Acoustic Memory, Phonological Memory and Attention Control in L2 Speech Perception

Elena Safronova and Joan C. Mora

March 24, 2012
Introduction

Cross-linguistic phonetic similarity affects:
• L2 perception
• accurate formation of L2 categories
• L2 production

Why?
• L1 phonetic prototypes function as perceptual “magnets”
• Perceptual assimilation to the most similar L1 phoneme

What to do?
• Phonetic differences between L1 and L2 must be discerned to form a new L2 phonetic category

Individuals vary in L2 speech learning and ultimate attainment in L2 pronunciation

Kuhl, 1994; Best, 1995; Flege, 1995
Factors affecting L2 speech learning

- **Age of L2 learning** (Long, 1990; Flege et al., 1995, 1999)
- **Quality and quantity of input** (Flege, 2009; Moyer, 2009)
- **Amount of L1/L2 use** (MacKay et al., 2001; Piske et al., 2001)
- **Individual differences** (Gardner, 1985): affective, personality-related AND

**Individual differences in cognitive ability:**
- **aptitude** (Skehan, 1998; Robinson, 2005)
- **phonetic talent** (Jilka et al., 2007)
- **musical ability** (Gottfried, 2007; Isaacs & Trofimovich, 2011)
- **working memory** (Kormos & Safar, 2008)
- **phonological short-term memory** (Cerviño-Povedano & Mora, 2011; MacKay et al., 2001)
The Present Study: Focus

- Focus
- Individual differences in Phonological memory (PM), Acoustic memory (AM) and Attention Control (AC)
- Inter-subject variability in weighting of phonetic cues in speech perception
- Spanish/Catalan EFL learners difficulty in target-like perception of English /iː/ and /ɪ/ contrast
  - Single Category assimilation
    English /iː/ - /ɪ/ to the Native /i/ category
  - Overreliance on duration when perceiving /iː/ and /ɪ/ contrast

Best, 1995; Cerviño-Povedano & Mora, 2009; Escudero & Boersma, 2004; Flege, 1991; Mora & Fullana, 2007
The Present Study: Aim and RQ

AIM:

- To investigate the extent to which individual differences in PM, AM and AC are related to L2 vowel perception.

RQ:

- Are L2 learners with higher cognitive ability (PM, AM and AC) better able to rely on spectral information than lower ability learners in the perception of the English tense-lax /iː/-/ɪ/ contrast?

Hypothesis

- Differing individual capacities in these cognitive skills involved in speech processing might partly explain inter-subject variability in L2 vowel discrimination scores.
The Present study: Method

- **Participants:**
  - 113 Spanish/Catalan EFL learners
  - No speech disorders or hearing problems
  - Self-estimated proficiency level: from intermediate to advanced

- **Tasks and Procedure:**
  - Linguistic Background Questionnaire
  - PM (Serial nonword recognition task - SNWR)
  - AM (SNWR Spectrally rotated nonwords)
  - AC (attention-shift task)
  - Cue-weighting in L2 vowel perception (Vowel Discrimination Task)

DmDx display software (Foster & Foster, 2003)
Phonological short-term memory (PM)

- Temporary storage of verbal–acoustic information
- Subvocal rehearsal of encoded information

Predicts:
- L1&L2 vocabulary knowledge
- L2 grammatical development
- L2 oral fluency
- L2 speech perception

ASSUMPTION: Individuals with greater PM capacity
- use more native-like L2 cue-weighting in the perception of L2 vowel contrasts
- attend to both durational and spectral information of a sound.

Baddeley & Hitch, 1974; Gathercole at al., 1997; Masoura & Gathercole, 2005; French & O’Brien, 2008; O’Brien et al., 2006; Cerviño-Povedano & Mora, 2011; MacKay et al., 2001
The Present Study: Materials

- **Phonological memory: SNWR task**

  **Stimuli:**
  - 144 one-syllable Danish CVC nonwords (AVOID Language Dominance)
  - organized into 5-, 6- and 7-item length sequences,
  - 8 pairs of sequences at each item length (24 pairs in total)
  - Participants had to decide (within 5 seconds) whether the sequences were the same or different.
  - **Measure:** weighted score (out of 144) (O’Brian *et al*., 2007)

<table>
<thead>
<tr>
<th>Danish</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tys</td>
<td>dam</td>
<td>rød</td>
<td>mild</td>
<td>fup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>vul</td>
<td>bend</td>
<td>sids</td>
<td>påk</td>
<td>ryd</td>
<td>ham</td>
<td>jøb</td>
</tr>
</tbody>
</table>
Acoustic Memory (AM)

- Memory capacity for temporary storing of non-verbal acoustic information at pre-phonological level, i.e. prior to phonological encoding

<table>
<thead>
<tr>
<th>PSTM</th>
<th>AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operates at the phonological level</td>
<td>acoustic/phonetic details of the speech signal</td>
</tr>
<tr>
<td>Between-category discrimination</td>
<td>Within-category vowel discrimination L2-L1 sound mapping</td>
</tr>
</tbody>
</table>

- Spanish/Catalan EFL: /iː/-/i/ - within-category discrimination

**ASSUMPTION:** Individuals with greater AM capacity more sensitive to differences in acoustic information between perceptually similar L2 vowels and perceptually close L2-L1 vowel pairs.

Baddeley, 2003; Isaacs & Trofimovich, 2011; Darwin & Baddeley, 1974; Pisoni, 1973; Cowan & Morse, 1986
The Present Study: Materials

- **Acoustic memory task:**
  - SNWR task adapted:
    - Same or in a different
  - Stimuli
    - Catalan nonwords (144 nonwords) spectrally rotated using Praat
    - 3-, 4-, 5- and 6-item length sequence pairs
    - 32 testing trials and 2 initial warm-up trials

Spectrally Rotated Speech
- temporal and spectral complexity of ordinary speech, but NOT intelligible
- cannot be phonologically encoded, understood, repeated or sub-vocally rehearsed (Scott *et al.*, 2000)
- subjects are forced to rely only on the acoustic information
- **Measure:** weighted score (out of 144) (O’Brian *et al.*, 2007)
The Present Study: Materials

[map]

Rotated [map]
Sequences of 3 – 4 – 5 – 6 “Rotated Nonwords”:

Cognitive ability and L2 speech perception
Attention Control (AC)

- L1/L2 processing – complex cognitive skill
  - L1: efficient and flexible AC (automatic processing)
  - L2: controlled processing (requires greater attentional resources)

**AC** ability to shift attention efficiently among different sets of linguistic relationships
  - foregrounding/backgrounding of relevant/irrelevant linguistic info

**ASSUMPTION:** Individuals with higher AC capacity might be better able to rely on spectral information in the categorization of English /iː/ and /ɪ/ because they would be more successful at bringing segmental duration to the background.

(Isaacs & Trofimovich, 2011; Segalowitz, 2010; Talmy, 1996)
The Present Study: Materials

- **Attention Control: Attention-shift task**
  
  A speech-based version of the alternating runs procedure
  
  (Rogers & Monsell, 1995; Segalowitz & Frenkiel-Fishman, 2005)

  **Stimuli:** 7 Catalan vowels /i e ɛ a ɔ o u/
  
  - Dimension 1: segmental duration short (200ms) / long (500ms)
  - Dimension 2: voice quality male / female
  - 3 Practice blocks = 48 trials; Test block = 224 trials
  - Picture of a ‘loudspeaker’
  - Auditory feedback ‘beep’

- **Measures:**
  
  - Shift Cost = Shift RT (longer) – Repeat RT (shorter)
  - Error rates: Overall, Repeat, Shift
A person’s ability to shift focus of attention from one speech-based attention-directing function to another.

Foregrounding of duration vs. backgrounding of (partial) closure voicing in word-final obstruents.

(Safronova 2011, Safronova & Mora forthcoming)
Attention-shift task

Cognitive ability and L2 speech perception
Attention-shift task

Duration

Quality

Long  Left key  Female

Short  Right key  Male

[Keyboard images with key instructions]
Attention-shift task

Duration

Quality

Long  Left key  Female

Short  Right key  Male
Attention-shift task

Duration

Quality

Long  Left key  Female

Short  Right key  Male
Attention-shift task

<table>
<thead>
<tr>
<th>Duration</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td>Female</td>
</tr>
<tr>
<td>Short</td>
<td>Male</td>
</tr>
</tbody>
</table>

Left key: Female
Right key: Male
Attention Control
descriptives $N=83$

Error Rate (% ER)
- $S$ trials = 8.00
- $R$ trials = 5.23
- Overall = 6.01

RTs
- $S$ RTs = 1117 ms
- $R$ RTs = 923 ms
- $R$ RTs - $S$ RTs = 193 ms

Shift Cost (SC)

(Safronova & Mora, forthcoming)
The Present Study: Materials

- **Vowel Discrimination Task**

  FC AXB Categorial Discrimination Test *(Moya-Galé & Mora, 2011)*

  - /iː/ and /i/ in 6 CVC minimal pairs /b_d/, /d_d/, /s_d/, /b_t/, /p_k/, /p_t/
  - 6 native English speakers (3 males, 3 females)
  - 72 natural and 72 duration manipulated stimuli
  - Different tokens within trial
  - Different speakers within trial

  **bead – bid – bid**
  *(male¹ - female³ - male²)*
## Results: Correlations

<table>
<thead>
<tr>
<th></th>
<th>PM</th>
<th>AM</th>
<th>AC (ERR)</th>
<th>AC (SC)</th>
<th>DIS (NAT)</th>
<th>DIS (MAN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>.361**</td>
<td>-.410**</td>
<td>-.065</td>
<td>.198</td>
<td>.194</td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td></td>
<td>-.466**</td>
<td>.048</td>
<td>.502**</td>
<td>.435**</td>
<td></td>
</tr>
<tr>
<td>AC (ERR)</td>
<td></td>
<td></td>
<td>.176</td>
<td>-.431**</td>
<td>-.476**</td>
<td></td>
</tr>
<tr>
<td>AC (SC)</td>
<td></td>
<td></td>
<td></td>
<td>-.039</td>
<td>-.159</td>
<td></td>
</tr>
<tr>
<td>DIS (NAT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.760**</td>
</tr>
</tbody>
</table>

** Correlation is significant at .001 level
Results: Phonological Memory

**AXB Discrimination**

**Pearson r**  
PM  
DIS Nat n.s. .198  
DIS Man n.s. .194

**ANOVA**  
within: Nat/Man \( p < .001 \)  
between: Low/High n.s.

**Group differences:**  
Low PM (\( N=33 \))  
High PM (\( N=28 \))  
Nat: \( p = .435 \)  
Man: \( p = .098 \)

Cognitive ability and L2 speech perception  
(Safronova 2011, Safronova & Mora forthcoming)
Previous Research:
Greater PM capacity may provide learners with an advantage (Mora & Cerviño-Povedano, 2010)

Present Study: NON-significant

Methodological issues:
• nature of the SNWR task.
• may be more directly involved in ID (than in DIS)
• language knowledge effects.
• cross-language differences in vowel and consonant inventory size.
Results: Acoustic Memory

**AXB Discrimination**

<table>
<thead>
<tr>
<th></th>
<th>Low AM</th>
<th>High AM</th>
<th>Low AM</th>
<th>High AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Pearson r**
- DIS Nat: 0.502***
- DIS Man: 0.435**

**ANOVA**
- within: Nat/Man *p* < .001
- between: Low/High *p* = .009

**Group differences:**
- Low AM (*N* = 27): Nat: *p* = .007, Man: *p* = .010
- High AM (*N* = 39): Nat: *p* = .007, Man: *p* = .010

Cognitive ability and L2 speech perception
Greater AM capacity may provide learners with an advantage in perceptual cue-weighting in L2 speech.

Methodological issues:
• sequence item length and ISI probably needs adjusting
• may be more directly involved in DIS (than in ID)
Results: AC Error Rate

**Pearson r**  
AC ER  
DIS Nat - .431**  
DIS Man - .476**

**ANOVA**s  
within: Nat/Man  *p* < .001  
between: Low/High  *p* < .001

**Group differences:**  
Low AC ER (*N*=32)  
Nat:  *p* = .002  
Man:  *p* < .001  
High AC ER (*N*=28)  
Nat:  *p* = .002  
Man:  *p* < .001

Cognitive ability and L2 speech perception  
(Safronova & Mora, *forthcoming*)
Results: AC Shift Cost

**AXB Discrimination**

**Pearson r**

<table>
<thead>
<tr>
<th>AC SC</th>
<th>DIS Nat</th>
<th>DIS Man</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n.s.</td>
<td>-.039</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>-.159</td>
</tr>
</tbody>
</table>

**ANOVA**

within: Nat/Man  *p*<.001
between: Low/High n.s.

**Group differences:**
Low AC SC (*N*=30)
High AC SC (*N*=30)
Nat:  *p*=.572
Man:  *p*=.209

(Cafronova & Mora, *forthcoming*)
Greater AC may provide learners with an advantage

- in perceptual cue-weighting in L2 speech
- in phonetic training involving backgrounding and foregrounding of L2-specific use of acoustic features

Methodological issues:

- speech dimensions need to be operationalized more accurately.
- participants focused mainly on accurate performance (feedback)
Results: Regression

- $R^2 = .286$ (28.6%); $p = .001$ (Nat)
- $R^2 = .285$ (28.5%); $p = .001$ (Man)

<table>
<thead>
<tr>
<th>PM</th>
<th>% Unique variance explained</th>
<th>$p$</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
<td>.945</td>
<td>NAT</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>.881</td>
<td>MAN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AM</th>
<th>% Unique variance explained</th>
<th>$p$</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11.3</td>
<td>.007</td>
<td>NAT</td>
</tr>
<tr>
<td></td>
<td>4.9</td>
<td>.070</td>
<td>MAN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AC (Error Rate)</th>
<th>% Unique variance explained</th>
<th>$p$</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.5</td>
<td>.123</td>
<td>NAT</td>
</tr>
<tr>
<td></td>
<td>9.5</td>
<td>.013</td>
<td>MAN</td>
</tr>
</tbody>
</table>
Results: Regression

- $R^2 = 0.258$ (25.8%); $p = 0.002$ (Nat)
- $R^2 = 0.236$ (23.6%); $p = 0.004$ (Man)

<table>
<thead>
<tr>
<th></th>
<th>% Unique variance explained</th>
<th>$p$=</th>
<th>AXB Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>0.23</td>
<td>0.696</td>
<td>NAT</td>
</tr>
<tr>
<td></td>
<td>0.55</td>
<td>0.552</td>
<td>MAN</td>
</tr>
<tr>
<td>AM</td>
<td>20.4</td>
<td>0.001</td>
<td>NAT</td>
</tr>
<tr>
<td></td>
<td>14.6</td>
<td>0.003</td>
<td>MAN</td>
</tr>
<tr>
<td>AC</td>
<td>0.7</td>
<td>0.510</td>
<td>NAT</td>
</tr>
<tr>
<td>(Shift Cost)</td>
<td></td>
<td>4.5</td>
<td>.092</td>
</tr>
</tbody>
</table>
Conclusions

AM and AC
• involved in the processing of L2 speech
• may facilitate target-like cue-weighting
• may explain inter-learner variation in L2 phonological attainment

Future research:
• Solve methodological issues
• Other cognitive abilities: E.g. ability for oral mimicry
• Focus on both: L2 speech perception and production
Thank you!

Acknowledgments:

Participants in this study
Eva Cerviño-Povedano
James E. Keidel
Pavel Trofimovich
Victor Kravchenko
GRAL research assistants