Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 95 (2013) 587 – 593

Procedia
Social and Behavioral Sciences

5th International Conference on Corpus Linguistics (CILC2013)

CIVIL Corpus: Voice Quality for Speaker Forensic Comparison

Eugenia San Segundo*, Helena Alves, Marianela Fernández Trinidad

Phonetics Laboratory, Spanish National Research Council (CSIC), Madrid 28037, Spain

Abstract

The most frequent way in which criminals disguise their voices implies changes in phonation types, but it is difficult to maintain them for a long time. This mechanism severely hampers identification. Currently, the CIVIL corpus comprises 60 Spanish speakers. Each subject performs three tasks: spontaneous conversation, carrier sentences and reading, using modal, falsetto and creak(y) phonation. Two different recording sessions, one month apart, were conducted for each speaker, who was recorded with microphone, telephone and electroglottography. This is the first (open-access) corpus of disguised voices in Spanish. Its main purpose is finding biometric traces that remain in voice despite disguise.

Keywords: Corpus; CIVIL; disguise; voice; phonation types; forensics; falsetto; creak(y); phonetics

1. Introduction

Firstly, we will provide a definition of Forensic Phonetics and then a description of phonation types relevant to this research.

1.1. What is Forensic Phonetics?

We can find many possible definitions of Forensic Phonetics (e.g. Künzel 1987; Eriksson 2005; Watt 2010). Yet the main idea which underpins all of them is basically the same: Forensic Phonetics is the application of general phonetic knowledge to legal problems, for example contributing to the identification of a speaker. Other possible

* Corresponding author. Tel.: +34-916022940
E-mail address: eugeniasansegundo@gmail.com

doi:10.1016/j.sbspro.2013.10.686

tasks which a forensic phonetician can carry out include the design of voice line-ups, the creation of a speaker's vocal profile, or activities related to speech enhancement, disputed utterances, etc. However, speaker voice comparison is by far the most typical task associated with Forensic Phonetics. In such a case, the expert must compare one or several speech samples of an unknown speaker (the offender) with one or several samples of known origin (the suspects). The objective of this comparison is to answer the question "How much more likely are the observed properties of the known and questioned samples under the hypothesis that the questioned sample has the same origin as the known sample than under the hypothesis that it has a different origin?" (Morrison 2013: 265).

1.2. Voice Quality: types of phonation

Abercrombie (1967:91) defines voice quality as "those characteristics which are present more or less all the time that a person is talking [...] a quasipermanent quality running through all the sound that issues from the mouth". The term "voice quality" can be broadly understood as the acoustic-perceptive result of laryngeal and supralaryngeal behavior or, more narrowly, considering only laryngeal behavior. The possible laryngeal configurations (tension, elongation, longitudinal closing degree and compression on the midline of the vocal folds during phonation) produce different voice types. We describe below the different voice types resulting from these possible glottic configurations.

According to Hollien (1974), Laver (1980) and Titze (2000) there are three possible registers: modal voice, creak/pulse register and falsetto. See Table 1 and Figure 1.

Table 1. Different types of phonation and their vocal folds configuration.

Types of phonation	Adductive tension	Medial compression	Longitudinal tension
Falsetto	- adducted	+ tense	elongated
Modal voice	adducted	tense	-------------
Creak	+ adducted	- tense	shortened

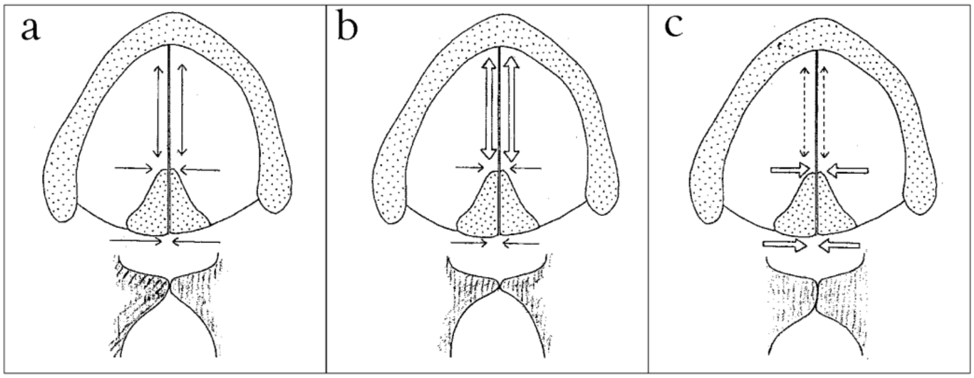

Figure 1. Diagrammatic view of vocal folds configuration in a) modal voice, b) falsetto and c) creak [Adapted from Hewlett, N., & Beck, J. 2006)]

Modal voice is the most common and spontaneous register. The vibration of the vocal folds is periodic and the undulation is wide. The pressure (longitudinal, adductive and medial) applied by the vocal folds is moderate on this voice type.

The phonation type known as creak (or pulse register) is the lowest of the three. In this configuration the vocal folds are longitudinally lax and shortened, so their mass increases and they vibrate with a fundamental frequency lower than the modal voice, generally below 70-80 Hz. It has been described in perceptual terms as, for example, "a rapid series of taps, like a stick being run along a railing" (Catford 1964:32). If creak is combined with modal voice, the result is creaky voice.

In the case of falsetto, the mass of the vocal folds intervening in the phonation is greatly reduced. The vocal folds are on their longitudinal maximum tension, while the medial and adductive tensions are moderate. As a consequence of the drastic reduction in vocal fold thickness and their high longitudinal tension rate, vibration is very fast and therefore fundamental frequency rises, resulting in a voice of higher pitch (Hollien 1974, Laver 1980). Falsetto is perceived as a "piping voice" and this involves a reduction in voice timbre, that is, a loss of sonorous energy.

2. Literature Review

2.1. Disguise as a problem in Forensic Phonetics

There are several non-electronic procedures through which the voice may be disguised. The most frequent ones are:

- Changes in the type of phonation, which imply that the sound source is modified. See Orchard & Yarmey (1995), Yarmey et al. (2001), Evans & Foulkes (2009) for whisper; Endres, Bambach & Floss (1971) and Wagner & Köster (1999) for falsetto; Hirson & Duckworth (1993) and Moosmüller (2001) for creak. The experiments in Künzel (2000) and Alves et al. (2012) include both falsetto and creak registers.
- Modification of some prosodic features, such as: habitual pitch, intonation or speech rate. For a review, see Dellwo, Ramyead & Dancovicova (2009) and Dellwo, Kolly & Leemann (2012).
- Modification of supraglottic cavities through the introduction of foreign bodies (Molina de Figueiredo & Souza Britto 2000 and Horga 2002) or techniques that interfere with normal transmission of speech, such as pinching one's nose (Gil & San Segundo 2013) and other methods (Rose &Simmons 1996, Llamas et al. 2008).
- Foreign accent disguise, this is, segments or suprasegments of a given language are deliberately pronounced so that the speakers are perceived as foreigners or so that they can hide certain features of their mother-tongue. For more information on this topic, see Zhang & Tan (2008), Tate (1979), Markham (1999), Storey (1996), Moosmüller (2006), Simpson & Neuhauser (2009, 2010).

Most criminals do not tend to combine these techniques, at least in a conscious way (cf. Masthoff, 1996). The most frequent form of disguise involves laryngeal modifications. Through this procedure individual phonic features can be masked. Results from the literature review show that changes within types of phonation make successful speaker identification most difficult. However, as it has been previously acknowledged (Künzel 2000), they are especially difficult to maintain for long periods of time.

We have carried out a literature review of the existing oral corpora designed for forensic-phonetic purposes and we have found a lack of databases including disguised voice samples. Among the different corpora reviewed, speakers of different languages and varieties have been recorded: Chinese (Zhang & Morrison, 2011), Standard Southern British English (Nolan et al. 2009), Scottish English (Watt & Yurkova, 2007), Dutch (Papp et al. 2011) Swiss German (Dellwo & Leemann 2012) and native and non-native German speakers imitating foreign accents (Neuhauser, 2011). Although most of these corpora are large-scale databases fulfilling the most common forensic criteria, such as providing non-contemporaneous recordings, only Papp et al. (2011) and Neuhauser (2011) have actually been designed in order to tackle the issue of voice disguise. The former have created a corpus of drug-influenced Dutch speakers, whereas the latter has collected a corpus within a project investigating foreign accent imitation from a production/perception perspective.

3. Corpus design

On completion, the CIVIL oral corpus will consist of a hundred recorded voices (50 women and 50 men) of which the 80 samples most representative of the distinct phonation types will be selected. The speakers are aged between 20 and 35 years, and their mother tongue is Standard European Spanish. These recruitment criteria have been established in order to achieve a more homogeneous corpus and to avoid isolated samples of different Spanish linguistic varieties.

Each speaker is required to perform three tasks: semi-spontaneous conversation reading of words on carrier sentences and reading of texts for each of the three phonation types (modal, falsetto and creak/y). During the first task, a conversation is held between the researcher and the speaker about topics related to cinema, music, work, etc., with the aim of eliciting as natural a sample as possible. The duration of this task ranges between three and four minutes depending on the disguise capacity of the speaker. The second task requires the speaker to read 33 words in turn, within the carrier sentence "Diga 'CV.CV.CV despacio varias veces", with a pause between each sentence. The words are mostly made up of three consonant-vowel syllables and are stressed on the third-to-last syllable. This list is shown on the Table 2. These words have been chosen in order to perform, in a further step, an acoustic analysis of the [a] vowel. The analyzed vowels must be in the stressed syllable and must be between two voiced, non-nasal consonants, as the nasal consonants can produce nasalization of the contiguous vowels, and this could modify laryngeal behavior. Finally, subjects read the two phonetically balanced texts presented in Table 3. The total duration of the recordings is about 10 minutes for each phonation type. Two different recording sessions, one month apart, were conducted for each speaker in order to take into account the temporal variability of their voice, such as pitch change depending on the time of the day, the emotional state of the speaker or the conversation topic among many other factors.

Table 2. List of words.

Bala	Gálica	Rábano
Bálago	Gálico	Rábica
Bálamo	Gállara	Rábico
Bálano	Gáraba	Rábida
Bávara	Gárrula	Rábido
Bávaro	Gárrulo	Rábula
Dádiva	Gávilos	Váguido
Gábata	Lábaro	Válida
Gádido	Ládano	Válidamente
Gálata	Lávala	Válido

Table 3. Texts.

Text 1	Text 2
Hay algo ahí, en el aire, que cambia el sentido de las cosas. Ese viento suave vuela, te toca la cara, mientras cuentas las hojas de los árboles. El agua corre buscando los campos. Al abrir las puertas de mi casa pienso: este país, una mañana más. A mi edad, comienzan a faltarme las fuerzas, ya casi no soy joven, y la muerte de mi mujer en la guerra me pesa mucho. Cuando el cuerpo llega a esa hora, la ciencia de los doctores no logra detener el paso del tiempo. De niño, allá en mi tierra, solía pasarme los días revolviendo de un lado a otro. Poco a poco, los coches de la ciudad fueron llamando mi atención. Mi madre decía que tuviera cuidado, pero yo me creía muy mayor, así que no tenía ni interés ni tiempo para mi propio signo. Pero sigo, es cierto, cuántas cosas buenas encontré entre su gente. Si cuento los queridos veranos de entonces, no son siete, ni nueve, ni veinte. Debe ser que soy niño de nuevo en este cuerpo triste.	El joyero Federico Vanero ha sido condenado por la Audiencia de Santander a ocho meses de arresto mayor y cincuenta mil pesetas de multa por un delito de compra de objetos robados. La vista oral se celebró el miércoles pasado y, durante ella, uno de los fiscales, Carlos Valcárcel, pidió para el joyero tres años de prisión menor y una multa de cincuenta mil pesetas. Gracias a las revelaciones de Vanero de hace dos años y medio se a descubrir la existencia de una sospechosa mafia policial en España, parte de la cual se vio envuelta en el llamado "caso El Nani".

From each session we obtained six audio files, two for each phonation type, containing both the voice signal and electroglottographic signal. Specifically three files were obtained containing the voice signal through a microphone (left channel) and through a telephone (right channel), and three more files with the microphone signal (left channel) -to be used as reference- and the electroglottographic signal (right channel). Importantly, all recordings were carried out in the recording booth of the Phonetics Laboratory at the CSIC (Spanish National Research Council) with the same equipment (Table 4) of sufficient quality for further analysis. These files are in an uncompressed format (wav with PCM conversion type) with a 44100 Hz sample rate and 16 bit resolution.

Table 4. Equipment.

Equipment	Models	Manufacturer
Condenser microphone	E6i Omnidirecctonal	Countryman
Audio Interface	UA-25EX	Roland
Software	Adobe Audition 1.0 for Windows	
Telephone 1	IP Phone 7912 Series	CISCO
Telephone 2	Galaxy 5500	Samsung
Electroglottograph	EG2-PCX2	Glottal Enterprises

Each of the collected signals will be used for the study of particular voice features. For example, the electroglottographic signal allows the measurement of the vocal folds' opening and closing times. As this is a forensic-phonetic corpus, and most criminal offenses are obtained from wiretapping recordings, the influence of the telephone distortion will be studied. Information considered irrelevant to the intelligibility of the message is deleted from the telephone signal, producing variation in the formant position, as explained by Künzel (2001), Nolan (2002), Byrne and Foulkes (2004), Chen et al. (2009), and Rosas and Sommerhoff (2009). In addition, telephone communication is characterised by an increase in fundamental frequency, as studied by Summers et al. (1989), Hirson, French and Howard (1995) or French (1998).

Currently, the corpus comprises 60 speakers, 32 women and 28 men with an average age of 26 years old, with an average separation between sessions of 32 days; therefore, there are a total of 120 samples of spontaneous speech and reading for each phonation type (modal, falsetto and creak/y). See Table 5.

Table 5. Corpus data.

Sex	Number of speakers	Age [years]	Separation between sessions [days]
Men	28	25.1	34.3
Women	32	26.5	30.7
Total	60	25.9	32.4

4. Final remarks

This is the first corpus of disguised voices in Spanish. When it is finished (with a total of 100 speakers) it will be an open-access research resource. We have designed this corpus primarily to identify the features which remain despite a speaker's attempts to disguise his or her voice. Our main finding so far is that speaker recognition is significantly easier under falsetto than under creaky condition, at least in female voices (Alves et al. 2012). The results for male voices can be found in Fernández Trinidad, Infante & Alves (2013). Our next objectives are: to test the influence of the telephone channel on speaker recognition and to investigate those laryngeal characteristics which cannot be disguised.

Acknowledgements

This research has been made possible by grants awarded by the Spanish Ministry of Science and Innovation (*FFI10-21690/Cualidad de voz*), Spanish Ministry of Education (*Beca FPU – Programa Nacional de Formación de Profesorado Universitario, BOE 11-07-2009*) and the Spanish National Research Council- CSIC (*Jae-Predoctoral-Junta para la Ampliación de Estudios, BOE 03-06-2011*). Thank you very much also to all the colleagues in the Laboratory of Phonetics (CSIC) directed by Juana Gil.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Alves, H., Fernández Trinidad, M., Gil Fernández, J., Infante, P., Lahoz, J.M., Pérez Sanz, C. & San Segundo, E. (2012). Disguised voices: a perceptual experiment, 3rd European Conference of the International Association of Forensic Linguistics. Oporto, 15-18 October 2012.

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

Catford, J.C. (1964). Phonation types: the classification of some laryngeal components of speech production. In D. Abercrombie et al. (Eds.), *Honour of Daniel Jones* (pp. 26-37). London: Longmans.

Chen, N. F., Shen, W., Campbell, J. & Schwartz, R. (2009). Large scale analysis or formant frequency estimation variability in conversational telephone speech, Proceedings of Interspeech 2009, Brighton.

Dellwo, V., Ramyead, S. & Dancovicova, J. (2009). The influence of voice disguise on temporal characteristics of speech, International Association for Forensic Phonetics and Acoustics Conference. University of Cambridge, UK.

Dellwo, V., Kolly, M.J & Leemann, A. (2012). Speaker identification based on temporal information: a forensic phonetic study of speech rhythm and timing in the Zurich variety of Swiss German, International Association for Forensic Phonetics and Acoustics Conference. Santander, Spain.

Dellwo, V., Leemann, A., & Kolly, M-J. (2012). Speaker idiosyncratic rhythmic features in the speech signal, Proceedings of Interspeech. Portland (OR), USA.

Endres, W., Bambach, W. & Flosser, G. (1971). Voice spectrograms as a function of age, voice disguise and voice imitation, *Journal of the Acoustical Society of America*, 49, 1842-1848.

Eriksson, A. (2005). Tutorial on forensic speech science. Part I: Forensic phonetics. In INTERSPEECH-2005. Eurospeech. Proceedings of the 9th European Conference on Speech Communication and Technology. Lisboa.

Evans, I. & Foulkes, P. (2009). Speaker identification in whisper, International Association for Forensic Phonetics and Acoustics conference. University of Cambridge, UK.

Fernández Trinidad, M., Infante, P. & Alves, H. (2013). Falsetto as a disguise method in male voices, 31st International Conference AESLA. Universidad de La Laguna, Tenerife, Spain.

French, J. P. (1998). Mr Akbar's nearest ear versus the Lombard reflex: a case study for forensic phonetics. *Forensic Linguistics*, 5(2), 58–68.

Gil, J. & San Segundo, E. (2013). El disimulo de la cualidad de voz en fonética judicial: un estudio perceptivo para un caso de hiponasalidad. In A. Palacios (Ed.) *Panorama de la Fonética Española Actual* (In Press).

Hewlett, N., & Beck, J. (2006). *An introduction to the science of phonetics.* Mahwah, NJ: Lawrence Erlbaum.

Hirson, A. & Duckworth, M. (1993). Glottal fry and voice disguise: A case study in forensic phonetics. *Journal of Biomedical Engineering*, 15, 193-208.

Hirson, A., French, J. P. y Howard, D. (1995). Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics. In J. Windsor Lewis (Ed.), *Studies in General and English Phonetics in Honour of Professor J. D. O'Connor* (pp. 230-240). Londres: Routledge.

Hollien, H. (1974). On vocal registers, *Journal of Phonetics*, 2, 125-143.

Horga, D. (2002). The influence of bite-blocks on continuous speech production. In A. Braun y H. R. Masthoff (Eds.), *Phonetics and its applications* (pp 143-152). Stuttgart: Steiner.

Künzel, H.J. (2000). Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics*, 7 (2),149-179.

Künzel, H. J. (2001). Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguistics*, 8, 80–99.

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

Llamas, D., Harrison, P., Donnelly, D. & Watt, D. (2008). Effects of different types of face coverings on speech acoustics and intelligibility. *York Papers in Linguistics*, 9, 80-104.

Markham, D. (1999). Listeners and disguised voices: The imitation and perception of dialectal accent, *International Journal of Speech, Language and the Law*, 6 (2), 289-299.

Masthoff, H. R. (1996). A report on a voice disguise experiment. *Forensic Linguistics*, 3(1), 160-167.

Morrison, G. (2013). Vowel inherent spectral change in forensic voice comparison. In G.Morrison & Assmann, P.(Eds.) *Vowel inherent spectral change* (pp. 263-283). Heidelberg, Germany: Springer-Verlag.

Moosmüller, S. (2001). The influence of creaky voice on formant frequency changes. *International Journal of Speech, Language and the Law*, 8 (1), 10-112.

Moosmüller, S. (2006). Articulatory avoidance. *Proc. IAFPA 2006*. Gotemburg.

Molina de Figuereido, R. and Souza Britto, H. (1996). A report on the acoustic effects of one type of disguise. *Forensic Linguistics*, 3 (1): 168-175.

Neuhauser, S. (2011). FAIC - Foreign Accent Imitation Corpus. *Proc. 20th Annual Conference of the IAFPA* 2011. Vienna, Austria.

Nolan, F. (2002). The 'telephone effect' on formants: A response. *Forensic Linguistics,* 9, 74–82.

Nolan, F., McDougall, K. de Jong, G. & Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *The International Journal of Speech, Language and the Law*, 16.1, 31-57.

Orchard, T. & Yarmey, A.D. (1995). The effects of whispers, voice sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology*, 9, 249-260.

Papp, V., Schreuder, M., Theunissen, E. & Ramakerers, J. (2011). Reference corpus of Dutch drug users I: MDMA/Ecstasy, 20th International Association for Forensic Phonetics and Acoustics Annual Conference. Vienna, Austria, 24-28 Jul 2011.

Rosas, C. & Sommerhoff, J. (2009). Efectos acústicos de las variaciones fonopragmáticas y ambientales. *Estudios filológicos*, 44, 195-210.

Rose, Ph. & Simmons, A. (1996). F-pattern variability in disguise and over the telephone. Comparisons for forensic speaker identification. In *Proceedings of the 6th Australian International Conference on Speech Science and Technology* -SST'96 (pp.127-132). Adelaida: ASSTA

Simpson, A. P. & Neuhauser, S. (2009). Enduring nature of epiphenomenal non-pulmonic sound production under disguise – a preliminary study. *Proceedings IAFPA 2009*. Cambridge, UK.

Simpson, A. P. & Neuhauser, S. (2010). The persistence of ephiphenomenal sound production in foreign accent disguise. *Proceedings IAFPA 2010*.Trier.

Summers, W. V., Johnson, K, Pisoni, D. B. & Bernacki, R. H. (1989). An addendum to "Effects of noise on speech production: acoustic and perceptual analyses". *Journal of the Acoustical Society of America*, 86, 1717–1721.

Storey, K. C. J. (1996). Constants in auditory and acoustic voice analysis in forensic speaker identification in cases of disguised voice. In H. Kniffka and S. Blackwell (Eds.). *Recent Developments in Forensic Linguistics* (pp. 203-216). Frankfurt: Lang.

Tate, D.A. (1979). Preliminary data on dialect in speech disguise. Current Issues in the Phonetics Sciences: Proceedings of the IPS-77 Congress, 9, 847-850. Amsterdam: John Benjamins.

Titze, I. R. (2000). *Principles of Voice Production,* 2nd Printing. Iowa City, National Center for Voice and Speech.

Wagner, I. & Köster, O. (1999). Perceptual recognition of familiar voices using falsetto as a type of voice disguise. *Proc. of the 14th International Congress of Phonetic Sciences*, 2, 1381–1384. San Franciso.

Watt, D. & Yurkova, J. (2007). Voice Onset Time and the Scottish Vowel Length Rule in Aberdeen English. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1521-1524. Saarbrücken, August 2007.

Watt, D. (2010). The identification of the individual through speech. In C. Llamas y D. Watt (Eds.), *Language and Identities* (pp. 76-85). Edinburgh: Edinburgh University Press.

Yarmey, D.A., Yarmey, A.L., Yarmey, M.J., & Parliament, L. (2001). Applied Cognitive Psychology, 15, Issue 3, 283-299.

Zhang, C. ,& Tan, T. (2008). Voice disguise and automatic speaker recognition. *Forensic Science International*, 175, 118-122.

Zhang, C., & Morrison, G.S. (2011). Forensic database of audio recordings of 68 female speakers of Standard Chinese. International Association for Forensic Phonetics and Acoustics Research Grant 2010.