CIVIL Corpus: Voice Quality for Forensic Speaker Comparison

Eugenia San Segundo
Helena Alves
Marianela Fernández Trinidad

Phonetics Lab. CSIC

CILC2013 – Alicante – 15 Marzo 2013
CIVIL Project

Cualidad Individual de la Voz en la Identificación de Locutores

• 2010
• Phonetics Lab CSIC

• Laryngeal settings modification

FORENSIC PHONETICS
Types of Voice Transformation (non electronic)

1) Phonation disguise:
   • whisper (Orchard y Yarmey 1995 & Yarmey et al. 2001, Evans & Foulkes 2009)

2) Prosody disguise:
   pitch, intonation, speech rate (Dellwo, Ramyead & Dancovicova 2009 & Dellwo, Kolly & Leemann 2012)
Types of Voice Transformation
(non electronic)

3) Supraglottal disguise:
• Through objects (Molina de Figueiredo & Souza Britto 2000; Horga, 2002)
• Techniques that interfere within the habitual speech transmission (Rose & Simmons 1996, Llamas et ál. 2008, Gil & San Segundo 2013)

4) Phonological system disguise:
Disguise as a Challenge in Forensic Phonetics

Most criminals do not combine all these disguising techniques (Masthoff 1996).

✓ The most frequently used is the voluntary modification of the phonation types.
✓ This kind of disguise is specially difficult to maintain for a long stretch of time (Künzel 2000).
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

Types of Phonation

- Different states of the vocal folds produce different types of phonation

<table>
<thead>
<tr>
<th></th>
<th>Falsetto</th>
<th>Modal</th>
<th>Creak/y</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>-adducted</td>
<td>adducted</td>
<td>+adducted</td>
</tr>
<tr>
<td>Tense</td>
<td>+tense</td>
<td>tense</td>
<td>-tense</td>
</tr>
<tr>
<td>Length</td>
<td>elongated</td>
<td>--------------------</td>
<td>shortened</td>
</tr>
</tbody>
</table>
Corpus CIVIL

• 31 female speakers and 27 male speakers
• Standard European Spanish
• 20-35 years old → mean 25.6 years old
• Two recording sessions → mean 29.8 days

Why? → Forensically realistic

Non-contemporaneous speech samples

( - ) Within-speaker variation
(+ ) Between-speaker variation
Corpus CIVIL

• Three tasks:
  – 3-4 minutes of conversation
  – 33 carrier sentences
  – 2 texts

• Voice Signal:
  – Microphone
  – Telephone
  – EGG

• Three Types of Phonation:
  – Modal
  – Falsetto
  – Creak/y
Electroglottograph

→ Measures the time variation of the degree of contact between the vibrating vocal folds

Pérez Sanz, C. Ajustes laríngeos y estilos de habla en radio y televisión (Ph.D.)
Recording Equipment & Settings

• **Equipment**
  – Recording booth of the CCHS Phonetics Lab
  – Condenser microphone ➔ E6i Omnidirectional Earset Audio Interface ➔ UA-25EX by Roland
  – PC with the software Adobe Audition 1.0 for Windows
  – Telephones ➔ CISCO IP Phone as emitter & Samsung Galaxy as receiver
  – Electroglottograph ➔ Glottal Enterprises EG2-PCX2

• **Settings:**
  – Sample Rate: 44100
  – Resolution: 16-bits
  – Channels for voice: Stereo (L-microphone & R-telephone)
  – Channels for EGG: Stereo (L-microphone & R-EGG)
Results so far


- Listeners’ recognition of disguised voices is above chance ($p < 0.001$ ***)
- Speakers are worse recognized when using creak than when using falsetto.
- No performance differences between experts and naïve listeners in disguised voice recognition.

FEMALE VOICES !!
Future directions…

MALE VOICES?

-Hypothesis:
Worse recognition results when using falsetto

- Expectations not met:
  - Creak less expected for female voice prototype
  - Falsetto less expected for male voices prototype
CIVIL Corpus: Voice Quality for Forensic Speaker Comparison

Thank you for your attention!

eugenia.sansegundo@cchs.csic.es
helena.alves@cchs.csic.es
marianela.fernandez@cchs.csic.es

CILC – Alicante – 15 Marzo 2013