

Albalá, M. J., Battaner, E., Gil, J., Llisterri, J., Machuca, M., Marrero, V., . . . Ríos, A. (2009). Vowel formant structure and speaker identification. A perceptual study.

CIP 2009. 3ª Conferência Ibérica de Percepção.

Guimarães, Portugal, 8-10 Julho 2009.

http://liceu.uab.cat/~joaquim/phonetics/VILE/VILE_CIP09.pdf

CIP - Guimarães 2009

3ª Conferência Ibérica de Percepção

CIP
09

Vowel formant structure and speaker identification. A perceptual study



Victoria Marrero, UNED
vmarrero@flog.uned.es



Elena Battaner
Universidad Rey Juan Carlos



Juana Gil, M.^a José Albalá
Consejo Superior de Investigaciones Científicas



Joaquim Llisterri, María Machuca,
Montserrat Marquina, Carme de-la-Mota, Antonio Ríos
Universitat Autònoma de Barcelona



Acoustic and Perceptual Study of Inter and
Intra-speaker Variation in Spanish (VILE-II)



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- **VILE Projects (*Inter-and-Intra-Speaker-Variation-in-Spanish*)**

- **1st. stage: 2001-2004**

BFF2001-2551 VILE I: Estudio **acústico** de la variación inter e intralocutor en español (2001-2004)

Acoustic Study of Inter and Intra-Speaker Variation in Spanish

- **2nd. stage: 2005-2009**

HUM2005-06980/FILO VILE II: Estudio **perceptivo** de la variación inter e intralocutor en español (2005-2009)

Perceptual Study of Inter and Intra-Speaker Variation in Spanish



UAB
Universitat Autònoma
de Barcelona

Acoustic and Perceptual Study of Inter and
Intra-speaker Variation in Spanish (VILE-II)




Universidad
Rey Juan Carlos

Vowel formant structure and speaker identification. A perceptual study

CIP
09

- Introduction
- Experimental Procedure
- Results
- Discussion
- Concluding remarks



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- **Introduction**
- Experimental Procedure
- Results
- Discussion
- Concluding remarks



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- Acoustic parameters of vowels

- F1, F2, F3, F4
- F0

Individual or linguistic?

- Literature

- High formants (F3 - F4) → Individual information
- Low formants (F1 - F2) → Vowel quality
- F0 → Vowel quality (intrinsic F0)

→ Suprasegmental variations (intonation, tone, stress)

→ Speaker identification



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- Introduction
- **Experimental Procedure**
- Results
- Discussion
- Concluding remarks



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- AHUMADA database [Ortega, González & Marrero (2000)]
 - 30 male speakers
 - Reading a phonetically balanced text
 - 3 different recording sessions
 - 1890 vowel utterances
 - Automatic extraction (Praat) + manual supervision
 - F0, F1, F2, F3 and F4
 - /i, e, a, o/
 - Context: [p, t, k, s] _ [p, t, k, s]
 - Lexically stressed / unstressed



Vowel formant structure and speaker identification. A perceptual study

CIP
09

Pairs selection:

- **Main parameter** → **maximum difference**
- **Secondary parameters** → **minimum difference**









































| Speaker | F0 | F1 | F2 | F3 | F4 |
|---------|--------|--------|--------|---------|---------|
| 42 | 87,06 | 435,48 | 838,17 | 2538,44 | 3681,13 |
| 25 | 156,82 | 454,27 | 819,61 | 2453,17 | 3680,12 |



Vowel formant structure and speaker identification. A perceptual study

Corpus: 40 pairs of short but natural stimuli

- Very different in one acoustic cue
- Very similar in the other

| | F1 | F2 | F3 | F4 | F0 |
|-----|---|---|---|---|---|
| [i] |  |  |  |  |  |
| [í] |  |  |  |  |  |
| [e] |  |  |  |  |  |
| [é] |  |  |  |  |  |
| [a] |  |  |  |  |  |
| [á] |  |  |  |  |  |
| [o] |  |  |  |  |  |
| [ó] |  |  |  |  |  |



Vowel formant structure and speaker identification. A perceptual study

CIP
09

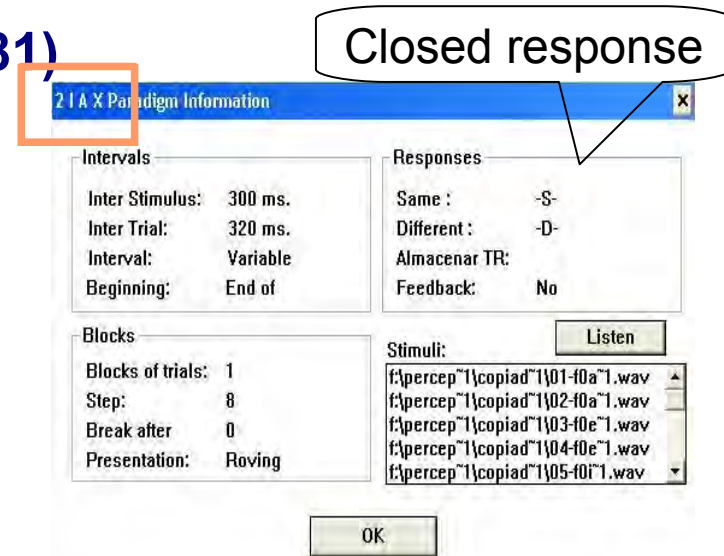
- **Software: *Paradigm*** (López Bascuas, Carrero & Serradilla, 1999)
 - Stimulus duration: 70 ms
 - Response time: free
 - Total duration: 40 min. +-

- **Procedure: 2IAX paradigm (Noreen, 1981)**

Four comparisons for each pair

- | | | |
|-----|---|-------------------------------------|
| 1-1 | } | Same stimuli |
| 2-2 | | |
| 1-2 | } | +difference in main parameter |
| 2-1 | | |
| | | -difference in secondary parameters |

- **Subjects: 46**
 - 34 W / 12 M
 - 18 - 35 years old
- **Earphones**
 - Radiobooth



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- Introduction
- Experimental Procedure
- **Results**
- Discussion
- Concluding remarks



Vowel formant structure and speaker identification. A perceptual study

CIP
09

A) General results

- Same stimuli
 - Correct rejection: 94,54%
 - False alarm: 5,46%
- Different stimuli
 - Hits: 72,97%
 - Miss: 27,03%



Vowel formant structure and speaker identification. A perceptual study

B) Results by acoustic cue (%):

| Cue | Same Stimuli | | Different Stimuli | |
|-----|--------------|------|-------------------|-------|
| | C.R. | F.A. | Hits | Miss |
| F0 | 96,11 | 3,89 | 86,27 | 13,73 |
| F1 | 94,29 | 5,71 | 68,91 | 31,09 |
| F2 | 94,40 | 5,60 | 85,30 | 14,70 |
| F3 | 94,27 | 5,68 | 64,65 | 35,35 |
| F4 | 93,61 | 6,39 | 60,03 | 39,97 |

Student t
Errors

F0/F1 0,00000000
F0/F3 0,00000000
F0/F4 0,00000000
F1/F2 0,00000000
F1/F4 0,00000006
F2/F3 0,00000000
F2/F4 0,00000000
F1/F3 0,00560266
F0/F2 0,19852823
F3/F4 0,43122926

Student t
False alarms

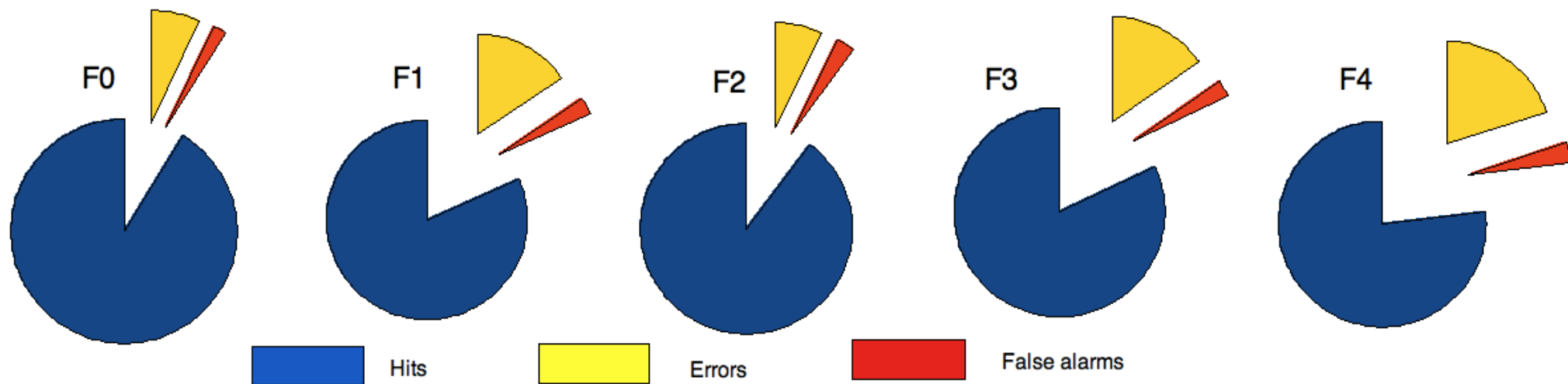
F0/F1 0,00063203
F0/F2 0,00092455
F0/F3 0,00078188
F0/F4 0,00078188
F2/F4 0,10335333
F1/F4 0,14303402
F1/F2 0,42924336
F3/F4 0,43444056
F2/F3 0,44699647
F1/F3 0,48265409



Vowel formant structure and speaker identification. A perceptual study

CIP
09

B) Results by acoustic cue (%):



Vowel formant structure and speaker identification. A perceptual study

CIP
09

C) Results by vowel (%):

| Vowel | Same Stimuli | | Different Stimuli | |
|-------|--------------|------|-------------------|-------|
| | C.R. | F.A. | Hits | Miss |
| /i/ | 94,52 | 5,48 | 73,86 | 26,14 |
| /e/ | 94,52 | 5,48 | 73,52 | 26,48 |
| /a/ | 95,37 | 4,63 | 72,99 | 27,01 |
| /o/ | 93,75 | 6,25 | 71,53 | 28,47 |

Student t

- False alarm
 - /a/-/o/: 0,0010
 - Others > 0,05
- Miss
 - No differences < 0,05

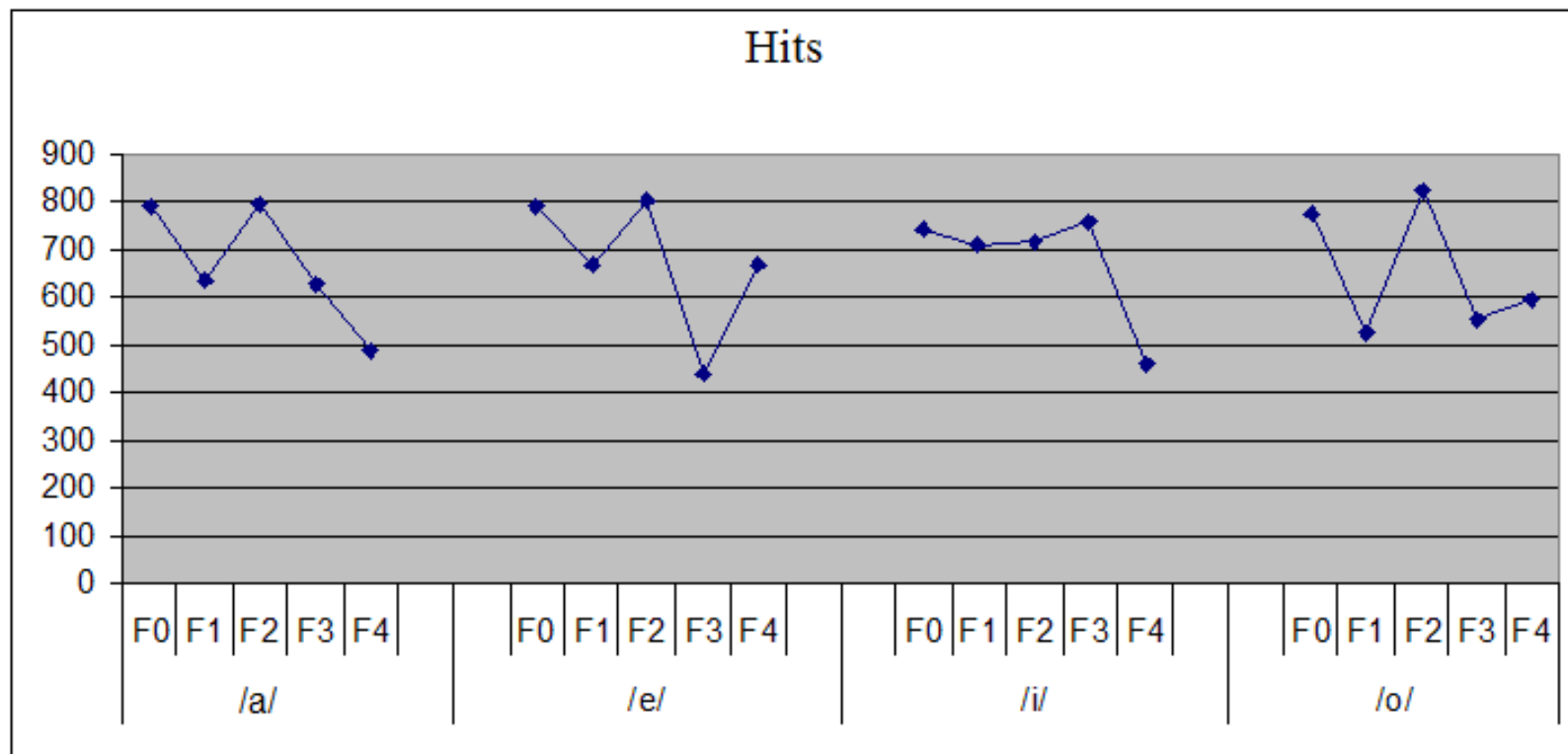


Vowel formant structure and speaker identification. A perceptual study

CIP
09

C) Results by vowel:

- Different relevance of the acoustic cues in each vowel



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- Introduction
- Experimental Procedure
- Results
- **Discussion**
- Concluding remarks



Vowel formant structure and speaker identification. A perceptual study

CIP
09

Previous studies

The analysed parameters in this study are usually used in forensic speaker identification and speaker verification.

- **F0** is one of the most frequently used parameter for voice identification (Compton 1963, Wolf 1972, Matsumoto et alii 1973, Brown 198, van Dommelen 1987, Rose 1991, Traunmüller & Eriksson 1995)
- **Formants** are also considered as main clues to identify speaker's voice (Shearme y Colmes 1959, Miller 1964, Itoh y Saito 1982, Carrell 1984, Kuwabara & Takagi 1991, Kuwabara & Ohgushi 1987, Kuwabara & Sagisaka 1995)
- **The most relevant formant can vary from one vowel to other** (van Heuvel, 1996, Nissen, 2004)
- **Vowel's F-pattern** (trajectories, distances...) are responsible for its auditory-phonetic quality; speaker-specific information too (Ingram et alii 1996, Rose 2002).
- **Other: LTAS** (Furui 1986, Pittam 1987, Pausewnag Gelfer 1989), **temporal parameters** (Brown 1981, Johnson et alii 1984), **tone contour** (Van Dommelen 1987)



Vowel formant structure and speaker identification. A perceptual study

CIP
09

Our results

- **Parameters:**
 - F0 is the most important perceptive cue
 - Coincident with most of literature
 - » Compton 1963, Wolf 1972, Matsumoto et alii 1973, Brown 1981, van Dommelen, 1987...
 - F2 has played also an important role in the identification of stimuli
 - Vowel quality over voice quality?
 - F3 and F4 seem to be less useful to identify speaker
 - F4: important acoustic cue, not so relevant perceptually
- (Marrero et al. 2008)



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- Vowels
 - /o/ presents the highest percentage of false alarms
 - /i/, /e/ /a/ are the preferred vowels for voice comparison in literature
- Parameters / vowels
 - F0 and F2
 - except /i/ → good rate of hits in F3
 - F4:
 - /e/ - /o/ show a higher rate of hits *versus* /a/ - /i/



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- Introduction
- Experimental Procedure
- Results
- Discussion
- **Concluding remarks**



Vowel formant structure and speaker identification. A perceptual study

CIP
09

- F0: best parameter in identification of speakers
- F2: best parameter in identification of vowels
 - Interaction *pitch* $\leftarrow \rightarrow$ *vowel quality*
 - New perception test with synthetic stimuli
- Perceptual weight of parameters differs across the vowels
 - F3 is particularly important for /i/
 - F4 \rightarrow /e, o/ - /a, i/
 - Opening degree?
- Perceptual correlates of physical signal
 - Important acoustic cues can be non significant from a perceptual point of view



Vowel formant structure and speaker identification. A perceptual study

CIP
09

<http://liceu.uab.es/~joaquim/VILE.html>



UAB
Universitat Autònoma
de Barcelona

Acoustic and Perceptual Study of Inter and
Intra-speaker Variation in Spanish (VILE-II)



U
Universidad
Rey Juan Carlos