Disguised voices: a perceptual experiment

Helena Alves
Marianela Fernández
Juana Gil
Patricia Infante
José María Lahoz
Carolina Pérez
Eugenia San Segundo

CIVIL Project

Cualidad Individual de la Voz en la Identificación de Locutores

- 2010

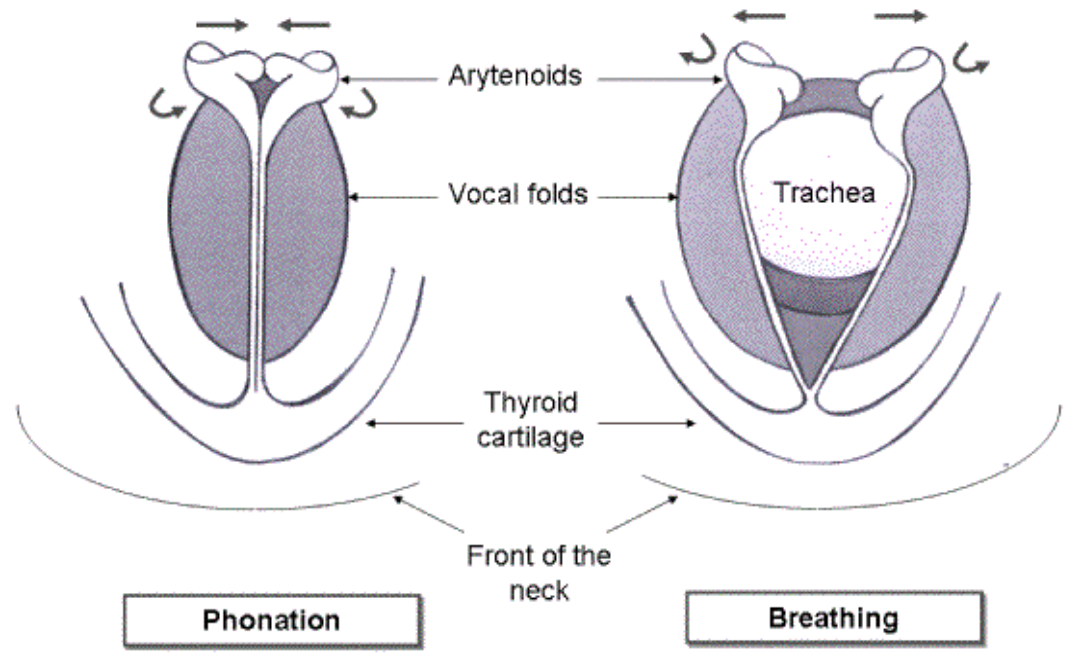
- Phonetics Lab CSIC, Madrid
- Laryngeal settings modification

CIVIL: hypotheses

- Changes in phonation = harmful for speaker recognition
- Idiosyncratic phonetic features (biometric traces):
  - Remain despite the disguise attempts
  - Some laryngeal characteristics cannot be disguised

Types of Phonation

- Phonation = vocal folds vibration



From: http://www.phys.unsw.edu.au/jw/voice.html

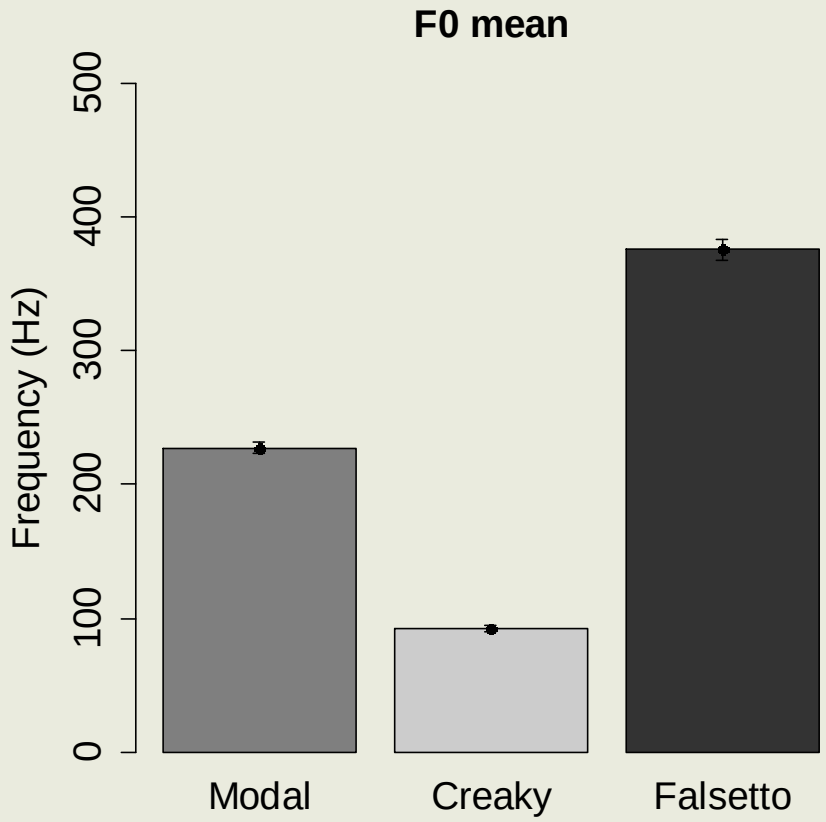
Types of Phonation

- Phonation = vocal folds vibration
- Different states of the vocal folds produce different types of phonation

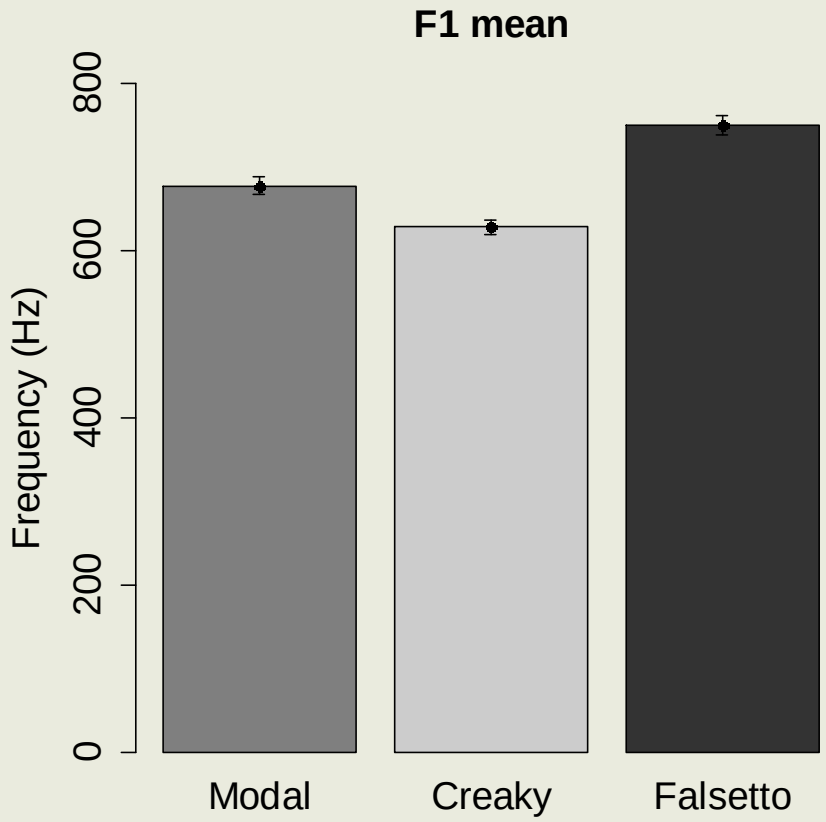
Falsetto	-adducted	+tense	elongated
Modal	adducted	tense	----------
Creak/y	+adducted	-tense	shortened



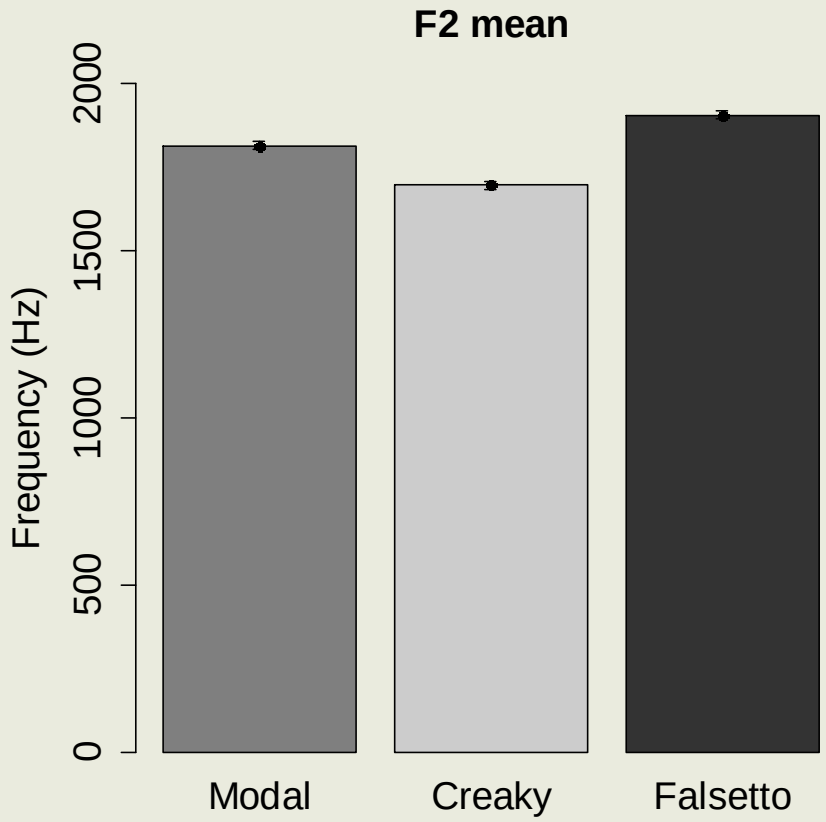
Acoustic differences of phonation types



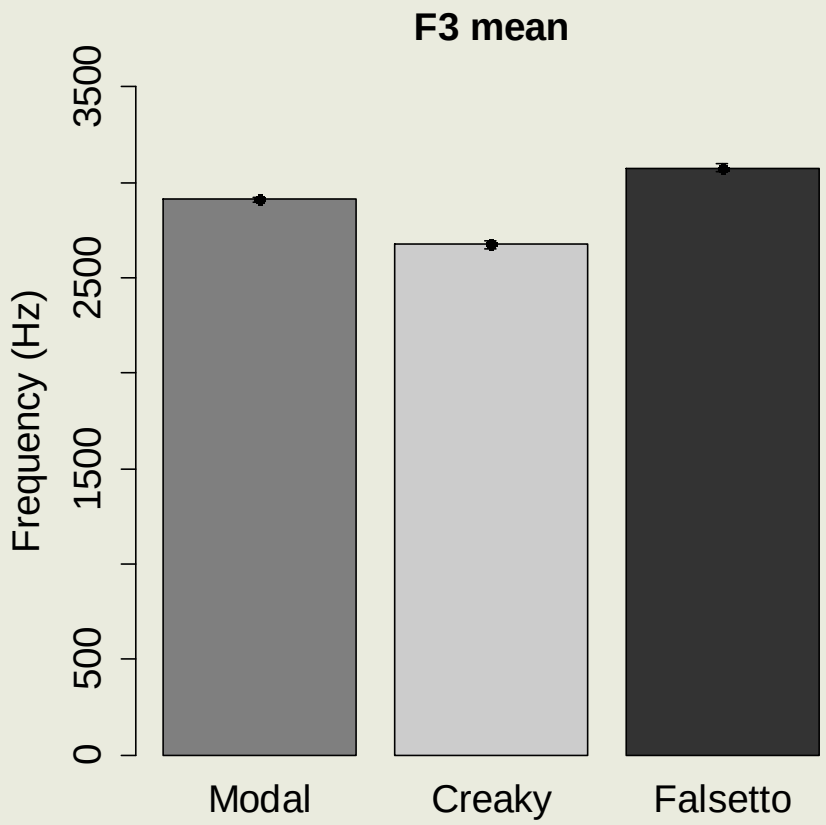
Acoustic differences of phonation types



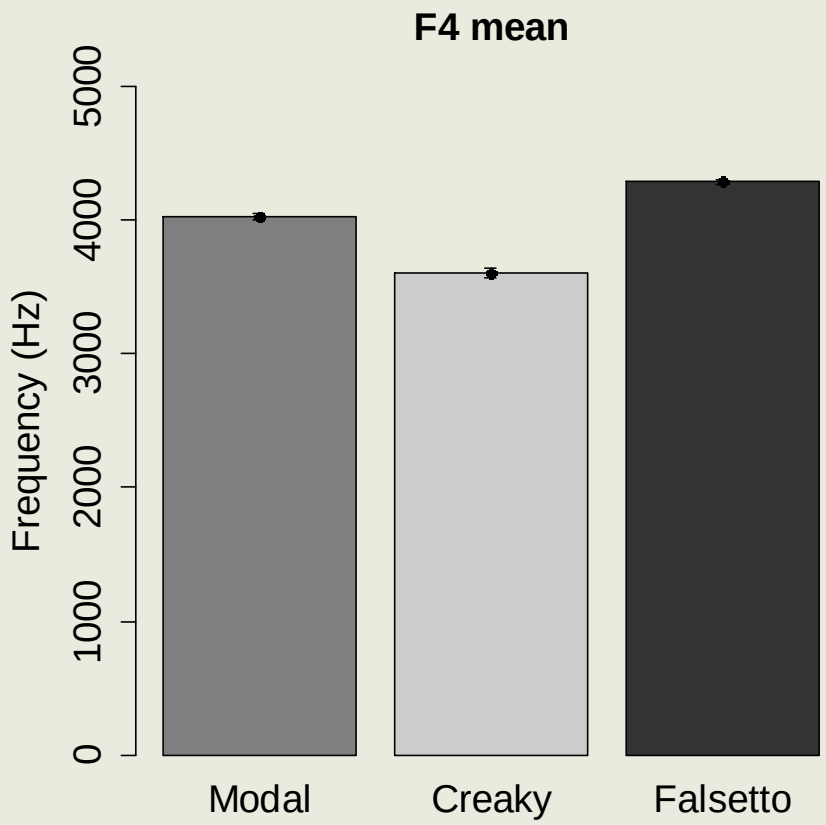
Acoustic differences of phonation types



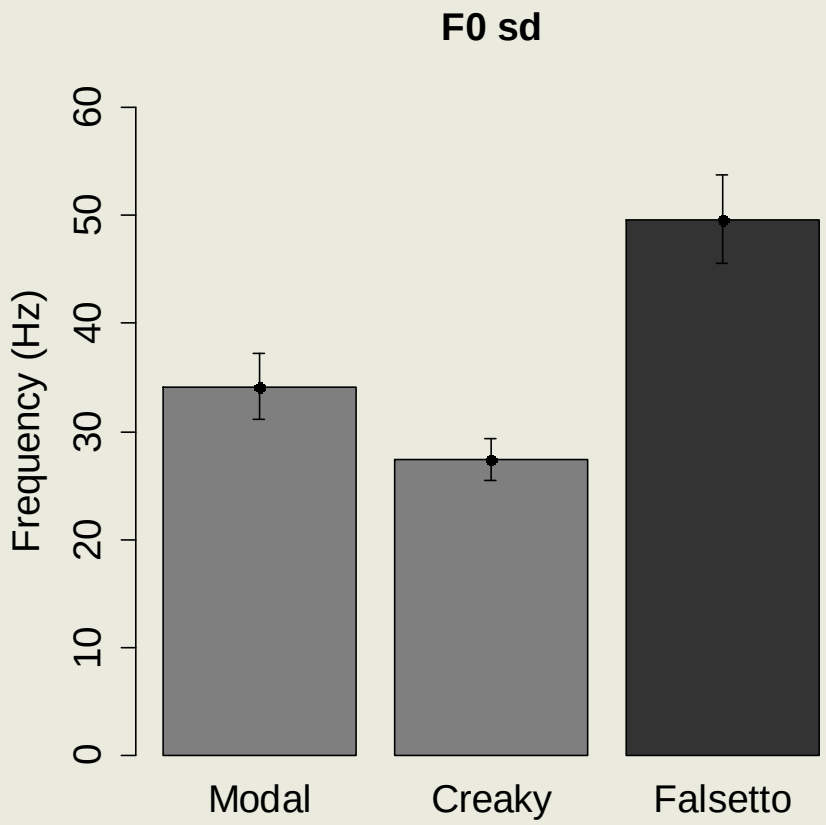
Acoustic differences of phonation types



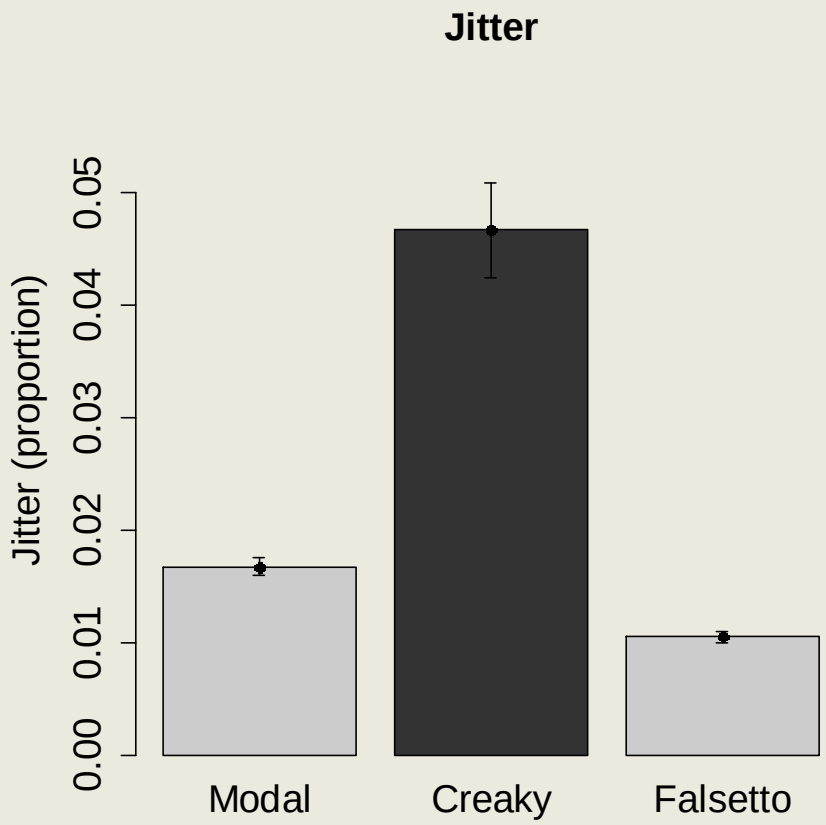
Acoustic differences of phonation types



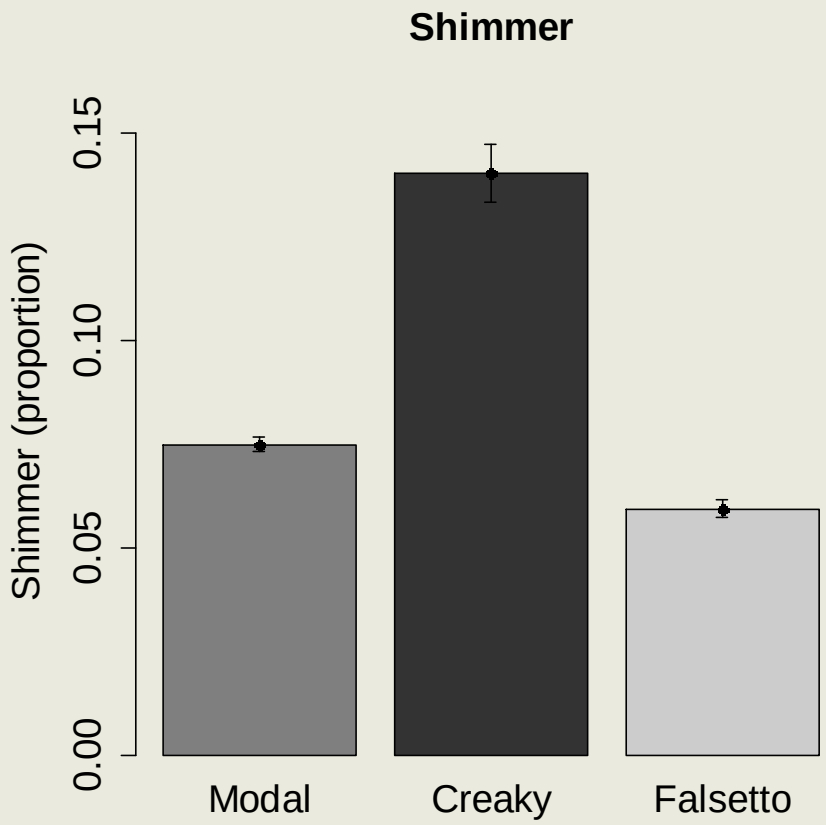
Acoustic differences of phonation types



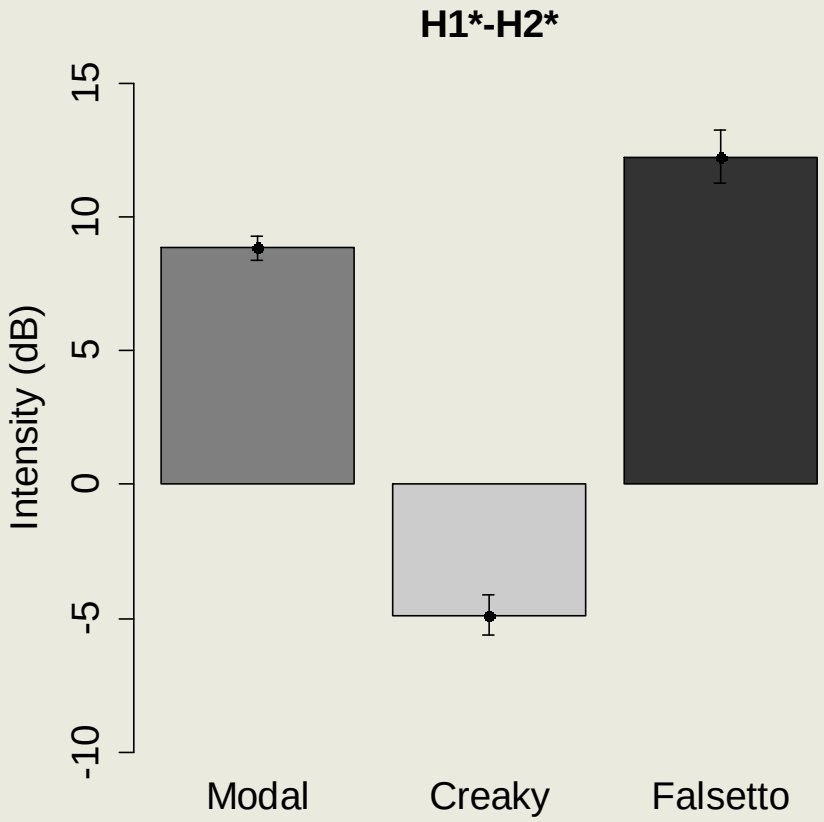
Acoustic differences of phonation types



Acoustic differences of phonation types



Acoustic differences of phonation types



Perceptual experiment RQs

Assuming that phonation types interfere with speaker recognition

- Do listeners recognise disguised voices above chance level? H1 = yes
- Is recognition more affected by a particular phonation type? H2 = falsetto
- Do expert listeners perform better than naïve listeners? H3 = no

Hypotheses

- H1: Listeners are capable of recognising disguised voices over chance level.
  - "Something" in the glottal source remains

Hypotheses

- H2: - Falsetto register is more harmful for voice recognition than creak

  Wagner & Koester (1999)

  Hirson & Duckworth (1993)

  Moosmueller (2001)

  – Different: methodology and results

  – While recognizing, better results with creak than falsetto

Hypotheses

- H3: No performance differences between experts and naïve listeners in disguised voices recognition
  - Voice perception is not analytical but holistic

Experiment design

- Corpus read sentences
- Frame sentence:

“Diga 'CV.CV.CV despacio”

- Open syllable: voice quality not affected

Speakers

- 6 female speakers Standard European Spanish
  - 25-35 years old
  - no speaking nor hearing disorders
  - mantain the phonation type

Recordings

- Equipment
  - Recording booth of the CCHS Phonetics Lab
  - Condenser microphone→ E6i Omnidireccional Earset Microphone
  - Audio Interface →UA-25EX by Roland
  - PC with the software Adobe Audition 1.0 for Windows
- Recording settings:
  - Sample Rate: 44100
  - Resolution: 16-bits
  - Channels: Mono

Test Description

- 120 triplets→ disguised voice (X) + modal voice (A) + modal voice (B)

6 speakers x 2 phonation registers x 2 listening orders x 5 distractors

- Random order of the 120 triplets
  - 0,5 seconds between voices
  - Duration of one triplet: 5-7 seconds
  - Duration of the test: 20-30 minutes

- Each item played up to two times

Subjects

- 61 (14 phoneticians + 47 non-phoneticians)
- Spanish L1
- No hearing disorders
- Not familiar with any of the speakers

Results H1

- One-sample t-test shows that:

  Recognition of disguised voices is above chance, i.e. hit rate > 0.5

  (99.9% confidence level)

  - Creaky: $t(60) = 10.04, p < 0.001$ ***
  - Falsetto: $t(60) = 10.41, p < 0.001$ ***

Results H2

HIT RATE	Creaky	Falsetto
mean	0.59	0.62
sd	0.07	0.09

Please note:

- 8% probability of getting hit rate = 0.59 by chance
- 3% probability of getting hit rate = 0.62 by chance

Results H2

- Two-sample t-test shows that:

  Hit rate is significantly better for falsetto than for creaky voice

  (95% confidence level)

  $t(112) = 2.11, p < 0.05$ *

Results H2

Results H3

HIT RATE	Phonetician	Non-phonetician
Creaky	0.60 (0.06)	0.59 (0.07)
Falsetto	0.62 (0.09)	0.62 (0.09)
No. subjects	14	47

Discussion

- Phonation types DO affect recognition:
  - Hit rates not very high (0.60)

- However, hit rates above chance

Discussion

- Speaker recognition is easier under falsetto than under creaky condition
  - Against H2
  - Possible explanation:
    - Previous studies: male voices
    - Falsetto (male) & creaky (female) introduce greater distortion with respect to normal voice than the other way round. Expectations not met:
      - Creak less expected for female voice prototype
      - Falsetto less expected for male voices prototype
    - Creak has no F0: loss of a crucial acoustic cue

Discussion

- Good performance rates might be due to:
  - musical education
  - ear training

- Rather than:
  - background in phonetics

Further research

- Enlarge corpus:
  - Male voices
  - Spontaneous speech
- Characterise phonation types (acoustically and physiologically)
- Find which traits are speaker-specific
- Check telephone signal influence
- Study temporal parameters

Bibliography

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

Gil, J. (2012). *Bibliografía temática de fonética judicial*. URL: http://www.estudiosfonicos.cchs.csic.es/fonetica/civil/Gil,J._2012_Bibliografia_Tematica_de_Fonetica_Judicial.pdf (05 junio 2012).

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

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

Hollien, H., Majewski, W. y Doherty, E. T. (1982). Perceptual identification of voices under normal, stress and disguise speaking conditions. *Journal of Phonetics, 10*, 139-148.

Köster, O., Jessen, M., Khairi, F. y Eckert, H. (2007). Auditory-perceptual identification of voice quality by expert and non-expert listeners. En *XVIth International Congress of Phonetic Sciences*, Saarbrücken, Alemania.

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

Künzel, H., Gonzalez, J. y Ortega, J. (2004). Effect of voice disguise on the performance of a forensic automatic speaker recognition system. En *ODYS-2004*, 153-156.

Ladefoged, J. y Ladefoged, P. (1980). The ability of listeners to identify voices. *UCLA Working Papers in Phonetics, 49,* 43-1

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

McGlone, R., Hollien, H. y Hollien, P. (1977). Acoustic analysis of voice disguise related to voice identification. En *Proceedings of the International Conference on Crime Countermeasures* (pp. 31-35). Kentucky: University of Kentucky.

Molina de Figueiredo, R. (2000). Algumas considerações sobre o disfarce de voz. *Estudos Lingüísticos, XXIX*, 543-548.

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

Moosmüller, S. (2001). The influence of creaky voice on formant frequency changes. *International Journal of Speech, Language and the Law, 8*(1), 100-112.

Perrot, P. y Chollet, G. (2012). Helping the Forensic Research Institute of the French Gendarmerie to identify a suspect in the presence of voice disguise or voice forgery. En A. Neustein y H. A. Patil(Eds.), Forensic Speaker Recognition. Law Enforcement and Counter-Terrorism (pp.469-503). New York: Springer

Perrot, P., Aversano, G. y Chollet, G. (2007). Voice disguise and automatic detection: Review and perspectives. En Y. Stylianou, M. Faundez y A. Esposito (Eds.), *Progress in Non-Linear Speech Processing. Lecture Notes in Computer Science* (pp. 101-117). Berlín: Springer Verlag.

Perrot, P., Preteux, C., Vasseur, S. y Chollet, G. (2007). Detection and recognition of voice disguise. Comunicación presentada en el Congreso de la *IAPFA 2007*. Plymouth.

Reich, A. (1975). *Certain Effect of Selected Vocal Disguises Upon Spectrographic Speaker Identification*. Tesis Doctoral. University of Iowa.

Reich, A. (1981). Detecting the presence of vocal disguise in the male voice. *Journal of the Acoustical Society of America, 69*(5), 1458-60.

Reich, A. y Duke, J. (1979). Effects of selected vocal disguises upon speaker identification by listening. *Journal of the Acoustical Society of America, 66*, 1023–1028.

Reich, A., Moll, K. y Curtis, J. (1976). Effects of selected vocal disguises upon spectrographic speaker identification. *Journal of the Acoustical Society of America, 60*, 919– 925.

Rodman, R. (1998). Speaker recognition of disguised voices. En M. Demirekler, A. Saranli, H. Altincay y A. Paoloni (Eds.), *Proceedings of the Consortium on Speech Technology Conference on Speaker Recognition by Man and Machine: Directions for Forensic Applications* (pp. 9-22). Ankara: COST 250 Publishing Arm.

Rodman, R. (2003). *Speaker Recognition of Disguised Voices: A Program for Research*. Raleigh, Carolina del Norte: North Carolina State University.

Rose, P. y Simmons, A. (1996). F-pattern variability in disguise and over the telephone – comparisons for forensic speaker identification. *Proceedings of the 6th Australian International Conference on Speech Science and Technology*, Adelaide.

Simpson, A. P. y Neuhauser, S. (2009). Enduring nature of epiphenomenal non-pulmonic sound production under disguise – a preliminary study. Comunicación presentada en *IAFPA-2009,* Cambridge, R.U.

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

Wagner, I. y Köster, O. (1999), Perceptual recognition of familiar voices using *falsetto* as a type of voice disguise. *Proceedings of the14th International Congress of Phonetic Sciences*, 1381–1384.

Zhang, C. (2005). *Acoustical Study on Disguised Voices*. Tesis Doctoral. Nankai University.

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

Many thanks!

Helena Alves
helena.alves@cchs.csic.es
Marianela Fernández
marianela.fernandez@cchs.csic.es
Juana Gil
juana.gil@cchs.csic.es
Patricia Infante
patricia.infante@cchs.csic.es
José María Lahoz
josemarialahoz@yahoo.es
Carolina Pérez
carolinaper@gmail.com
Eugenia San Segundo
eugenia.sansegundo@cchs.csic.es