs aligned with a background baseline fixed at $\lambda <$ -10. This is fulfilled in almost all the cases, being the only exception that found in speakers 27-28 (LLR = -9.7). While this value implies a rejection of the hypothesis if we strictly apply our decision threshold, it seems clear that the difference between -9.7 and -10 is almost irrelevant, especially when we are expressing the results in logarithmic figures. The degree of H5 corroboration is then very satisfactory, and this is particularly relevant, as it indicates that in a typical forensic scenario — when unrelated speakers are compared — our glottal source based system performs very well, with none of the speakers being misidentified (false alarms). Besides, with these results more evidence is gained in favor of our main hypothesis that glottal parameters are genetically influenced, as none of the unrelated speakers show any similarity, in comparison with the somehow genetically related DZ and B — with larger LLR values — and with the much genetically related MZ, with still larger LLR values.

Conclusions and directions for future research

We can conclude that the glottal parameters analyzed, considered as a whole set of 68 features, are genetically influenced. With few exceptions, the system performance for DZ and full siblings is similar ($\lambda >$ -10) while MZ twins obtain larger LLRs and the values of US gather homogenously around the baseline ($\lambda <$ -10). This is in agreement with our hypotheses, as we predicted that the LLR values of the forensic comparison would be distributed in a line going from the largest positive LLRs for the MZ twins, at one end of the line, and the largest negative LLRs for the US, at the other end of the line. The former share 100% of their genes while the latter share 0%. In between, there are the DZ twins and the B, sharing on average 50% of their genetic information. Furthermore, our results are in agreement with previous studies about twins, such as Loakes (2006), insofar as different results have been found for different twin pairs, indicating a lack of homogeneity in this speaker group. The idiosyncrasies in the relationship of each pair could be only studied on a case-by-case basis to find the causes for speech convergence or divergence, which probably indicates that the weight of external factors, such as psychological aspects, or educational and environmental influences (i.e. 'nurture') is more important than it could be a priori thought in this type of voice studies, or at least as important as 'nature' in many speaker comparisons.

All in all, this study has tried to show the relevance of applying the twin methodology to forensic voice investigations in order to find whether a parameter — or a set of parameters, as in this case — could be robust and hence useful for speaker comparison. It becomes also apparent that the study of glottal source features deserves an important position among the many possible phonetic parameters that can be considered in forensic casework, both for their easy extraction in natural speech and for their good discrimination results in several studies so far, not to mention that they are obtained from inverse filtering of the vocal tract. This makes them independent from traditional vocal-tract features, which opens great possibilities for their combination with such parameters; this fusion/combination being one of the advantages of the LR approaches.

Besides studying the degree of similarity in MZ, DZ, B and US intra-pair comparisons, we have also taken advantage to study intra-speaker variation in all the four types of speakers participating in this study. We have found that in more cases than desirable in a forensic context, the system performance is not completely good when two speaker

sessions are tested against each other. These *missed hits* represent a 16.6% of the intra-speaker comparisons, or targets, regardless of the fact that the speaker is MZ, DZ, B or US. There was an especially striking case of LLR = -42.2. While this is a clear exception, it was in-depth analyzed and a possible explanation for this large intra-speaker variation could be found in a hormonal disease suffered by this speaker. Yet the other cases of missed hits still represent large figures whose cause would deserve further research. Likewise, future studies could consider the study of the 68 glottal parameters independently, to test if some of the seven feature subgroups outperform the others for forensic purposes.

Acknowledgments

This research has been possible thanks to a grant awarded by the Spanish Ministry of Education (*Beca FPU – Programa Nacional de Formación de Profesorado Universitario*; reference AP2008-01524) and also thanks to a grant awarded by the IAFPA (International Association for Forensic Phonetics and Acoustics).

Notes

[1]Some terminology controversies have arisen in recent times in relation to the proper name that this specific application of Forensic Phonetics should receive. It seems that the term comparison is widespread nowadays, at least in the linguistic-phonetic realm. The Position Statement in French and Harrison (2007), signed by nine researchers and with several more co-signatories, accepts the replacement of identification by comparison: "It will be apparent from the arguments developed here that the term FSI should be replaced by FSC" (French and Harrison, 2007: 144). A summary of this controversy can be read in San Segundo (2014). For more details, see Coulthard and Johnson (2007); French and Harrison (2007); Morrison (2009); Rose and Morrison (2009) and French *et al.* (2010).

[2]Note however that *speaker profiling* does not necessarily involve LADO. The former simply consists in determining the phonetic profile of an unknown speaker on the basis of his voice and speech patterns; i.e. trying to derive as much information as possible about the speaker age, gender or dialect, among other characteristics.

[3]Most probably because providing a definition of 'voice quality' is clearly not the purpose of the investigation by Gold and French (2011). In fact, this is not an easy concept to define. In Gil and San Segundo (2014) we track the description of 'voice quality' in six of the most relevant works about Forensic Phonetics — including references to voice quality — to this date (Gil and San Segundo, 2014: 176-183). Namely, the reviewed works were, in chronological order: Nolan (1983), Hollien (1990), Künzel (1994), French (1994), Rose (2002) and Jessen (2008). An examination of those works allows the reader to see how far speech scientists are from arriving at a definition consensus. No wonder Hollien (1990) points to the occasional view of the label 'voice quality' as a 'wastebasket' used for those voice aspects that other categories fail to describe.

[4]If we strictly apply what we know about the genetic endowment of DZ and B (as explained above: same amount of shared genes per sibling pair, that is 50%), it could be thought that it would have been more coherent to establish this decreasing scale MZ > DZ ≥ B > US. Yet, two aspects should be taken into account: a) Although it is widely accepted that both DZ pairs and non-twin sibling pairs "share 50% of their genes, on average, by descent" (Pakstis *et al.*, 1972 in Segal, 1990: 612), a more realistic range seems to be 25% - 75% while this theoretical range can actually vary between 0% to 100% (Pakstis *et al.*, 1972 in Segal, 1990: 612). Therefore it should be highlighted that the 50% value is — to some degree – a convention; it can vary from one pair to another. b) The newly-developed scientific field of epigenetics (the study of the changes in gene expression caused by mechanisms other than changes in the underlying DNA sequence) has shown us that environmental factors do affect genes in ways that still need to be fully explored. As environmental and genetic aspects cannot be completely disentangled, we consider that DZ twins could be –although maybe only slightly — more genetically related than non-twin siblings because the former usually share more environmental experiences than the latter due to the fact that they are born on the same day whereas in the case of non-twin siblings their age gap makes them more susceptible for

environmental divergence. There are several arguments supporting that DZ cotwins are genetically more similar to one another than non-twin siblings. The interested reader is encouraged to read — for instance — Stromswold (2006), where she raises the case of transplant surgery, a field where "it has been known for decades that the incidence of graft rejection is lower between DZ cotwins than between non-twin full siblings, and this clinical observation has been used to argue that DZ cotwins are genetically more similar to one another than non-twin full siblings (see Geschwind, 1983)" (Stromswold, 2006: 338–9). As all these genetic aspects are not free of controversy, it seems prudent for us to maintain the hypothesized decreasing scale MZ > DZ > B > US while — at the time of fixing the thresholds for the corresponding H3 and H4 — establishing lambda at the value -10 in both cases. This is not in contradiction with the explanation of each hypothesis: (H3) *DZ intra-pair comparisons should yield large LRs although not as large as H1 or H2*, and (H4) *B intra-pair comparisons should yield LRs at least over the background baseline.*

[5]The LR formula has a numerator and a denominator. As explained in Morrison (2010: 17), "the numerator of the LR can be considered a *similarity* term, and the denominator a *typicality* term. In calculating the strength of evidence, the forensic scientist must consider not only the degree of similarity between the samples, but also their degree of typicality with respect to the relevant population. In fictional television shows, forensic scientists are often portrayed comparing two objects, finding no measurable differences between them, and shouting: 'It's a match!' Similarity alone, however, does not lead to strong support for the same-origin hypothesis".

[6]This does not mean that the speech material available for comparison had been telephone-filtered. The corpus used contains some speaking tasks which have undergone a filtering through real telephone transmission, but on this occasion we used studio-quality recordings. The recording set-up was such that the speakers were at different rooms and held a real telephone conversation but they were being recorded with high quality microphones (*Countryman E6i Earset* microphone). The recordings took place in the *Centro de Ciencias Humanas y Sociales* at *CSIC (Consejo Superior de Investigaciones Científicas)* in Madrid, Spain.

[7]In the first speaking task, the speakers were suggested some topics, including those described in Loakes (2006). For especially sparing speakers, other possible topics were raised. In order to minimize the "observer's paradox" (Labov, 1972), we followed the indications in Moreno (2011), particularly with regard to the use of "icebreakers" as conversational starting points.

[8]Note that Speaker 25 (only the value for his intra-speaker comparison, i.e. LLR = -42) was excluded from representation in the black line of figure because that LLR value was considered an outlier, i.e. being exceptionally low for the reasons which will be more thoroughly discussed in the section devoted to the discussion of the results.

[9]Note that a LLR of -10 (LR = -10, 000, 000, 000) means that it is 10, 000, 000, 000 times more likely that the observed differences between the speech samples of suspect and offender occur under the hypothesis that they come from different speakers than under the hypothesis that they come from the same speaker. According to the verbal equivalents for LRs proposed by Evett (1998), LRs larger than 1000 indicate a *very strong support* for the respective hypothesis (in this case, the hypothesis of the defense, as it is a negative logarithmic value). Nevertheless, it should be noted that not all scientists agree in using such verbal scales (for a summary of this controversy, see San Segundo, 2014, *cf. Introduction).*

[10]In the questionnaire, they rated their relationship closeness as "not especially close" and answered that they have liked to be independent and different since they were children. This compares with the most common situation for the rest of MZ twins participating in this study, who — on average — rated their relationship closeness as "very close" and stated that they like to be together and share leisure activities, group of friends, etc.

References

Ariyaeeinia, A., Morrison, C., Malegaonkar, A. and Black, S. (2008). A test of the effectiveness of speaker verification for differentiating between identical twins. *Science & Justice*, 48(4), 182–186.

BioMet®Soft, (2010). Biomet®soft. universidad politécnica de madrid. retrieved from httpÂů//www.biometrosoft.com.

Boersma, P. and Weenink, D. (2012). Praat: doing phonetics by computer (version 5.3.79).

Cambier-Langeveld, T. (2007). Current methods in forensic speaker identification: Results of a collaborative exercise. *International Journal of Speech, Language and the Law*, 14(2), 223–243.

Cicres, J. (2007). Análisis discriminante de un conjunto de parámetros fonético-acústicos de las pausas llenas para identificar hablantes. *Síntesis Tecnológica*, 3(2), 87–98.

Coulthard, M. and Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. New York: Routledge.

Debruyne, F., Decoster, W., Van Gijsel, A. and Vercammen, J. (2002). Speaking fundamental frequency in monozygotic and dizygotic twins. *Journal of Voice*, 16(4), 466–471.

Doddington, G., Liggett, W., Martin, A., Przybocki, M. and Reynolds, D. (1998). Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. In *Proceedings of the 5th International Conference on Spoken Language Processing*, 1–5.

Drygajlo, A., Meuwly, D. and Alexander, A. (2003). Statistical methods and bayesian interpretation of evidence in forensic automatic speaker recognition. In *Proceedings of the 8th European Conference on Speech Communication and Technology*, 689–692.

Evett, I. (1998). Towards a uniform framework for reporting opinions in forensic science casework. *Science & Justice*, 38(3), 198–202.

Evett, I. and Buckleton, J. (1996). Statistical analysis of str data. In A. Carraredo, B. Brinkmann and E. Bär, Eds., *Advances in Forensic Haemogenetics*. Heidelberg: Springer-Verlag.

French, P. (1994). An overview of forensic phonetics with particular reference to speaker identification. *International Journal of Speech, Language and the Law*, 1(2), 169–181.

French, P. and Harrison, P. (2007). Position Statement concerning use of impressionistic likelihood terms in forensic speaker comparison cases, with a foreword by Peter French & Philip Harrison. *The International Journal of Speech, Language and the Law*, 14(1), 137–144.

French, P., Nolan, F., Foulkes, P., Harrison, P. and McDougall, K. (2010). The UK position statement on forensic speaker comparison; a rejoinder to Rose and Morrison. *The International Journal of Speech, Language and the Law*, 17(1), 143–152.

French, P. and Stevens, L. (2013). Forensic speech science. In M. Jones and R. Knight, Eds., *The Bloomsbury Companion to Phonetics*. London: Bloomsbury.

Geschwind, N. (1983). Genetics: fate, chance, and environmental control. In C. L. . J. Cooper, Ed., *Genetics aspects of speech and language disorders*, 21–33. New York: Academic Press, 1 ed.

Gil, J. and San Segundo, E. (2014). La cualidad de voz en fonética judicial. In E. Garayzábal and M. J. M. Reigosa, Eds., *Lingüística Forense: la lingüística en el ámbito legal y policial*. Madrid: Euphonia Ediciones.

Gold, E. and French, P. (2011). An international investigation of forensic speaker comparison practices. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 1254–1257, Hong Kong, China.

Gómez-Vilda, P., Fernández-Baillo, R., Nieto, A., Díaz, F., Fernández-Camacho, F. J., Rodellar, V., Alvarez, A. and Martínez, R. (2007). Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters. *Journal of Voice*, 21(4), 450–476.

Gómez-Vilda, P., Fernández-Baillo, R., Rodellar-Biarge, M. V., Nieto-Lluis, V., Álvarez Marquina, A., Mazaira-Fernández, L. M. and Godino-Llorente, J. I. (2009). Glottal

source biometrical signature for voice pathology detection. *Speech Communication*, 51(9), 759–781.

Gómez-Vilda, P., Mazaira-Fernández, L. M., Martínez-Olalla, R., Álvarez Marquina, A., Hierro, J. and Nieto, R. (2012). Distance Metric in Forensic Voice Evidence Evaluation using Dysphonia-relevant Features. In *VI Jornadas de Reconocimiento Biométrico de Personas (JRBP)*, Las Palmas de Gran Canaria.

Gómez-Vilda, P., Álvarez Marquin, A., Mazaira-Fernández, L. M., Fernández-Baillo, R., Nieto-Lluis, V., Martínez-Olalla, R. and Rodellar-Biarge, M. V. (2008). Decoupling vocal tract from glottal source estimates in speaker's identification. *Language Design*, Special Issue, 111–118.

Gómez-Vilda, P., Álvarez Marquina, A., Mazaira-Fernández, L. M., Fernández-Baillo, R., Rodellar-Biarge, M. V. and Nieto-Lluis, V. (2010). Glottal biometric features: Are pathological voice studies appliable to voice biometry? In *I Workshop de Tecnologías Multibiométricas para la Identificación de Personas*, Las Palmas de Gran Canaria.

Hazen, K. (2002). The family. In J. Chambers, P. Trudgill and N. Schilling-Estes, Eds., *The Handbook of Language Variation and Change*. Malden, MA: Blackwell.

Hollien, H. (1990). *The acoustics of crime*. New York: Plenum Press.

Jessen, M. (1997). Speaker-specific information in voice quality parameters. *The International Journal of Speech, Language and the Law*, 4(1), 84–103.

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, 2(4), 671–711.

Künzel, H. (1994). Current approaches to forensic speaker recognition. In *Proceedings of the ESCA Workshop on Automatic Speaker Recognition, Identification and Verification*, 135–141.

Künzel, H. (2010). Automatic speaker recognition of identical twins. *The International Journal of Speech, Language and the Law*, 17(2), 251–277.

Künzel, H. and Köster, J. (1992). Measuring vocal jitter in forensic speaker recognition. In *Proceedings of the 44th Annual Meeting, American Academy of Forensic Sciences*, 113–114.

Labov, W. (1972). The transformation of experience in narrative syntax. In W. Labov, Ed., *Language in the Inner City*. Philadelphia: University of Philadelphia Press.

Laver, J., Wirz, S., Mackenzie, J. and Hiller, S. M. (1981). A perceptual protocol for the analysis of vocal profiles. *Edinburgh University Department of Linguistics Work in Progress*, 14, 139–155.

Loakes, D. (2006). *A forensic phonetic investigation into the speech patterns of identical and non-identical twins*. Doctoral dissertation, University of Melbourne.

Longo, D. and Fauci, A. (2011). Disorders of the thyroid gland. In D. Longo, A. Fauci, D. Kasper, S. Hauser, J. Jameson and S. Loscalzo, Eds., *Harrison's Principles of Internal Medicine*. New York: McGraw-Hill.

Mora, M. (2013). La policía francesa detiene a dos gemelos para aclarar una ola de ataques sexuales. *El País*, Retrieved from http://internacional.elpais.com/internacional/2013/02/10/actualidad/1360530132_840599.html, 10 February.

Moreno, F. (2011). La entrevista sociolingüística: esquemas de perspectivas. *Linred: lingüística en la Red*, 9, 1–16.

Morrison, G. (2009). Comments on Coulthard & Johnson's (2007) portrayal of the likelihood-ratio framework. *Australian Journal of Forensic Sciences*, 41(2), 155–161.

Morrison, G. (2010). Forensic voice comparison. In I. Freckelton and H. Selby, Eds., *Expert Evidence*. Sydney: Thomson Reuters.

Nolan, F. (1983). *The phonetic bases of speaker recognition.* Cambridge: Cambridge University Press.

Pakstis, A., Scarr-Salapatek, S., Elston, R. and Siervogel, R. (1972). Genetics contributions to morphological and behavioral similarities among sibs and dizygotic twins: Linkages and allelic differences. *Social Biology*, 19, 185–192.

Rose, P. (2002). *Forensic Speaker Identification.* London: Taylor & Francis.

Rose, P. (2006). Technical forensic speaker recognition: Evaluation, types and testing of evidence. *Computer Speech & Language*, 20(2), 159–191.

Rose, P. and Morrison, G. S. (2009). A response to the UK position statement on forensic speaker comparison. *The International Journal of Speech, Language and the Law*, 16(1), 139–163.

San Segundo, E. (2013). A phonetic corpus of Spanish male twins and siblings: Corpus design and forensic application. *Procedia-Social and Behavioral Sciences*, 95, 59–67.

San Segundo, E. (2014). *Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source parameters and cepstral characteristics.* Doctoral dissertation, CSIC/UIMP.

San Segundo, E. and Gómez-Vilda, P. (2013). Voice biometrical match of twin and non-twin siblings. In *Proceedings of the 8th International Workshop Models and analysis of vocal emissions for biomedical applications*, 253–256, Firenze, Italy.

San Segundo, E. and Gómez-Vilda, P. (2014). Forensic voice comparison using glottal parameters in twins and non-twin siblings. In *The 23rd Conference of the International Association for Forensic Phonetics and Acoustics*, Zürich, Switzerland.

Segal, N. (1990). The importance of twin studies for individual differences research. *Journal of Counseling & Development*, 68(6), 612–622.

Stevens, L. and French, P. (2012). Voice quality in Standard Southern British English: distribution of features, inter-speaker variability and the effect of telephone transmission. In *The 21st Conference of the International Association for Forensic Phonetics and Acoustics*, Santander, Spain.

Stromswold, K. (2006). Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition*, 101(2), 333–384.

Tippett, C., Emerson, V., Fereday, M., Lawton, F., Richardson, A., Jones, L. and Lampert, M. (1968). The evidential value of the comparison of paint flakes from sources other than vehicles. *Journal of the Forensic Science Society*, 8(2), 61–65.

Tomblin, J. and Buckwalter, P. (1998). Heritability of poor language achievement among twins. *Journal of Speech, Language, and Hearing Research*, 41(1), 188.

Wolf, J. (1972). Efficient acoustic parameters for speaker recognition. *The Journal of the Acoustical Society of America*, 51(6B), 2044–2056.

Zheng, N. (2005). *Speaker Recognition Using Complementary Information from Vocal Source and Vocal Tract.* Doctoral dissertation, The Chinese University of Hong Kong.

Using Dysphonic Voice to Characterize Speaker's Biometry

Pedro Gómez, Eugenia San Segundo, Luis M. Mazaira, Agustín Álvarez & Victoria Rodellar

Center for Biomedical Technology, Universidad Politécnica de Madrid & Universidad Internacional Menéndez Pelayo (UIMP), Madrid, Spain

Abstract. *Phonation distortion leaves relevant marks in a speaker's biometric profile. Dysphonic voice production may be used for biometrical speaker characterization. In the present paper phonation features derived from the glottal source (GS) parameterization, after vocal tract inversion, is proposed for dysphonic voice characterization in Speaker Verification tasks. The glottal source derived parameters are matched in a forensic evaluation framework defining a distance-based metric specification. The phonation segments used in the study are derived from fillers, long vowels, and other phonation segments produced in spontaneous telephone conversations. Phonated segments from a telephonic database of 100 male Spanish native speakers are combined in a 10-fold cross-validation task to produce the set of quality measurements outlined in the paper. Shimmer, mucosal wave correlate, vocal fold cover biomechanical parameter unbalance and a subset of the GS cepstral profile produce accuracy rates as high as 99.57 for a wide threshold interval (62.08-75.04%). An Equal Error Rate of 0.64 % can be granted. The proposed metric framework is shown to behave more fairly than classical likelihood ratios in supporting the hypothesis of the defense vs that of the prosecution, thus offering a more reliable evaluation scoring. Possible applications are Speaker Verification and Dysphonic Voice Grading.*

Keywords: *Phonation, Speaker Recognition, Voice Production, Speech Processing.*

Resumo. *A distorção de fonação deixa marcas relevantes no perfil biométrico de um falante. A produção de voz disfônica pode ser usada como caracterização biométrica. Neste artigo, propõe-se a utilização de aspectos de fonação derivados da parametrização da fonte glótica (FG), após a inversão do trato vocal, para caracterização de voz disfônica em tarefas de verificação de locutor. Os parâmetros derivados da fonte glótical são combinados em um sistema de avaliação forense para definir uma especificação métrica baseada em distância. Os segmentos de fonação utilizados no estudo são derivados de elementos de preenchimento, vogais longas e outros segmentos de fonação produzidos em conversas telefônicas espontâneas. Segmentos de fonação de um banco de dados telefônicos de 100 falantes nativos espanhóis do sexo masculino são combinados em uma tarefa de*

validação cruzada por 10 vezes para produzir o conjunto de medições de qualidade descrito neste artigo. Shimmer, correlato de onda mucosa, desequilíbrio de parâmetro biomecânico de cobertura da prega vocal e um subconjunto dos perfis de cepstrais de FG produzem taxas de precisão de até 99,57 para um largo intervalo (62,08-75,04%). Uma Taxa de Erros Iguais de 0,64% pode ser concedida. Demonstra-se que a estrutura métrica proposta comporta-se de forma mais justa do que a clássica razão de verossimilhança para apoiar a hipótese da defesa vs a do promotor, oferecendo assim um escore de avaliação mais confiável. As aplicações possíveis são Verificação de Locutor e Graduação de Voz Disfônica.

Palavras-chave: *Fonação, Reconhecimento de Falante, Produção de Voz, Processamento de Fala.*

Introduction

Voice Pathology has been profoundly studied and characterized in the past decade (Dejonckere, 2010; Hakkesteegt *et al.*, 2010; Roy *et al.*, 2013). Most of the advances produced in the detection and grading of pathology can be applied in other fields such as forensic speaker recognition. In this article phonation features derived from the parameterization of the glottal source after the vocal tract inversion is proposed for dysphonic voice characterization in speaker verification tasks (Gomez-Vilda *et al.*, 2012), where the glottal source can be seen as a correlate of pressure build up in the glottis.

Phonation is the activity of voice production as a consequence of vocal fold vibration. It is present in speech, in voiced sounds, although speech is composed of both voiced and voiceless sounds, and the latter sounds are not based on phonation. Phonation must be seen as a biometrical mark of the person, similar to other behavior-based activities, such as gait, or writing. It presents several advantages with respect to speech as a study signal, in the sense that the vocal tract transfer function in speech is interfering with phonation biometry by introducing articulation features, which increment intra-speaker variability.

Phonation may be classified into the following overlapping groups:

- Normophonic, which is defined by the presence of a stable fundamental frequency in sustained vowels, stable intensity and long phonation capability, absence of roughness, absence of breathiness, and effortless voice production. Besides, it is characterized by clear and precise open and closed phases of the vocal folds, large Maximum Flow Declination Rate, and good extension of harmonic spectrum, extending over 5 KHz. The instrumental exploration of the larynx must not reveal organic or anatomical defects or lesions.
- Dysphonic, non organic, which is defined by the presence of perceptual acoustical features related to unstable or asymmetric phonation, such as presence of roughness, air in voice or strain, showing an irregular or too short vocal fold closed phase. The extension of the harmonic spectrum may not reach 4 KHz. Nevertheless the instrumental exploration of the larynx does not reveal organic defects or lesions, although anatomical defects may be present, as a certain degree of asymmetry.
- Pathologic, organic, which is defined by perceptual phonation defects affecting stability of fundamental frequency and intensity, shorter phonation capability, and roughness, air in voice, weak voice, and affected short harmonic spectrum,

usually not extending over 2 kHz. Instrumental exploration of the larynx will reveal specific defects or lesions, as nodules, polyps, cysts, edemae, granulomae, sulci, carcinomae, etc.

- Pathologic, neurological, which is defined by perceptual phonation defects as in the organic case, but in this group the instrumental inspection of the larynx will not reveal specific organic defects or lesions, although vocal folds will not show a regular vibration pattern, and many times vocal fold vibration asymmetry is present, affecting one of the vocal folds (unilateral paresis), or both vocal folds. Other forms of irregularity may affect the stability of phonation (spasmodic dysphonia). Frequently the etiology of the irregularity remains unclear.

The burning question is to what extent dysphonic voice may be present in a given speaker. In other words, to what extent normophonic voice is the norm in a sample of a general population. This extent is difficult to assess, and depends on how strict the specification for the term *normophonic* is established. Besides, the phonation capability of a speaker will vary strongly during a lifetime, progressively degrading with age to become a *presbyphonic* voice during the third age in most of the population, characterized by an increment in roughness, breathiness and asthenia, depicting a creaky phonation condition. It must be taken into account that many people suffer from a higher degree of phonation deterioration due to specific habits such as smoking, drinking or drug abuse, or to the consequence of larynx inflammatory processes (flu, cough, and other respiratory diseases), or simply from voice abuse (contact center professionals, actors, speakers, dealers, etc.). Thus, it may be said that phonation conditions are better during youth, and start to degrade with age. Therefore, it is really hard to establish the population percentage corresponding to each group.

It is very important to determine the characteristics of normophonic voice production, since even in that case, small irregularities may be expected in the main features mentioned, as stability in frequency and intensity, regular and symmetric fold vibration, perfect and complete open and closed phases, and timbre spectrum, making phonation a specific personal print. Even under perfect phonation conditions population differences exist, opening the possibility to use phonation features as biometrical marks.

The main phonation features resulting from biometrical differences are due to very specific physiological causes, and can be grouped into these two classes (Gómez *et al.*, 2013):

- Vocal fold vibration asymmetry
- Deficient glottal closure during the closed phase (contact phase)

The physiological reasons conditioning phonation features are summarized in Figure 1.

The template in Figure 1.a shows the vocal folds as two vertical bands united in the anterior side of the cricoid process (upper part of each sketch), separated in the posterior side (lower part of each sketch), leaving a space for the free flow of air to and from lungs. In Figure 1.b the vocal folds are shown together closing the glottis (contact phase), due to the action of the transversal and oblique laryngeal and crico-arytenoid muscles. The flow of air is stopped. In Figure 1.c the vocal folds are still united in the posterior part of the glottis under the action of the laryngeal muscles, but the pressure built up in the lungs has taken them apart (abduction), leaving a glottal space through with air can

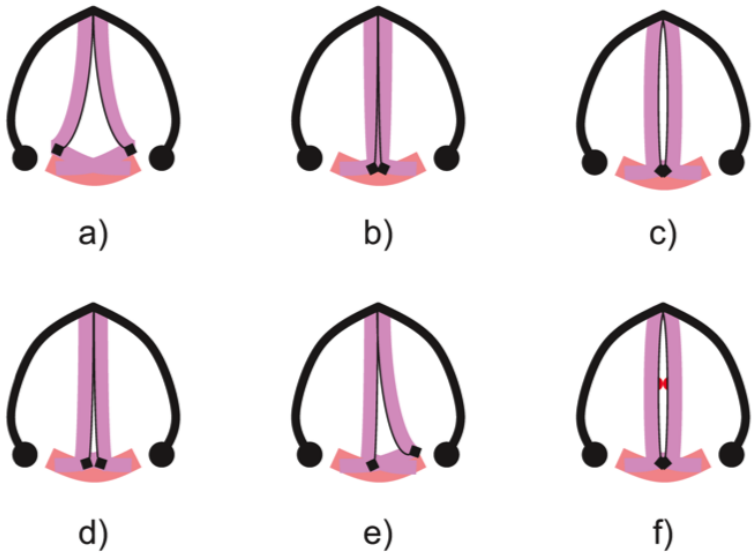

Figure 1. Vocal fold simplified situations: a) Open glottis in breathing; b) closed phase (contact phase) as part of the phonation cycle; c) open phase as part of the phonation cycle; d) deficient closure in the posterior third of the glottis, showing a permanent gap; e) asymmetric contact defect; f) deficient closure in the medial third, due to a bilateral lesion (nodules). Contact defects during the contact phase may be produced by other lesions (unilateral or bilateral). In all the plots the anterior part of the glottis is depicted upwards. (Figures produced by authors.)

flow from lungs to pharynx (open phase). The situations described in a), b) and c) are considered normal in the behavior of a healthy larynx. In the lower row some defects are described related with the contact phase. For instance, in Figure 1.d both vocal folds are not completely closed at the posterior side, therefore an air escape is to be expected. In Figure 1.e the incomplete closure is due to an asymmetry affecting mainly one of the vocal folds (unilateral paralysis). In Figure 1.f the contact is compromised by a bilateral lesion in the contact surface of the vocal folds, as in the case of nodules, for instance. The closure is not perfect and an escape of air is to be expected. Pictures of these contact defects from actual endoscopic recordings taken during the contact phase are presented in Figure 2.

The situations described in d), e) and f) produce observable correlates in the air flow and pressure build up in larynx, and propagate to the signal recorded by a microphone as phonated speech. Therefore, the contact defects will leave a biometrical mark in the phonation of a speaker if any of these defects is present to a greater or lesser extent. The behavior of the biometrical mark may be inferred from Fant's source-filter model illustrated in Figure 3 (Fant, 1997).

Voiced speech (phonation) is produced by a glottal excitation model, resulting from vocal fold vibration. The pressure build up in the vocal folds (glottal source) propagates through the vocal tract (or more properly, the oro-naso-pharyngeal tract) to reach the mouth or nostrils (depending on nasalization) to be radiated as a signal $S_r(n)$ reaching a microphone or other recording device. Voiceless speech is produced by frictional air turbulence (turbulent source) resulting from fast airflow in specific parts of the vocal tract (vocal folds, pharynx, tongue, teeth, lips…). Either glottal source, or turbulent

Figure 2. Pictures illustrating contact defects: Left picture: deficient closure in the posterior third of the glottis as a result of bilateral nodules. Middle picture: Unilateral contact defect due to a right vocal fold Reinke's edema. Right picture: bilateral contact defect in an hourglass pattern showing anterior and posterior gaps. Anterior section of larynx upwards (Photos provided by the ENT Services of Hospital Universitario Gregorio Marañón of Madrid.)

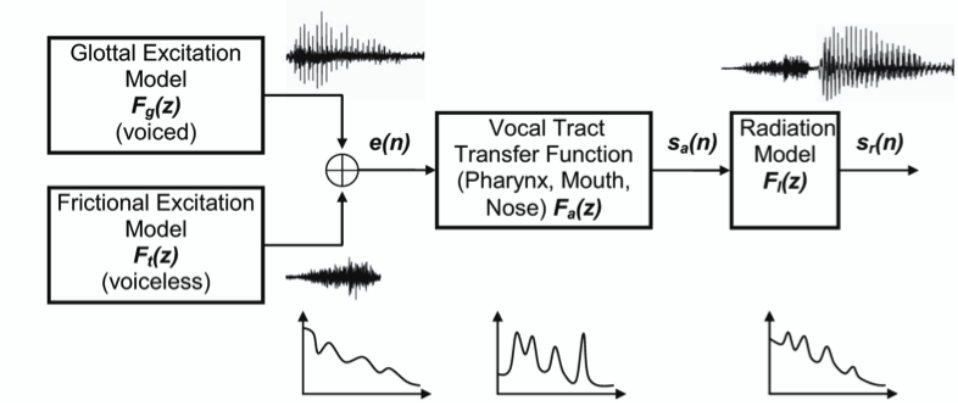

Figure 3. Fant's source-filter model to explain speech production. (Figure produced by authors)

flow, or both, will be the cause of the speech signal radiated. The resulting spectrum of the radiated signal (Figure 3, low row, right) will be the consequence of the application of the vocal tract transfer function (Figure 3, low row, middle) on the source spectrum (Figure 3, low row, left). Fant's model inspires the methodology to reconstruct the glottal source from phonated speech. The methodology consists in removing the influence of the radiation model and the vocal tract transfer function by inverse filtering by different methods. The one used in the present study is described in Gómez-Vilda *et al.* (2009), and is summarized in Figure 4.

The speech signal s(n) is first processed (1) to eliminate the influence of radiation and other undesirable effects due to channel characteristics. The radiation-compensated signal s_l(n) is filtered by a lattice-ladder mirror filter (2) which is designed to remove partially the influence of a hypothesized glottal source, generating a signal $s_v i$(n) which is mainly characterized by the vocal tract. This signal is modeled (4) to obtain the inverse signature of the vocal tract, which will be applied to the radiation-compensated signal s_l(n) to remove the influence of the vocal tract (5). The resulting signal $s_r i$(n) will be

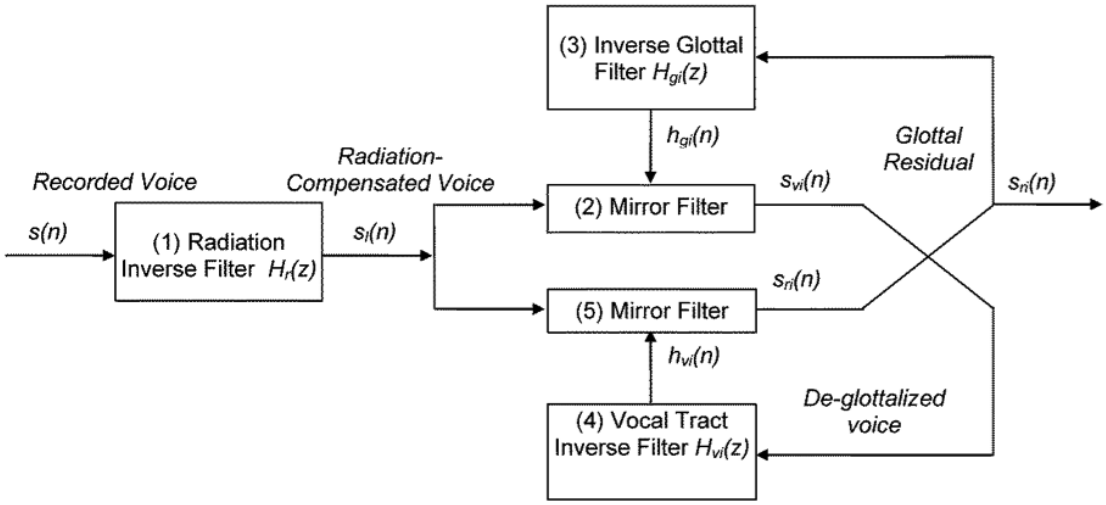

Figure 4. Methodology for the reconstruction of the glottal source from segments of phonated speech by recursive inverse adaptive filtering. (Figure produced by authors.)

dominated by the glottal features, and may be modeled (3) to produce a better inverse estimate of the glottal features, and injected in (2) to produce also a better estimate of $s_v i$(n). The recursion is iterated a low number of times, and the glottal residual $s_r i$(n) will be used to produce the glottal source by numerical integration. An example of the glottal source reconstruction is shown in Figure 5.

It may be seen that the reconstructed glottal residual $s_r i$(n) in Figure 5.b is the result of removing vocal tract resonances found in the original speech signal s(n). In particular the presence of the first resonance (formant) may be seen as a ringing (successive oscillations) taking place during each of the 17 pseudo-periodical glottal cycles extending over slightly more than 180 ms in Figure 5.a. The residual $s_r i$(n) is numerically integrated to produce the glottal source in Figure 5.c, which shows the main features of the pressure build up in the glottis. The main feature as far as the harmonic spectral contents of speech are concerned, is the maximum flow declination rate (MFDR), which is the negative drop of pressure signaled by red asterisks due to the closing phase. The glottal source is restored to its quiescent value (0) following a recovery pattern to reach a plateau, marking the duration of the contact phase. During the open phase, a pressure increment can be appreciated to reach a maximum, after which a sharp drop to reach the MFDR may be appreciated (closing phase). Finally in Figure 5.d a series of patterns showing the successive glottal flow cycles may be seen.

Once the glottal source has been reconstructed it is being parameterized according to different techniques in the time as well as in the frequency domain. The parameters are evaluated for each of the phonation cycles in the speech segment being analyzed (typically between 50 and 200 ms long). For male voice, between 5-20 glottal cycles are to be found in such an interval. Cycle-synchronous estimations of each parameter are stored in an array, average values and standard deviations are also evaluated. In what follows a brief description of these techniques and the resulting parameters is given:

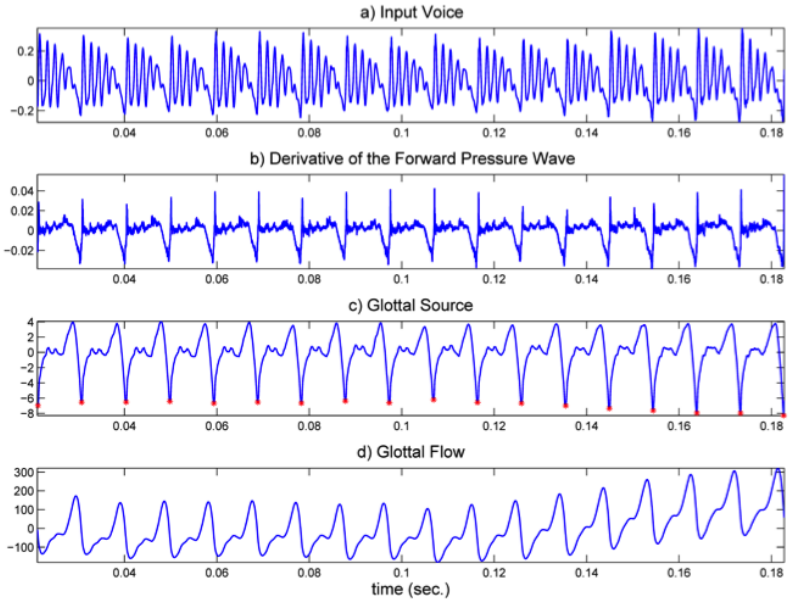

Figure 5. Example of glottal source and flow reconstruction from phonated speech: a) original speech signal (s); b) glottal residual $s_r i$(n), or derivative of the forward pressure wave; c) reconstructed glottal source (correlate of the pressure build up in the glottis); d) reconstructed glottal flow. (Figure produced by authors.)

- Perturbation parameters. These are a group of time-domain parameters related with voice quality, as the fundamental frequency f0, the jitter (relative fluctuations of the glottal source period), the shimmer (relative fluctuations of the glottal source amplitude for each glottal cycle), the absolute minimum sharpness (value of the MFDR), the noise to harmonic energy contents (HNR), or the ratio between the higher glottal source components to the first-order glottal source component (MAE). These parameters are given in Table 1.

Perturbation parameters
1. Absolute Pitch
2. Abs. Norm. Jitter
3. Abs. Norm. Ar. Shimmer
4. Abs. Norm. Min. Sharp (MFDR)
5. Noise-Harm. Ratio (NHR)
6. Muc./AvAc. Energy (MAE)

Table 1. Perturbation parameters.

- Cepstral parameters. This group consists in a collection of 14 parameters directly estimated from the cepstral description of the glottal source. The estimation process consists in generating the Fourier power spectrum of the glottal source. The cosine transform is applied to the logarithm of this spectrum and the first 14 resulting parameters are selected. Some of these parameters are extremely sensitive to certain factors such as gender or age (Muñoz, 2014). The parameters are listed in Table 2.

Cepstral Parameters
7. MWC Cepstral 1
8. MWC Cepstral 2
9. MWC Cepstral 3
10. MWC Cepstral 4
11. MWC Cepstral 5
12. MWC Cepstral 6
13. MWC Cepstral 7
14. MWC Cepstral 8
15. MWC Cepstral 9
16. MWC Cepstral 10
17. MWC Cepstral 11
18. MWC Cepstral 12
19. MWC Cepstral 13
20. MWC Cepstral 14

Table 2. Cepstral parameters.

- Spectral parameters. The spectral profile of the glottal source is conditioned by the biomechanical behavior of the vocal folds, especially the visco-elastic link between the fold body (*musculus vocalis*) and the epithelial cover and conjunctive tissues in Reinke's space. The envelope of the harmonic spectrum of the glottal source shows peaks and valleys which are influenced by this biomechanical behaviour. Anomalous relations among these peaks and valleys may serve as biometrical markers. The first group of parameters given in Table 3 are amplitude estimates of the peaks and valleys (21-27). The second group give their relative positions in frequency (28-32). Parameters 33 and 34 give the depth of the two first valleys relative to their frequency span (slenderness).

Spectral Parameters
21. MW PSD 1st Max. ABS.
22. MW PSD 1st Min. rel.
23. MW PSD 2nd Max. rel.
24. MW PSD 2nd Min. rel.
25. MW PSD 3rd Max. rel.
26. MW PSD End Val. rel.
27. MW PSD 1st Max. Pos. ABS.
28. MW PSD 1st Min. Pos. rel.
29. MW PSD 2nd Max. Pos. rel.
30. MW PSD 2nd Min. Pos. rel.
31. MW PSD 3rd Max. Pos. rel.
32. MW PSD End Val. Pos. rel.
33. MW PSD 1st Min NSF
34. MW PSD 2nd Min NSF

Table 3. Spectral parameters.

- Biomechanical parameters. The spectral behavior of the glottal source is directly related to the distribution of mass and visco-elasticity of the vocal fold body and cover. A methodology to estimate the distribution of mass and stiffness of each structure is possible using spectral matching techniques (Gómez-Vilda *et al.*, 2007). The most significant estimates are the mass and stiffness of the vocal fold body and cover, the ratio of energy losses due to viscid and turbulent flow behavior, and their respective unbalances. These are estimated using relative comparisons of mass, stiffness and losses from neighbor glottal cycles. The list of estimated parameters is given in Table 4.

Biomechanical Parameters
35. Body Mass
36. Body Losses
37. Body Stiffness
38. Body Mass Unbalance
39. Body Losses Unbalance
40. Body Stiffness Unbalance
41. Cover Mass
42. Cover Losses
43. Cover Stiffness
44. Cover Mass Unbalance
45. Cover Losses Unbalance
46. Cover Stiffness Unbalance

Table 4. Biomechanical parameters.

- Temporal parameters. The glottal cycle is divided into a closed phase and an open phase. The time instants associated with the start of the closed and open phase, as well as the time required to reach the quiescent pressure (recovery) and the maximum amplitude of the glottal source relative to the MFDR are estimated as important parameters in the time domain. Due to irregularities in the glottal source time profile, the recovery and open instants are estimated twice to produce more robust results. The open and closed instants, as well as the start of the closing phase are also estimated on the flow signal. The list of temporal parameters is given in Table 5.

- Glottal gap parameters. This set of parameters is designed to evaluate the contact defects, directly on the flow, calculating the ratio of air escape during the contact phase relative to the air escape during the open phase (59), or on the glottal source, in which case the defects are differentiated as contact, adduction or permanent ones, depending to which phase of the glottal cycle they affect. The list of the parameters is given in Table 6.

- Tremor parameters. The stiffness of the vocal fold body (*musculus vocalis*) is directly influenced by the neuromotor action of the laryngeal muscles, therefore, many neurological pathologies may be characterized from the estimates of this stiffness (parameter 37). Hypo-tonic or hyper-tonic deviations of this parameter are important correlates in Parkinson's Disease, for instance, as well as tremor.

Temporal Parameters
47. Rel. Recov. 1 Time
48. Rel. Recov. 2 Time
49. Rel. Open 1 Time
50. Rel. Open 2 Time
51. Rel. Max. Ampl. Time
52. Rel. Recov. 1 Ampl.
53. Rel. Recov. 2 Ampl.
54. Rel. Open 1 Ampl.
55. Rel. Open 2 Ampl.
56. Rel. Stop Flow Time
57. Rel. Start Flow Time
58. Rel. Closing Time

Table 5. Temporal parameters.

Glottal GAP Parameters
59. Val. Flow GAP
60. Val. Contact GAP
61. Val. Adduction GAP
62. Val. Permanent GAP

Table 6. Glottal GAP parameters.

A set of six parameters is devoted to track this disease. The first three give a description of the tremor in terms of its autoregressive modeling (63-65). The last ones give the tremor frequency in cycles/s (66), the reliability of this estimate (67), or the tremor amplitude in root mean square relative to the vocal fold body average amplitude (68). The list of tremor parameters is given in Table 7.

Tremor Parameters
63. 1st. Order Cyc. Coeff.
64. 2nd. Order Cyc. Coeff.
65. 3rd. Order Cyc. Coeff.
66. Tremor Frequency
67. Estimation Reliability
68. Tremor rMS Amplitude

Table 7. Tremor parameters.

The interested reader can find a more detailed description of each parameter meaning and distribution in Gómez *et al.* (2013).

Materials and methods

The purpose of the present research was to describe a methodology to parameterize the glottal source in terms of dysphonic voice and to study how to apply these parameters in speaker verification tasks. For this purpose a database of GSM-quality recordings from telephone conversations by 100 male speakers was used. Speech was recorded

at 8 KHz 16 bits and mu-law. Each conversation lasted between 5 and 30 min., fillers and long vowels were extracted from them. These long vowels were samples of vowels [a], [æ], [ɛ] and [e]. For classification purposes, the first two groups were labelled as /a/, whilst the last two groups were labelled as /e/. This last group covers most of the fillers which may be found in Spanish, consisting in lengthening of words as "de" or "que", or spontaneous insertions of /e/. An average of 6-8 of these fillers may be found in recordings of hesitating statements along a duration of 1-2 minutes. Fillers and long vowels were segmented as 100 ms fragments, and 68 parameters were obtained from each glottal cycle in the fragment. The resulting feature database is a matrix referred to as $\mathbf{Z}_t$.

Three experiments are described in this paper, the first oriented to provide full compatibility of parameter distributions of phonations from /a/ against phonations from /e/. This experiment is described in this section. The second experiment is designed to select a database of normative speakers from telephone quality recordings based in /e/ by contrasting the available telephone recordings with a normative database from high quality recordings. The selected normative speakers will be used as a control group in future work. The third experiment is designed to match telephone-quality /e/ recordings from the normative speakers against themselves to test the forensic matching capability of the methodology and to produce sensitivity and specificity estimates for the matching protocol.

The first experiment consisted in confronting the distributions of each parameter in $\mathbf{Z}_t$=**[**$\mathbf{Z}_{ta}$ $\mathbf{Z}_{te}$**]** from the /a/-group $\mathbf{Z}_{ta}$ and the /e/-group $\mathbf{Z}_{te}$ to check their degree of equivalence. The null hypothesis consisted in assuming the equivalence of distributions. The histograms for the fundamental frequency f0, jitter, shimmer, body mass and stiffness, and cover mass and stiffness are given in Figures 6 to 9.

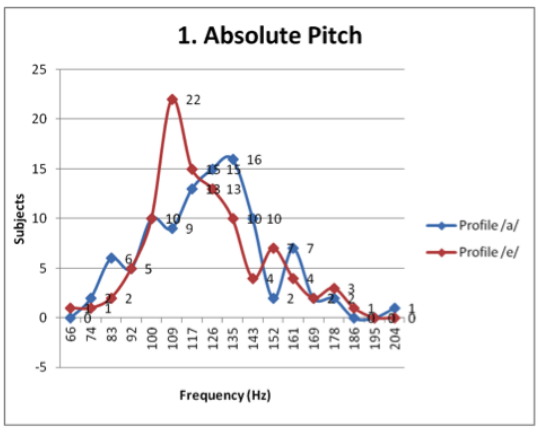

Figure 6. Comparison of the histograms of f0 from the /a/-group vs the /e/-group: The null hypothesis cannot be rejected given the distribution overlap (Figure produced by authors.)

The second experiment consisted in dividing the speakers in the database $\mathbf{Z}_t a$ in two subsets of 50 speakers each ($\mathbf{Z}_{tan}$ and $\mathbf{Z}_{tad}$) according to the degree of dysphonia present in their phonations confronting the whole speaker set with a normative set of 50 normophonic speakers selected and inspected at the ear, neck and throat service of Hospital Gregorio Marañón in Madrid. Normophonic speakers were inspected by video-

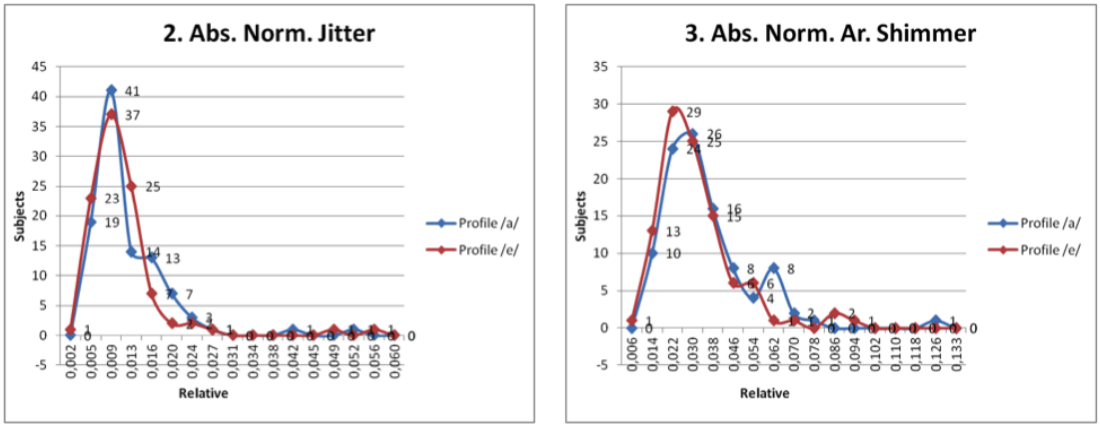

Figure 7. Comparison of the histograms of jitter and shimmer from the /a/-group vs the /e/-group: The null hypothesis cannot be rejected given the distributions overlap. (Figure produced by authors.)

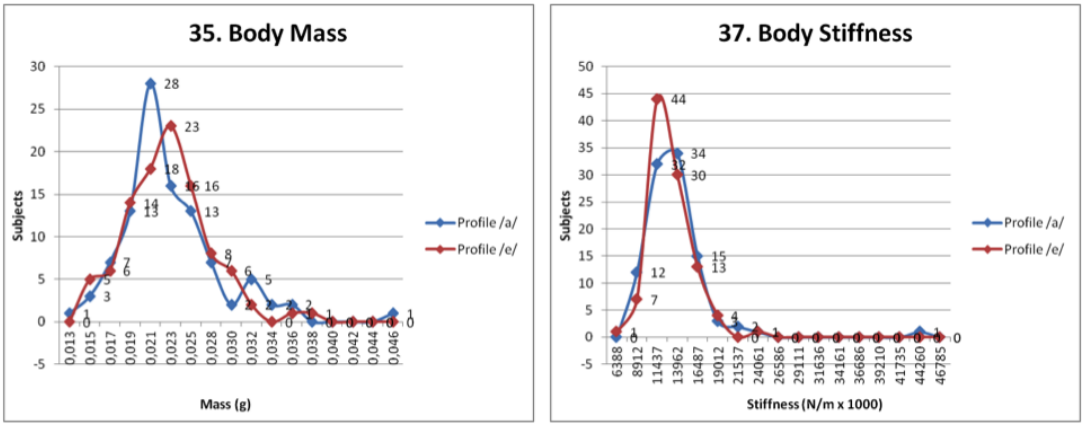

Figure 8. Comparison of the histograms of body mass and stiffness from the /a/-group vs the /e/-group: The null hypothesis cannot be rejected given the distributions overlap. (Figure produced by authors.)

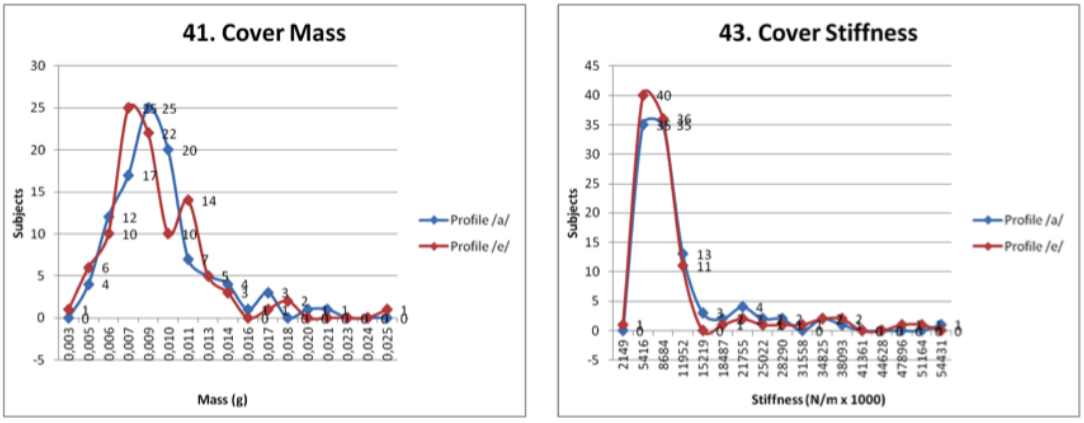

Figure 9. Comparison of the histograms of cover mass and stiffness from the /a/-group vs the /e/-group: The null hypothesis cannot be rejected given the distributions overlap. (Figure produced by authors.)

endoscopy to discard any organic problem in their larynx, and their non-dysphonic condition was assessed by the GRBAS test (Hirano, 1981). Fragments of phonations of vowel /a/ lasting 200 ms from the normative set of speakers taken at 44100 Hz and 16 bits were parameterized and used as a normative model ($\mathbf{Z}_{man}$) in the task of grading the /a/-group from GSM-quality recordings.

The third experiment was to match the features from each speaker in the subset of 50 normophonic males of the /e/-grou and telephone quality ($\mathbf{Z}_{ten}$) against his own feature set as target, and against the other 49 as imposters using the matching methodology to be described in what follows. The fillers from each speaker used in the matching as questioned tokens and the target set used as suspects' set were generated from two different recording sessions.

For the second experiment the membership of each speaker to the normophonic or dysphonic group was assessed using the log likelihood ratio between the conditioned probability of membership of a specific speaker s_i with feature set $\mathbf{z}_tai$ relative to the normative Gaussian mixture model (GMM) defined as $\Gamma_{man} = \mathbf{w}_{man}, \mu_{man}, C_{man}$ built on the normative feature dataset $\mathbf{Z}_{man}$, where $\mathbf{w}_{man}$, μ_{man} and $\mathbf{C}_{man}$ are the mixture weights, the average vector and the covariance matrix of the dataset. The definition of the normophonic membership log likelihood may be estimated as:

$$\lambda_{tai} = \log\{Pr(\mathbf{z_{tai}} \mid \mathbf{\Gamma_{man}})\} - \log\{1 - Pr(\mathbf{z_{tai}} \mid \mathbf{\Gamma_{man}})\} \quad (1)$$

where the conditioned membership probability will be given as:

$$Pr(\mathbf{z_{tai}} \mid \mathbf{\Gamma_{man}}) = \sum_k w_k^{man} \frac{1}{(2\pi)^{m_k/2}\left|\mathbf{C}_k^{man}\right|^{m_k}} e^{-1/2\left(\mu_{tai}-\mu_k^{man}\right)^T\left[C_k^{man}\right]^{-1}\left(\mu_{tai}-\mu_k^{man}\right)} \quad (2)$$

where k is the order of the GMM, and mk is the size of each Gaussian cluster.

In turn, the speaker matching methodology used in the third experiment was designed to estimate to which extent acoustic evidence from speaker s_i ($\mathbf{z}_{tei}$, considered the questioned evidence) against acoustic evidence from speaker s_j (Γ_{tej}, built on the suspect's evidence $\mathbf{z}_{tej}$) can modify the degree of conviction (gain of belief) in favour or against the suspect in relation with the case. This gain of belief is formulated as a log likelihood between the conditioned probability of $\mathbf{z}_{tei}$ being produced by the GMM model Γ_{tej} relative to the conditioned probability of $\mathbf{z}_{tei}$ being produced by any foil speaker from a line-up set characterized by the GMM model Γ_{ten}. This log likelihood ratio is a rephrasing of the balanced reasons method established by C. S. Peirce (1878), formulated as the conditioned probability of the prosecutor's hypothesis vs the defender's hypothesis (see Taroni *et al.* 2006; Gomez-Vilda *et al.* 2012:

$$V_{pt} = L_{pd} \times V_{pr} = \frac{Pr(E \mid H_p, I)}{Pr(E \mid H_d, I)} \times V_{pr} \quad (3)$$

where E is the evidence (questioned), H_p is the prosecutor's hypothesis (questioned evidence being produced by the suspect), and H_d is the defender's hypothesis (questioned

evidence being produced by any other speaker). In this way the a priori probability V_{pr} in favour of H_p will be amplified or attenuated by the gain of belief L_{pd} (likelihood ratio) to produce the a posteriori probability V_{pt}. The log likelihood ratio may be estimated as:

$$\lambda_{pd} = \lambda_{teij} = \log\{\Pr(\mathbf{z}_{tei} \mid \mathbf{\Gamma}_{tej})\} - \log\{\Pr(\mathbf{z}_{tei} \mid \mathbf{\Gamma}_{ten})\} \tag{4}$$

and the conditioned probabilities evaluating the prosecutor's and defender's hypotheses are given as:

$$\Pr(\mathbf{z}_{tei} \mid \mathbf{\Gamma}_{tej}) = \sum_k w_k^{tej} \frac{1}{(2\pi)^{m_k/2} \left|\mathbf{C}_k^{tej}\right|^{m_k}} e^{-1/2\left(\mu_{tai}-\mu_k^{tej}\right)^T \left[\mathbf{C}_k^{tej}\right]^{-1}\left(\mu_{tai}-\mu_k^{tej}\right)} \tag{5}$$

$$\Pr(\mathbf{z}_{tei} \mid \mathbf{\Gamma}_{ten}) = \sum_k w_k^{ten} \frac{1}{(2\pi)^{m_k/2} \left|\mathbf{C}_k^{ten}\right|^{m_k}} e^{-1/2\left(\mu_{tai}-\mu_k^{ten}\right)^T \left[\mathbf{C}_k^{ten}\right]^{-1}\left(\mu_{tai}-\mu_k^{ten}\right)} \tag{6}$$

It must be noted that in the third experiment the questioned and the suspect evidence were derived from individual speakers in the /e/-group normative feature set $\mathbf{Z}_{ten}$, whereas the line-up feature set was generated using the whole feature set $\mathbf{Z}_{ten}$. The results of the second and third experiments will be commented on in the section entitled "Validation and Sample Matching Results".

Another relevant aspect has to do with the selection of the parameters considered most relevant for dysphonia assessment or speaker matching. This procedure will be a premise to be incorporated into any of these procedures prior to the conditional probability estimation. The feature selection carried out was based on the evaluation of Fisher's discriminant ratios (Kim *et al.*, 2005), defined as:

$$C_{Fi} = \frac{\mu_{ki} - \mu_{kj}}{\sqrt{\frac{\zeta_{ki}^2}{n_i} + \frac{\zeta_{kj}^2}{n_j}}} \tag{7}$$

where μ_{ki} and μ_{kj} are the sample averages of subsets i and j for parameter k, ζ_{ki} and ζ_{kj} are the sample standard errors of subsets i and j, also for parameter k, and n_i and n_j are the respective subset sample sizes. To select the most relevant features a comparison of subset distributions is carried out, and only the most relevant features are included in the posterior analysis. An example is given in Figure 10.

Finally the issue of speaker match metrics is to be addressed. When estimating log likelihood ratios following (4), (5) and (6), if feature datasets can be grouped in a low number of clusters, log likelihood ratios can be expressed in terms of normalized distances among the questioned (test), suspect (control) and line-up (model) centroids, as shown in Figure 11.

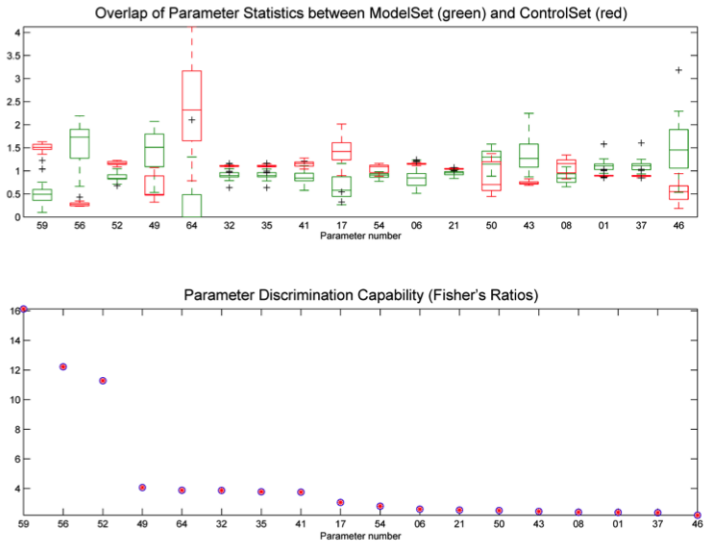

Figure 10. Parameter selection based on Fisher discriminant analysis: Upper template: boxplots of the most relevant parameters comparing the normophonic feature subset (green) against the dysphonic one (red). It may be seen that most of the distributions show low overlap, and small extent (being the conditions to produce a large Fisher's ratio as given by (7)). Lower template: Values of Fisher's ratios. (Figure produced by authors.)

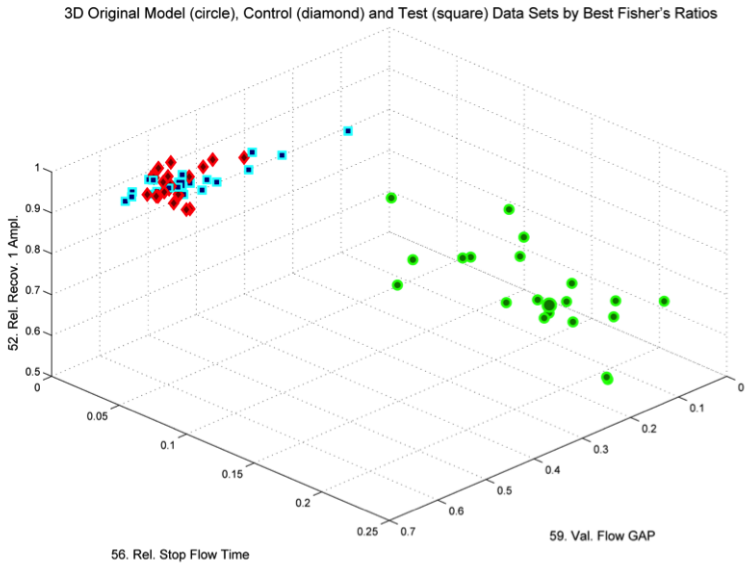

Figure 11. 3D description of evidence matching from a practical case in terms of the three most relevant features derived from Fisher's analysis in Figure 10: The questioned evidence is grouped as the test subset (blue squares). The suspect's evidence is grouped as the control subset (red diamonds). The line-up data is grouped as the model subset (green circles). Each subset centroid is signaled by a larger circle, diamond or square. A simple visual inspection allows inferring that the clusters of questioned and suspect evidence are much closer between themselves than to the line-up cluster. (Figure produced by authors.)

The 3D plot may be seen as the expression of the projection from an 18-dimensional vector space defined in terms of the 18 selected features to a 3-dimensional subspace in terms of the 3 most relevant ones. Reducing clusters to centroids allows defining the log likelihood as a normalized distance balance given by:

$$\lambda_{pd} = \frac{D^2_{TM} - D^2_{TC}}{2} \tag{8}$$

where D_TM is the normalized distance between the centroids of the questioned evidence set to the model set, and D_TC is the distance between the centroids of the questioned and suspect evidence. The centroids of the three sample sets (test, control and model) define the match triangle CTM as depicted in Figure 12.

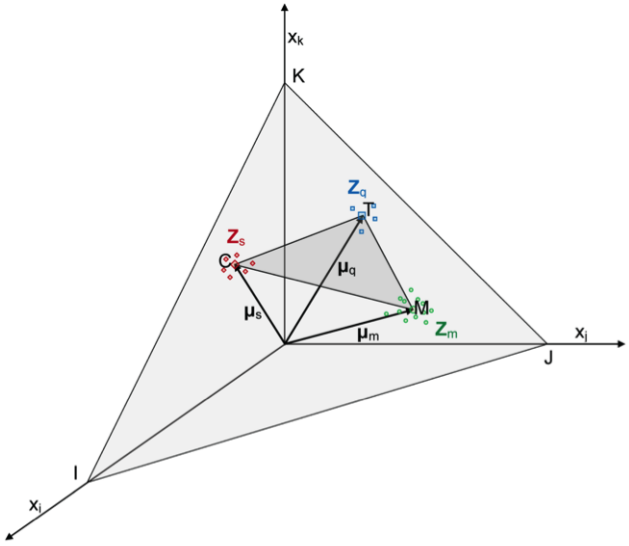

Figure 12. Match triangle defined by the test, control and model centroids on the 2D plane projection of a 3D description of evidence matching in similar terms to the one given in Figure 11. (Figure produced by authors.)

It may be seen that the centroids of clusters T (questioned), C (suspect) and M (line-up) define a plane intersecting the three feature axes x_i, x_j and x_k at the points I, J and K. This property allows summarizing the matching results in a balanced chart as the one given in Figure 13.

The Mahalanobis normalized distances between each two centroids C, T and M defining the match triangle, MDTC, MDTM and MDCM as seen in Figure 13 can be used to establish the relationship between questioned, suspect and model evidence. It is clear that the vertical axis in the figure is the place of all possible solutions which share the condition of D_{TC}=D_{TM}, for which the log likelihood will be null (λ_{pd}=0: neutral decision). The right hand plane defined by the vertical axis will define the place of all possible solutions where D_{TC}>D_{TM}, therefore the log likelihood will be negative (λ_{pd}<0: decision favoring the defender's hypothesis). The left hand plane defined by D_{TC}<DT_M will correspond to positive log likelihood ratios (λ_{pd}<0: decision favoring the prosecutor's hypothesis). Nevertheless, the decision cannot be based on just crossing the vertical line

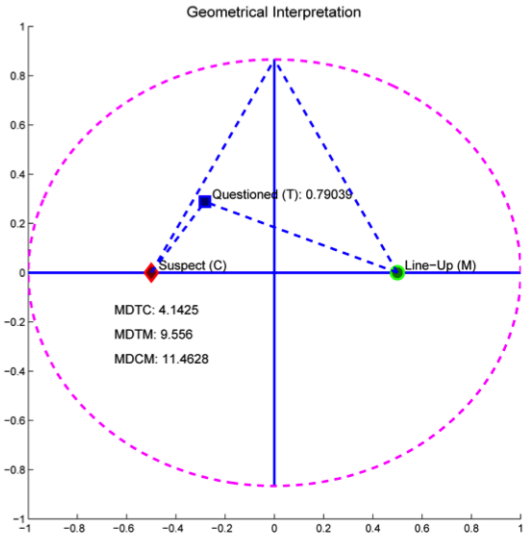

Figure 13. Balanced chart summarizing the match between questioned and suspect evidence relative to the line-up model. (Figure produced by authors.)

to accept the prosecutor's hypothesis, as this threshold would be unfair with respect to the guarantees due to the legal defense of the suspect. A more conservative threshold decision should be used. Accordingly with Daubert rules (see U.S. Supreme Court, 1993) accepting and evaluating the strength of evidence should be left to the Court. However, it is generally accepted by the European Network of Forensic Science Institutions that this kind of scale would be of useful application to grade the strength of the evidence to help the decision of the Court. A reasonable scale can be found in Lucy (2005), and is reproduced in Table 8.

There is another important detail regarding expression (8) and Figure 13, which concerns situations where $D_{TC} >> D_{CM}$ and $D_{TM} >> D_{CM}$. This ill-conditioned case happens when questioned and suspect evidence are far apart from the line-up data, and would indicate a bad selection of the line-up. In this unfair situation accepting results in the left hand side ($D_{TC} < D_{TM}$) would break the guarantee of a fair evaluation, helping to produce a decision in favor of the prosecutor's hypothesis although the line-up has not been well selected. For this reason, the boundary signaled by the pink dash ellipse corresponding to the place of the points meeting the condition:

$$\frac{D_{TM} + D_{TC}}{2} \leq D_{CM} \tag{9}$$

has been defined as a protection boundary. No match should be accepted as valid if the questioned centroid appears beyond the limits of the guarantee boundary thus defined.

Detection and Matching Results

In the present section an account of the results obtained for the second and third experiments, as described in the above section will be given. The summarized characteristics and objectives of each experiment are given below.

Second experiment description:

Range (decimal log)	Range (natural log)	Statement
$\vartheta_{lg} < 0$	$\vartheta_{ln} < 0$	Likelihood **unconditionally** supports the hypothesis that the questioned and the suspect evidence have not been produced by the same speaker (favoring defender's hypothesis)
0 $\vartheta_{lg} < 1$	0 $\vartheta_{lg} <$ 2,3026	Likelihood **weakly** supports the hypothesis that the questioned and the suspect evidence have been produced by the same speaker (favoring prosecutor's hypothesis)
1 $\vartheta_{lg} < 2$	2,3026 $\vartheta_{lg} <$ 4,6052	Likelihood **mildly** supports the hypothesis that the questioned and the suspect evidence have been produced by the same speaker (favoring prosecutor's hypothesis)
2 $\vartheta_{lg} < 3$	4,6052 $\vartheta_{lg} <$ 6,9078	Likelihood **moderately** supports the hypothesis that the questioned and the suspect evidence have been produced by the same speaker (favoring prosecutor's hypothesis)
3 $\vartheta_{lg} < 4$	6,9078 $\vartheta_{lg} <$ 9,2103	Likelihood **strongly** supports the hypothesis that the questioned and the suspect evidence have been produced by the same speaker (favoring prosecutor's hypothesis)
$\vartheta_{lg} \geq 4$	$\vartheta_{lg} \geq$ 9,2103	Likelihood **very strongly** supports the hypothesis that the questioned and the suspect evidence have been produced by the same speaker (favoring prosecutor's hypothesis)

Table 8. Strength of evidence according to Lucy (2005).

- Splitting the 100 male speakers into two equal-sized subsets according to their normophonic condition.
- Using a normative database validated by Hospital Gregorio Marañon in Madrid with samples of /a/ (50 male speakers).
- Log likelihood ratios according to (1) and (2) estimate the conditional probability of a given sample being normophonic or dysphonic (10-fold cross-validation, taking 47 subjects, leaving 3 in each set of normophonics and dysphonics per run).

Second experiment objectives:

- Estimate the discrimination accuracy of the methodology and the most relevant parameters.

- Produce two reference subsets from GSM quality from the /e/-group of use in Spanish.

Second experiment results:

- The normophonic vs dysphonic cumulants, sensitivity, specificity and accuracy, and Detection-Error Trade-off plots are given in Figure 14.

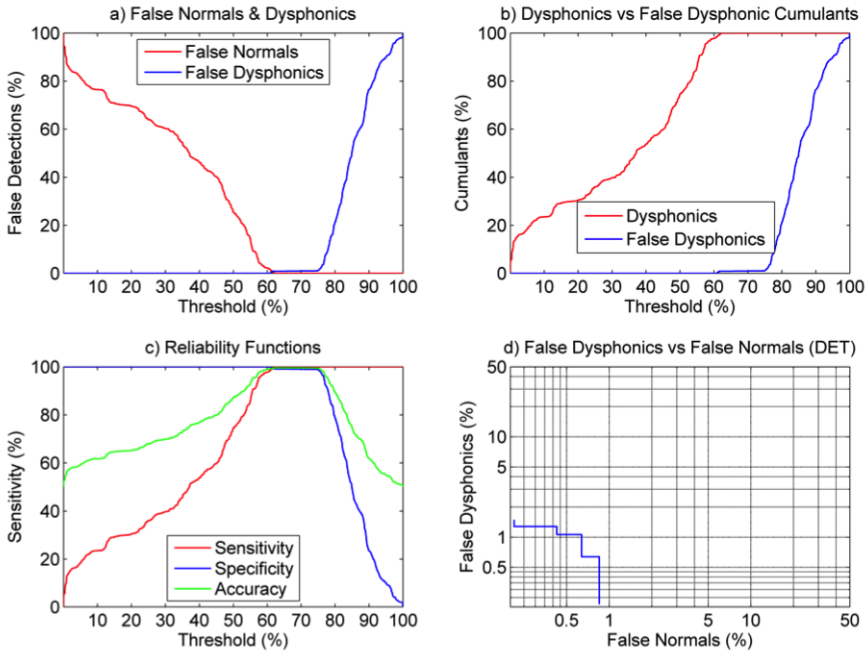

Figure 14. a) False normal vs false dysphonic cumulants. b) Associated Tippet plots. c) Sensitivity, specificity and accuracy for the second experiment. d) Detection-error trade-off curve. (Figure produced by authors.)

The detection procedure consists in generating a vector with the log likelihood ratios generated for each sample, and their assumed condition of normophonic or dysphonic. The log likelihood span is normalized as a percentage, and a moving threshold scans it from 0 to 100%. For each value scanned the number of false normophonics (samples annotated as dysphonic but quoted as normophonic because their log likelihood is over the threshold) and false dysphonics (samples annotated as normophonic but quoted as dysphonic because their log likelihood is under the threshold) is annotated and plotted. See that the number of false normophonics diminishes as the threshold moves rightwards in Figure 14.a to reach the point 1, where the number of false normophonics is very low (only 3 cases out of 470 possible ones), whereas the number of false dysphonics is still 0. At point 8 this number starts raising to 4 out of 470, whereas the number of false normophonics has decreased to 0, indicating that the optimal detection conditions are somewhere between 1 and 8, keeping both false detections at a minimum value simultaneously. This fact indicates that there are two different distributions for each population (false normophonics and false dysphonics), whose accumulated distributions are given in Figure 14.b, known as Tippett plots. Based on these distributions, the plots in Figure 14.c give the well-known variables of sensitivity, specificity and accuracy, according to the following relations:

where TP, FP, TN and FN are the number of true dysphonics, false dysphonics, true normophonics and false normophonics, respectively. These three functions are plotted in

$$\begin{aligned} Sn &= \frac{TP}{TP + FN}; \\ Sp &= \frac{TN}{TN + FP}; \\ Ac &= \frac{TP + TN}{TP + FN + TN + FP} \end{aligned} \qquad (10)$$

Figure 14.c, where the optimum detection point is the one where the accuracy is the maximum. If the number of false dysphonics is plotted vs the number of false normophonics the result is the template in Figure 14.d, which is known as the detection-error trade-off plot, because the specific situations combining false positives vs false negatives is confronted for a set of critical threshold values. The number of these situations is 8, and they are signaled in the plot. In fact, apart from the two points already analyzed (1 and 8), the rest of the cases is as follows:

2. False dysphonics jump up to 1/470, false normophonics do not change.
3. False dysphonics do not change, false normophonics drop to 2/470.
4. False dysphonics jump to 2/470, false normophonics do not change.
5. False dysphonics do not change, false normophonics drop to 1/470.
6. False dysphonics jump to 3/470, false normophonics do not change.
7. False dysphonics do not change, false normophonics drop to 0.

The optimal case is point 3, where the rates of false dysphonics and normophonics are equal to 0.638% (equal error rate). The detection accuracy function is at its maximum value of 99.57% at this point, for a threshold range between 62.08% and 75.04%, which implies a reasonably wide noise margin.

Third experiment description:

- Matching each normative speaker's sample (questioned) against every other normative sample (suspect: one target sample vs 49 non-target samples). Eliminating repetitions, these settings imply 50 target detections vs 1,225 non-target detections.
- Using as model set (line-ups) the set of 50 normative speakers, to grant condition (9) as much as possible.

Third experiment objective:

- Estimate the discrimination accuracy of the sample matching methodology in target vs non-target detection tasks.

Third experiment results:

- False target vs false non-target detection cumulants, sensitivity, specificity and accuracy functions, and detection-error trade-off plots given in Figure 15.

As before, the detection procedure consists in generating a vector with the log likelihood ratios (LLR) generated for each sample, and their assumed condition of target or non-target. No normalization of the threshold span has been carried out in this case.

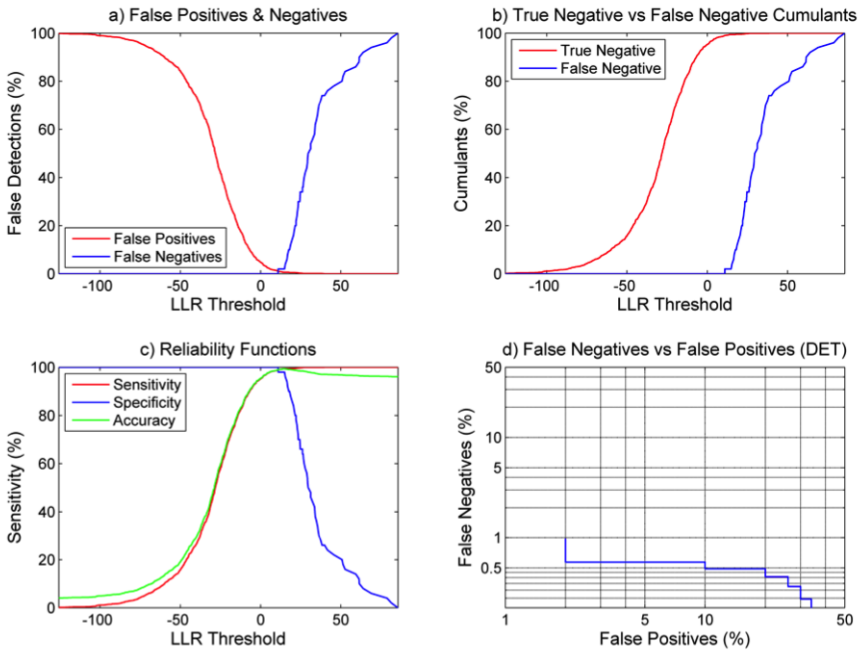

Figure 15. a) False positive vs false negative cumulants. b) Associated Tippet plots. c) Sensitivity, specificity and accuracy for the third experiment. d) Detection-error trade-off curve. (Figure produced by authors.)

The information provided by Figure 15 once the experimental conditions are fixed, can be summarized as follows:

a. The rate of false positives (in red) gives the evolution of the non-target cases detected equivocally as targets, as the detection threshold for the log likelihood ratio is moving from left to right. Given the relatively large number of non-target cases (1,225) the evolution of this curve is a smooth decay (inverted sigmoid), expressing that its distribution function will be bell-shaped. On the contrary the low number of target cases (50) given by the blue curve shows slight jumps as the threshold is moving, incorporating new targets as if they were non-targets (false negatives). Both curves cross at the threshold value of 10.88. This is the point of maximal accuracy, on the sixth interval reflected in 0, in the margin where the evidence supports strongly the prosecutor's vs the defender's hypothesis. The detection methodology is maximally accurate just at the beginning of that interval, availing the guarantee of the test. The value of the accuracy function at that point is 99.29%.
b. The graphics given in template a) are given now as Tippett plots. They do not provide any more information to what has been commented up to now, except stressing the fact that the overlap between the two accumulated distributions is very low, granting that at the cross-point the residual tail probabilities (p-values) are under 0.02 and 0.0057, well below the significance level of 0.05.
c. The sensitivity (number of non-targets detected as targets over the total non-targets), specificity (number of targets detected as non targets over the total targets) and accuracy (number of the total targets and non-targets detected as such over the total cases) are plotted as a function of the threshold. The accuracy is very large for the margin of strong support of the prosecutor's hypothesis vs the defender's, with a maximum at 99.29% and not being below 96.08% in any case.

d. The detection-error trade-off curve is shown with a staircase pattern given the specific design of the test. The log likelihood ratios and the corresponding false positive and negative rates are given in Table 9.

Points in the intersection interval between the false positive and false negative curves			
Point	False Positive Rate (%)	False Negative Rate (%)	Log Likelihood Ratio
1	2.0	0.98	10.92
2	2.0	0.90	11.85
3	2.0	0.82	13.06
4	2.0	0.73	13.35
5	2.0	0.65	13.67
6	2.0	0.57	14.84
7	4.0	0.57	15.36
8	6.0	0.57	16.41
9	8.0	0.57	16.79
10	10.0	0.57	17.53
11	10.0	0.49	17.98
12	12.0	0.49	18.61
13	14.0	0.49	19.49
14	16.0	0.49	20.46
15	18.0	0.49	20.91
16	20.0	0.49	21.48
17	20.0	0.41	21.96
18	22.0	0.41	22.05
19	24.0	0.41	22.59
20	26.0	0.41	22.88
21	26.0	0.33	22.94
22	28.0	0.33	23.33
23	30.0	0.33	23.45
24	30.0	0.24	24.33
25	32.0	0.24	24.55
26	34.0	0.24	24.66

Table 9. List of points in the intersection interval between the false positive and false negative curves.

The equal-error-rate is not easily determined in this case due to the abrupt staircase behavior of the transition interval in the case of false positive rate. Nevertheless several merit figures may be inferred, for instance, it will be possible to sustain a rate of 2% false positives with a rate of 0.57% false negatives (point 6). This means that accepting an error of one negative in 50 taken as positive grants an error of one positive in 175 taken as negative. The merit figures of both the second and third experiments are given in Table 10.

Conclusions

The process of speaker recognition from speech is a complex matter, as far as the co-articulation involved in message coding expands the limits of intra-speaker variability.

Experiment	Samples	No. Tests	Accuracy (%)	LLR	EER	p-values
First	50N + 50D	90 Samples vs Model x 10 times cross-val. = 900	99.57	NA	0.638 (3)	0.00638, 0.00638
Second	50N + 50D	50 Samples vs each: 51∗50/2 = 1275 (50 target + 1225 non-target)	99.29	10.88	NA	0.02, 0.0057

Table 10. Summary of results for the second and third experiments.

This problem can be alleviated if biometrical markers are defined in relation with phonation, as this phenomenon is less variable for a given speaker, depending only on phonatory settings (creaky, modal, pressed, falsetto, etc). Phonation may experience changes from aging as well as from hormonal status, tobacco, drugs or alcohol consumption, vocal abuse, infections, allergies, other health status conditions, and even circadian cycles (phonation late in the evening is not the same as during the first hours after waking up). It must be assumed that no forensic voice analysis system can realistically manage all this variability, as most of the times the questioned evidence is just a segment of poor quality conversation, and not much more. Regarding the modeling of suspect's evidence, it would be possible sometimes to obtain speech samples under different conditions and in different sessions, but this is not possible most of the time. Our group has conducted multisession tests in very specific collaborative situations such as twins' voice studies (San Segundo and Gómez, 2013; San Segundo, 2014) trying to simulate various possible forensic scenarios. Furthermore, for the current study session variability has been taken into account as far as parameter selection is concerned. Our study is based on 68 phonation parameters, from which some are very variable with phonation modality and condition, while others are almost invariant to the alterations described. The parameters used in the forensic phonation match have been previously selected according to prior knowledge: for instance jitter, shimmer, noise-harmonic ratios, certain cepstral parameters, glottal source spectral profile, closure and contact defects, and low order tremor are not very sensitive to temporal alterations, and can be safely used in these studies. Focussing on phonation biometrical markers does not necessarily reduce the recognition capability of the methodology, as happens with fingerprints. It is well known that fingerprint matching does not use the whole information available in a fingerprint image; on the contrary, only specific biometrical markers, known as minutias are involved in pattern matching. In this way the process of fingerprint matching becomes more efficient, accurate, robust and less computationally expensive (Jain *et al.*, 1997). The application of this deconstructive methodology to speech implies focussing on phonated speech, rather than in the whole set of voiced and unvoiced patterns. Furthermore, from phonated speech only long vowels close to the axis /a/-/e/ were considered in the present study. These are some of the conclusions derived from the experimental setup used in the study:

- The detection of dysphonic voicing from normophonic seems viable using parameterizations of phonation based on the reconstruction of the glottal source.
- The sensitivity, specificity and accuracy in detecting dysphonic phonation are large enough to grant using phonated segments of speech as long vowels and fillers in forensic voice matching over sufficiently wide detection spans.
- The parameterizations of /a/ and /e/ groups of vowels are interchangeable to a paired test extent, to be used in cross-matching tests with no significant statistical differences.
- The accuracy of target vs non-target sample phonation matches grants the applicability of these tests to real forensic cases.
- The margin of optimal log likelihood ratios granting the strength of phonation evidence over 4 in Lucy's scale (Lucy, 2005) allows its applicability under robust conditions.
- The matching of questioned vs suspect's evidence in reference to line-ups may be summarized in meaningful 2D plots of simple and easy interpretation, granting the reliability and security of the procedure regarding court standards.
- Hybridizing scores from speech and phonation standards as MFCC's and glottal source derived parameters may attain competitive low equal error rates over telephone-quality speech (Khoury *et al.*, 2013).

The proposed methodology for voice pathology detection and monitoring, as well as for forensic voice inspection is being used by police services in Spain and other academic and private institutions (Gomez-Vilda *et al.*, 2012).

Acknowledgements

This work is being funded by grant TEC2012-38630-C04-04 from Plan Nacional de I+D+i, Ministry of Economy and Competitivity of Spain.

References

Dejonckere, P. H. (2010). Assessment of voice and respiratory function. In M. Remacle and H. E. Eckel, Eds., *Surgery of Larynx and Trachea*, 11–26. Berlin: Springer-Verlag.

Gomez-Vilda, P., Olalla, R. M., Fernandez, L. M. M., Biarge, M. V. R., Mulas, C. M., Marquina, A. A., Hierro, J. A. H. and Salinero, R. N. (2012). Distance metric in forensic voice evidence evaluation using dysphonia-relevant features. In *Proceedings of the VI Meeting of Biometric Recognition of Persons*, 169–178: Ed. Universidad de Las Palmas de Gran Canaria.

Gómez, P., Rodellar, V., Nieto, V., Martínez, R., Alvarez, A., Scola, B., Ramírez, C., Poletti, D. and Fernández, M. (2013). BioMet®Phon: A system to Monitor Phonation Quality in the Clinics. In *Proceedings of the 5th International Conference on e-Health, Telemedicine and Social Medicine*, 253–258, Nice, France.

Gómez-Vilda, P., Fernández-Baillo, R., Rodellar-Biarge, V., Lluis, V. N., Álvarez Marquina, A., Mazaira-Fernández, L. M., Martínez-Olalla, R. and Godino-Llorente, J. I. (2009). Glottal source biometrical signature for voice pathology detection. *Speech Communication*, 51(9), 759–781.

Gómez-Vilda, P., Fernández-Baíllo, R., Nieto-Altuzarra, A., Díaz-Pérez, F., Fernández-Camacho, F. J., Rodellar-Biarge, V., Álvarez Marquina, A. and Martínez-Olalla, R. (2007). Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters. *Journal of Voice*, 21(4), 450–476.

Hakkesteegt, M. M., Brocaar, M. P. and Wieringa, M. H. (2010). The applicability of the dysphonia severity index and the voice handicap index in evaluating effects of voice therapy and phonosurgery. *Journal of Voice*, 24(2), 199–205.

Hirano, M. (1981). *Psycho-acoustic evaluation of voice.* New York: Springer-Verlag.

Jain, A., Hong, L. and Bolle, R. (1997). On-line fingerprint verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4), 302–314.

Khoury, E., Vesnicer, B., Franco-Pedroso, J., Violato, R., Boulkcnafet, Z., Mazaira Fernandez, L. M., Diez, M., Kosmala, J., Khemiri, H., Cipr, T., Saeidi, R., Gunther, M., Zganec-Gros, J., Candil, R. Z., Simoes, F., Bengherabi, M., Alvarez Marquina, A., Penagarikano, M., Abad, A., Boulayemen, M., Schwarz, P., Van Leeuwen, D., Gonzalez-Dominguez, J., Neto, M. U., Boutellaa, E., Gomez Vilda, P., Varona, A., Petrovska-Delacretaz, D., Matejka, P., Gonzalez-Rodriguez, J., Pereira, T., Harizi, F., Rodriguez-Fuentes, L. J., El Shafey, L., Angeloni, M., Bordel, G., Chollet, G. and Marcel, S. (2013). The 2013 speaker recognition evaluation in mobile environments. In *Proceedings of the 6th IAPR International Conference on Biometrics*, Madrid, Spain.

Kim, S. J., Magnani, A. and Boyd, S. (2005). Robust fisher discriminant analysis. In *Advances in Neural Information Processing Systems*, 659–666. MIT Press.

Lucy, D. (2005). *Introduction to Statistics for Forensic Scientists.* Hoboken, NJ: Wiley.

Muñoz, C. (2014). *Speech signals Feature Extraction Model for a Speaker's Gender and Age Identification System.* Phd thesis, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid.

Peirce, C. S. (1878). The probability of induction. *Popular Science Monthly*, 12, 705–718.

Roy, N., Barkmeier-Kraemer, J., Eadie, T., Sivasankar, M. P., Mehta, D., Paul, D. and Hilman, R. (2013). Evidence-based clinical voice assessment: A systematic review. *American Journal of Speech-Language Pathology*, 22, 212–226.

San Segundo, E. (2014). *Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source parameters and cepstral characteristics.* Phd thesis, Universidad Internacional Menéndez Pelayo.

San Segundo, E. and Gómez, P. (2013). Voice biometrical match of twin and non-twin siblings. In *Proceedings of MAVEBA 2013*, 253–256, Florence, Italy: Firenze University Press.

Taroni, F., Aitken, C., Garbolino, P. and Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science.* Hoboken, NJ: Wiley.

U.S. Supreme Court, (1993). *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 509 US 579, 589".

Considerações sobre o papel da sociofonética na comparação forense de locutores

Cintia Schivinscki Gonçalves & Cláudia Regina Brescancini

Instituto-Geral de Perícias/SSP-RS, Laboratório de áudio e Fonética Acústica-LAFA/PUCRS
&
PUCRS/CNPq, Faculdade de Letras, Laboratório de áudio e Fonética Acústica-LAFA/PUCRS

Abstract. *This study examines the field of Forensic Linguistics, from the perspective of Sociophonetics. In the paper we present sociolinguistic and phonetic considerations which are relevant to the work in Forensic Phonetics, with an emphasis on what happens at the official level, specifically with reference to the expertise of speaker comparison. The objective is to lay the foundations for the legitimate linguistic descriptions that are used in this type of investigation confrontation aimed at determining the origin of a speaker from his/her voice/speech. Issues concerning the relevance of the concept of 'community of practice' in the context of forensics are approached, as well as the application of sociolinguistic interviews, the stylistic variation present in the material for analysis, and linguistic elements of a sociophonetic nature, commonly used as technical and comparative parameters. This analysis can help improve the practice of speaker comparison, and contribute to the quality of the resulting technical production, namely the technical report.*

Keywords: *Forensic linguistics, forensic phonetics, speaker comparison, sociolinguistics, sociophonetics.*

Resumo. *Este estudo é do escopo da Linguística Forense, sob a perspectiva da Sociofonética. Nele são apresentadas considerações sociolinguísticas e fonéticas pertinentes ao trabalho pericial em Fonética Forense, com ênfase ao que ocorre no âmbito oficial, especificamente as que sejam aplicáveis à perícia de Comparação de Locutores. Objetiva-se estabelecer os fundamentos que legitimam a descrição linguística utilizada nesse tipo de confronto voltado à determinação de origem a partir da voz/fala. São abordadas questões relativas à pertinência do conceito de comunidade de prática no contexto da abordagem pericial forense, à aplicação da entrevista sociolinguística, à variação estilística presente no tipo de material habitualmente analisado e aos elementos linguísticos de natureza sociofonética comumente utilizados como parâmetros técnico-comparativos. Estima-se que os apontamentos aqui feitos possam colaborar para o aprimoramento da prática empregada na perícia de Comparação de Locutores, no sentido de contribuir para a*

qualificação da produção técnica resultante, a saber, o laudo pericial e/ou o parecer técnico.

Palavras-chave: *Linguística forense, fonética forense, comparação de locutores, sociolinguística, sociofonética.*

A linguística forense e a perícia de comparação de locutores

A Linguística Forense, entendida como o estudo científico da linguagem dirigido aos objetivos e contextos forenses (McMenamin, 2002), abarca uma ampla gama de tipos de perícias relacionadas à linguagem oral e escrita. Especificamente quanto às análises em registros de áudio, de competência da área de Fonética Forense, são atualmente realizadas pelos órgãos periciais oficiais brasileiros a Análise de Conteúdo, cujo enquadramento como perícia não é consensual, e as perícias de Verificação de Edição e de Comparação de Locutores (Morrison *et al.*, 2009).

Destaca-se que no Brasil, diferentemente do que ocorre em outros países, como no Reino Unido, profícuo em publicações na área de Fonética Forense, a perícia criminal está a cargo exclusivamente do Estado, cabendo à iniciativa privada somente a participação autónoma como Perito Judicial ou Assistente Técnico da parte. Nesse caso, são possíveis solicitantes desse tipo de perícia a Autoridade Policial (Delegado), o Policial Militar (Oficial responsável por Inquéritos Policiais Militares), a Autoridade Judiciária (Juiz de Direito), a Defensoria Pública (Defensor Público) e o Ministério Público (Promotor de Justiça)[1].

Conceitualmente, o reconhecimento de um indivíduo a partir de sua voz e fala é intitulado Reconhecimento de Locutor (*Speaker Recognition*) (Hollien, 2002; Nolan, 1983; Rose, 2002), podendo envolver a identificação ou a verificação, sendo ambas, segundo Nolan (1983), um processo de decisão que confirma ou nega que duas amostras de voz foram produzidas pelo mesmo aparato vocal.

A relevância desses tipos de tarefas na área criminal é pontuada por Braid (2003), para quem "um exame de Verificação de Locutor também é capaz de desvincular o envolvimento de um inocente num crime que lhe possa estar sendo imputado, o que talvez seja até mais importante do que incriminar um culpado." (p. 6)

Na Identificação de Locutor procede-se à comparação da amostra de fala de um indivíduo desconhecido com um grupo de amostras de fala pertencentes a locutores de identidade sabida, reunidas para fins de confronto ou pertencentes a um determinado banco de dados de produções orais, ainda inexistente de forma consistente no Brasil. Na literatura internacional, o trabalho de alinhamento de vozes é referido como *voice line-up*, sendo utilizado para reconhecimento por testemunha, profissional da área ou sistema automático (Hollien, 2002).

Atualmente, as perícias em registro de áudio em âmbito nacional mostram-se imersas em um contexto no qual prevalece a tarefa de Verificação de Locutor. Nesta, são confrontadas propriedades de, via de regra, duas amostras, tendo sido uma delas, a amostra relativa ao locutor que se deseja saber a autoria, por exemplo, obtida através de interceptação telefónica e a outra, a amostra relativa ao locutor de identidade conhecida, obtida pelos próprios peritos em procedimento de coleta técnica de padrão vocal. Tal contraponto situacional compreende a chamada Comparação de Locutores (doravante CL).

Embora a CL envolva a comparação de uma amostra de fala teste (amostra questionada) com uma amostra de referência (amostra padrão) de um único locutor (Nolan, 1983), na prática, contudo, observa-se, por vezes, a ocorrência de pequenas variações em tal delineamento de confronto, encontrando-se, por exemplo, a indicação de um locutor de identidade sabida e uma gravação (ou mais) com duas vozes passíveis de lhe serem atribuídas, ou seja, um locutor-padrão *versus* dois locutores-questionados. Pode-se encontrar ainda a indicação de uma amostra de fala de autoria desconhecida e dois ou três suspeitos de a terem produzido, ou seja, um locutor-questionado *versus* dois ou mais locutores-padrão, não pertencentes a alinhamento pré-estabelecido ou banco de dados, mas apontados como suspeitos pelo solicitante da perícia.

Independente do delineamento da investigação, o objetivo é definir a autoria de falas armazenadas em uma determinada mídia, avaliando-se se essas, de fato, foram produzidas (ou não) pelo aparelho fonador de um determinado indivíduo (suspeito, indiciado ou réu). Os perfis de voz e de linguagem expressiva oral do locutor questionado e do locutor de identidade sabida são, portanto, cotejados, identificando-se quais parâmetros, dos elencados, são indicativos de convergência e de divergência entre as amostras.

Atualmente já se tem consolidada a percepção de que no confronto forense de voz e fala não existe parâmetro que possa, isoladamente, ser utilizado como referência individualizante indelével, de forma que as conclusões sobre a autoria das emissões orais consideram não apenas um, mas um conjunto de parâmetros técnico-comparativos, sendo o comportamento vocal e linguístico dos locutores do cotejo escrutinado em suas características gerais e particularizantes.

French *et al.* (2010) apontam como comumente admitidos na CL os seguintes parâmetros:

- configuração vocal (Laver, 1980, 1994), qualidade vocal e *pitch*, obtido através da média e da variação da frequência fundamental (f0);
- taxa de articulação;
- entonação e traços rítmicos;
- processos da fala encadeada, como padrões de assimilação e de elisão;
- traços consonantais (por exemplo, o *locus* de energia das fricativas; a soltura das plosivas; a duração de nasais, das líquidas e das fricativas em contexto fonológico específico; o tempo de início de vozeamento (VOT) das plosivas; a presença ou a ausência de pré-vozeamento em plosivas átonas e variáveis sociolinguísticas discretas);
- traços vocálicos, incluindo configuração formântica, frequência central, densidade, largura de banda e qualidade auditiva de variáveis sociolinguísticas;
- informações linguísticas de níveis mais altos, como o uso e padrão de marcadores discursivos, escolhas lexicais, variantes morfológicas e sintáticas, comportamento pragmático como os encontrados na tomada de turno de fala e no atendimento de ligações telefónicas, comportamentos multilíngues (como *codeswitching*);
- evidências de comprometimento de fala, de patologia de voz ou de fala e de traços não linguísticos característicos do falante (por exemplo, respiração audível, limpeza de garganta, cliques linguais e marcadores de hesitação).

Segundo Nolan (1983), os parâmetros selecionados devem preferencialmente ter alta variabilidade interfalante e baixa variabilidade intrafalante; ser resistentes à tentativa de

disfarce; ser usualmente observados, mesmo em pequenas amostras; ser robustos a diferenças na transmissão, ou seja, não variar suas propriedades se advindos, por exemplo, de gravação telefónica ou de gravação ambiental, e ser facilmente mensuráveis.

O método de análise prevalentemente empregado no desenvolvimento da CL na Perícia Oficial brasileira está em consonância com o apregoado pela comunidade científica internacional. Segundo panorama apresentado por Gold e French (2011), baseado no depoimento de 36 peritos de 13 países distintos, há predomínio da utilização do método combinado(constituído das análises perceptivo-auditiva e acústica), o qual também é referido por Byrne e Foulkes (2004), Kuwabara e Sagisaka (1995), McDougall (2005), Nolan (2001), Rodman *et al.* (2002), Romito e Galatá (2004), Rose (2002) e Watt (2010). Complementarmente, a perícia oficial de alguns estados brasileiros considera ainda resultados provenientes de sistemas de reconhecimento automático de locutor.

Na análise perceptivo-auditiva são investigados elementos vocais de caracterização geral do indivíduo, assim como os relativos à variedade linguística dos locutores confrontados. São observados fatores indicativos do sexo, da fase do ciclo de vida (se aparente infância, adolescência, fase adulta ou velhice), do estado de saúde dos órgãos fonoarticulatórios e da provável condição sociocultural e intelectual; características referentes à qualidade vocal e aos ajustes musculares utilizados na fonação; hábitos vocais típicos (pigarro, estalos, cliques, etc.); forma de articulação; presença de desvios fonéticos (distorções) e/ou fonológicos (problemas no sistema de contrastes da língua); alterações temporais, de ritmo ou fluência da fala; padrão entonacional empregado; coordenação pneumofonoarticulatória; idioleto e dialeto[2], entre outros.

Já na análise acústica (por vezes referida como "instrumental") são extraídas medidas físicas que documentam a condição e o comportamento de fatores segmentais e suprassegmentais, resultantes de configurações específicas do aparelho fonador, objetivando-se corroborar ou refutar os achados perceptivos. As informações são obtidas com a aplicação de recursos de análise disponibilizados em *softwares* de edição de áudio, sendo comumente utilizados oscilogramas (formas de onda), espectrogramas (em banda larga e estreita de frequência), curvas de formantes e de variação da f0 e espectros FFT (*Fast Fourier Transform*), LPC (*Linear Predictive Code*) e LTAS (*Long-term Average Spectrum*). Tais recursos de análise são ilustrados na Figura 1 a seguir.

No cotejo entre as amostras podem ser consideradas sentenças, palavras, sílabas ou segmentos (fones), atentando-se para que os segmentos confrontados sejam pares quanto ao acento (ao menos lexical) e imersos em ambiente fonético antecedente e seguinte maximamente análogos.

Os resultados obtidos a partir das análises realizadas são apresentados tanto qualitativamente, especialmente no que se refere ao comportamento linguístico manifesto, quanto quantitativamente, por meio de estatística descritiva, restrita normalmente à exposição das medidas extraídas durante o exame acústico e à caracterização da diferença percentual existente entre as amostras questionada e padrão, no que se refere a um determinado parâmetro quantitativo (por exemplo, a frequência dos formantes).

De uso mais recente, a razão de verossimilhança (*Likelihood Ratio*) constitui um novo paradigma na confecção e exposição dos resultados da perícia de CL (Morrison, 2009; Rose, 2002, 2006). Quanto à interpretação dos resultados, as escalas verbais são utilizadas tanto para indicação direta da conclusão, a exemplo da escala de nove pontos apresentada

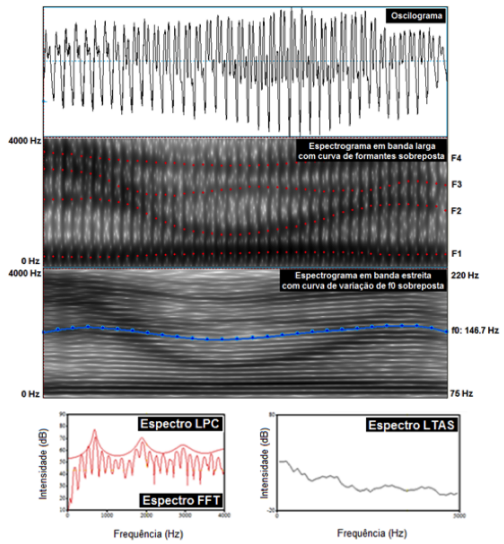

Figura 1. Recursos de análise comumente utilizados no exame acústico da CL
Fonte: Gonçalves e Petry (2014: 251).

por Eriksson (2012), a qual varia entre os extremos "os resultados suportam a hipótese (de unicidade das amostras) com quase certeza" e "os resultados contrariam a hipótese com quase certeza", quanto para o enquadramento *a posteriori* de resultados quantitativos, como no caso da razão de verossimilhança, cuja classificação verbal tem associada uma escala de escores baseada no grau de suporte à hipótese assumida.

Devido ao fato de os referidos tipos de perícia adotarem a fala espontânea como material de análise, compreendendo uma investigação que se debruça sobre a fala efetivamente produzida, de considerarem amostras de fala com características estilísticas próprias e de empregarem procedimentos de verificação acústica, este estudo admite a relevância da abordagem sociofonética à perícia de CL. Nas seções que seguem, pretende-se apresentar os argumentos que sustentam essa afirmação.

Considerações sociofonéticas pertinentes à perícia de CL

Segundo Foulkes *et al.* (2010), a Sociofonética é um campo de investigação linguística que faz uso dos princípios e técnicas da Sociolinguística e da Fonética a fim de identificar e, por fim, explicar a variação socialmente estruturada da fala. Seu escopo de atuação envolve questões referentes ao aprendizado da variação sociolinguística (a compreensão de seu armazenamento cognitivo e a avaliação subjetiva) e de seu processamento, tanto na fala quanto na percepção. Nesse sentido, considera-se como sociofonético qualquer aspecto da variação fonética sistemática na qual o fato indexado é ao menos em parte o produto da construção social.

Tal noção fundamenta-se na premissa de que os falantes ajustam-se aos contextos (sociais) através de modificações em suas produções orais (Thomas, 2011), ou seja, as línguas, por um número específico de fontes de variação, que conduzem à realização de padrões sistemáticos, possibilitam ao falante, frente às inúmeras situações de interação comunicativa a que se expõe, constantemente adaptar-se e acomodar-se.

Por ser de caráter híbrido, a Sociofonética estabelece interface com diversos campos relacionados, como a Psicolinguística, a Linguística Clínica, a Aquisição de L1 e L2, a Linguística Computacional e a Fonética Forense.

A aplicação forense da Sociofonética caracteriza-se, segundo Foulkes e Docherty (2006), pela descrição de parâmetros de variação individuais ou de grupo, assim como pela identificação das fontes adicionais de variação, entre elas, o meio de transmissão sonora e as influências externas, como estresse e drogas. Nesse sentido, Foulkes *et al.* (2010) afirmam que:

> (…) a sociofonética tem um papel central no crescimento da área de Fonética Forense, pois a compreensão da variação intrassujeito e intersujeitos é essencial para a Comparação de Locutor, na qual a voz gravada de um criminoso é comparada com a de um suspeito[3] (p. 737)

A relevância da aplicação do conhecimento sociovariacionista e do instrumental fonético-acústico no trabalho de detecção de formas próprias de falar, alvo da investigação forense, já havia sido apontada por Nolan (2001), para quem

> (…) a linguagem e a fala formam um imenso e complexo sistema plástico, e o entendimento dos meios através dos quais a identidade do falante mostra-se na fala requer uma consistente fundamentação ao menos em linguística, dialetologia, sociolinguística, fonética e acústica[4] (p. 17)

É a partir dessa perspectiva que, nas seções que se seguem, procurar-se-á identificar os aspectos relevantes do campo para a investigação forense, a partir da noção de comunidade de prática, das características da entrevista sociolinguística e da variação estilística.

Sobre a comunidade de prática

Na perspectiva sociolinguística, entende-se comunidade de fala como um grupo de falantes que compartilham um conjunto de normas e de regras referentes ao uso da linguagem. Sua composição considera fatores demográficos, como gênero, idade, raça, etnia, classe social, local de trabalho ou, ainda, observa uma população geograficamente definida (local de residência, região, etc.). Ainda, compreende o uso convergente de determinadas variáveis pelos membros da comunidade, em razão de motivação social comum, essa considerada na organização dos significados linguísticos. Assim, para Romaine (2000), uma comunidade de fala é:

> (…) um grupo de pessoas que não necessariamente compartilham a mesma língua, mas que compartilham um conjunto de normas e regras a serem consideradas no emprego da linguagem. As fronteiras entre as comunidades de fala são essencialmente sociais e não linguísticas.[5] (p. 23)

Há ao menos duas outras importantes abordagens relativas às comunidades de falantes em Sociolinguística: as redes sociais[6] e as comunidades de prática (Mullany, 2007).

A abordagem de comunidade conhecida como redes sociais (Milroy, 1987) foca as relações sociais que falantes específicos mantêm entre si, examinando como tais relações afetam o comportamento linguístico desses locutores. A unidade de referência nessa abordagem é a força da rede, avaliada como densa ou leve, em decorrência da sucessão de contatos interativos entre os membros do grupo, e como única ou múltipla, em decorrência do número de ambientes de interação existente entre os pares.

No entanto, é a comunidade que se constitui a partir do exercício de uma atividade regular conjunta, a intitulada "comunidade de prática", que melhor exprime a relação sociolinguística encontrada na realidade pericial. Segundo Eckert (2006), comunidade de prática refere-se a um grupo de pessoas que recorrentemente se envolvem em algum empreendimento comum. Para a autora, as comunidades de prática formam-se em razão de interesses ou posições sociais comuns aos falantes, os quais assumem um *modus operandi*, visões, valores, relações de poder e uso da língua próprios.

No contexto de realização da CL considera-se um grupo particular de falantes: indivíduos por alguma razão associados à prática delituosa, especificamente a relacionada ao tráfico de entorpecente e/ou a (tentativa de) homicídio, apresentando normalmente histórico de aprisionamento, seja em unidade de reabilitação socioeducativa ou em unidade prisional convencional, para cumprimento de pena ou de prisão preventiva.

Tais indivíduos solidarizam-se no emprego de comportamentos linguísticos tipicamente associados ao contexto de execução de crimes, entre eles, por exemplo, o referente ao uso de expressões rotineiramente observadas em gravações de diálogos mantidos entre criminosos, cujo significado é de entendimento restrito ao grupo de usuários de drogas e/ou traficantes e a quem os investiga ou pericia, como "partir o cara" (assassinar alguém); "tá baixado" (estar foragido); "fazer uma caminhada" (fazer algo determinado por um superior, por exemplo, uma entrega de drogas ou uma execução), "escolher um bicho" (escolher um veículo, que será furtado ou roubado para cometimento de outros delitos); "escama" significando maconha.

A prática linguística acaba por diretamente refletir a identidade que o falante constrói enquanto membro do grupo, identidade essa que guia o comportamento linguístico manifesto e que faz com que esse falante seja reconhecido como um falante típico, cujo padrão de fala é associado, no âmbito da criminalística, como sendo, por exemplo, do traficante e/ou do homicida, o que para Caldeira (2000) corresponde à "fala do crime".

A importância da linguagem verbal como um dos elementos marcadores do grupo também é apontada por Da Hora (2008) que, ao analisar a problemática social da droga, destaca a autenticidade e os modos peculiares de construção da realidade encontrados nesse contexto social e as interações locais socialmente construídas.

Sobre a entrevista sociolinguística

A entrevista sociolinguística é basicamente uma sequência dialógica, uma vez que envolve ao menos dois locutores que se alternam na produção de turnos conversacionais. Esse tipo de entrevista é projetado para angariar grande quantidade de dados de falantes em condição tão casual e natural quanto possível, embora apresente certa estruturação, obtendo do informante a produção oral na forma vernacular, definida por Labov (2008) como sendo aquela em que o mínimo de atenção é dado ao monitoramento da fala.

Em uma entrevista sociolinguística invariavelmente tem-se tanto a fala monitorada, que conforme Labov (2008) compreende uma fala moldada à presença de um observador externo, quanto a fala espontânea, equivalente à fala casual, contudo, de ocorrência reservada aos contextos formais, a partir do momento em que se têm superado os constrangimentos próprios do formalismo.

No contexto da perícia de CL, entende-se que a fala obtida através da gravação desavisada (interceptação telefónica feita sem a ciência dos locutores) é prevalentemente casual. Já a gravação avisada, provinda da coleta técnica de padrão vocal para fins de

perícia, é admitida, em razão de compreender uma entrevista semidirigida, contexto enquadrado por Labov (1972) como formal, e de se ter excertado os dois ou três primeiros minutos iniciais da conversação, como sendo prevalentemente espontânea.

Segundo Görski (2011), no tipo de entrevista sociolinguística as perguntas do entrevistador funcionam como gatilhos para a evocação de diferentes sequências textuais por parte do entrevistado. Durante o procedimento de entrevista incentivam-se as narrativas de experiência pessoal, gênero discursivo em que os informantes, por estarem emocionalmente envolvidos com a elaboração do relato, despendem menos atenção à fala (Tarallo, 1986).

As entrevistas utilizadas para levantamento de dados são normalmente monólogos curtos elaborados em resposta a perguntas genéricas do entrevistador que se esforça para interferir o menos possível nos relatos, fazendo-o somente quando há exaustão do tópico em curso, a fim de formular novo questionamento. Para Llisterri (1992), a entrevista, mesmo semidirigida, não corresponde à conversação, pois nela há violação, ao menos, das tomadas espontâneas de turno. Segundo o autor, o mais próximo de uma conversação natural seria o diálogo mantido entre o falante em observação e alguém conhecido, contexto encontrado em um dos tipos de gravação considerados na CL, a saber, o da gravação desavisada.

Para fins forenses, a gravação da entrevista com fins de confronto de voz e fala, efetuada, geralmente, por peritos em registros de áudio, não difere, em essência, da descrita entrevista sociolinguística. Preocupam-se os peritos, da mesma forma, em extrair do falante a fala mais espontânea (menos monitorada) possível, propondo-lhe tópicos de conversação emocionalmente fortes, como os relacionados ao seu histórico familiar, experiência pessoal pregressa, envolvimento no crime, condição carcerária a que está exposto (se recluso em unidade prisional), perspectiva de vida futura, entre outros.

Sobre a variação estilística

Segundo Labov (1972), a variação estilística refere-se às alternâncias através das quais um falante adapta sua linguagem ao contexto imediato do ato de fala. Conforme o autor, na pesquisa de campo, destacam-se cinco princípios fundamentais relativos ao(s) estilo(s) de fala, a saber:

- alternância de estilo: o falante altera o estilo de sua fala durante a produção oral, ou seja, não há falante de estilo único;
- atenção: os estilos de fala variam em um *continuum* que se estende da fala casual (vernacular), fala cotidiana usada em situações informais (Labov, 2008), à fala padrão (monitorada), sendo escalonado pelo grau de atenção dispensado pelo falante à própria fala, o que é associado à formalidade da situação comunicativa;
- vernáculo: o vernáculo é um estilo de fala que se destaca pela sistematicidade, sendo considerado o mais regular quanto à estrutura e na relação com a evolução da língua;
- formalidade: a observação sistemática do falante a um contexto formal, em que certo grau de atenção é dispensado à fala;
- bons dados: a entrevista individual é uma observação sistemática que pretende garantir a obtenção de dados de fala em quantidade suficiente.

Os princípios elencados conduzem a um importante paradoxo metodológico, intitulado Paradoxo do Observador, que caracteriza um problema aparentemente insolúvel:

como sistematicamente observar os falantes em momento de interação comunicativa natural sem a introdução da formalidade tipicamente associada à presença do entrevistador?

No que se refere à caracterização dos estilos de fala, Labov (1970) afirma que existem mais estilos e dimensões estilísticas do que um analista pode constatar, sendo possível, no entanto, avaliar o estilo de fala em razão do grau de atenção prestado à mesma.

O problema na investigação da variação estilística em Sociolinguística é, segundo Labov (1972), controlar adequadamente os contextos e definir os estilos que ocorrem em cada um deles. O autor aponta a existência de ao menos cinco níveis estilísticos que afetam diferentemente a fala, a saber, fala casual, fala monitorada, leitura oral, lista de palavras e pares mínimos.

Foulkes *et al.* (2010) entendem que a classificação laboviana de níveis estilísticos, baseada no grau de atenção dispensado à fala, é um tanto simplista, pois os falantes fazem ajustes fonéticos não só em razão do automonitoramento, mas também em decorrência de fatores externos, como condição física, tópico conversacional e audiência, os dois últimos também referidos por Bell (1984).

Schilling-Estes (2002) aponta como limitações da concepção de variação estilística baseada na atenção dispensada à fala as dificuldades encontradas quando se tenta separar, em um registro de entrevista, a fala casual da monitorada e as relativas à quantificação do nível de atenção prestado à fala, assim como o fato de que há estilos que não se ajustam à proposta de condicionamento em razão da formalidade e/ou da atenção. A autora ressalva ainda que, sendo admitida a existência de variações estilísticas intencionais, o nível de atenção dispensado à fala não estaria rigorosamente associado à formalidade da situação comunicativa.

A alternância de estilo é discutida por Schilling-Estes (2002, 2008), que a considera uma variação de fala de ordem intrassujeito, passível de acometer a forma de uso de língua de um grupo (dialeto), a forma prevista para uma situação de uso de língua em particular (registro) ou um gênero linguístico específico (sendo o último, para a autora, variedades rotineiras, bastante ritualizadas). Destaca-se que tal alternância pode se dar em qualquer um dos níveis de organização da língua: fonológico, morfossintático, lexical, semântico, pragmático ou discursivo.

Nessa perspectiva, os estilos de fala variam tanto na produção oral de um mesmo indivíduo (variação intrassujeito) quanto entre grupos de falantes (variação intersujeitos). Para a autora, o falante pode incorrer na alteração do estilo durante a emissão oral tanto de forma voluntária quanto inconsciente, por período temporalmente curto ou extenso e em uma escala que vai do estilo conversacional ao estilo formal.

A variação estilística tem sido caracterizada a partir de três abordagens sociolinguísticas principais, conforme ilustra a Figura 2 a seguir. A primeira, já mencionada, concernente ao grau de atenção dispensado à fala (Labov, 1972, 2001); a segunda, de cunho interacional, que privilegia o papel da audiência e considera as relações interpessoais envolvidas (intitulada *audience design*, conforme Bell, 1984, 2001); e a terceira, referente à questão da identidade social (intitulada *speaker design*, conforme Coupland, 1996).

De acordo com a Figura 2, cada abordagem de variação estilística tem associada uma fonte de variação particular, de natureza psicolinguística (função cognitiva da atenção,

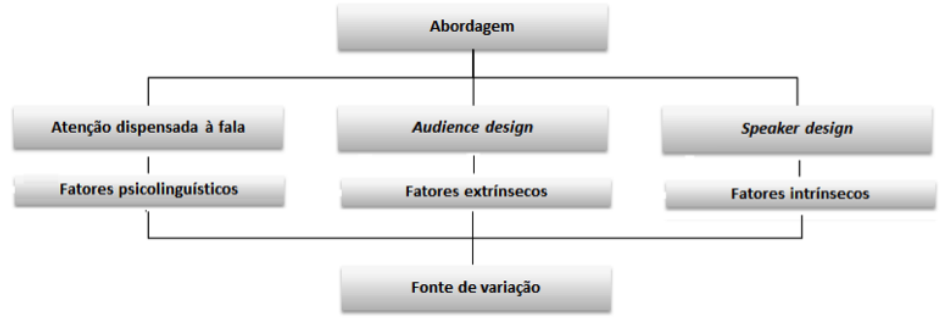

Figura 2. Variação estilística: principais abordagens e respectivas fontes de variação
Fonte: Gonçalves (2013: 54).

especialmente), extrínseca (a audiência, o tópico conversacional e o contexto situacional) ou intrínseca (a intenção comunicativa).

A diferenciação entre as duas primeiras abordagens (a baseada na atenção dispensada à fala e a projetada na audiência) decorre essencialmente das mencionadas motivações. Enquanto para Labov (1972) a variação no estilo é resultado da alteração no grau de automonitoramento da fala, para Bell (1984) refere-se à adequação do estilo de fala ao interlocutor, visando o falante a aproximação ou o distanciamento em relação aos membros da audiência.

Bell (1984, 2001) atribui as variações estilísticas a adaptações do locutor à composição e características da audiência. Segundo o autor, os falantes ajustam sua fala na intenção de expressar solidariedade (visando parecença com a(s) pessoa(s) com quem está falando) e intimidade em relação ao interlocutor (presente ou não), assim como, sendo a intenção, o distanciamento. Em tal proposta, aplicável não só à situação de entrevista, mas também a outras formas de interação conversacional (por exemplo, à conversa naturalística entre pares e colegas), o impacto dos membros constituintes da audiência sobre o resultado da fala é proporcional ao nível de consciência que o falante tem da presença desses no contexto da fala.

Já entre as duas primeiras abordagens e a terceira (projetada no falante) há, conforme Schilling-Estes (2002), uma mudança conceitual envolvendo o papel do falante na variação estilística em curso (admitido como passivo nas duas primeiras e como ativo na última) e o fato daquelas serem unidimensionais (por considerarem basicamente, nessa ordem, ou o grau de atenção dispensado à fala ou a composição/ propriedades da audiência) enquanto a última é multidimensional (contemplando um maior número de fatores como possíveis fontes de variação, entre eles, os relacionados aos níveis de organização da língua, ao tópico conversacional, ao contexto situacional, ao humor, ao meio de transmissão da linguagem, além de fatores paralinguísticos e não linguísticos). Para Schilling-Estes (2002), são pontos principais da abordagem projetada no falante o fato de esses não alternarem o estilo somente em resposta a elementos da situação de fala e de usarem suas falas para moldar e remoldar as situações externas, os relacionamentos interpessoais e, especialmente, suas identidades pessoais.

Na abordagem projetada no falante destaca-se a questão da intencionalidade (*agency*) e da prática social (Eckert, 2000). Tal perspectiva, mais condizente com o que é comumente encontrado no contexto da perícia forense de áudio, prevê que a variação

estilística não é determinada pela reatividade ao meio externo e sim provocada pelo próprio falante, não só para marcar afiliação a grupo social específico ou o atributo social desejado, mas também para significar diferentes propósitos conversacionais na interação em curso (Schilling-Estes, 2002).

Com o exposto, depreende-se que os estudos variacionistas, no que se refere ao entendimento acerca das variações no estilo de fala, evoluíram da ideia de adaptação de variáveis individuais à formalidade da situação comunicativa à noção de uso de modos distintos de fala, vinculados à identidade social e à intencionalidade do discurso (Camacho, 2010).

Na CL aborda-se a questão dos estilos de fala em razão de serem consideradas gravações realizadas em situações comunicativas distintas, por exemplo, conversação ao telefone celular entre um locutor-alvo e algum indivíduo conhecido, sem que ambos tivessem, à ocasião, ciência de que estavam sendo gravados, e diálogo presencial com interlocutor até então desconhecido do locutor-alvo, que transcorre com o consentimento dos presentes, em ambiente onde constam visíveis os equipamentos utilizados na gravação. É prevalente, portanto, o confronto de amostras de fala produzidas em situações comunicativas distintas, sendo comuns em laudos periciais ressalvas acerca da possível interferência, entre outros, das propriedades do contexto situacional e do estado emocional do locutor em relação ao tópico de conversação desenvolvido. Assim, entende-se como mais adequada à prática forense a proposta de Schilling-Estes (2008), concernente à intencionalidade e à multidimensionalidade, do que as propostas baseadas na atenção dispensada à fala e projetadas na audiência, que apregoam a reatividade e a unidimensionalidade.

Parâmetros técnico-comparativos utilizados na CL

Na análise contrastiva das amostras de fala são apontados, entre outros, elementos que caracterizam, ressalvada a costumeira limitação quantitativa especialmente do áudio questionado, o comportamento linguístico dos locutores confrontados.

Uma tentativa de formalizar a observação sistemática do material de fala compreende o Protocolo Forense para Análise Perceptivo-Auditiva de Amostras de Fala (Gonçalves e Petry, 2014), integralmente reproduzido no Anexo 1. O referido protocolo, em seu 2º bloco, intitulado "Parâmetros de Fala", relaciona como itens de investigação os que seguem:

- organização do raciocínio (coerência, manutenção do tema, etc.);
- continuidade (quantidade e distribuição das pausas silenciosas e preenchidas, assim como manifestações de disfluência não patológica);
- prosódia (acento, entoação e ritmo);
- tempo de fala (estimado a partir do cálculo da taxa de articulação e/ou taxa de elocução);
- léxico (presença de item delator, compatibilidade com o nível de instrução, uso de recursos de apoio discursivo, uso de linguagem de grupo, gírias, termos regionais e/ou de formas de baixo prestígio, presença de itens lexicalizados, etc.);
- forma de referência ao interlocutor (expressões de tratamento, tomadas de turno, forma de anuência, etc.);
- distanciamento das produções em relação à norma culta (erros de concordância, construções sintáticas irregulares ou atípicas, etc.);

- tipo de articulação (no que se refere a precisão da mesma, independente de quais sejam as variantes linguísticas eleitas);
- extensão de articuladores (capacidade de movimentação dos lábios, da ponta/lâmina/corpo da língua e da mandíbula);
- estado particular dos articuladores (se se destaca a ação, em contextos não previstos, de lábios, ponta/lâmina/corpo de língua e/ou mandíbula);
- desvio de fala fonético, fonológico e/ou de fluência;
- aplicação de processos fonético-fonológicos de variação linguística (supressões, inserções, modificações, transposições e/ou processos envolvendo acento);
- outros elementos linguísticos destacáveis (disfarce, imitação, simulação, etc.).

Dessa forma, com relação especificamente à fala, são observados, entre outros, fatores caracterizadores do idioleto dos sujeitos e do provável dialeto a que pertencem. Tais fatores são confrontados a fim de que se possa afirmar se são ou não páreos entre si.

Reitera-se que a análise acústica é utilizada no sentido de objetivamente documentar os fenómenos segmentais e suprassegmentais específicos, que ilustram a paridade/disparidade encontrada, sendo preferível o emprego de um número significativo de tokens de um dado evento (por exemplo, frequência dos formantes vocálicos) e medidas de longo termo (por exemplo, f0 habitual a partir de recorte de sinal de áudio superior a 60 segundos). O objetivo tenderá a ser normalmente o de corroborar ou refutar os achados perceptivos.

Considerações finais

O presente artigo procurou identificar os aspectos que aproximam a pesquisa científica em Linguística e a aplicação forense. Nele preocupou-se em discorrer sobre as questões linguísticas que perpassam a geração do material de fala coletado a ser utilizado como padrão no confronto de voz/fala, as possíveis interferências dos estilos de fala observados nas amostras, as especificidades da fala produzida por uma comunidade de prática típica e os elementos linguísticos passíveis de utilização no confronto da perícia de CL.

Espera-se ter, com as informações expostas, contribuído para a reflexão acerca da pertinência da utilização da descrição linguística na CL, bem como justificado a importância do desenvolvimento de estudos sociolinguísticos que apontem especificidades relativas à linguagem oral empregada por comunidades de prática alvos desse tipo de perícia.

Notas

[1]Um detalhado panorama do desenvolvimento relativo à Fonética Forense no Brasil e em outros países é apresentado por Gomes e Carneiro (2014), que também informam sobre impasses terminológicos e grupos de pesquisas atuantes na área.

[2]*Dialeto* é entendido aqui como "The pronunciation, lexis and grammar of a language variety, associated with a particular geographical area or social group." Llamas *et al.* (2007: 211)

[3]"(…) sociophonetics plays a central role in the growing field of forensic phonetics. Understanding cross-speaker and within-speaker variation is essential in the process of speaker comparison, in which the recorded voice of a criminal is compared with that of a suspect."

[4]"Language and speech form an immensely complex and plastic system, and understanding the ways in which a speaker's identity imprints itself on a sample of speech requires a firm foundation in linguistics, dialectology, sociolinguistics, phonetics, and acoustics, at the very least."

[5]"(…) is a group of people who do not necessarily share the same language, but share a set of forms and rules for the use of language. The boundaries between speech communities are essentially social rather than linguistic."

[6]*Redes sociais* referem-se neste texto aos laços sociais estabelecidos entre falantes e que afetam, em graus variados, o uso da língua por esses falantes.

Referências

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

Bell, A. (2001). Back in style: reworking audience design. In P. Eckert e J. Rickford, Orgs., *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.

Braid, A. (2003). *Fonética Forense*. Campinas: Millennium, 2 ed.

Byrne, C. e Foulkes, P. (2004). The 'mobile phone effect' on vowel formants. *The International Journal of Speech, Language and the Law*, 11(1), 83–102.

Caldeira, T. (2000). *A cidade dos muros*. São Paulo: Edusp.

Camacho, R. (2010). Uma reflexão crítica sobre a teoria sociolinguística. *D.E.L.T.A.*, 26(1), 141–162.

Coupland, N. (1996). Language, situation, and the relational self: Theorising dialect-style in sociolinguistics. In *Paper presented at Stanford Workshop on Stylistic Variation*, Stanford.

Da Hora, L. (2008). *Nem repressão nem educação: uma droga de cenário*. Tese de doutoramento, UFRJ – Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Eckert, P. (2000). *Linguistic Variation as Social Practice*. Malden: Blackwell.

Eckert, P. (2006). Communities of practice.

Eriksson, A. (2012). Aural/ Acoustical vs. Automatic Methods in Forensic Phonetic case Work. In A. Neustein e H. Patil, Orgs., *Forensic Speaker Regognition: Law Enforcement and Counter-terrorism*. New York: Springer-Werlag.

Foulkes, P. e Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34, 409–438.

Foulkes, P., Scobbie, J. e Watt, D. (2010). Sociophonetics. In W. Hardcastle, J. Laver e F. Gibbon, Orgs., *The handbook of Phonetic Sciences*. Oxford: Wiley-Blackwell, 2 ed.

French, P., Nolan, F., Foulkes, P., Harrison, P. e McDougall, K. (2010). The UK position statement on forensic speaker comparison: a rejoinder to Rose and Morrison. *The International Journal of Speech, Language and the Law*, 17(1), 143–152.

Gold, E. e French, P. (2011). International practices in forensic speaker comparison. *International Journal of Speech, Language and the Law*, 18(2), 293–307.

Gomes, M. e Carneiro, D. (2014). A fonética forense no Brasil: cenários e atores. *Language end Law/ Linguagem e Direito*, 1(1), 22–36.

Gonçalves, C. (2013). *Taxa de elocução e de articulação em corpus forense do português brasileiro*. Tese de doutoramento, PUCRS, Porto Alegre.

Gonçalves, C. e Petry, T. (2014). Comparação Forense de Locutores no Âmbito da Perícia Oficial dos Estados. In M. Rehder, L. Cazumba e M. Cazumba, Orgs., *Identificação de Falantes: Uma Introdução à Fonoaudiologia Forense*, chapter 15, 241–264. Rio de Janeiro: Revinter.

Görski, E. (2011). A variação estilística na ótica da sociolinguística laboviana: (re)dimensionando o papel do contexto.

Hollien, H. (2002). *Forensic Voice Identification*. London: Academic Press.

Kuwabara, H. e Sagisaka, Y. (1995). Acoustic characteristics of speaker individuality: Control and conversion. *Speech Communication*, 16, 165–173.

Labov, W. (1970). The Study of Language in Its Social Context. *Studium Generale*, 23, 30–87.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (2001). The anatomy of style-shifting. In P. Eckert e J. Rickford, Orgs., *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.

Labov, W. (2008). *Padrões sociolinguísticos*. São Paulo: Parábola.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Llamas, C., Mullany, L. e Stockwell, P. (2007). *The Routledge companion to sociolinguistics*. London: Routledge.

Llisterri, J. (1992). Speaking Styles in Speech Research. In *ELSNET/ ESCA/ SALT, Workshop on Integrating Speech and Natural Language, Dublin*, 1–28, Dublin.

McDougall, K. (2005). *The Role of Formant Dynamics in Determining Speaker Identity*. Phd dissertation, University of Cambridge, Cambridge.

McMenamin, G. (2002). *Forensic Linguistics: Advances in Forensic Stylistics*. New York: CRC Press.

Milroy, L. (1987). *Language and Social Networks*. Oxford: Blackwell, 2 ed.

Morrison, A., Sampaio, J. e Ribeiro, J. (2009). Exames de registro de áudio e imagens: recomendações técnicas para a padronização de procedimentos e metodologias. In D. Tochetto e A. Espindula, Orgs., *Criminalística: Procedimentos e Metodologias*. Porto Alegre: s.n., 2 ed.

Morrison, G. (2009). Forensic voice comparation and the paradigm shift. *Science and Justice*, 49, 298–308.

Mullany, L. (2007). Speech communities. In C. Llamas, L. Mullany e P. Stockwell, Orgs., *The Routledge companion to sociolinguistics*. London: Routledge.

Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

Nolan, F. (2001). Speaker Identification Evidence: Its Forms, Limitations, and Roles. In *The conference "Law and language: Prospect and retrospect"*, 12–15, Levi.

Rodman, R., McAllister, D., Bitzer, D., Cepeda, L. e Abbitt, P. (2002). Forensic speaker identification based on spectral moments. *Forensic Linguistics*, 9(1), 22–43.

Romaine, S. (2000). *Language in society: an introduction to sociolinguistics*. London: Blackwell.

Romito, L. e Galatá, V. (2004). Towards a protocol in speaker recognition analysis. *Forensic Science International*, 146, S107–S111.

Rose, P. (2002). *Forensic Speaker Identification*. London: Taylor & Francis.

Rose, P. (2006). Technical forensic speaker recognition: Evaluation, types and testing of evidence. *Computer Speech & Language*, 20(2-3), 159–191.

Schilling-Estes, N. (2002). Investigating stylistic variation. In J. Chambers, P. Trudgill e N. Sschilling-Estes, Orgs., *Handbook of Language Variation and Change*. Oxford: Blackwell Publishing Ltd.

Schilling-Estes, N. (2008). Stylistic variation and the sociolinguistic interview: a reconsideration. In R. Monroy e A. Sánchez, Orgs., *25 Años de Lingüística Aplicada en España: Hitos y Retos (25 Years of Applied Linguistics in Spain: Milestones and Challenges; proceedings from AESLA 25)*, Murcia, Spain: Servicio de Publicaciones de la Universidad de Murcia.

Tarallo, F. (1986). *A pesquisa socio-lingüística*. São Paulo: Editora Ática, 2 ed.
Thomas, E. (2011). *Sociophonetics: an introduction*. New York: Palgrave Macmillian.
Watt, D. (2010). The identification of the individual through speech. In C. Llamas e D. Watt, Orgs., *Language and Identities*. Edinburgh: Edinburgh University Press.

Anexo 1

Protocolo Forense para Análise Perceptivo-Auditiva de Amostras de Fala

Protocolo/ Solicitante: _____/ _____ **Data do Quest.(Q):** _____/ _____/ _____

Locutor do padrão: _____ **Data do Padrão (P):** _____/ _____/ _____

Idade na coleta de padrão: _____ ***Gap* temporal entre Q e P:** _____

Escolaridade: _____ **Tempo de fala exclusiva Q/P:** _____

AMOSTRA QUESTIONADA	AMOSTRA PADRÃO
Parâmetros de Voz	
1. Caracterização geral provável do indivíduo	
Sexo ( ) M ( ) F Faixa etária: ( ) adolescente ( ) adulto jovem ( ) adulto ( ) idoso	Sexo: ( ) M ( ) F Faixa etária: ( ) adolescente ( ) adulto jovem ( ) adulto ( ) idoso
2. Tipo de voz	
( ) eufônica ( ) disfônica ( ) rugosa ( ) soprosa ( ) tensa ( ) outros: _____	( ) eufônica ( ) disfônica ( ) rugosa ( ) soprosa ( ) tensa ( ) outros: _____
3. Elementos fonatórios	
Modo de fonação: ( ) modal ( ) falsete ( ) crepitância/ vocal fry ( ) voz crepitante Fricção laríngea: ( ) SED - Sem Elementos Destacáveis ( ) escape de ar ( ) voz soprosa Irregularidade laríngea: ( ) SED ( ) voz áspera Ocorrências de curto-termo: ( ) SED ( ) quebras ( ) instabilidades ( ) diplofonia ( ) tremor	Modo de fonação: ( ) modal ( ) falsete ( ) crepitância/ vocal fry ( ) voz crepitante Fricção laríngea: ( ) SED - Sem Elementos Destacáveis ( ) escape de ar ( ) voz soprosa Irregularidade laríngea: ( ) SED ( ) voz áspera Ocorrências de curto-termo: ( ) SED ( ) quebras ( ) instabilidades ( ) diplofonia ( ) tremor
4. Tensão muscular	
Do trato vocal: ( ) SED ( ) hiperfunção ( ) hipofunção Laríngea: ( ) SED ( ) hiperfunção ( ) hipofunção	Do trato vocal: ( ) SED ( ) hiperfunção ( ) hipofunção Laríngea: ( ) SED ( ) hiperfunção ( ) hipofunção
5. Respiração	
( ) não evidente ( ) evidente	( ) não evidente ( ) evidente

( ) com inspiração ruidosa
( ) com bloqueio
( ) profunda
( ) com reposição súbita
( ) com reposição irregular
( ) com uso do ar de reserva
( ) outros: _____

( ) incoordenada durante a fala

( ) com inspiração ruidosa
( ) com bloqueio
( ) profunda
( ) com reposição súbita
( ) com reposição irregular
( ) com uso do ar de reserva
( ) outros: _____

( ) incoordenada durante a fala

6. Tipo de ressonância

( ) equilibrada
( ) com foco predominante
- ( ) hipernasal
- ( ) hiponasal
- ( ) oral
- ( ) faríngeo
  - ( ) por constrição
  - ( ) por expansão
- ( ) laringofaríngeo

( ) com escape de ar nasal audível

( ) equilibrada
( ) com foco predominante
- ( ) hipernasal
- ( ) hiponasal
- ( ) oral
- ( ) faríngeo
  - ( ) por constrição
  - ( ) por expansão
- ( ) laringofaríngeo

( ) com escape de ar nasal audível

7. Pitch

Habitual:
- ( ) SED
- ( ) elevado
- ( ) abaixado

Extensão:
- ( ) SED
- ( ) diminuída
- ( ) aumentada

Variabilidade:
- ( ) SED
- ( ) diminuída
- ( ) aumentada

Habitual:
- ( ) SED
- ( ) elevado
- ( ) abaixado

Extensão:
- ( ) SED
- ( ) diminuída
- ( ) aumentada

Variabilidade:
- ( ) SED
- ( ) diminuída
- ( ) aumentada

8. Loudness

Habitual:
- ( ) SED
- ( ) diminuído
- ( ) aumentado

Extensão:
- ( ) SED
- ( ) diminuída
- ( ) aumentada

Variabilidade:
- ( ) SED
- ( ) diminuída
- ( ) aumentada

Habitual:
- ( ) SED
- ( ) diminuído
- ( ) aumentado

Extensão:
- ( ) SED
- ( ) diminuída
- ( ) aumentada

Variabilidade:
- ( ) SED
- ( ) diminuída
- ( ) aumentada

9. Psicodinâmica vocal **(estado físico ou emocional, discrepância de gênero e/ou compleição, alteração de muda vocal, etc)**

( ) SED
Obs.:

( ) SED
Obs.:

10. Elementos vocais intervenientes **(estalo comunicativo, pigarro não produtivo, clique velar, clique labial, clique nasal, etc)**

( ) SED
Obs.:

( ) SED
Obs.:

11. Outros elementos vocais destacáveis (risada, tosse, etc)

( ) SED
Obs.:

( ) SED
Obs.:

Parâmetros de Fala

1. Organização do raciocínio **(coerência, manutenção do tema, etc)**

( ) SED
Obs.:

( ) SED
Obs.:

2. Continuidade

Pausas silenciosas e preenchidas (quant. e distribuição):
( ) SED
Obs.:

Manifestações de disfluência não patológica:
( ) SED
Obs.:

Pausas silenciosas e preenchidas (quant. e distribuição):
( ) SED
Obs.:

Manifestações de disfluência não patológica:
( ) SED
Obs.:

3. Prosódia **(acento, entoação e ritmo)**

( ) SED
Obs.:

( ) SED
Obs.:

4. Tempo de fala

Taxa de ( ) articulação ( ) elocução:
ref.TA local média=6,20 (±0,5) síl/s
ref.TE local média=5,47 (±0,7) síl/s
( ) normal ( ) lenta ( ) rápida
Obs.:

Taxa de ( ) articulação ( ) elocução:
ref.TA local média=6,20 (±0,5) síl/s
ref.TE local média=5,47 (±0,7) síl/s
( ) normal ( ) lenta ( ) rápida
Obs.:

5. Léxico **(item delator; compatibilidade com o nível de instrução; uso de RADs; uso de linguagem de grupo, gírias, termos regionais e/ou de formas de baixo prestígio; presença de itens lexicalizados, etc)**

( ) SED
Obs.:

( ) SED
Obs.:

6. Referência ao interlocutor **(expressões de tratamento, tomadas de turno, forma de anuência, etc)**

( ) SED
Obs.:

( ) SED
Obs.:

7. Distanciamento em relação à norma culta **(erros de concordância, construções sintáticas irregulares ou atípicas, etc)**

( ) SED
Obs.:

( ) SED
Obs.:

8. Tipo de articulação

( ) precisa ( ) imprecisa

Extensão de articuladores:
( ) SED
( ) aumentada
( ) diminuída
( ) lábio
( ) ponta/lâmina de língua
( ) corpo de língua
( ) mandíbula

( ) precisa ( ) imprecisa

Extensão de articuladores:
( ) SED
( ) aumentada
( ) diminuída
( ) lábio
( ) ponta/lâmina de língua
( ) corpo de língua
( ) mandíbula

9. Estado particular dos articuladores **(em contextos não previstos)**

Lábios:	Lábios:
( ) SED	( ) SED
( ) arredondados/ protraídos	( ) arredondados/ protraídos
( ) estirados	( ) estirados
( ) labiodentalizando	( ) labiodentalizando
Mandíbula:	Mandíbula:
( ) SED	( ) SED
( ) protraída	( ) protraída
( ) com excursão lateral acentuada	( ) com excursão lateral acentuada
Ponta da língua:	Ponta da língua:
( ) SED	( ) SED
( ) avançada	( ) avançada
( ) recuada	( ) recuada
Corpo de língua:	Corpo de língua:
( ) SED	( ) SED
( ) avançado	( ) avançado
( ) recuado	( ) recuado
( ) elevado	( ) elevado
( ) abaixado	( ) abaixado
( ) lateralmente interposta	( ) lateralmente interposta

10. Desvios de fala

( ) SED	( ) SED
( ) fonéticos	( ) fonéticos
( ) imprecisão de alveolares	( ) imprecisão de alveolares
( ) dorsalização de /r/	( ) dorsalização de /r/
( ) ceceio anterior ou lateral	( ) ceceio anterior ou lateral
( ) outro	( ) outro
( ) fonológicos	( ) fonológicos
( ) apagamentos	( ) apagamentos
( ) substituições	( ) substituições
( ) inserções	( ) inserções
( ) transposições	( ) transposições
( ) da fluência	( ) da fluência
( ) bloqueios	( ) bloqueios
( ) alongamentos	( ) alongamentos
( ) falsos começos	( ) falsos começos
( ) repetições	( ) repetições

11. Aplicação de processos fonético-fonológicos de variação linguística

Supressões:	Supressões:

Inserções:	Inserções:
Modificações:	Modificações:
Transposições:	Transposições:
Processos envolvendo acento:	Processos envolvendo acento:

12. Outros elementos linguísticos destacáveis **(disfarce, imitação, simulação, etc)**

( ) SED Obs.:	( ) SED Obs.:

OBSERVAÇÕES ADICIONAIS:

FECHAMENTO DOS RESULTADOS DA ANÁLISE PERCEPTIVO-AUDITIVA:
(grau de suporte/contradição da hipótese de mesma origem)

( ) +4, o resultado suporta muito fortemente a hipótese
( ) +3, o resultado suporta fortemente a hipótese
( ) +2, o resultado suporta moderadamente a hipótese
( ) +1, o resultado suporta levemente a hipótese
( ) 0, o resultado nem suporta nem contradiz a hipótese
( ) -1, o resultado contradiz levemente a hipótese
( ) -2, o resultado contradiz moderadamente a hipótese
( ) -3, o resultado contradiz fortemente a hipótese
( ) -4, o resultado contradiz muito fortemente a hipótese

Data (mês/ ano): _____/ _____

Perito(a): _____

Fonoaudiologia: Contribuições nos estudos forenses de comparação de locutores

Paloma Alves Miquilussi, Marilisa Exter Koslovski & Denise de Oliveira Carneiro

Universidade Tuiuti do Paraná, Estado do Paraná
& Universidade Tecnológica Federal do Paraná

Abstract. *This research was motivated by the growing demand for speech-pathologists to act as forensic experts, particularly in cases of speaker comparison. A literature search discovers very little about speech-language pathologists working as forensic experts. Thus, the search was widened to include the science of Speech-Language Pathology, the national curricula for Bachelor degrees in Speech-Language Pathology and the work of forensic experts in the area of speaker comparison. The results showed that human communication is one of the objects of study of the speech-language pathologist, and that work in forensic speaker comparison uses parameters standardly analyzed by speech-language pathologists and that forensic expertise in this area is enriched through multidisciplinary. Hence, it follows that the inclusion of speech-language pathologists in multidisciplinary teams working on forensic speaker comparison is amply justifiable.*

Keywords: *Speech-Language Science, Forensic Speaker Recognition, Forensic Speaker Comparison, Forensic Studies.*

Resumo. *A elaboração deste trabalho foi motivada pela demanda em crescimento dos profissionais fonoaudiólogos atuando como peritos, em particular na área de comparação de locutores. A revisão bibliográfica sobre fonoaudiologia no âmbito da perícia na área de comparação de locutores mostrou-se escassa. Assim, buscou-se por referências sobre a Fonoaudiologia enquanto ciência, sobre as diretrizes curriculares nacionais dos cursos de graduação em Fonoaudiologia e sobre a perícia forense na área de comparação de locutores. O resultado dos estudos apontou que um dos objetos de estudo da Fonoaudiologia é a comunicação humana, que o exame de comparação de locutores utiliza-se em grande parte de parâmetros analisados por fonoaudiólogos e que a perícia forense nesta área beneficia-se da multidisciplinaridade. Assim, a inserção do fonoaudiólogo em equipes multidisciplinares de estudos forenses de comparação de locutor seria amplamente justificável.*

Palavras-chave: *Fonoaudiologia, Verificação de Locutor, Comparação de Locutores, Estudos Forenses.*

Introdução

Como ciência relativamente nova (cuja regulamentação profissional ocorreu somente na década de 80), a Fonoaudiologia evoluiu da mera aplicação de técnicas simples a um profundo conhecimento da comunicação humana e seus distúrbios, dentre outras áreas de atuação e pesquisa.

Atualmente, até mesmo áreas não relacionadas à comunicação, mas que de alguma forma têm relação estreita com os órgãos produtores da fala, são objetos de estudo da Fonoaudiologia.

Em muitos casos, para que os operadores do direito e autoridades judiciárias possam atuar com propriedade em assunto do qual não sejam especialistas, são designados profissionais denominados peritos. Assim, sendo a Fonoaudiologia uma profissão regulamentada, os profissionais foram naturalmente inseridos neste cenário.

Dentre as várias formas de atuação que esta ciência pode abarcar no âmbito forense, pode-se destacar o estudo forense de comparação de locutores, cujo resultado é o exame que visa apontar se a autoria de uma determinada fala pertence ou não a um determinado indivíduo.

Neste exame, são analisadas características linguísticas, vocais e articulatórias que fazem parte dos conhecimentos de um fonoaudiólogo. No entanto, estaria o fonoaudiólogo realmente apto a realizar tal exame?

Este estudo busca, por meio da revisão da literatura, responder a tal problemática, pesquisando qual a formação acadêmica desses profissionais, como se desenvolve, resumidamente, um exame forense de comparação de locutores e em que termos encontra-se a atuação do fonoaudiólogo especificamente nesta área.

A formação acadêmica do fonoaudiólogo parece englobar aspectos bastante analisados no exame de comparação de locutores, como a qualidade vocal, a prosódia, a análise acústica e articulatória, entre outros.

A revisão bibliográfica sobre Fonoaudiologia no âmbito da perícia na área de comparação de locutores mostrou-se bastante escassa. Assim, buscou-se por referências sobre a Fonoaudiologia enquanto ciência, sobre as diretrizes curriculares dos cursos de graduação em Fonoaudiologia e sobre a perícia forense, a fim de encontrarmos aspectos em comum que estas áreas possuam, tornando assim justificável a inserção do profissional fonoaudiólogo em equipes multidisciplinares de estudos forenses de comparação de locutor.

Revisão de Literatura

Perícia Forense

Perícia pode ser definida como um exame de algo ou alguém, realizado por técnicos ou especialistas em determinados assuntos, podendo fazer afirmações ou extrair conclusões pertinentes a um determinado processo (Rodrigues *et al.*, 2010).

A perícia forense é fundamental para a materialização de provas. Silva (2010: 13) afirma que "A prova pericial, sob o aspecto objetivo, é meio pelo qual a verdade chega ao espírito de quem aprecia, sendo o de demonstração da verdade dos fatos sobre os

quais versa a ação", expressando, em seu entendimento, a relevância da perícia no âmbito forense.

Da mesma forma, mas em outra visão: "perícia é a capacidade teórica e prática para empregar, com talento, determinado campo do conhecimento, alcançando sempre os mesmos resultados" (Alcântara, 2006: 3).

> A prova pericial é definida como sendo uma prova técnica, pois, representa algo que se objetiva certificar acerca da existência de fatos, a partir de conhecimentos específicos. Menciona-se, ainda, que a prova pericial através de sua materialização instrumental, isto é, do laudo pericial, demonstra a peculiaridade de ser uma função estatal destinada a fornecer dados instrutórios (Dias, 2010: 2).

A perícia envolve, portanto, análises que exigem conhecimentos específicos para a sua realização, materializando a prova pericial, por meio do laudo.

"O laudo pericial é essencial para que os promotores de justiça peçam o arquivamento do inquérito policial, ou ofereçam denúncia contra alguém, pois a denúncia depende da prova de que o crime existiu (materialidade) e indícios de autoria" (Rodrigues *et al.*, 2010: 848).

Dorea *et al.* (2005), *apud* Schaefer *et al.*, 2012: 229, refere que "os profissionais designados para as atividades periciais necessitam de conhecimentos técnico-científicos especializados que os possibilitem compreender e distinguir os fatos investigados".

Rodrigues *et al.* (2010: 846) refere que a perícia criminal "deve ser realizada por perito oficial (vinculado ao Estado), que ao final emite um laudo pericial".

"Na ausência de *perito oficial*, ou se a instituição pública não dispuser de serviço próprio para o exame que se pretende realizar, o juiz poderá nomear duas pessoas idôneas de nível superior para a realização da perícia" (Del-Campo, 2008: 48).

Outra figura de fundamental relevância para que seja respeitada a ampla defesa e o contraditório é a do assistente técnico, introduzida recentemente no Código de Processo Penal Brasileiro, embora já fosse admitida sua atuação na área cível. O assistente técnico, também dotado de conhecimentos técnico-científicos sobre determinada área, pode emitir pareceres técnicos que se contrapõem aos laudos oficiais (Brasil, 2008). Concluindo a atuação pericial de forma geral e sua importância:

> a competência profissional é o recurso principal, enquanto que os artefatos tecnológicos são auxiliares e servem de suporte à competência no processo de produção do serviço e para entregar valor. O perito não é mero expectador, nem coadjuvante da tecnologia; é ele quem dirige a sua aplicação no caso concreto (Rodrigues *et al.*, 2010: 854).

Exame de comparação de locutores

Em relação à perícia de comparação de locutores, não é incomum que uma gravação de áudio, seja ela ambiental ou produto de uma interceptação telefônica legal, seja a única evidência possível de se tornar uma prova para a resolução de uma situação levada à justiça, seja esta um crime, infração penal ou outra. Muitas vezes, o único elemento disponível em uma interceptação telefônica seria a identificação do falante através de técnicas confiáveis e reconhecidas (Louis, 2000).

Essa identificação do falante envolvendo conhecimentos técnico-científicos almejando resultados fidedignos é exatamente do que se trata a perícia em questão no pre-

sente trabalho: o exame de comparação de locutores. Apresentado sob diferentes nomenclaturas, o exame recebe diversas denominações, de acordo com cada autor, tais como verificação de locutor, identificação de falante e comparação de locutores.

Scatena (2010: 12–13) traz um exemplo disso ao citar que o exame de Verificação de Locutor "Investiga se as falas gravadas em uma mídia (fita K7, CD, DVD, VHS), provêm ou não do aparelho fonador de uma pessoa em questão". Ainda reafirma a importância desse exame quando não se tem outros tipos de provas a serem utilizadas e ainda fundamenta a base de como esse procedimento é realizado:

> Muitas vezes nos processos de investigação policial, a única maneira de atribuir a autoria de um crime ou desvincular uma pessoa dele é determinar se a voz contida em uma mídia é ou não da pessoa em questão (...). Este tipo de perícia é feita por meio de comparação entre dois arquivos de voz levando em conta vários parâmetros acústicos e várias realizações articulatórias do falante. O resultado da investigação dá origem a um laudo técnico, que apresentado por perito qualificado é considerado como prova material (p. 12).

Ao longo do tempo, a nomenclatura deste exame teve algumas variações, conforme anteriormente já citado, por isso encontramos com frequência a terminologia "Verificação de Locutores", entre outras.

Braid (2003: 96) é um autor que usa a terminologia supracitada e reafirma a metodologia do exame ao definir a Verificação de Locutores como "um procedimento em que são comparadas duas falas, com o objetivo de determinar se ambas foram produzidas pelo mesmo falante".

Gomes *et al.* (2012: 8), da mesma forma, afirmam a importância desse exame na área forense dizendo que ele "(...) contempla diversas fases e é primordial que se parta da garantia da legalidade do material questionado encaminhado à perícia". Na visão de Morisson (2003: 19), o exame de Verificação de Locutor: "é o braço da fonética forense que busca determinar se as falas armazenadas em uma mídia de gravação provêm ou não do aparelho fonador de determinada pessoa".

O perito que atua nessa área deve ficar atento a todas as características presentes no material a ser analisado, sabendo reconhecê-las e explorar suas significâncias nos determinados contextos.

Para atingir o objetivo desse exame acerca da conclusão sobre a autoria de uma voz, o perito recorre a uma composição de análises técnicas sobre o comportamento vocal (Braid, 2003). Nesse sentido, são levantadas diversas características do falante alvo pericial, a partir da realização de um complexo estudo em torno do perfil de voz e fala do locutor examinado. Nessa busca pela identidade a partir da voz ouvida, qualquer aspecto da fala pode ser decisivo, dependendo de sua particularidade.

Múltiplos fatores constituem a identidade de um sujeito. Na sua dimensão de um ser falante, que tanto produz sons de acordo com a configuração do seu trato vocal, como emite traços característicos de regiões em que frequenta, há que se ponderar cada um desses aspectos. Dessa forma, cada pessoa deve ser considerada não somente sob a ótica individual, mas também como um ser social (Braid, 2003).

É possível, por exemplo, caracterizar fonoaudiologicamente um locutor, considerando suas condições relacionadas desde à qualidade vocal até à sua psicodinâmica vocal. Essa identidade do falante é composta por diversos fatores, sendo que a Fonoaudiologia

pode interpretá-la à luz de vários aspectos, como a caracterização linguística e vocal do indivíduo, considerando sotaque, pronúncia, articulação, recursos figurativos de linguagem, velocidade e ritmo de fala, frequência, intensidade, ressonância, frequência fundamental, entre outros (Buriti e Batsita, 2009).

Etapas do exame de comparação de locutores

Para se dar início a uma comparação de locutores são necessários pelo menos dois materiais de áudio a serem analisados: um material questionado (gravação com locução de autoria desconhecida, produto de uma interceptação, por exemplo) e um material padrão (geralmente obtido por meio de coleta de voz, doravante pormenorizada). Para isso é preciso que estes materiais estejam em condições que permitam a realização do exame de forma fidedigna.

Segundo Braid (2003: 96),"os métodos a serem utilizados e os tipos de análises a serem efetuadas irão depender das condições e características do material questionado e, também, do padrão a ser comparado". As condições que permitem a realização do exame referem-se a determinados requisitos a que os materiais de áudio devem atender, os quais são citados por Tonaco (2003): autenticidade (origem certa), contemporaneidade (padrão e questionado devem ser contemporâneos entre si), adequabilidade (repertório contendo locuções análogas entre os áudios, bem como fala natural) e quantidade (registros suficientemente numerosos).

Ao discorrer sobre o material questionado, a autora ressalta a importância do mesmo apresentar "qualidade técnica adequada à realização do confronto" (Tonaco, 2003: 24). Nesse sentido, ela destaca que há fatores que podem dificultar ou impedir a determinação e obtenção de elementos preciosos ao exame. Tais fatores contemplam desde o tipo inadequado de mídia ou a falta de ajuste do equipamento gravador, até seu posicionamento ou acondicionamento inadequado durante a gravação, ou ainda o ambiente em que ocorreu a gravação (com reverberação ou ruídos).

Figueiredo (1994: 13) refere isso em seu trabalho, citando que "fatores como a qualidade da gravação, duração do material gravado, marcas particulares de um determinado falante, etc, fazem com que cada caso deva ser examinado à luz de suas próprias características".

Ainda analisando esse contexto, o mesmo autor refere que, por mais experimentado que seja o profissional responsável por essa análise, o resultado satisfatório será compatível às condições dos materiais por ele recebido (Figueiredo, 1994).

Cumprindo os requisitos de adequabilidade, o perito realizará minuciosa análise das duas ou mais amostras, a fim de verificar se as falas contidas nos materiais de áudio foram emitidas pelo mesmo falante. Para isso, o perito necessita de grande habilidade, treinamento na área e equipamentos que o auxiliem efetivamente para que ele consiga realizar esse procedimento da forma mais confiável possível (Braid, 2003).

Esse mesmo autor descreve em seu livro duas etapas importantes desse exame, que são as análises perceptual e acústica. Segundo ele, a análise perceptual leva em consideração a poderosa ferramenta natural que é a audição humana. Por meio dela, somos capazes de perceber características peculiares de cada falante. Assim, enfatiza que esse método é muito eficiente no auxílio da comparação de locutores. Essa análise fará a seleção dos aspectos das falas dos materiais a serem analisados, como articulação de fonemas, entonação melódica e ritmo de fala, qualidade vocal, sotaques, gírias, sexo, faixa

etária, estado emocional no momento da gravação e presença de alterações na fala (Braid, 2003).

Apesar da grande importância da análise perceptual, essa não consegue mensurar algumas características fonatórias. Por isso, a análise acústica torna-se tão importante quanto, pois viabiliza quantificar medidas de características da fala. Assim, essas duas análises se tornam uma a complementação da outra, de forma a agregar confiabilidade ao exame de comparação de locutores. Braid (2003: 102) pontua isso ao dizer: "Assim como não é indicado como método de verificação de locutor a aplicação apenas da análise perceptual, o uso da análise acústica isoladamente também não é aconselhável".

De acordo com um estudo realizado por Gold e French (2011), o qual considerou as diferentes metodologias empregadas na realização do exame de comparação de locutores em pesquisa abrangendo diversos países, dentre as metodologias empregadas, a análise perceptual combinada com a análise acústica é a mais utilizada, considerando-se os países abordados no trabalho.

Coleta do padrão de voz

Para que seja possível a obtenção de uma fala que se tenha absoluta certeza ter sido produzida pelo suspeito em questão, ou seja, a fala padrão, faz-se necessária a coleta de voz.

Tal procedimento pode ser dividido em três etapas: Na primeira etapa, Braid (2003) sugere uma entrevista, na qual o falante é levado a falar sobre sua vida cotidiana, da forma mais espontânea possível. Na segunda etapa, o examinador pode influenciar o falante a utilizar palavras específicas que o auxiliem na comparação com o outro material objeto da análise (material questionado, produto de interceptação, por exemplo). E, na terceira etapa, pode-se pedir ao entrevistado que repita algumas palavras contidas no material questionado, a fim de se comparar as duas emissões. Todas as etapas utilizam procedimentos padronizados, como, por exemplo, o posicionamento do microfone e o controle de interferências de ruídos externos, dentre outros. Seguindo esses três passos, o autor refere que se inicia a comparação, reconhecendo a espontaneidade do indivíduo, bem como aspectos linguísticos e fonéticos.

Fonoaudiologia

A Fonoaudiologia é a ciência que tem como um dos objetos de estudo a comunicação humana em suas diversas formas. E, como toda ciência, ao longo dos anos vem passando por várias etapas de desenvolvimento e aperfeiçoamento atingindo os mais complexos domínios relacionados à linguagem humana (Amorim, 1982).

Enfatizando o constante aperfeiçoamento e amplitude dessa ciência, Pittioni (2001: 5) refere que: "A Fonoaudiologia como ciência aplicada encontra-se em um processo de expansão do campo de estudos e práticas que vem se mostrando pelo rápido surgimento de áreas específicas de atuação e pesquisa".

A idealização da profissão de fonoaudiólogo no Brasil data da década de 1930, oriunda da preocupação da medicina e da educação com a profilaxia, bem como com a correção de erros de linguagem apresentados pelos escolares (CFFa, 2001).

Nos anos 60, foi criado o primeiro curso de graduação, até então formando Tecnólogos em Fonoaudiologia, na Universidade de São Paulo. O primeiro currículo mínimo exigido para a formação em Fonoaudiologia foi elaborado somente na década de 70, por

meio da resolução n 54/76 do Conselho Federal de Educação, a qual focava basicamente na profissão do fonoaudiólogo como o organizador da linguagem. Esse mesmo documento trazia ainda informações sobre as aptidões do fonoaudiólogo, como a atuação na promoção da saúde, prevenção, pesquisa, terapia, avaliação e aperfeiçoamento em suas áreas de competência (CFFa, 2007).

Somente em 09 de dezembro de 1981, a profissão do fonoaudiólogo foi regulamentada a partir da Lei n 6965/81. Essa dispõe em parágrafo único que: "Fonoaudiólogo é o profissional, com graduação plena em Fonoaudiologia, que atua em pesquisa, prevenção, avaliação e terapia fonoaudiológicas na área da comunicação oral e escrita, voz e audição, bem como em aperfeiçoamento dos padrões da fala e da voz" (Brasil, 1981: 1).

Quanto às áreas de especialidade do fonoaudiólogo, são reconhecidas as seguintes: Audiologia, Linguagem, Motricidade Orofacial, Voz e Saúde Coletiva, havendo, no entanto, uma tendência à criação de novas especialidades (Ferigotti e Nagib, 2009). Nesse sentido, essas mesmas autoras dissertam sobre o processo dinâmico pelo qual passa a Fonoaudiologia, sugerindo haver um movimento de progressão, desenvolvimento e continuidade, contribuindo para que propostas quanto a novas áreas de especialidades possam vir a ser reconhecidas.

Além desse dinamismo atrelado à profissão, o fonoaudiólogo pode ser conceituado como:

> (…) um profissional de atuação autônoma e independente, que exerce suas funções nos setores público e privado, é responsável pela promoção da saúde, avaliação, diagnóstico, orientação, terapia (habilitação e reabilitação) e aperfeiçoamento dos aspectos fonoaudiológicos da função auditiva periférica e central, função vestibular, linguagem oral e escrita, voz, fluência, articulação da fala, sistema miofuncional orofacial, cervical e deglutição. (Pimentel *et al.*, 2010: 40).

Destaca-se a comunicação humana em todas as suas dimensões como foco da Fonoaudiologia, que encontra nesse objeto de estudo a essência e a especificidade da profissão Ferigotti e Nagib (2009).

Diretrizes Curriculares Nacionais

Com a evolução das diretrizes nacionais para a regulamentação de cursos de nível superior, determinadas pelo Ministério da Educação, deixou de ser exigido um "currículo mínimo" para os cursos superiores, dentre os quais o de Fonoaudiologia, e foram instituídas as "Diretrizes Curriculares Nacionais".

Tais diretrizes definiram os princípios, fundamentos, condições e procedimentos à formação do fonoaudiólogo, definindo habilidades gerais, com a intenção de tornar o fonoaudiólogo apto a atuar: (I) na atenção à saúde, tanto na prevenção quanto na promoção, na reabilitação e na proteção individual ou coletiva; (II) nas tomadas de decisões, avaliando e executando condutas condizentes com suas aptidões da maneira mais eficaz e ética possível; (III) na comunicação, acessibilidade a outros profissionais e confidencialidade das informações de posse do fonoaudiólogo; (IV) na liderança frente a equipes multidisciplinares, visando o bem estar do paciente e da sociedade; (V) na administração e no gerenciamento, tornando o profissional apto ao empreendedorismo e; (VI) na educação permanente, visando à contínua manutenção dos conhecimentos, tornando-o sempre aberto a novas ideias e atualizações.

Já no âmbito das habilidades específicas do fonoaudiólogo, as referidas diretrizes buscam nortear a formação acadêmica de modo que o profissional seja apto a: (I) conhecer a fundamentação teórica das áreas de sua responsabilidade; (II) compreender o ser humano nos aspectos físicos, psicológicos e sociais; (III) ter consciência da complexidade dos processos fonoaudiológicos; (IV) avaliar, diagnosticar e prevenir alterações do âmbito fonoaudiológico; (V) capacitar o profissional a intervir com eficácia face às demandas fonoaudiológicas; (VI) dominar o conhecimento na atuação fonoaudiológica; (VII) garantir o direito à saúde do indivíduo; (VIII) participar efetivamente nas equipes multi, inter e transdisciplinares; (IX) dotar-se de recursos científicos em sua atuação; (X) ter autonomia para empreender uma contínua formação profissional; (XI) atuar de forma fundamentada e crítica frente a situações profissionais de sua competência; (XII) moldar sua atuação ao contexto social, visando contribuir socialmente; (XIII) conhecer métodos de elaboração de trabalhos acadêmicos; (XIV) acompanhar as inovações científicas no seu campo de atuação. Ainda nessas diretrizes, relacionam-se os conteúdos essenciais para o curso, enfatizando que esses devem estar relacionados com o processo saúde-doença dos indivíduos. Os conteúdos são: (I) Ciências Biológicas e da Saúde; (II) Ciências Sociais e Humanas; (III) Ciências Fonoaudiológicas. Além disso, as Diretrizes fazem referência à importância e à obrigatoriedade dos estágios supervisionados, a fim de que o graduando adquira experiência prática nas áreas de competência do fonoaudiólogo, e também versam em relação às atividades complementares.

Por fim, mencionam a necessidade de avaliação constante do documento, visando seu aprimoramento em relação ao conteúdo, permitindo os ajustes necessários (Brasil, 2002).

Perícia Forense e Fonoaudiologia

Somente no ano de 1998, o Conselho Federal de Fonoaudiologia criou a resolução n 214, a qual gabarita o fonoaudiólogo na atuação como perito nas áreas de sua competência, quando cita que: "Art. 1 - É permitido ao Fonoaudiólogo atuar judicial ou extrajudicialmente como perito em assuntos de sua competência" (Brasil, 1998: 1).

A perícia fonoaudiológica na prática forense não é algo, no âmbito mundial, tão novo quanto possa parecer. Buriti e Batsita (2009: 15) referem que: "Por se tratar de Fonoaudiologia, a atuação deste profissional na prática forense por vezes parece objeto novo de estudo, mas vale ressaltar que até o presente momento, a Fonoaudiologia forense é largamente utilizada mundialmente". Essas mesmas autoras referem ainda haver uma tímida atuação do fonoaudiólogo na área forense no Brasil, por tratar-se de uma ciência relativamente nova no país. Por fim, afirmam ainda que acreditam que, em poucos anos, esta área estará muito mais difundida e valorizada.

É plausível admitir que essa difusão e valorização crescerá muito ao se observar as considerações realizadas por autores que dissertam sobre os requisitos que um perito deve ter para a execução de determinados exames forenses. Descrevendo as aptidões que o perito deve ter nessa área de comparação de locutores, por exemplo, Ribeiro *et al.* (2008), *apud* Scatena 2010: 30, referem que: "Os principais conhecimentos para um perito trabalhar com a verificação de locutor e edição são os de processamento digital de sinais, física acústica, fonética articulatória, fonética acústica e fonologia do português". Fundamenta-se, portanto, a atuação do profissional fonoaudiólogo como perito, já que assuntos como física acústica, fonética articulatória, fonética acústica e fonologia do por-

tuguês estão também inseridos em suas áreas de conhecimento, uma vez que são parte essencial na análise da comunicação, da fala, da voz humana.

Acerca do domínio desse assunto sobre a voz humana na escala forense, ao discorrer sobre a intersecção entre Fonoaudiologia e biodireito, é citado pelas autoras Buriti e Batsita (2009):

> (...) quando tratamos de Fonoaudiologia forense, muitos estudantes e pesquisadores se assustam a associação destas duas profissões: o fonoaudiólogo em favor das práticas legais do biodireito, mas a determinação do perito criminalista, está na dominância de determinado assunto, para os fonoaudiólogos, o domínio dos parâmetros de análise da voz humana que independe por sua vez das circunstâncias, sejam estas clínicas ou jurídicas, a voz humana na escala forense é contemplada como objeto de estudo, na determinância de provas. (p. 16)

Discussão

Motivado pelo aumento do número de fonoaudiólogos atuando na área pericial, especificamente nos exames de comparação de locutores, este trabalho ancorou-se na fundamentação teórica por meio de levantamento bibliográfico. Sendo ainda escassos os trabalhos relacionando a Fonoaudiologia com a área forense, optou-se por discorrer sobre um e outro tema, abordando a ambos separadamente.

Pode-se destacar como relevante o fato de a Fonoaudiologia, conforme as citações de Amorim (1982) e de Ferigotti e Nagib (2009), ter como objeto de estudo a comunicação humana.

Apesar da maior parte do material encontrado ainda relacionar essa ciência aos distúrbios da comunicação, os autores enfatizam que a Fonoaudiologia, como tantas outras áreas de conhecimento, encontra-se em um constante processo de expansão. Esse desenvolvimento dinâmico e os constantes progressos levam a avanços diversos, abrindo a possibilidade da profissão se recriar, bem como de se inter-relacionar com demais áreas que, em conjunto, podem contribuir para além do que já contribuem individualmente.

As diretrizes curriculares sugeridas pelo Ministério da Educação indicam que, para tornar-se Bacharel em Fonoaudiologia, o então graduando deve ter, dentre outros, o conhecimento dos processos biológicos e sociais complexos que envolvem a comunicação e seus distúrbios. Isso já denota a relação próxima da Fonoaudiologia vinculada à comunicação.

No entanto, a Fonoaudiologia ainda é bastante relacionada aos distúrbios da comunicação humana, não ficando claro, muitas vezes, que, para que se conheça o distúrbio e se reabilite uma função, o profissional deve, necessariamente, ser conhecedor da normalidade.

Por isso, sendo a comunicação humana composta também pela fala, e sendo esta um dos objetos de estudo da Fonoaudiologia, torna-se clara a relação entre a perícia de comparação de locutores e esta ciência. Buriti e Batsita (2009) e Morisson (2003) referiram com clareza o importante papel da fonética forense na área de comparação de locutores, bem como o do fonoaudiólogo enquanto profissional com qualificação para trabalhar nas áreas de comunicação humana.

A comunicação humana dentro da Fonoaudiologia é profundamente essencial como objeto a ser estudado. O fonoaudiólogo dedica-se ao estudo da comunicação humana de forma ampla. Afinal, sua visão abarca desde atuações terapêuticas objetivando uma

melhora na comunicação, por exemplo, como atuações acadêmicas visando à expansão dos estudos acerca do seu funcionamento. Tal amplitude também compreende a extensão que tal tema pode abranger, desde aspectos relacionados à audiologia que também podem permear a comunicação, como aqueles relacionados a voz e fala, itens mais abordados no tema disposto neste estudo.

É importante ressaltar que em nenhum momento se descartou, no entanto, a importância de demais profissionais que contribuem com outras áreas de conhecimento correlatas ao exame, como, por exemplo, o processamento digital de sinais e os aspectos linguísticos. A intenção é enriquecer as equipes responsáveis pelos setores de comparação de locutores, com profissionais capazes de agregar conhecimentos nos aspectos da comunicação humana.

A partir disso, observou-se que os peritos forenses são detentores de saberes aprofundados sobre uma determinada área, como Rodrigues *et al.* (2010) referiu com clareza. Observou-se, ainda, ser a Fonoaudiologia a ciência que tem como um dos seus objetos de estudo a comunicação humana. Fica evidente, portanto, o quanto o fonoaudiólogo pode contribuir com efetividade em ramos das ciências forenses, como o de comparação de locutores, sendo um profissional de grande importância para compor tais equipes multidisciplinares.

Considerações Finais

Este trabalho tratou da contribuição da Fonoaudiologia como ciência, indicando a forma como esta pode colaborar na área forense, especificamente no exame de comparação de locutores, agregando conhecimentos. Observou-se que a literatura sobre a junção desses temas é escassa, o que não ocorre com a pesquisa distinta dos assuntos, separadamente.

Ao confrontar os objetos de análise da Fonoaudiologia e da perícia de comparação de locutores, fica evidenciada a maior intersecção de ambas: fundamentalmente a comunicação. Não existe, no entanto, a pretensão de pleitear a exclusividade do fonoaudiólogo para a realização de tal perícia, mas destacar a importância de seu papel em uma equipe multidisciplinar, incentivando os órgãos oficiais a admitirem esses profissionais em seus quadros de peritos. Pretende-se, ainda, estimular que os fonoaudiólogos agreguem aos seus conhecimentos saberes necessários ao exame em questão, encorajando-os a atuar com propriedade na perícia forense.

Dessa forma, cabe aos profissionais interessados no domínio em questão realizar as devidas conexões para a própria atuação, além de, a partir disso, desenvolver pesquisas a fim de aprofundar a relação entre essas áreas.

Sugere-se, ainda, que novos estudos sejam realizados, e que este trabalho, trazendo o subsídio da literatura, seja apenas o início de publicações unindo as duas atuações, envolvendo inclusive experimentos que conduzam à parte prática.

Referências

Alcântara, H. R. (2006). *Perícia Médica Judicial.* Rio de Janeiro: Guanabara Koogan.

Amorim, A. (1982). *Fonoaudiologia Geral.* São Paulo: Enelivros.

Braid, A. C. M. (2003). *Fonética Forense.* São Paulo: Millennium.

Brasil, (1981). *Lei nr. 6.965, de 9 de dezembro de 1981 - Dispõe sobre a regulamentação da Profissão de Fonoaudiólogo, e determina outras providências.* Online.

Brasil, (1998). *Resolução CFFa nr. 214 - Dispõe sobre a atuação do Fonoaudiólogo como perito em assuntos de sua competência e dá outras providências.* Online.

Brasil, (2002). *Diretrizes Curriculares Nacionais do Curso de Fonoaudiologia.* Ministério da Educação online.

Brasil, (2008). *Lei nr. 11.690 de 09 de junho de 2008 - Dispõe das alterações operadas no Código de Processo Penal quanto à prova pericial.* Online.

Buriti, A. K. L. e Batsita, F. S. R. (2009). A fonoaudiologia forense e o biodireito: Limites entre a lei da interceptação telefônica versus crime organizado. In *Anais online do II Encontro Nacional de Bioética e Biodireito - III Encontro de Comitês de Ética em Pesquisa da Paraíba.*

CFFa, (2001). *Exercício Profissional do Fonoaudiólogo.* Conselho Federal de Fonoaudiologia online.

CFFa, (2007). *Áreas de Competência do Fonoaudiólogo no Brasil.* Conselho Federal de Fonoaudiologia - 8º Colegiado - Gestão 2004/2007 Documento Oficial online.

Del-Campo, E. R. A. (2008). Exame e levantamento técnico pericial de locais de interesse à justiça criminal: Abordagem descritiva e crítica. Dissertação de Mestrado, Universidade de São Paulo.

Dias, F. C. (2010). A prova pericial no direito processual penal brasileiro. *Âmbito Jurídico*, XIII(80).

Dorea, L. E. C., E., S. V. P. e Quintela, V. (2005). *Criminalística.* Campinas: Millenium, 3 ed.

Ferigotti, A. C. M. e Nagib, L. (2009). Fonoaudiologia: reabertas as discussões sobre especialidades. *Revista da Sociedade Brasileira de Fonoaudiologia*, 03(14).

Figueiredo, R. M. (1994). *Identificação de Falantes: Aspectos Teóricos e Metodológicos.* Tese de doutorado, Universidade Estadual de Campinas.

Gold, E. e French, P. (2011). International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law*, 18(2), 293–307.

Gomes, M. L. C., Richert, L. C. e Malakoski, J. (2012). Identificação de locutor na área forense: A importância da pesquisa interdisciplinar. In *Anais do X Encontro do CELSUL - Universidade Estadual do Oeste do Paraná, online. Disponível em http://www.celsul.org.br/Encontros/10/completos/xcelsul_artigo%20(149).pdf , Acesso em março de 2013.*, Cascavel-Paraná.

Louis, J. B. (2000). Forensic voice identification in France. *Speech Comunication*, 31, 205–224.

Morisson, A. L. C. (2003). Verificação de locutor. *Perícia Federal*, 16, 19–23. Online. Disponível em http://www.apcf.org.br/Portals/0/revistaAPCF/16.pdf, Acesso em março de 2013.

Pimentel, A. G. L., Lopes-Herrera„ A., S. e Duarte, T. F. (2010). Conhecimento que acompanhantes de pacientes de uma clínica-escola de fonoaudiologia tem sobre a atuação fonoaudiológica. *Revista da Sociedade Brasileira de Fonoaudiologia*, 15(1), 40–46.

Pittioni, M. E. M. (2001). Fonoaudiologia hospitalar: Uma realidade necessária. Monografia de especialização.

Ribeiro, J. F., Morisson, A. L. d. C., Ricardo, J. d. L. e Sampaio, J. F. (2008). Exames periciais em fonética forense: Recomendações técnicas para a padronização de procedimento em metodologias. *Disponível em http://www.abcperitosoficiais.org.br/hotsites/seminariopara/Criminal-12-fonetica.pdf . Acesso em 29/05/2010.*

Rodrigues, C. V., Silva, M. T. e Truzzi, O. M. S. (2010). Perícia criminal: uma abordagem de serviços. on-line.

Scatena, H. J. (2010). A física aplicada à perícia criminal: Fonética forense. Trabalho de Conclusão de Curso de Licenciatura.

Schaefer, L. S., Rossetto, S. e Kristensen, C. H. (2012). Perícia psicológica no abuso sexual de crianças e adolescentes. *Psicologia: Teoria e Pesquisa*, 28(16), 227–234.

Silva, A. A. G. (2010). A perícia forense no brasil. Dissertação de Mestrado, Escola Politécnica da Universidade de São Paulo.

Tonaco, N. L. A. (2003). Cuidados com a gravação de material sonoro. *Perícia Federal*, 16(24).

Uso de técnicas acústicas para verificação de locutor em simulação experimental

Aline de Paula Machado & Plínio Almeida Barbosa

Universidade Estadual de Campinas (UNICAMP)

Abstract. *The aim of this research was to investigate the efficiency of a set of acoustic measures to identify one randomly pre-selected individual, named "the criminal", out of a group of ten speakers of Brazilian Portuguese. Among the measures used were the first two formants of vowels, fundamental frequency, the duration of syllables and vowels, formant movement rate, intensity and the standard deviation of consonantal interval durations (ΔC). We analyzed all the Brazilian Portuguese vowels. All the subjects were speakers of Brazilian Portuguese from the states of São Paulo, Rio Grande do Sul, Pará and Bahia. All the samples were extracted from interviews recorded in low quality acoustic environments. The samples were divided into two groups i) interviews and ii) recorded telephone conversations (mobile phone to mobile phone). The parameters that were most robust were rhythm and timing, such as duration, speech rate, ΔC and formant movement rate for the second formant. Because of their interspeaker variability, timing measures proved to be highly discriminative. The statistical tests showed that the speech of three of the subjects contained similarities to that of "the criminal".*

Keywords: *Forensic phonetics, speaker verification, acoustic phonetics.*

Resumo. *Este artigo tem como objetivo investigar a eficácia de um conjunto de medidas acústicas no que concerne ao reconhecimento da fala de um indivíduo em um grupo de dez falantes do português brasileiro. Um sujeito desse grupo foi sorteado e nomeado o "criminoso". Entre as medidas usadas na pesquisa estão as frequências dos dois primeiros formantes, a frequência fundamental média, a duração de unidades do tamanho da sílaba e da vogal, a dinamicidade dos formantes e o desvio padrão de durações de intervalos consonânticos (ΔC). Analisamos todas as vogais do português brasileiro. Todos os entrevistados eram falantes do português brasileiro de regiões dos estados de São Paulo, Rio Grande do Sul, Pará e Bahia. Todas as amostras foram extraídas de entrevistas gravadas em ambientes acústicos de baixa qualidade. Os trechos escolhidos para essa análise foram divididos em dois grupos, (i) entrevistas e (ii) gravações telefônicas (de celular para celular). Os parâmetros mais robustos foram ritmo e tempo, tais como duração, taxa de elocução, ΔC, e taxa de movimento do segundo formante. As medidas*

temporais da pesquisa, por serem as mais variáveis intersujeitos, tiveram grande poder discriminador. Os testes estatísticos apontaram que três dos indivíduos estudados, apresentavam semelhanças com o "criminoso".

Palavras-chave: *Fonética forense, verificação de locutor, fonética acústica.*

Introdução

O reconhecimento de locutor se caracteriza por "qualquer atividade pela qual uma amostra de fala é atribuída a uma pessoa com base em suas propriedades fonético-acústicas ou perceptuais" (Jessen, 2008: 671). Entra em jogo a fonética forense: a aplicação de técnicas de análise fonética em contextos policiais jurídicos. Essa é uma área que vem crescendo desde a década de 1960 no Reino Unido e que tem tido sua importância disseminada para todo o globo desde então (French, 1994).

No Brasil, essa subárea da fonética não é demasiadamente promovida nas universidades, e suas técnicas de análise pela polícia não são, de modo geral, semelhantes às usadas em outros países cujo sistema judicial demanda esse tipo de análise. No exterior, normalmente o especialista que faz as análises das amostras de fala trazidas pela polícia é um foneticista ou um profissional com extenso *background* fonético-linguístico. Nota-se, então, a relação estreita que existe entre departamento policial e a universidade, o que facilita essa troca de serviços. No Brasil, a análise das gravações colhidas é feita por um perito (não necessariamente um foneticista) utilizando métodos auditivos e de análise acústica. O profissional também tem como opção o uso de método automático de análise utilizando um *software* de computador, por exemplo, o Batvox (Gold e Fench, 2011). Nos países da Europa, como Inglaterra, Suécia e Alemanha, entre outros, o uso de sistemas automáticos é acompanhado por *insights* de um profissional com conhecimentos em fonética ou em linguística, tal como ocorre na Universidade de Gotemburgo, onde o *software* utilizado é o ALIZE SpkDet e os resultados obtidos pelo programa são combinados com análise acústico-auditiva tradicional (Eriksson, 2012).

Diante da necessidade de trabalhos de pesquisa no Brasil, este trabalho teve como objetivo reconhecer um indivíduo através de sua fala dentro de um grupo de dez falantes do português brasileiro, assinalando quais parâmetros acústicos são relevantes para a análise desse reconhecimento. Com relação ao método de análise, usou-se o método acústico semiautomático. Ressalta-se que o método auditivo não é aplicado neste estudo, pois: (1) os sujeitos, excetuando-se o "criminoso", não são desconhecidos dos pesquisadores e (2) não apresentam grandes diferenças de sotaque e/ou outras características importantes para a discriminação nesta análise (i.e. patologia na fala).

Verificação de locutor e método de análise

Escolhemos utilizar em nossa pesquisa o termo "verificação", em vez de "identificação de locutor". Segundo Hollien (2002), na verificação de locutor é a identidade da pessoa que está em questão, ou seja, a voz é utilizada para acessar uma conta de banco por telefone ou alguma informação privilegiada. Essa análise é controlada por analistas e feita por computadores que comparam a voz questionada com uma voz já armazenada, o que permite que a verossimilhança seja verificada. O falante a ser avaliado, portanto, é cooperativo: ele produz várias amostras de sua fala para a comparação de voz sem, provavelmente, adotar algum tipo de disfarce ou variações em sua voz. Embora a fonética forense seja associada à tarefa de identificação de locutor, ou seja, à identificação

de uma única pessoa (desconhecida) em uma população — reconhecimento indireto de um sujeito —, na prática, a identificação de locutor acaba sendo verificação, já que o trabalho forense, na maioria das vezes, toma um número finito de suspeitos para o reconhecimento de um criminoso a partir da comparação entre gravações questionada e de referência.

As bases para a pesquisa

A Fonética Forense é constituída tanto pela aplicação de conhecimentos, teorias e métodos da fonética geral em tarefas práticas em contexto de trabalho policial ou de apresentação de evidência em tribunal, quanto pelo constante desenvolvimento de novos métodos, teorias e conhecimentos (Jessen, 2008).

Em uma situação forense há geralmente o seguinte cenário: uma gravação que pode vincular ou desvincular um indivíduo a uma atividade criminosa a partir da comparação com uma gravação de referência. A primeira, ou gravação questionada, costuma ser feita por interceptação telefônica, situação em que o indivíduo tende a falar espontaneamente, não sabendo que está sendo gravado. A segunda gravação geralmente é realizada em ambiente acusticamente tratado e os peritos pedem ao suspeito que leia um texto de forma clara para um microfone posicionado a sua frente. Porém, essa situação é apenas ideal em papel, não acontecendo, em grande parte, como metodologia adotada no Brasil. O sujeito pode, por exemplo, apresentar um nível de alfabetização muito baixo, não tornando viável a tarefa de leitura de texto. Esse tipo de técnica de análise acaba se tornando um ponto que dificultará o trabalho de perícia, pois são gravações em contextos diferentes, (1) uma situação de fala espontânea, com o discurso fluente e (2) em laboratório, com material lido. Com isso, palavras que foram encontradas na primeira gravação podem não estar presentes na segunda. O nível de estresse e a naturalidade da fala também afetam a produção de palavras, o que prejudica a precisão necessária para realizar a comparação das análises de cada indivíduo (Byrne e Foulkes, 2004). Outro motivo que pode dificultar a análise são os efeitos que o telefone celular pode causar na gravação.

O efeito do celular

Em muitas situações forenses, cientistas têm em mãos como material de avaliação escutas telefônicas que são, em sua grande maioria, de péssima qualidade, (embora geralmente sejam essas as únicas fontes sonoras para a extração de parâmetros acústicos) por meio das quais eles devem apresentar algum resultado substancial para o júri. Por isso, trazemos tal situação para a pesquisa, simulando casos de interceptação telefônica.

Primeiramente, escolhemos o celular, em lugar do telefone fixo, pois aquele aparelho é de grande uso pelos criminosos; além disso, sabe-se que no Brasil, há mais de 271[1] milhões de linhas de telefone celular. Foi evidenciado também que a gravação por telefone fixo, em comparação ao telefone celular, apresenta resultados mais robustos, principalmente para o primeiro formante (Kunzel, 2001; Byrne e Foulkes, 2004). Kunzel (2001) considera os efeitos do telefone fixo no sinal de fala para calcular quais consequências a diferença de canal de transmissão (no caso, telefone celular) pode causar nas frequências dos formantes nas gravações. Um dos fatores que influenciam a análise de dados a partir de gravação telefônica é a questão da perda do sinal, além de ruído ambiental — no caso do celular, a distorção do próprio aparelho (pela compressão provocada pelo *vocoder*) é o elemento mais crítico para análise fonética.

Alguns efeitos causados pelo telefone celular foram evidenciados por Byrne e Foulkes (2004). Esperava-se encontrar durante a pesquisa uma degradação do sinal da fala das gravações coletadas advinda da combinação desses efeitos.

i. Efeitos do ambiente: Um dos efeitos mais comuns que telefones podem causar no sinal de fala está ligado ao ambiente físico, isto é, por vezes ligações telefônicas podem acontecer em ambiente de alto nível de ruído de fundo, como no trânsito. Assim, esse tipo de efeito deve ser levado em consideração em análises forenses, pois os ruídos podem afetar informações cruciais no sinal da fala.
ii. Efeito dos falantes: Os próprios falantes influenciam nos dados da conversação telefônica, dado que eles tendem a modificar seu comportamento ao falar por telefone e a tornarem-se, por exemplo, mais formais (no caso do inglês britânico, segundo os autores). Já em interceptações telefônicas de crimes brasileiros, pode-se ocorrer o oposto, i. g. conversas entre integrantes de quadrilhas, apresentando tons informais, gírias etc. O registro telefônico da voz muda consciente ou subconscientemente influenciando na taxa de elocução, na qualidade da voz e, como dito anteriormente, na pronúncia. Um dos efeitos mais comuns que pode ser mencionado é o aumento do volume da voz do indivíduo enquanto esse fala ao telefone, o que afeta diretamente a frequência fundamental do falante (F0).
iii. Efeitos técnicos: Ocorre o que é chamado de "distorção espectral", isto é, o aumento das frequências que se encontram acima do filtro passa-baixa (300 Hz) e a diminuição das frequências que se encontram ligeiramente abaixo do filtro passa-alta (3500 Hz). Em outras palavras, as frequências que estão abaixo de 300 Hz e acima de 3.500 Hz são "apagadas" pelo filtro do telefone celular. Um outro exemplo de efeito técnico é o fenômeno conhecido como "deslocamento de frequências": quanto menor a frequência (por exemplo, o primeiro formante), mais atenuada ela fica pelo canal telefônico em comparação a uma gravação direta (Kunzel, 2001; Byrne e Foulkes, 2004). O contrário também acontece, causando a perda dos componentes de alta frequência, algo que é destrutivo para a identificação forense de falante, já que um grande número de informações (qualidade de voz, por exemplo) é codificado em faixa de frequências mais altas das vogais.

Um dos motivos que nos leva a acreditar que ocorra expressiva diferença entre telefone fixo e celular é que os telefones celulares estão sujeitos a um maior alcance de influências ambientais que os telefones fixos. Pelo fato de os primeiros poderem ser usados em qualquer lugar, muitos tipos diferentes de ruído de fundo serão encontrados nas gravações. Além disso, a telefonia celular utiliza taxas de transmissão inferiores, com compressão e codificação maiores do que a de telefonia fixa.

Parâmetros acústicos estudados para a pesquisa

Frequência fundamental e frequência *baseline*

A frequência fundamental é o correlato acústico da frequência de vibração das pregas vocais na produção de voz (Jessen, 2008). Ela é um parâmetro útil para a comparação interfalantes no ambiente forense. Suas medidas de distribuição de longo-termo, como sua média, são sempre sugeridas por pesquisadores da área (Eriksson, 2012; Rose, 2002). Segundo Eriksson (2012), o seu cálculo depende diretamente da duração da amostra de fala, ou seja, é necessário um tempo mínimo de trecho de fala para a extração de seu valor. Enquanto alguns autores sugerem durações de 14 segundos (Horii, 1975 *apud* Eriksson,

2012), outros sugerem de 60 segundos (Nolan, 1997) ou até de 2 minutos (Baldwin e French, 1990). Nesta pesquisa extraímos a frequência fundamental global do trecho com gravações que tiveram duração mínima de 50 segundos.

Algumas causas podem influenciar na variação da frequência fundamental, como fatores fisiológicos e emocionais do falante (idade, tabagismo, doença, intoxicação, estresse etc.), além de elementos externos, como ruído na amostra de gravação (Braun, 1995 *apud* Eriksson, 2012). Um outro fator que pode influenciar na variação desse parâmetro é o disfarce, pois indivíduos tendem a aumentar ou diminuir sua frequência fundamental para disfarçar a voz em algumas situações de crime (Kunzel, 2001).

Em meio a essa variação, que pode causar uma distorção na medida de F0, Lindh e Eriksson (2007) desenvolveram uma forma de representação para a frequência fundamental chamada de *baseline*. A frequência *baseline* se fundamenta na proposta de um nível de frequência fundamental neutro. Esse nível é um ponto estável estimado como 1,43 desvios-padrão de F0 abaixo da média de F0. Ela foi testada em diferentes materiais de fala que variavam quanto ao estilo de fala, esforço vocal e qualidade de gravação. Esta última condição consistia em gravações usando diferentes canais de transmissão, gravador digital e também telefone celular. Os resultados foram robustos para todos os contextos de gravação.

Frequência de formantes

Formantes são frequências de ressonância no trato vocal. Eles são constituídos por formas e volumes de diferentes cavidades do trato vocal (Fant, 1960).

É possível observar uma nova situação no que concerne o uso de canais de transmissão e sua relação com formantes: sabe-se que, hoje em dia, a maioria das chamadas telefônicas que têm conexão com crimes são feitas usando telefones celulares. Da mesma maneira, investigadores indicam que um número substancial de casos envolvendo fala gravada em celular está crescendo vertiginosamente (Öhman *et al.*, 2010). Assim, Byrne e Foulkes (2004) mostram como a transmissão por celular tem um efeito significativo nos formantes e, de maneira similar, Kunzel (2001) mostrou grandes efeitos causados pelo telefone fixo nos primeiros formantes.

Kunzel (2001) realizou um experimento no qual os participantes — 10 homens e 10 mulheres com idade de 20 a 59 anos – faziam uma leitura do texto "The North Wind and the Sun" em alemão, com taxa de elocução e altura de fala normais. As leituras duraram entre 35 a 40 segundos. O sinal de fala foi gravado simultaneamente em gravador e telefone. Foram analisados cerca de 25 contextos fonológicos de 13 vogais. O autor revelou que encontrou problemas com a própria metodologia do seu experimento, uma vez que o algoritmo usado ocasionava erros e, por exemplo, um formante mais alto do que deveria acabava sendo escolhido, situação que ocorreu principalmente nos dados telefônicos. Os resultados do experimento mostraram que todos os sujeitos apresentaram diferenças significativas para o primeiro formante em gravação telefônica, embora não houvesse diferenças significativas para o segundo formante. Outro dado expressivo foi que o valor da frequência do primeiro formante de cada vogal foi maior na transmissão telefônica do que por gravação direta. A diferença é maior para vogais fechadas como [i] e [u], média para vogais como [e] e [o] e menor ou zero para vogais abertas como [ɔ, a]. Com sua pesquisa, Künzel pôde concluir que os valores das frequências dos formantes

baixos das vogais de falantes masculinos e femininos são deslocados para cima (*formant shifted upwards*), causando erros de medidas.

Alguns anos depois, Byrne e Foulkes decidiram testar o efeito do telefone celular no sinal de fala como resposta ao experimento de Künzel, comparando os achados deste com a realidade da nova transmissão usada (via celular). O experimento consistia em 12 voluntários falantes do inglês, seis homens e seis mulheres, entre 20 e 39 anos. Enquanto esses sujeitos liam o texto "The story of Arthur the rat", duas gravações ocorriam simultaneamente. As gravações diretas foram realizadas por um microfone posicionado diretamente na frente do locutor, conectado em um gravador. Um segundo gravador foi conectado com o propósito de interceptar a chamada recebida na sala do experimentador. Os dados foram armazenados em um computador para análise acústica. Os resultados obtidos por Byrne e Foulkes indicam que: devido ao efeito de filtro da transmissão telefônica as frequências de F1 para a maioria das vogais foram maiores que seus homólogos nas gravações diretas; já as frequências do primeiro formante foram maiores do que as por telefone fixo apresentadas por Kunzel (2001); e as frequências do segundo formante não foram afetadas significativamente pelo canal telefônico. A frequência fundamental também foi comparada entre os dois contextos e obteve-se um aumento de 217Hz para a transmissão por telefone celular em relação ao fixo.

Dinamicidade de formantes de parâmetros do domínio de tempo

Um outro exemplo de estudo de formantes, só que relacionado a sua dinamicidade, foi proposto por Nolan *et al.* (2006). Os autores sugerem que as diferenças individuais em movimentos articulatórios podem ser usadas para a comparação de locutor. Seu experimento mostrou que esse parâmetro acústico apresenta informações idiossincráticas dos locutores, sendo calculado entre a diferença da frequência no contorno do formante e da sua área de transição até o centro do formante. Em seu experimento, valores ligados ao movimento das frequências do segundo formante apresentaram resultados determinantes para a discriminação de locutores. A medida foi feita da seguinte maneira: a partir do segmento de uma vogal, por exemplo, /u:/, foram feitas medidas do ponto médio dos contornos das frequências do primeiro e segundo formantes de cada segmento de /u:/ a partir do "formant tracker" do PRAAT. Um script foi usado para calcular a duração de cada segmento, dividindo-o em dez intervalos iguais. Um outro script mediu o centro das frequências dos formantes a cada passo, normalizando cada contorno formântico.

Outra medida de duração que também foi estudada com o objetivo de comparação de locutor é o ΔC, ou seja, o desvio padrão da duração de intervalos consonânticos. Dellwo e Koreman (2008), em estudo que consistia na gravação de dez falantes do alemão, avaliaram dados de diferentes taxas de elocução ao gravar seus sujeitos variando tais taxas de normal até rápida. O teste mostrou que parâmetros de tempo como o ΔC conseguiam capturar informações idiossincráticas dos sujeitos, mantendo-se robusto em diferentes condições de fala.

Ênfase espectral

Traunmüller e Eriksson (2000) tratam a ênfase espectral como a diferença entre a intensidade acústica do sinal integral e a intensidade do sinal submetido a um filtro passa-baixa com um limite de banda superior definido pela expressão 1, 5 F0, em que F0 é a média da frequência fundamental na vogal sendo analisada. Esperamos desse parâmetro uma

grande variação para o canal telefônico devido ao ruído e ao filtro. Segundo Constantini (2014), a ênfase espectral, em seu experimento, apresentou um aumento de 156% em gravações com ruído, inserido artificialmente pelo PRAAT, em relação às gravações originais.

Taxa de elocução

A taxa de elocução (*speech rate*), é o número de unidades da fala produzidas por minuto ou por segundo. As notações mais comuns são palavras por minuto e sílabas por segundo (Eriksson, 2012). Neste trabalho, ela é medida a partir da média da duração das unidades V-V, unidade do onset de uma vogal até o *onset* da vogal imediatamente seguinte. Pode ser medida automaticamente — no caso de boa qualidade na amostra de fala estudada — ou manualmente, quando há baixa qualidade na gravação. Em outras palavras, a ideia deste parâmetro é contar quantas unidades existem em um determinado trecho, medir a duração deste mesmo trecho e dividir o primeiro número pelo segundo. Esse cálculo resulta em uma taxa, um número x de unidades de fala (sílabas, V-V etc.) por unidade de tempo (em geral, segundos).

Segundo Eriksson (2012), a taxa de elocução apresenta um baixo poder de discriminação interfalantes, apresentando uma variação intrafalante alta. Testaremos nesta pesquisa como ela é afetada pelo canal telefônico, uma vez que a detecção do início da vogal pode ser prejudicada pelo canal. Neste trabalho, levando em consideração que a média da duração de unidade do tamanho da sílaba é o inverso da taxa de elocução e que, portanto, diferenças entre essas médias assinalam diferenças nas taxas, tomaremos a duração média da unidade V como medida de taxa de elocução.

O procedimento e os resultados da pesquisa

Este artigo é resultado de uma pesquisa de mestrado que teve como objetivo identificar um indivíduo pela voz em um grupo de dez falantes do português brasileiro divididos em quatro estados, São Paulo, Rio Grande do Sul, Bahia e Pará. Coletamos gravações de seis participantes do estado de São Paulo (três da capital, um de Jundiaí, um de Campinas e um de Cordeirópolis); dois sujeitos da Bahia, ambos de Salvador; um sujeito de Santarém no estado do Pará; e, por fim, um de Pelotas no Rio Grande do Sul. Os sujeitos tinham uma faixa etária de 18 a 28 anos, com nível de educação mínimo de ensino superior incompleto (completando a Graduação) e moraram a maior parte da vida (mais do que a metade) em suas respectivas cidades natais.

Para realizar essa tarefa, analisamos os seguintes parâmetros acústicos, de todas as vogais do português brasileiro, de cada falante: frequência dos dois primeiros formantes, frequência fundamental média, taxa de elocução, frequência *baseline*, ênfase espectral, a dinamicidade dos formantes e o desvio padrão de durações de intervalos consonânticos (ΔC).

As amostras de todos os indivíduos foram gravadas em dois canais de gravação, gravação direta e gravação por telefone celular; essa última simula a dificuldade encontrada pelos peritos ao analisar gravações de baixa qualidade, tal como pode ser observado, por exemplo, em uma interceptação telefônica, cujo áudio apresenta ruído e deterioração. Além disso, o indivíduo escolhido, ao qual nos referimos como "criminoso", teve também sua fala gravada em ambiente acusticamente tratado para uma análise comparativa mais robusta. Simulamos um caso forense habitual, de crime, tendo como objetivo prin-

cipal o reconhecimento do "criminoso" dentro do grupo de falantes, além de mostrar qual método de análise estatística e parâmetros acústicos são mais eficazes para essa tarefa.

Nesta pesquisa, a gravação em estúdio da qual o indivíduo participa não foi feita através da leitura de um texto, mas em formato de entrevista conduzida pelos pesquisadores. Isso se deu com o objetivo de inserir os mesmos assuntos discutidos na primeira gravação e de deixar o entrevistado o mais à vontade possível para que sua fala fosse fluente e espontânea. Tentando atingir um grau mais próximo de fala espontânea, foi feita uma gravação de cada locutor simulando uma conversa corriqueira na qual eram abordados assuntos do cotidiano, como trabalho, plano para as férias etc.

Foram feitas vinte e uma gravações, dez usando o Mini Gravador Coby Cx-r190 ao ar livre, dez por telefone celular e uma gravação direta em ambiente acusticamente tratado. Nas gravações telefônicas, utilizou-se um celular *Samsung Galaxy Young* pela rede da operadora TIM. O experimentador, permanecendo em um ambiente com nível mínimo de ruído de fundo, fazia a ligação para o participante, que se encontrava em sua respectiva cidade natal. O aparelho de interceptação foi uma placa de áudio, U-Control UCA222, conectado ao telefone celular que, por sua vez, também se conectava com o *desktop*, e cada conversa foi gravada pelo *software* Audacity. Os arquivos de áudio coletados foram do formato *.wav* e a frequência de amostragem de 8.000 Hz.

Todas as gravações foram segmentadas manualmente via *software* PRAAT e as medidas de interesse extraídas automaticamente pelo *script* ForensicDataTracking, desenvolvido e disponibilizado por Barbosa (2013).

Apresentamos na Tabela 1, os dados das gravações dos sujeitos participantes da pesquisa, assim como a origem de cada um.

O *script* automaticamente extraiu medidas para frequência do segundo formante (F2) das vogais, taxa de movimento de formante para o segundo formante, frequência *baseline*, média da frequência fundamental, duração das vogais, inverso da taxa de elocução (média da duração de unidade do tamanho da sílaba), ênfase espectral e ΔC.

Métodos de análise estatística e resultados

Para este experimento decidimos utilizar os testes estatísticos ANOVA e teste de Duncan. A seguir, explicaremos os resultados obtidos nas gravações através deles.

ANOVA

Todos os testes estatísticos utilizados nesta pesquisa foram feitos a partir do *software* R^2. O teste estatístico de ANOVA, ou análise de variância, é a técnica estatística que permite avaliar afirmações sobre as médias de populações. Ele verifica se existe uma diferença significativa entre as médias e se os fatores exercem influência em alguma variável dependente (Dowdy *et al.*, 2004).

Para a pesquisa, estudamos a ANOVA com o seguinte intuito: (i) determinar se os parâmetros acústicos analisados permaneciam robustos com a mudança de canal de transmissão, ou seja, de uma gravação direta por gravador digital para uma gravação por telefone celular, e (ii) se algum desses parâmetros acústicos conseguiriam determinar qual dos sujeitos analisados é o "criminoso". Para a realização desse teste é preciso seguir algumas condições, tal como verificar se os resíduos compõem uma distribuição normal, o que pode ser identificado pelo teste estatístico Shapiro-Wilk, assim como verificar a

Sujeito	Naturalidade	Canal de comunicação	Duração (min)	Número de segmentos (vogais)
1	Bahia	Ar livre	2:15	229
		Celular	3:26	461
2	São Paulo	Ar livre	3:40	515
		Celular	2:21	279
3	São Paulo	Ar livre	1:50	152
		Celular	1:50	193
4	São Paulo	Ar livre	1:05	102
		Celular	2:07	185
5	São Paulo	Ar livre	01:40	180
		Celular	00:56	50
6	São Paulo	Ar livre	03:10	405
		Celular	01:36	245
7	São Paulo	Ar livre	01:27	148
		Celular	02:38	207
8	São Paulo	Ar livre	02:53	297
		Celular	02:57	296
9	Pará	Ar livre	01:55	217
		Celular	01:40	174
10	Rio Grande do Sul	Ar livre	02:10	250
		Celular	02:24	245
Criminoso	Desconhecida	Estúdio	9:40	2181

Tabela 1. Lista com os sujeitos participantes da pesquisa, contexto de gravação (celular ou não) cidade natal, duração de cada gravação e número de vogais estudadas de cada um.

homogeneidade das variâncias dos grupos através do teste Fligner-Killeen. Em seguida, é feita a análise de Kruska-Wallis, que é o correspondente não-paramétrico da ANOVA.

As tabelas 2 e 3 mostram os parâmetros acústicos estudados na pesquisa para o contexto de gravação telefônica e de gravação direta. Neste caso, se o parâmetro acústico apresentou um valor de $p > 5\%$ significa que ele não sofreu variação de canal de transmissão, portanto se mostrando um parâmetro robusto para a pesquisa. Em outras palavras, este é um bom parâmetro acústico para a comparação de trechos por diferentes canais. Podemos perceber através dos testes que os seguintes parâmetros acústicos aceitaram a hipótese nula, apresentando-se robustos para a transmissão telefônica: duração das vogais, taxa de elocução, ΔC e taxa de movimento do segundo formante (F2).

Celular x Gravação direta	MeanV	MeanVV	ΔC
Shapiro-Wilk	p-value = 0.9108	p-value = 0.9515	p-value = 0.822
Fligner-Killeen	p-value = 0.4227	p-value = 0.5611	p-value = 0.2825
ANOVA	p-value = 0.245	p-value = 0.36	p-value = 0.05265

Tabela 2. Valor de p para testes de condições de uso da ANOVA (normalidade e homogeneidade de variâncias) e do teste ANOVA, para $\alpha = 0{,}05$, para a condição de gravações por celular e direta. Resultados para a média da duração das vogais (MeanV), taxa de elocução (MeanVV) e ΔC.

Celulax x Gravação direta	F2	Taxa de F2	Taxa de transição de F2	F0	Baseline	Ênfase Espectral
Fligner-Killeen	p-value = 0.05298	p-value = 9.707e-05	p-value = 0.7776	p-value = 1.833e-13	4.435e-10	p-value < 2.2e-16
Kruskal-Wallis	p-value = 1.3e-09	p-value = 0.5911	p-value = 0.6792	p-value < 2.2e-16	p-value < 2.2e-16	p-value < 2.2e-16

Tabela 3. Valor de p para testes de condições de uso da ANOVA (normalidade e homogeneidade de variâncias) e do teste ANOVA, para $\alpha = 0{,}05$, para a condição de gravações por celular e direta para transição de F2; e Kruskal-Wallis, para $\alpha = 0{,}05$, para os valores de segundo formante (F2), taxa de F2, frequência fundamental (F0), frequência *baseline* e ênfase espectral.

Em seguida, analisamos quais dos parâmetros acústicos tiveram ou não variação em relação aos sujeitos. Ou seja, se um parâmetro acústico de um sujeito não apresentou variação com o "criminoso", poderemos dizer, a princípio, que são a mesma pessoa. Como podemos ver nas tabelas 4 e 5, os seguintes parâmetros acústicos apresentaram baixa variação em entre os sujeitos e o criminoso: taxa de movimento do segundo formante, ΔC, taxa de elocução e frequência *baseline*.

Sujeitos x Criminoso	MeanV	MeanVV	ΔC
Shapiro-Wilk	p-value = 1	p-value = 0.9744	p-value = 0.7885
Fligner-Killeen	p-value = 0.02925	p-value = 0.02925	p-value = 0.02925
Kruskal-Wallis	p-value = 0.06432	p-value = 0.1736	p-value = 0.5828

Tabela 4. Kruskal-Wallis, para $\alpha = 0{,}05$, para a variação interfalante. Resultados para a média da duração das vogais (MeanV), taxa de elocução (MeanVV) e ΔC.

Sujeitos x Criminoso	F2	Taxa de F2	Taxa de transição de F2	F0	Baseline	Ênfase Espectral
Fligner-Killeen	p-value = 3.117e-15	p-value < 2.2e-16	p-value < 2.2e-16	p-value < 2.2e-16	p-value < 2.2e-16	p-value < 2.2e-16
Kruskal-Wallis	p-value < 2.2e-16	p-value = 0.0002058	p-value < 2.2e-16	p-value < 2.2e-16	p-value < 2.2e-16	p-value < 2.2e-16

Tabela 5. Kruskal-Wallis, para $\alpha = 0{,}05$, para a variação interfalante. Resultado para os valores de segundo formante (F2), taxa de F2, transição de F2, frequência fundamental (F0), frequência *baseline* e ênfase espectral.

Os parâmetros acústicos que apresentaram um valor de p>0,05 foram duração média das vogais, taxa de elocução e ΔC. Através da análise por boxplots, o sujeito 4 foi o que mais apresentou semelhanças, em relação aos demais indivíduos, em seus parâmetros acústicos — os parâmetros de taxa de movimento do segundo formante, ΔC, taxa de elocução e frequência baseline — com o criminoso.

Teste de Duncan

Este teste faz um agrupamento de valores semelhantes baseado nas médias de cada parâmetro analisado. Se duas médias não são estatisticamente diferentes, elas ficarão no mesmo grupo.

De acordo com essa análise estatística, os sujeitos 5 e 7 apresentaram um número maior de médias semelhantes com as do "criminoso". O primeiro para os parâmetros de frequência fundamental, taxa de movimento do segundo formante, taxa de transição do segundo formante, frequência *baseline*, média da duração das vogais, taxa de elocução e ΔC; já o sujeito 7 apresentou semelhança com o "criminoso" nos seguintes parâmetros, frequência fundamental, frequência do segundo formante, taxa de movimento do segundo formante, frequência *baseline*, média de duração das vogais, taxa de elocução e ΔC. Em seguida, os sujeitos 1, 2, 3 e 10 apresentaram seis parâmetros acústicos com média semelhante ao do "criminoso"; logo após, os sujeitos 4 e 9, com cinco semelhantes; e, por fim, os sujeitos 6 e 8, com apenas 3 combinações, foram os que menos se assemelharam com o "criminoso".

Discussão e conclusão

Segundo os resultados analisados, os parâmetros acústicos que mais se mostraram robustos em relação à mudança de canal de transmissão foram: média da duração das vogais, taxa de elocução, ΔC e taxa de movimento de segundo formante. Com base na literatura da área, parâmetros de tempo conseguem capturar informações idiossincráticas do falante (Dellwo e Koreman, 2008) e foi isso que os resultados confirmaram.

A taxa de elocução, por sua vez, embora tenha se apresentado como um parâmetro acústico sem variação entre os sujeitos — reiterando Eriksson (2012) de que essa apresentaria um baixo poder de discriminação interfalantes —, foi um dos parâmetros que se manteve robusto na mudança de canal de transmissão, não tendo variação para o canal telefônico em relação à gravação direta.

A frequência fundamental também obteve um resultado esperado ao ser afetada pelo telefone celular. Esse parâmetro teve um aumento de seu valor de 4% em relação à gravação direta, valor estatisticamente pequeno para a variação. Conforme apontam Byrne e Foulkes (2004), através da transmissão GSM (2G) há um aumento de até 217 Hz em relação a gravação direta em seu experimento.

Outros parâmetros acústicos, como a frequência do segundo formante, também sofreram influência da mudança de canal de transmissão. Considera-se que as frequências formânticas são parâmetros que devem ser evitados ao realizar uma tarefa de comparação de voz (Kunzel, 2001; Byrne e Foulkes, 2004) por serem suscetíveis à variação. Nos resultados deste trabalho, a frequência do segundo formante teve uma diminuição de 7% em seu valor. Este é um efeito curioso para a transmissão telefônica, pois, segundo alguns autores (Kunzel, 2001; Byrne e Foulkes, 2004), formantes mais baixos, como os três primeiros, tendem a sofrer um fenômeno de "deslocamento para cima", ou seja, ao passarem pelo canal telefônico, os valores de suas frequências tendem a aumentar.

A frequência *baseline*, segundo Lindh e Eriksson (2007) se manteria robusta em diferentes tipos de canais de transmissão, incluindo canal telefônico. Porém, de acordo com nossos resultados essa frequência sofreu o impacto do efeito do celular, tendo uma diminuição de 4% em seu valor, mesma porcentagem que a frequência fundamental.

De acordo com o teste estatístico ANOVA, através de uma comparação para determinarmos quais parâmetros não apresentam variação interfalantes, é possível dizer que aqueles que se caracterizaram como menos variáveis entre os sujeitos foram média de duração das vogais, taxa de elocução e ΔC.

Já os demais parâmetros apresentaram uma variância entre os sujeitos. O sujeito 4, por exemplo, através da taxa de movimento do segundo formante, ΔC, taxa de elocução e frequência *baseline*, mostrou-se o mais semelhante com o "criminoso". Acreditamos, apoiados na literatura (Eriksson, 2012), que parâmetros de tempo como o ΔC, e um parâmetro que analisa a dinamicidade formântica, como a taxa de movimento para o segundo formante, são parâmetros que conseguem capturar informações idiossincráticas dos falantes. Com isso, o resultado da análise por meio dos *boxplots* apontaria o sujeito 4 como um possível candidato ao "criminoso", seguido pelos sujeitos 5, 7, 1, 2, 3, 10 e 9.

O "criminoso" deste experimento foi escolhido pelo orientador da pesquisa, sendo revelado somente após a análise de resultados. Soube-se, então, que o sujeito 4 era o "criminoso". No teste de Duncan, o sujeito 4 teve médias semelhantes às do "criminoso" para cinco parâmetros acústicos, a saber, a frequência fundamental, a taxa de movimento do segundo formante, a média de duração das vogais, a taxa de elocução e o ΔC. Isso nos mostra que os mesmos parâmetros que capturam informações idiossincráticas de falantes, também apontaram o sujeito 4 como sendo o "criminoso".

Os sujeitos 5 e 7, de acordo com o mesmo teste estatístico, apresentaram um total de sete médias de parâmetros acústicos similares ao "criminoso".

O que podemos concluir da pesquisa é que nenhum dos parâmetros acústicos foi definidor para a identificação precisa do "criminoso", objetivo principal do experimento. Porém, conseguimos demonstrar que os parâmetros acústicos que mais se caracterizam como robustos pela literatura internacional para a identificação interfalante, também apresentaram valor significativo para o trabalho, já que ΔC e a dinamicidade dos formantes foram essenciais para mostrar traços idiossincráticos dos indivíduos.

Também verificamos a robustez dos nove parâmetros acústicos analisados na mudança de canal de transmissão da fala. Obtendo resultados sólidos através do teste ANOVA, pode-se dizer que a média da duração das vogais, a taxa de elocução e a taxa de movimento do segundo formante foram os que não apresentaram variação do canal de gravação direta para o telefone celular.

A taxa de movimento do segundo formante foi o parâmetro acústico que apresentou melhores resultados na pesquisa. Sendo assim, sugerimos a utilização do mesmo para as pesquisas em fonética forense que caminham com metodologia análoga a nossa. É um parâmetro que será usado e melhor explorado em futuras pesquisas.

Assim como para Kunzel (2001), os nossos resultados para as demais frequências de formantes, incluindo a frequência fundamental, apresentaram grande variação para o canal de telefone celular. Assim como o autor, sugere-se evitar o uso das frequências dos formantes como formantes discriminadores para a comparação interfalante no contexto telefônico.

Agradecimentos

Gostaria de agradecer ao apoio do meu orientador, Plínio Almeida Barbosa, pela parceria e ensinamentos.

Notas

[1]http://www.anatel.gov.br/

[2]http://www.r-project.org/

Referências

Baldwin, J. e French, P. (1990). *Forensic Phonetics*. London: Pinter.

Barbosa, P. A. (2013). Forensic data tracking. programa de computador.

Braun, A. (1995). Fundamental frequency - how speaker specific i it? In A. Braun e J.-P. Koster, Orgs., *Studies in forensic phonetics*, 9–23. Trier: WVT Wissenschaftlicher.

Byrne, C. e Foulkes, P. (2004). The "mobile phone effect" on vowel formants. *The International Journal of Speech, Language and the Law*, 11(1), 83–102.

Constantini, A. C. (2014). *Caracterização prosódica de sujeitos de diferentes variedades de fala do português brasileiro em diferentes relações sinal-ruído*. Tese de doutorado em linguística, Unicamp, Campinas, SP.

Dellwo, V. e Koreman, J. (2008). How speaker idiosyncratic is measurable speech rhythm? In *Anais de IAFPA 2008*, Lausanne: Swiss Federal Institute of Technology Lausanne Disponível em http://www.hf.ntnu.no/isk/koreman/Publications/2008/IAFPA2008abstract_DellwoKoreman.pdf, último acesso em 30/09/2014.

Dowdy, S., Wearden, S. e Chilko, D. (2004). *Statistics for Research*. New York: John Wiley & Sons.

Eriksson, A. (2012). Aural/acoustic vs. automatic methods in forensic phonetic case work. In A. Neustein e H. Patil, Orgs., *A Forensic Speaker Recognition: Law Enforcement and Counter-terrorism*, 41–69. New York: Springer-Verlag.

Fant, G. (1960). *Acoustic Theory of Speech Production*. Haia: Mouton.

French, P. (1994). An overview of forensic phonetics with particular reference to speaker identification. *Forensic Linguistics*, 1(1), 169–181.

Gold, E. e Fench, P. (2011). International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law*, 18(2), 293–307.

Hollien, H. (2002). *Forensic Voice Identification*. London: Academic Press.

Horii, Y. (1975). Some statistical characteristics of voice fundamental frequency. *Journal of speech and hearing research*, 18(1), 192–201.

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, 2(4), 671–711.

Kunzel, H. J. (2001). Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguistics*, 8(1), 80–99.

Lindh, J. e Eriksson, A. (2007). Robustness of long time measures of fundamental frequency. In *Proceedings of INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association*, 2025–2028, Antwerp, Belgium.

Nolan, F. (1997). Speaker recognition and forensic phonetics. In W. Hardcastle e J. Laver, Orgs., *A Handbook of Phonetic Science*. Oxford: Blackwell.

Nolan, F., McDougall, K., de Jong, G. e Hudson, T. (2006). A forensic phonetic study of 'dynamic' sources of variability in speech: The DyViS project. In P. Warren e C. Watson, Orgs., *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, 13–18, Auckland: Australasian Speech Science and Technology Association.

Öhman, L., Eriksson, A. e Granhag, P. A. (2010). Overhearing the planning of a crime: do adults outperform children as earwitnesses? *Journal of Police and Criminal Psychology*, 26(2), 118–127.

Rose, P. (2002). *Forensic-Phonetic parameters*. New York: Taylor and Francis.
Traunmüller, H. e Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America*, 107(6), 3438–3451.

Resenha de tese

Taxa de Elocução e de Articulação em Corpus Forense do Português Brasileiro

Cintia Schivinscki Gonçalves

Instituto-Geral de Perícias do RS

Perita Criminal
Departamento de Criminalística/ Seção de Perícias em Áudio e Imagens (SEPAI)
Instituto-Geral de Perícias do RS
Porto Alegre/RS, Brasil

Instituição que conferiu o grau:
Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre/RS – CEP: 90619-900.

Data da colação do grau: 2013

Palavras-chave: Taxa de elocução, taxa de articulação, tempo de fala, Linguística Forense, Fonética Forense, comparação de locutor

Esta tese tem por tema a taxa de elocução (TE) e de articulação (TA) em fala espontânea do português brasileiro (PB), ambas calculadas a partir de dois tipos de amostras de fala: uma gravada sem a ciência dos interlocutores (registros de interceptações telefônicas judicialmente autorizadas), intitulada "gravação desavisada" e outra gravada de forma sabida e consentida (registro de entrevista semidirigida), intitulada "ravação avisada" contraponto situacional comumente encontrado na área de Fonética Forense, especificamente na perícia de Comparação de Locutor (CL). Trata-se, o **problema** em questão, da tentativa de uso criterioso de medidas temporais de fala no confronto forense a partir de amostras de produção oral.

As **hipóteses** que guiaram a investigação foram: a TA, devido a sua maior estabilidade é, das taxas temporais de fala estudadas, a mais indicada para utilização no confronto de voz forense, sendo capaz de, em algum grau, distinguir falantes; não há diferença significativa entre os valores globais obtidos via divisão do número total de sílabas pela duração total do intervalo de fala dos obtidos via média das mensurações locais; o aumento da idade do sujeito conduz ao uso de menores valores de taxa; os falantes com taxas mais altas são os do sexo masculino; empregam-se valores de taxa maiores em intervalos de fala com maiores durações, observando-se a progressiva redução na duração média das sílabas (encurtamento antecipatório); o indivíduo menos escolarizado apresenta valores mais elevados de taxa; o intervalo de tempo transcorrido entre as gravações do sujeito (desavisada e avisada) impacta as taxas; a ciência da gravação leva o locutor a um maior controle sobre a própria fala. Assim, o **objetivo** principal do estudo foi estabelecer o potencial individualizante da TE e da TA, visando incorporar medidas de tempo de fala ao conjunto de parâmetros técnico-comparativos utilizados na perícia de CL.

Secundariamente, objetivou-se verificar se havia diferença significativa entre os tipos de taxa e as formas de mensuração, bem como a relação existente entre as taxas investigadas e as variáveis independentes idade, sexo, escolaridade, gap temporal entre as gravações desavisada e avisada, tipo de gravação e tamanho do intervalo de fala, intentando-se prever, ainda, o comportamento das taxas em razão das mencionadas variáveis. A **metodologia** contemplou o cálculo da TE e da TA na fala de sete sujeitos (sendo cinco do sexo masculino e dois do sexo feminino), todos estabelecidos no estado do Rio Grande do Sul (Brasil), tendo o PB como língua materna e dialetos indefinidos devido à ocasional(is) aprisionamento(s). Os sujeitos selecionados integram o banco de dados do Instituto Geral de Perícias (IGP), órgão da Secretaria de Segurança Pública (SSP) do referido estado, figurando nele como alvo da perícia de CL, assegurada a existência de resultado positivo para o confronto de perfil de voz e fala, outrora efetuado com vistas ao atendimento de requisição departamental. Os tempos de fala foram mensurados global e localmente, tendo sido avaliados 539 turnos de fala (no caso da TE) e 748 intervalos interpausais (no caso da TA).

Os **preceitos teórico-conceituais** que sustentam a configuração do estudo e que fundamentam a discussão dos resultados são referentes à Linguística/Fonética Forense, à Fonética Segmental e Suprassegmental e à Sociofonética, sendo oportunamente abordados os conflitos metodológicos que costumeiramente permeiam a concepção de um estudo sobre parâmetros temporais de fala (entre eles o tratamento a ser dado à pausa e à fala disfluente), os critérios considerados na composição do corpus, a forma de organização dos dados dos sete sujeitos selecionados como participantes do estudo, o procedimento de verificação acústica e o tratamento estatístico utilizado. Os **resultados** obtidos, referentes às amostras avaliadas, indicaram que: (i) quanto ao potencial individualizante das taxas, exclusivamente na TA a variância intersujeitos é superior à variância intrassujeito, obtendo-se para tal tipo de taxa um Coeficiente de Correlação Intraclasse indicativo de satisfatório poder discriminatório de falante (com CCI em torno de 0,7 nas duas formas de mensuração) enquanto que para a TE um coeficiente associado à pobre poder discriminatório (com CCI em torno de 0,2 nas duas formas de mensuração); (ii) não há diferença significativa entre os dois tipos de taxa, destacando-se, contudo, que a TA mostra-se menos variável do que a TE, especialmente na mensuração local; (iii)

há diferença significativa entre as formas de mensuração empreendidas (global e local média) somente no que se refere à TE; (iv) na análise da variabilidade por sexo e por tipo de gravação, restou significativo somente o tipo de gravação na TE; (v) quanto à variável sexo especificamente, observou-se que os sujeitos de ambos os sexos tendem a diminuir as taxas quando têm ciência da gravação (diminuição mais expressiva nos sujeitos do sexo masculino) e que na fala classificada como casual (gravação desavisada) são prevalentemente os homens os falantes com as maiores taxas; (vi) quanto à variável tipo de gravação especificamente, encontrou-se diferença significativa entre os fatores desavisada e avisada na TE; (vii) das variáveis independentes escalares (idade, escolaridade e *gap* temporal), considerando-se as duas taxas estudadas, foi evidenciada correlação significativa exclusivamente entre a TA e a variável *gap* temporal; (viii) há correlação significativa entre a TE (a partir das múltiplas tomadas locais) e a variável tamanho do intervalo de fala; (ix) considerando-se as 1.287 taxas locais, são significativos preditores do aumento da TE e da TA o fator masculino da variável sexo e o avanço tanto na escolaridade quanto no *gap* temporal, enquanto que significativos preditores da diminuição da TE e da TA o avanço na idade e a ciência de gravação.

Dessa forma, é possível concluir pela indicação da incorporação da TA local média (tipo de taxa e forma de mensuração menos variáveis) ao rol de parâmetros técnico-comparativos utilizados na perícia de CL, ressalvando-se a importância da máxima contemporaneidade possível entre as gravações confrontadas e a necessidade de adoção de providências que visem minimizar o impacto causado tanto pela ciência da gravação (e demais diferenças entre os tipos de gravação) quanto por eventual incremento na escolarização ocorrido no gap temporal existente entre os áudios do cotejo.

PhD Abstract

Legal Translation: A Study of *Rogatory Letters* and its Implications

Luciane Reiter Fröhlich

Universidade Federal de Santa Catarina

Postdoctoral Researcher in Legal Translation
Centro de Comunicação e Expressão – CCE
Universidade Federal de Santa Catarina
Florianópolis-SC, Brazil

Awarding Institution:
Universidade Federal de Santa Catarina – UFSC
Florianópolis-SC, Brazil

Date of award: 2014

Keywords: Legal Translation; Forensic Linguistics; Letter Rogatory; Plain Language; Legalese; Translator Training.

Legal Translation has assumed a key role within the communicative purposes of our globalized reality, in a wide variety of multilingual and multicultural settings, receiving substantial attention from researchers in the field. Indeed, because of the special nature of the law, the languages, and the legal systems involved, legal translation is known as the most complex and demanding of all areas of specialized translation (Cao, 2007; Šarčević, 2012). This complexity requires advanced technical knowledge of the translator, not simply proficiency in the source (L1) and target (L2) languages, but also knowledge of the peculiarities of two legal languages and legal systems involved. Moreover, the translator needs to understand not only what individual words, phrases and

sentences mean, but also the legal effect they may have (Šarčević, 2010). Considering all this, the thesis explores this field of research, and is inserted into the interface between Translation Studies and Forensic Linguistics, focusing on the skills profile of beginning legal translators.

The motivation for this research was an observation of the peculiarities associated with the translation of Brazilian legal texts, especially rogatory letters. These texts are linguistically very complex and require maximal precision, since they usually carry great social, legal and also economic consequences. According to J. Gibbons (2004), the complexity of sentences and phrasal structures, as well as the use of grammatical metaphors, complicate the understanding of legal texts. This complexity, combined with the idea that legal translations are considered "the ultimate linguistic challenge" (Cao, 2007), makes the work of the legal translator extremely difficult. Therefore, the professional education of translators and the supervision of their translations must be undertaken in a systematic and stringent way. As a result, translator training and evaluation of professional performance should be even more stringent.

However, in this research we observed several worrying aspects that deserve careful consideration. It is not unusual for translators to be left to their own devices. There are no undergraduate or postgraduate courses in Brazil that adequately prepare students for working as legal translators; the profession is not properly regulated; there is no official supervision, and not even a regulatory standard. The combination of these factors makes the translation process more error prone and directly impacts the quality of translations. The situation is made even worse by the increasing demand for translations from the international legal communities.

In this context, we propose some methodological guidelines for legal translation designed to help legal translators to improve the quality of their translations, and to collaborate in the establishment and consolidation of the research field in Legal Translation in Brazil. In addition, we suggest specific translation solutions (to/from the German language) for many of the complex terms and key-expressions typically found in rogatory letters as well as for sentences that feature recurring "legalese" patterns (Andrade, 2009).

It is therefore an exploratory study, mainly based on theories of Forensic Linguistics, with contributions from Malcolm Coulthard (2007; 2010), John Gibbons (2004; 2005), Lawrence Solan (1998; 2010), Peter Tiersma (1999), among others; and theories of Legal Translation Studies, based on publications by Malcolm Coulthard (1991), Deborah Cao (2007; 2010) and Susan Šarčević (2010; 2012).

Thus, for the implementation of the research and subsequent analysis of the particularities of Brazilian legal texts, six sample of rogatory letters, considered highly complex legal documents (OAB-Paraná, 2011), were chosen. All of them had been issued by the Judiciary of the State of Santa Catarina, between the years 2004 and 2011, and submitted for public translation into German, which is the specialized language of the author of this thesis, who worked with these letters as an ad hoc sworn translator.

Beyond this analysis, we suggest ways to simplify legal language based partly on the assumptions of the Plain Language movement (http://www.plainlanguageaustralia.com) and partly on research from Tiersma (1999) and others. As a result of linguistic simplification, legal texts will become more "democratic" and the duties of translators less arduous.

Finally, as an extension of this research, we proposed an "Estudos da Tradução Forense" research line based on translation competence and aimed at assisting beginning legal translators. This resulted in a set of modules or "disciplinas" combined into a prototype curriculum. To support this issue, we used the research on translation competence carried out by the Spanish group PACTE (2000; 2011), by Amparo Hurtado Albir and Fabio Alves (2009), by Maria Lúcia Vasconcellos (2012) and by Jeremy Munday (2009), among others.

References

Albir, A. H. and Alves, F. (2009). Translation as a cognitive activity. In *The Routledge Companion to Translation Studies*, 54. London and New York: Routledge, 2009.

Andrade, V. (2009). O juridiquês e a linguagem jurídica: O certo e o errado no discurso.

Cao, D. (2007). *Translating law*, volume 33. Multilingual Matters.

Cao, D. (2010). Legal translation: Translating legal language. In *The Routledge Handbook of Forensic Linguistics*, 78–91. London: Routledge.

Coulthard, M. (1991). Tradução: Teoria e prática. In *A Tradução e seus Problemas*, 1–15. Florianópolis: Florianópolis, Editora da UFSC.

Coulthard, M. and Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. Routledge.

M. Coulthard and A. Johnson, Eds. (2010). *The Routledge Handbook of Forensic Linguistics*. London: Routledge.

Gibbons, J. (2004). Taking legal language seriously. In J. Gibbons, V. Prakasam, K. V. Tirumalesh and H. Nagarajan, Eds., *Language in the Law*. Hyderabad, India: Orient Longman.

Gibbons, J. (2005). *Forensic Linguistics: An Introduction To Language In The Justice System (Language In Society)*. Malden, USA: Blackwell Publishing.

J. Munday, Ed. (2009). *The Routledge Companion to Translation Studies*. Routledge.

OAB-Paraná, (2011). *Cartilha da Carta Rogatória*. Ordem dos Advogados do Brasil, Curitiba, Paraná.

PACTE, (2000). Acquiring translation competence. In *Investigating Translation: Selected papers from the 4th International Congress on Translation, Barcelona, 1998*, volume 32, 99, Amsterdam: John Benjamins Publishing.

PACTE, (2011). Results of the validation of the pacte translation competence model: Translation project and dynamic translation index. In S. O'Brien, Ed., *Cognitive Explorations of Translation.*, 30–53. London: Continuum.

Šarčević, S. (2010). Legal translation in multilingual settings. In I. Alonso Araguár, J. Baigorri Jalón and H. J. L. Campbell, Eds., *Translating Justice*. Granada: Comares.

Šarčević, S. (2012). Challenges to the legal translator. In L. M. Solan and P. M. Tiersma, Eds., *The Oxford Handbook of Language and Law*. Oxford: Oxford University Press.

Solan, L. M. (1998). Linguistic experts as semantic tour guides. *Forensic Linguistics*, 5(2), 87–106.

Solan, L. M. (2010). The expert linguist meets the adversarial system. In M. Coulthard and A. Johnson, Eds., *The Routledge Handbook of Forensic Linguistics*. London: Routledge.

Tiersma, P. M. (1999). *Legal Language*. Chicago: University of Chicago Press.

Vasconcellos, M. L. (2012). Dimensões da interdisciplinaridade na formação de tradutores: Competência tradutoria, enfoque por tarefas, tipologia textual baseada em contexto. In *I seminário Interdisciplinar das Ciências da Linguagem no Cariri, Ceará*.

PhD Abstract

Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source parameters and cepstral characteristics

Eugenia San Segundo Fernández

University of York, UK

Postdoctoral Research Assistant
Department of Language and Linguistic Science
University of York
UK

Awarding Institution:
Laboratorio de Fonética – Centro de Ciencias Humanas y Sociales
Consejo Superior de Investigaciones Científicas (CSIC), Spain

Date of award: 2014

Keywords: Forensic phonetics; twins; siblings; likelihood ratio; formant trajectories; glottal source.

Research problem

From a forensic phonetic perspective, the voice characteristics of twin pairs and non-twin sibling pairs have frequently awoken special interest because these speakers represent extreme examples of similarity. Distinguishing their voices poses a well-recognized challenge in this discipline (e.g. Künzel, 2010), as can be concluded from the literature

review carried out for this thesis. Nevertheless, the *per se* research interest in these sibling pairs is deeply rooted in the nature-nurture dichotomy, which transcends any possible forensic application. Therefore, in this investigation a comparative study has been undertaken between genetically identical speakers (monozygotic twin pairs, *MZ*) and non-genetically identical speakers (dizygotic twins, *DZ*; non-twin siblings; and unrelated speakers) in an attempt to shed some light on a relevant but under-researched question: to which extend is our voice determined by our genes (*nature*) and to which extent is it due to behavioral, environmental or social aspects (*nurture*)?

Objective

The general objective of this thesis has been investigating the phonetic characteristics of three main speaker groups: MZ twins, DZ twins and non-twin siblings (24, 10 and 8 participants, respectively). Their phonetic-acoustic similarities have been studied by also taken into account a reference population of unrelated speakers (12 subjects). For the 54 male Spanish speakers (North-Central Peninsular variety) recorded *ad hoc* for this study, three different analyses were carried out. On the one hand, we labeled and analyzed the F1-F3 formant trajectories of 19 Spanish vocalic sequences. Secondly, several naturally sustained [e] tokens were extracted from the speakers' spontaneous vowel fillers and their glottal source characteristics were analyzed. These two approaches were complemented with an automatic speaker recognition analysis carried out with the software *Batvox*, based on cepstral parameters.

Hypothesis

The research hypotheses have been established according to what is known so far about twin and non-twin sibling pairs—their shared genetic endowment and the environmental influences possibly affecting their voice. Accepting the *equal environment assumption* traditionally associated to the classic twin method (Segal, 1990), MZ cotwins and DZ cotwins were expected to share the same environmental influences, while DZ twin pairs would share only half the genetic information than MZ twin pairs. Siblings would share the same genetic endowment as DZ twins but, on average, less environmental factors, mainly because of the age gap between them. Finally, unrelated speakers would share neither nature nor nurture.

Accordingly, five working hypotheses have been established for this thesis. Firstly, it was considered that a speaker's voice would be similar to itself, i.e. from one recording session to another. This assumption is made for all speaker types (H1). Secondly, accepting that MZ twin pairs are the most similar speakers that can exist (because of their shared genes and shared environmental influences), we hypothesized (H2) that MZ intra-pair comparisons would yield matching scores similar to those obtained in intra-speaker comparisons. The third hypothesis (H3) suggested that DZ intra-pair comparisons would yield relatively large matching scores but not as large as in the case of MZ twins. In the fourth hypothesis (H4), we stated that the intra-pair comparisons in the case of brothers would yield matching scores over the background baseline. That means that brothers should be more similar than unrelated speakers because they share 50% of their genes, exactly the same as DZ twins, and they usually have environmental influences in common, although to a lesser degree than DZ twins. Finally, we hypothesized (H5) that a background baseline should exist for the matching scores obtained by the unrelated speakers.

Methodology / Theoretical Framework

The first chapter describes the main current methodologies in Forensic Speaker Comparison (FSC). As a result of this review, it was concluded that adopting a hybrid perspective, which combines traditional and automatic analyses, is the most comprehensive approach to speaker comparison. For that reason, the three-folded approach of this thesis combines (a) traditional phonetic-acoustic parameters with (b) not only features but also techniques which are characteristic of automatic methods. In chapter three, the methodological details for carrying out this investigation are described. This includes a description of the main characteristics (age, dialect, etc.) of the recruited participants. An *ad hoc* corpus has been designed and collected for this thesis, including five speaking tasks and a vocal control technique. Some details about the recording procedure are also presented in the third chapter, such as the material and technical characteristics of the recordings, as well as the data collection set-up. Finally, a description follows of the likelihood-ratio approach within which the results of the different analyses are offered.

Results

All the parameters tested for this investigation have proved to be genetically conditioned —to a greater or lesser extent— since the hypothesized decreasing scale $MZ > DZ >$ *non-twin siblings* $>$ *unrelated speakers* was observed regardless of the analysis approach and for most speaker comparisons (the rare discordant results were thoroughly discussed in the corresponding analysis chapter). Therefore, the proposed parameters would be useful for comparing speech samples of known and unknown origin, as found in legal cases. Moreover, as different features were tested depending on the type of analysis conducted, we could indicate separately which parameters (or combination of parameters) were found more useful in the formant-trajectory analysis, on the one hand, and in the glottal-source study, on the other hand.

Future studies could explore the fusion possibilities of the three different systems tested for this investigation. The independence of glottal features from vocal-tract characteristics makes them specially promising for an improvement of an overall forensic system performance. Besides, future research focusing on twins' voices should pay more attention to the concept of *epigenetics*, which was briefly described in chapter two. We have continuously referred throughout this thesis to two basic forces which would intermingle to explain the (dis)similarities in twins and non-twins' voices, namely, genetic and environmental factors. The often-neglected third factor, i.e epigenetics —or the study of the alteration in the expression of specific genes caused by mechanisms other than changes in the underlying DNA sequence— could be behind the striking dissimilarities found for certain twin pairs.

References

Künzel, H. (2010). Automatic speaker recognition of identical twins. *The International Journal of Speech, Language and the Law*, 17(2), 251–277.

Segal, N. (1990). The importance of twin studies for individual differences research. *Journal of Counseling & Development*, 68(6), 612–622.

Notes for Contributors

1. The editors of **Language and Law / Linguagem e Direito** (**LL/LD**) invite original contributions from researchers, academics and practitioners alike, in Portuguese and in English, in any area of forensic linguistics / language and the law. The journal publishes articles, book reviews and PhD abstracts, as well as commentaries and responses, book announcements and obituaries.

2. Articles vary in length, but should normally be between 4,500 and 8,000 words. All other contributions (book reviews, PhD abstracts, commentaries, responses and obituaries) should not exceed 1,200 words. Articles submitted for publication should not have been previously published nor simultaneously submitted for publication elsewhere.

3. All submissions must be made by email to the journal's email address llldjournal@gmail.com. Authors should indicate the nature of their contribution (article, book review, PhD abstract, commentary, response, book announcement or obituary).

   Before submitting an article, visit the journal's webpage (http://llld.linguisticaforense.pt) to access further information on the submission process, authors' guidelines and journal templates.

4. Contributions must be in English or Portuguese. Authors who are not native speakers of the language of submission are strongly advised to have their manuscript proofread and checked carefully by a native speaker.

5. All articles submitted for publication will be refereed before a decision is made to publish. The journal editors will first assess adherence both to the objectives and scope of the journal and to the guidelines for authors, as well as the article's relevance for and accessibility to the target audience of the journal. Articles will subsequently be submitted to a process of double blind peer review. For this reason, the name of the author(s) should not appear anywhere in the text;

self-referencing should be avoided, but if used the author(s) should replace both their own name and the actual title of their work with the word 'AUTHOR'.

6. The articles should be accompanied with a title and an abstract of no more than 150 words in the language of the article and, if possible, in the journal's other language as well. The abstract should also include up to five keywords. Contributions should indicate in the body of the accompanying email the name, institutional affiliation and email address(es) of the author(s).

7. The author(s) may be required to revise their manuscript in response to the reviewers' comments. The journal editors are responsible for the final decision to publish, taking into account the comments of the peer reviewers. Authors will normally be informed of the editorial decision within 3 months of the closing date of the call for papers.

8. Articles should be word-processed in either MS Word (Windows or Mac) – using one of the templates provided – or LaTeX. The page set up should be for A4, with single spacing and wide margins using only Times New Roman 12 pt font (also for quotations and excerpts, notes, references, tables, and captions). PDF files are not accepted. Where required, the following fonts should be used for special purposes:

    - Concordances and transcripts should be set in courier;
    - Phonetics characters should be set in an IPA font (use SIL IPA93 Manuscript or Doulos);
    - Special symbols should be set in a symbol font (as far as possible, use only one such font throughout the manuscript);
    - Text in a language which uses a non-roman writing system (e.g. Arabic, Mandarin, Russian) may need a special language font;
    - Italics should be used to show which words need to be set in italics, NOT underlining (underlining should be used as a separate style in linguistic examples and transcripts, where necessary).

9. The article should be divided into unnumbered sections, and if necessary subsections, with appropriate headings. Since the journal is published online only, authors can include long appendices, colour illustrations, photographs and tables, as well as embed sound files and hyperlinks.

10. Figures, tables, graphics, pictures and artwork should be both inserted into the text and provided as separate files (appropriately named and numbered), in one of the main standard formats (JPEG/JPG, TIFF, PNG, PDF). They should have a resolution of at least 300 dpi, be numbered consecutively and contain a brief, but explanatory caption. Captions should be placed after each table, figure, picture, graphic and artwork in the body of the text, but not in the artwork files. Where applicable, tables should provide a heading for each column.

11. Transcript data should be set in a Courier typeface, numbered by turns, rather than lines, and should be punctuated consistently. Where elements need to be aligned with others on lines above or below, use multiple spaces to produce alignment. Transcripts should be provided as separate image files (e.g. JPEG/JPG, TIFF, PNG, PDF), named according to the transcript number.

12. Abbreviations should be explained in the text, in full form. They should be presented consistently, and clearly referred to in the text. Times New Roman 12 pt should be used whenever possible, unless a smaller size font is necessary.

13. Endnotes are preferred to footnotes but even so should be kept to a minimum. When used, they should be numbered consecutively and consistently throughout the article, starting with 1, and listed at the end of the article, immediately before the References.

14. Manuscripts should clearly indicate the bibliographic sources of works cited. The authors must ensure that the references used are accurate, comprehensive and clearly identified, and must seek permission from copyright holders to reproduce illustrations, tables or figures. It is the responsibility of the author(s) to ensure that they have obtained permission to reproduce any part of another work before submitting their manuscript for publication. They are also responsible for paying any copyright fees that may be charged for the use of such material.

15. Citations in the text should provide the surname of the author(s) or editor(s), year of publication and, where appropriate, page numbers, immediately after the quoted material, in the following style: Coulthard and Johnson, 2007; Coulthard and Johnson (2007); Coulthard and Johnson (2007: 161). When a work has two authors, both names should be referenced each time they are cited. When there are more than two authors, only the first author followed by *et al.* should be used (Nolan *et al.* (2013)). The author, date and page can be repeated, if necessary, but 'ibid.' and 'op. cit.' must **not** be used. When citing information from a particular work, the exact page range should be provided, e.g.: Caldas-Coulthard (2008: 36–37), NOT Caldas-Coulthard (1996: 36 ff.).

16. Quotations should be clearly marked using quotation marks. Long quotations should be avoided. However, when used, quotations of over 40 words in length should be set as a new paragraph; the extract should be left and right indented by 1 cm and set in a smaller font size (11 pt). The citation should follow the final punctuation mark of the quotation inside brackets. No other punctuation should be provided after the citation, e.g.:

> The linguist approaches the problem of questioned authorship from the theoretical position that every native speaker has their own distinct and individual version of the language they speak and write, their own idiolect, and the assumption that this idiolect will manifest itself through distinctive and idiosyncratic choices in speech and writing. (Coulthard and Johnson, 2007: 161)

If author and date are used to introduce the quote, only the page number(s) preceded by 'p.' will appear at the end of the quotation:

As was argued by Coulthard and Johnson (2007):

> The linguist approaches the problem of questioned authorship from the theoretical position that every native speaker has their own distinct and individual version of the language they speak and write, their own idiolect, and the assumption that this idiolect will manifest itself through distinctive and idiosyncratic choices in speech and writing. (p. 161)

17. Quotations must be given in the language of the article. If a quotation has been translated from the original by the author(s), this should be indicated in an endnote where the original quotation should be provided.

18. A list of References should be placed at the end of the article. The References section should contain a list of all and only the works cited in the manuscript, and should be sorted alphabetically by the surname of the (first) author/editor. Multiple publications by the same author(s) should be sorted by date (from oldest to newest). If multiple works of one author in the same year are cited, these should be differentiated using lower case letters after the year, e.g. 1994a, 1994b, and not 1994, 1994a. Book publications must include place of publication and publisher. Page numbers should be provided for chapters in books and journal articles. In addition, the volume and issue number must also be given for journal articles, and the name of journals must not be abbreviated. Reference URLs should be provided when available. When cases and law reports are cited, these should be provided in a separate list following the References.

19. To summarise the following style guidelines should be followed, including the capitalisation and punctuation conventions:

    Books
    Coulthard, M. and Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. London and New York: Routledge.

    Mota-Ribeiro, S. (2005). *Retratos de Mulher: Construções Sociais e Representações Visuais no Feminino*. Porto: Campo das Letras.

    Chapter in a book
    Machin, D. and van Leeuwen, T. (2008). Branding the Self. In C. R. Caldas-Coulthard and R. Iedema (eds) *Identity Trouble: Critical Discourse and Contested Identities*. Basingstoke and New York: Palgrave Macmillan.

    Journal article
    Cruz, N. C. (2008). Vowel Insertion in the speech of Brazilian learners of English: a source of unintelligibility?. *Ilha do Desterro* 55, 133–152.

Nolan, F., McDougall, K. and Hudson, T. (2013). Effects of the telephone on perceived voice similarity: implications for voice line-ups. *The International Journal of Speech, Language and the Law* 20(2), 229–246.

Dissertations and Theses
Lindh, J. (2010). *Robustness of Measures for the Comparison of Speech and Speakers in a Forensic Perspective*. Phd thesis. Gothenburg: University of Gothenburg.

Web site
Caroll, J. (2004). Institutional issues in deterring, detecting and dealing with student plagiarism. *JISC online*, http://www.jisc.ac.uk/publications/briengpapers/2005/pub_plagiarism.aspx, Accessed 14 November 2009.

20. The main author of each contribution will receive proofs for correction. Upon receiving these proofs, they should make sure that no mistakes have been introduced during the editing process. No changes to the contents of the contribution should be made at this stage. The proofs should be returned promptly, normally within two weeks of reception.

21. In submitting an article, authors cede to the journal the right to publish and republish it in the journal's two languages. However, copyright remains with authors. Thus, if they wish to republish, they simply need to inform the editors.

Normas para apresentação e publicação

1. A direção da revista **Language and Law / Linguagem e Direito** (**LL/LD**) convida investigadores/pesquisadores, académicos e profissionais da área da linguística forense / linguagem e direito a apresentar trabalhos originais, em português ou em inglês, para publicação. A revista publica artigos, recensões de livros e resenhas de teses, bem como críticas e respostas, anúncios de publicação de livros e obituários.

2. A dimensão dos artigos pode variar, mas os artigos propostos devem possuir entre 4,500 e 8,000 palavras. As restantes contribuições (recensões, resenhas de tese, comentários, respostas e obituários) não deverão exceder 1200 palavras. Os artigos enviados para publicação não devem ter sido publicados anteriormente, nem propostos a outra publicação científica.

3. As propostas para publicação devem ser enviadas por email para o endereço de correio eletrónico da revista llldjournal@gmail.com. No corpo do email, os autores devem indicar a natureza do seu texto (artigo, recensão, resenha de tese, comentário, resposta, anúncio de publicação de livros ou obituário).

   Os autores devem consultar a página da revista na Internet (http://llld.linguisticaforense.pt) antes de enviarem os seus textos para obterem mais informações acerca do processo de submissão, instruções e modelos de formatação da revista.

4. São aceites textos para publicação em português ou em inglês. Aconselha-se os autores cujo texto se encontre escrito numa língua diferente da sua língua materna a fazerem uma cuidada revisão linguística do mesmo, recorrendo a um falante nativo.

5. Todos os textos enviados para publicação serão sujeitos a um processo de avaliação com vista à sua possível publicação. A direção da revista efetuará, em primeiro lugar, uma avaliação inicial da pertinência do texto face à linha

editorial da revista, do cumprimento das normas formais de apresentação estipuladas neste documento, bem como da relevância e acessibilidade do artigo para o público-alvo da revista. Posteriormente, os artigos serão submetidos a um processo de arbitragem científica por especialistas, em regime de dupla avaliação anónima. Por esta razão, o nome do(s) autor(es) não deverá(ão) ser apresentado(s) em qualquer parte do texto. Os autores devem evitar citar-se a si mesmos; porém, quando citados, devem substituir, quer o seu nome, quer o título do(s) trabalho(s) citado(s) pela palavra "AUTOR".

6. Os artigos devem ser acompanhados por um título e por um resumo até 150 palavras no idioma do artigo e, se possível, também no outro idioma da revista. Deve incluir, também, até cinco palavras-chave. Os textos enviados para publicação devem incluir, no corpo do email de envio, o nome, a afiliação institucional e o(s) endereço(s) de correio eletrónico do(s) autor(es).

7. Se necessário, aos autores poderá ser solicitada a revisão dos textos, de acordo com as revisões e os comentários dos avaliadores científicos. A decisão final de publicação será da responsabilidade da direção da revista, tendo em consideração os comentários resultantes da arbitragem científica. A decisão final sobre a publicação do texto será comunicada aos autores será comunicada até três meses após a data final do convite à apresentação de propostas.

8. Os artigos devem ser enviados em ficheiro MS Word (Windows ou Mac) – utilizando um dos modelos disponibilizados pela revista – ou LaTeX. Os textos devem ser redigidos em páginas A4, com espaçamento simples e margens amplas, tipo de letra Times New Roman 12 pt (incluindo citações e excertos, notas, referências bibliográficas, tabelas e legendas). Não é permitido o envio de ficheiros PDF. Sempre que necessário, em casos especiais, devem ser utilizados os tipos de letra seguintes:

    - Em concordâncias e transcrições deve utilizar-se Courier;
    - Os carateres fonéticos devem utilizar um tipo de letra IPA (SIL IPA93 Manuscript ou Doulos);
    - Os símbolos especiais devem utilizar um tipo de letra Símbolo (se possível, utilizar apenas um tipo de letra especial ao longo do texto);
    - No caso de textos escritos em idiomas com um sistema de escrita diferente do romano (e.g. Árabe, Mandarim, Russo), pode ser necessário um tipo de letra especial para essa língua;
    - Para assinalar palavras em itálico, deve utilizar-se itálico e NÃO sublinhados (os sublinhados estão reservados a exemplos e transcrições linguísticas).

9. O artigo deve ser organizado em secções e, se necessário, subsecções não numeradas, com títulos adequados. Uma vez que a revista é publicada apenas online, o(s) autor(es) pode(m) incluir anexos e apêndices longos, ilustrações, fotografias e tabelas a cores, e integrar ficheiros de som e hiperligações.

10. Figuras, tabelas, gráficos, imagens e desenhos devem ser inseridos no respetivo local no texto e enviados como ficheiro separado (utilizando o nome e o número correspondente como nome de ficheiro), num dos principais formatos de imagem existentes (JPEG/JPG, TIFF, PNG, PDF). Os ficheiros de imagem devem apresentar uma resolução de pelo menos 300 dpi, ser numerados sequencialmente e estar acompanhados por uma legenda curta, mas descritiva. As legendas devem ser colocadas a seguir às tabelas, figuras, imagens, gráficos ou desenhos correspondentes no corpo do texto, mas não devem ser incluídas no(s) ficheiro(s) em separado. Sempre que necessário, as tabelas devem apresentar os títulos das colunas.

11. As transcrições devem ser apresentadas em tipo de letra Courier, numeradas por turnos e não por linhas, e utilizar pontuação consistente. Sempre que for necessário alinhar elementos com outros elementos em linhas anteriores ou seguintes, deve utilizar-se vários espaços para efetuar o alinhamento. As transcrições devem ser fornecidas como ficheiros de imagem individuais (e.g. JPEG/JPG, TIFF, PNG, PDF), devendo o nome dos ficheiros corresponder ao número da transcrição.

12. As abreviaturas devem ser explicadas no texto, por extenso, apresentadas de modo consistente e mencionadas claramente no texto. Deve utilizar-se o tipo de letra Times New Roman 12 pt sempre que possível, exceto se for necessário um tipo de letra mais pequeno.

13. Deve evitar-se o recurso a notas; porém, quando utilizadas, é preferível utilizar notas de fim. Estas devem ser numeradas sequencialmente ao longo do artigo, começando por 1, e colocadas no final do artigo, imediatamente antes das Referências bibliográficas.

14. Os textos devem indicar claramente as fontes e as referências bibliográficas dos trabalhos citados. O(s) autor(es) deve(m) certificar-se de que as referências utilizadas são precisas, exaustivas e estão claramente identificadas, devendo obter a devida autorização dos respetivos autores para reproduzir ilustrações, tabelas ou figuras. O(s) autor(es) é(são) responsável(eis) pela obtenção da devida autorização para reproduzirem parte de outro trabalho antes de enviarem o seu texto para publicação. A **LL/LD** não se responsabiliza pelo incumprimento dos direitos de propriedade intelectual.

15. As referências no próprio texto devem indicar o apelido do(s) autor(es) ou organizador(es), ano de publicação e, sempre que necessário, os números de página imediatamente após o material citado, de acordo com o estilo seguinte: Coulthard e Johnson, 2007; Coulthard e Johnson (2007); Coulthard e Johnson (2007: 161). Sempre que um trabalho possuir dois autores, deve indicar-se os dois apelidos em todas as citações do mesmo. Os trabalhos com mais de dois autores citam-se indicando o apelido do primeiro autor, seguido de *et al.* (Nolan *et al.* (2013)). O autor, a data e o número de página podem ser repetidos, sempre que necessário, não devendo utilizar-se "ibid.", "ibidem" ou "op. cit.". Ao citar(em)

informações específicas de um determinado trabalho, o(s) autor(es) deve(m) indicar o intervalo de páginas respetivo, e.g.: Caldas-Coulthard (2008: 36–37), NÃO Caldas-Coulthard (1996: 36 ff.).

16. As citações devem ser claramente assinaladas, utilizando aspas. Deve evitar-se a utilização de citações longas; porém, quando utilizadas, as citações com mais de 40 palavras devem ser formatadas como um novo parágrafo, o texto deve ser indentado 1 cm à esquerda e à direita das margens, utilizando um tipo de letra mais pequeno (11 pt). A referência bibliográfica deve ser apresentada entre parênteses a seguir ao sinal de pontuação final da citação. Não deve utilizar-se qualquer pontuação após a citação, e.g.:

> As palavras usadas para expressar o direito, nas várias línguas indo-européias, têm sua formação na raiz "dizer". Dizer a verdade. Do ponto de vista da concepção de língua, que subjaz à concepção de direito, os profissionais do direito operam com uma noção de verdade fundada na relação entre a linguagem e o mundo, com base num conceito de seleção biunívoca e quase de especularidade ou, pelo menos, de correspondência. (Colares, 2010: 307)

Se o autor e a data forem apresentados na introdução à citação, deve apresentar-se apenas o(s) número(s) de página no final da citação, antecedidos de "p.":

Conforme descrito por Colares (2010):

> As palavras usadas para expressar o direito, nas várias línguas indo-européias, têm sua formação na raiz "dizer". Dizer a verdade. Do ponto de vista da concepção de língua, que subjaz à concepção de direito, os profissionais do direito operam com uma noção de verdade fundada na relação entre a linguagem e o mundo, com base num conceito de seleção biunívoca e quase de especularidade ou, pelo menos, de correspondência. (p. 307)

17. As citações devem ser apresentadas no idioma do texto enviado para publicação. Se a citação tiver sido traduzida do original pelo(s) autor(es), deverá apresentar-se a citação original numa nota de fim, com a indicação do tradutor.

18. As referências bibliográficas devem ser colocadas no final do texto. A secção de Referências deve incluir uma lista de todas as referências citadas no texto, e apenas estas, ordenadas alfabeticamente por apelido do (primeiro) autor/editor. Quando existirem várias publicações do mesmo autor, estas devem ser ordenadas por data (da mais antiga para a mais recente). Se forem citadas várias obras de um mesmo autor, publicadas no mesmo ano, estas devem ser diferenciadas utilizando letras minúsculas a seguir ao ano, e.g. 1994a, 1994b, e não 1994, 1994a. As referências a livros devem incluir o local da publicação e a editora. As referências a capítulos de livros e artigos publicados em revistas devem incluir os respetivos números de página. No caso de artigos publicados em revistas, deve indicar-se, ainda, o volume e o número, não devendo o nome das revistas ser abreviado. Sempre que aplicável, devem ser indicados os URL de referência.

As referências correspondentes a casos e boletins jurídicos devem ser indicadas numa lista própria, após as Referências.

19. Em suma, deverá observar-se os exemplos que se seguem, incluindo as convenções relativas a maiúsculas, minúsculas e pontuação:

    Livros
    Coulthard, M. e Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. Londres e Nova Iorque: Routledge.

    Mota-Ribeiro, S. (2005). *Retratos de Mulher: Construções Sociais e Representações Visuais no Feminino*. Porto: Campo das Letras.

    Capítulos de livros
    Machin, D. e van Leeuwen, T. (2008). Branding the Self. In C. R. Caldas-Coulthard e R. Iedema (org.) *Identity Trouble: Critical Discourse and Contested Identities*. Basingstoke e Nova Iorque: Palgrave Macmillan.

    Artigos de revistas
    Cruz, N. C. (2008). Vowel Insertion in the speech of Brazilian learners of English: a source of unintelligibility?. *Ilha do Desterro* 55, 133–152.

    Nolan, F., McDougall, K. e Hudson, T. (2013). Effects of the telephone on perceived voice similarity: implications for voice line-ups. *The International Journal of Speech, Language and the Law* 20(2), 229–246.

    Dissertações e Teses
    Lindh, J. (2010). *Robustness of Measures for the Comparison of Speech and Speakers in a Forensic Perspective*. Tese de doutoramento. Gotemburgo: Universidade de Gotemburgo.

    Websites
    Caroll, J. (2004). Institutional issues in deterring, detecting and dealing with student plagiarism. *JISC online*, http://www.jisc.ac.uk/publications/briengpapers/2005/pub_plagiarism.aspx, Acesso em 14 de novembro de 2009.

20. As provas para verificação e correção serão enviadas aos primeiros autores dos textos. Após a receção das provas, os autores deverão verificar a eventual existência de erros introduzidos durante o processo de edição. O conteúdo dos textos não deverá ser alterado nesta fase. As provas revistas devem ser enviadas tão brevemente quanto possível, normalmente no prazo de duas semanas após a receção.

21. Ao enviarem artigos para publicação, os autores cedem à revista o direito de publicar e republicar o texto nos dois idiomas da revista. Porém, os autores mantêm os direitos sobre o texto, pelo que, se desejarem republicar o artigo, terão apenas que informar a direção da revista.