

**After pronunciation training  
L2 learners converge prosodically to the native speaker**

**Elisa Pellegrino**

URPP Language and Space, University of Zurich  
Institute of Computational Linguistics, University of Zurich

## In collaboration with



Massimo Pettorino  
University of Naples L'Orientale



Deborah Vigliano  
University of Naples  
L'Orientale  
Tohoku University, Sendai



Philippe Boula  
de Mareuil  
LIMSI -Paris

# Introduction

- ▶ NNS speakers deviate both segmentally and suprasegmentally from NS
  - ▶ The proficiency in L2 is not fully attained if students have just interiorized the phonological, morpho-syntactic and lexical rules of the target language.
  - ▶ The prosody of a language, rather than individual vowels and consonants, can be more important for understanding and producing a language that is not L1 (Chun, 2013)
    - ▶ the semantic meaning of an utterance, but also of the intended meaning behind that utterance.

“even heavily accented speech is sometimes perfectly intelligible and [...] prosodic errors appear to be a more potent force in loss of intelligibility than phonetic errors” (Munro & Derwing, 1999)

## Why focusing on L2 prosody?

- ▶ Help L2 learners achieve native-like performance
  - ▶ teaching of pronunciation is not one of the top priorities for most language teachers
  - ▶ traditional focus on practicing difficult sounds
  - ▶ little attention on the acquisition of native-like prosodic performance
    - ▶ prosody involves many aspects of speech about which teachers do not themselves know enough (Derwin & Munro, 2015)

# Why focusing on L2 prosody?

- ▶ Avoid negative attitudes towards L2 accented speech
  - ▶ hiring decisions, employment, and social integration (Munro, 2003; Davila, Bohara and Saenz, 1993)
    - ▶ call centres: (Pickering et al. 2012; Wang et al. 2013)
    - ▶ language teaching: hiring bias against non-native speaker English teacher in UK, USA and in contexts of English as a Foreign Language (Clark & Paran, 2007; Lippi-Green, 2012; Moussu and Llorca, 2008)
- ▶ Improve communication between senior citizens and accented speakers with whom they interact daily
  - ▶ Alterations of speech stress and timing with accent are particularly difficult for elderly listeners to perceive (Gordon-Salant, 2017)

# How to teach/learn L2 suprasegmentals?

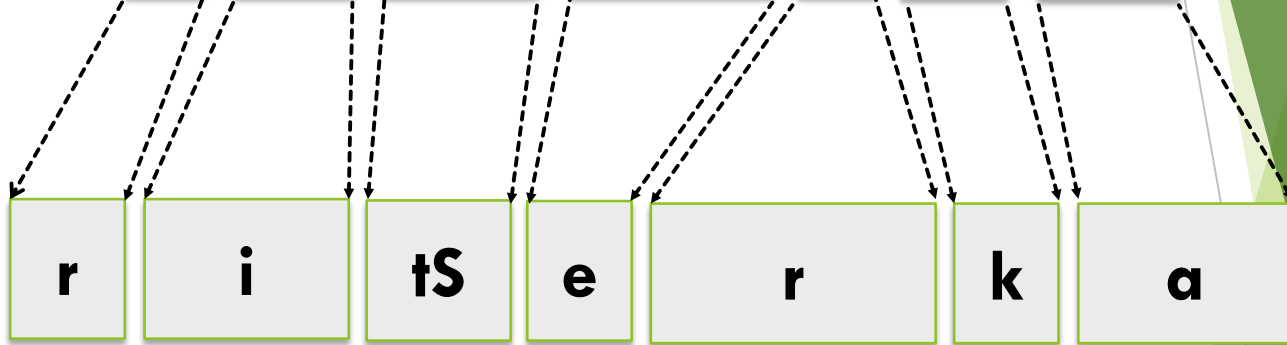
- ▶ **Speech technology** and the improvement of the **prosodic competence** in an L2 (Chun, 2013; Eskenazi, 2009; Levis, 2007)
- ▶ **Student/teacher voice similarity** for the **enhancement of pronunciation skills** (Bissiri, Pfitzinger & Tilmann 2006; Jilka & Möhler, 1998); Nagano & Ozawa 1990; Peabody & Seneff, 2006) Sundström (1998); Tang, Wang & Seneff (2001)
- ▶ **User-dependent golden speaker** (Probst, Ke & Eskenazi, 2002)
- ▶ The most effective technique is **self-imitation**: learners should imitate their own **voice** previously **modified to match the prosody** of the reference **native speaker** (Felps, Bortfeld & Guitierrez-Osuna, 2009)

# durations

L1 donor

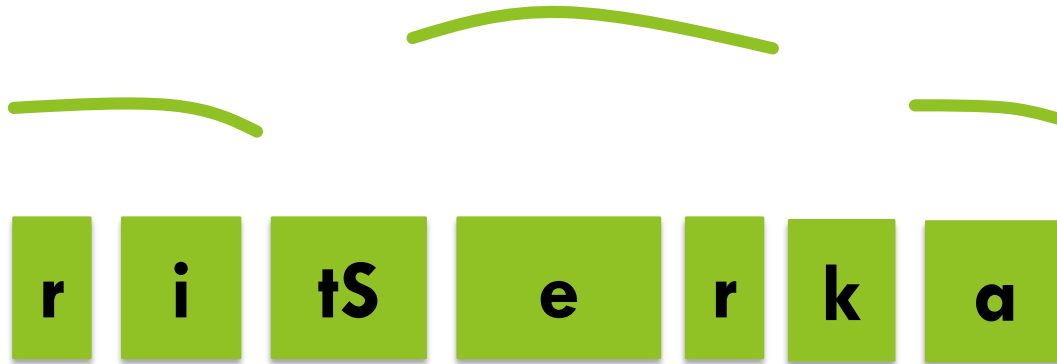


L2 receiver

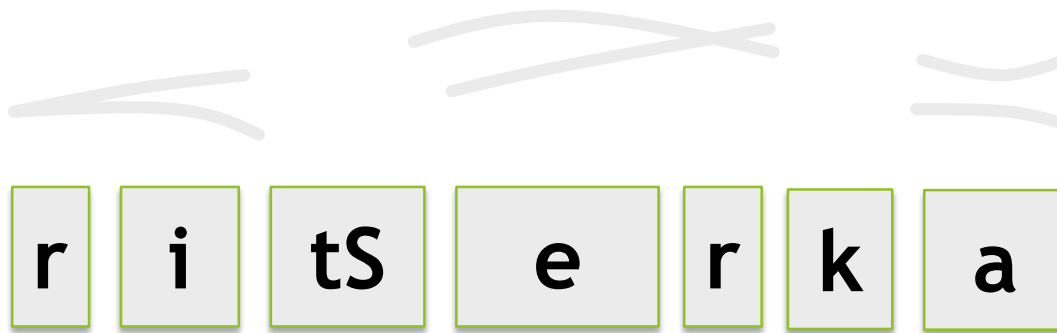


# pitch contour

L1 donor



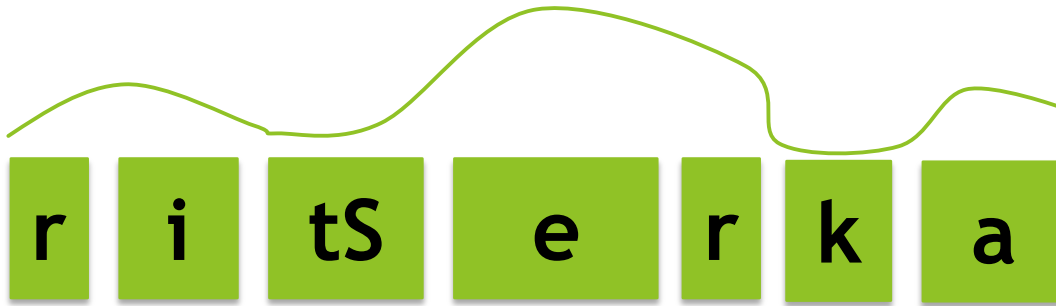
L2 receiver



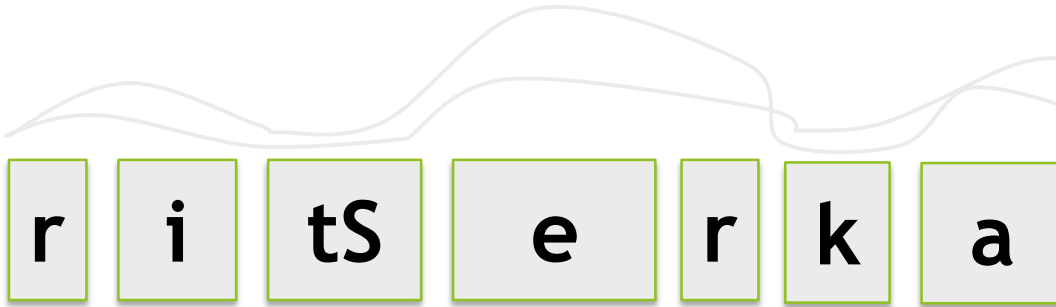


# intensity

L1 donor



L2 receiver



ITA L1 - 1



CHIN pre

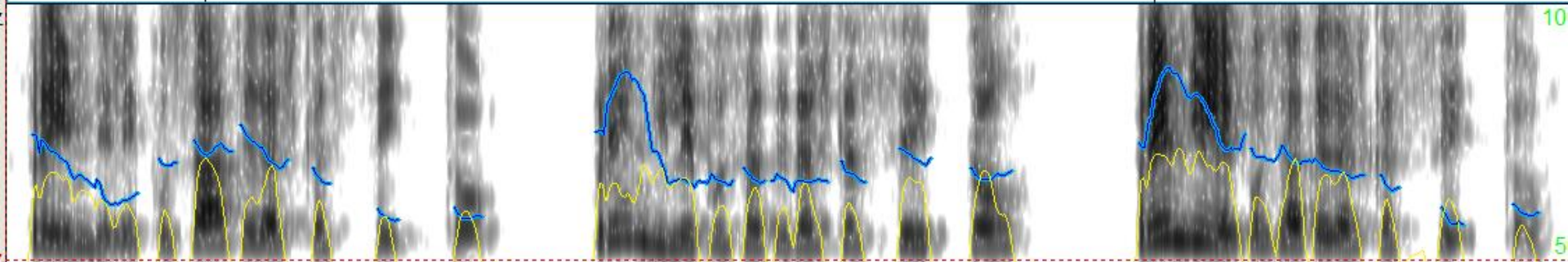
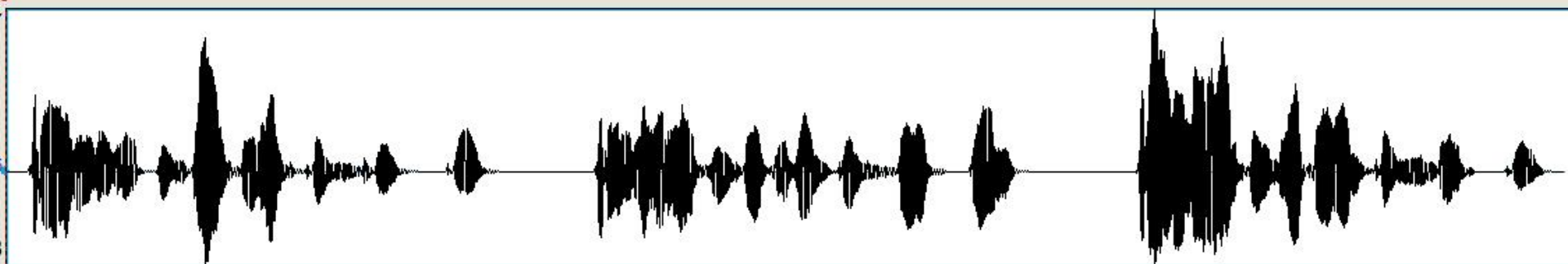


Transplanted utterance



# The study

- ▶ Given that:
  - ▶ prosody plays a crucial role in shaping the pragmatic function of an utterance
  - ▶ In Italian there are no morphological and syntactical means for distinguishing sentence modality (statements vs questions) (D'Imperio, 2002)
  - ▶ In Japanese, sentence modality is conveyed mainly by means of affixes that may vary based on how polite/(in)-direct talkers want to perform the speech acts
- ▶ We tested the pedagogical effectiveness of self-imitation for the acquisition of the prosodic features conveying different pragmatic functions in Italian



statement

request

command

2.033

1.861

1.807



Prende le posate dal cassetto  
(S)he takes the cutlery from the  
drawer

Prendi le posate dal cassetto?  
Can you take the cutlery from the  
drawer, please?

Prendi le posate dal cassetto!  
Take the cutlery from the drawer!

# The study

- ▶ **5 Japanese learners of L2 Italian**
  - ▶ Aged 21-28
  - ▶ Upper intermediate level (B2, CEFR)
  - ▶ 5-6 years of study
- ▶ **1 native Italian speaker**
  - ▶ Aged 25 F
- ▶ **Read Speech**
  - ▶ 2 sentences
  - ▶ 3 pragmatic functions
    - ▶ Request, Order, Grant

# Pre-test

- ▶ NS utterances were administered to students from different Italian regions, to check the correct match between the intended and perceived pragmatic functions

## Step 1: Pre-training Corpus

Recordings while reading two sentences  
conveying 3 pragmatic functions (Request, Order, Grant)

Corpus

**30 utterances in L2 Italian**

5 NNSs \* 2 sentences \* 3 functions

**6 utterances in L1 Italian**

1 NS \* 2 sentences \* 3 functions

# Corpus

## **Request**

✓ Accendi la radio? /eng. Can you turn on the radio?

## **Order**

✓ Accendi la radio! /eng. Turn on the radio!

## **Grant**

✓ Accendi la radio /eng. OK, you can turn on the radio

## **Request**

✓ Chiudi la finestra? /eng. Can you close the window?

## **Order**

✓ Chiudi la finestra! /eng. Close the window!

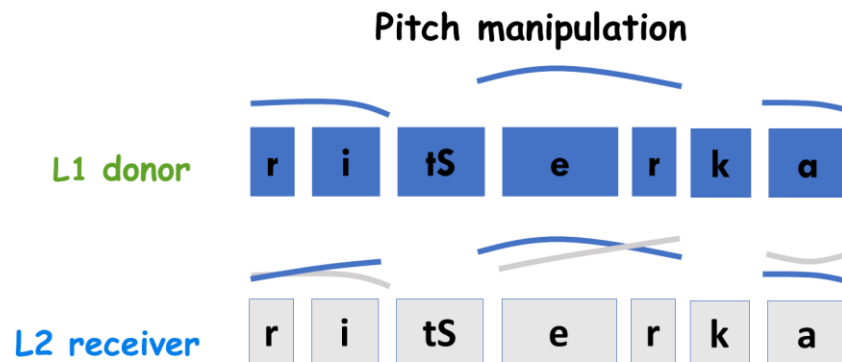
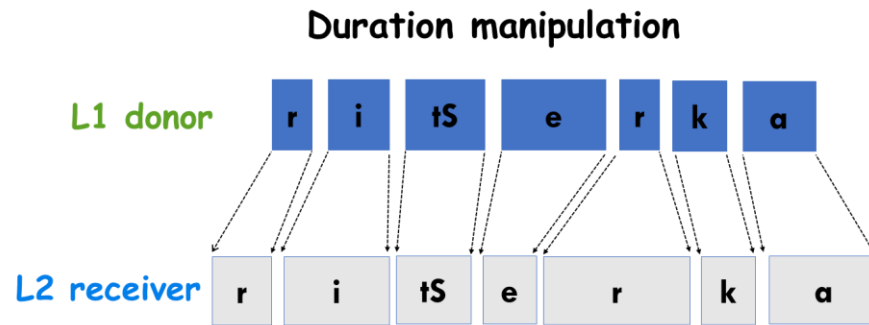
## **Grant**

✓ Chiudi la finestra /eng. OK, you can close the window



## Step 2: Corpus segmentation and manipulation

- Annotation in segments and syllables (Praat)



### Step 3: Self-imitation

- Learners listen to their **manipulated** utterances
- Individual training  
mimic utterances with native accents

### Step 4: Post Training Corpus

New recordings after self-imitation prosodic training



**Corpus**

**30 utterances in L2 Italian**  
5 NNSs \* 2 sentences \* 3 functions

# Acoustic analysis

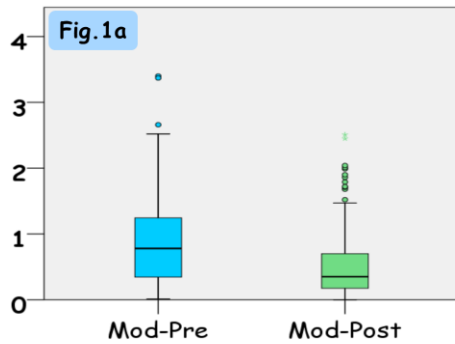
- **Step 1:** we measured duration, f0 mean and f0 max for each syllable produced by NS and by NNSs before and after training
- **Step 2:** we normalized syllable duration, f0 max and f0 mean using z-score transformation ( $z = (x - \mu) / \sigma$ )
- **Step 3:** we calculated the difference in duration, f0 mean and f0 max between the NS syllables and the corresponding NNS syllables before and after training (Mod-Pre and Mod-Post)

# Hypothesis

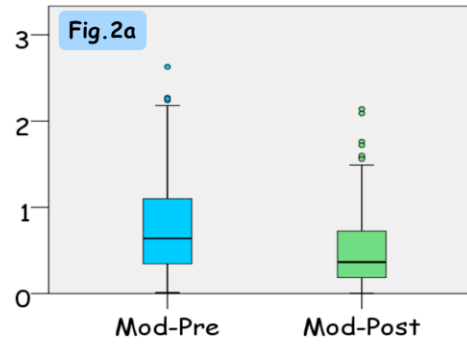
- ▶ If self-imitation is effective
- ▶ Production
  - ▶ L2 speakers converge prosodically towards the native model after self-imitation training
    - ▶ differences between L1 and L2 in duration and F0 decrease after the training

# Acoustic Analysis

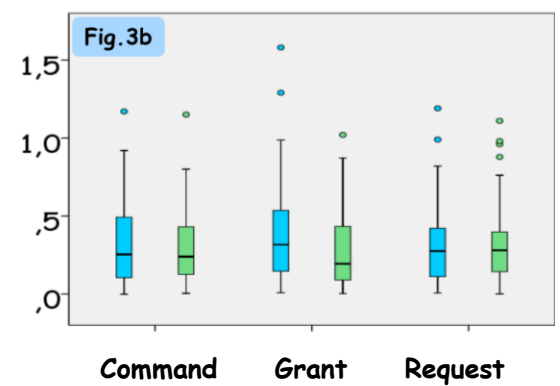
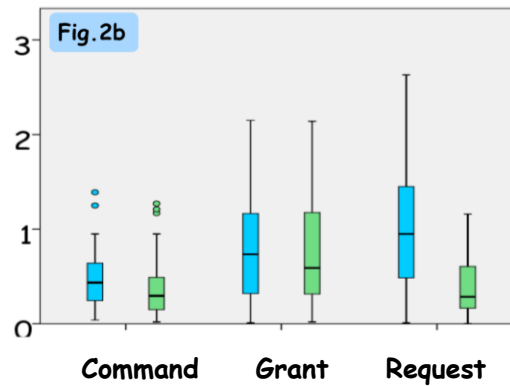
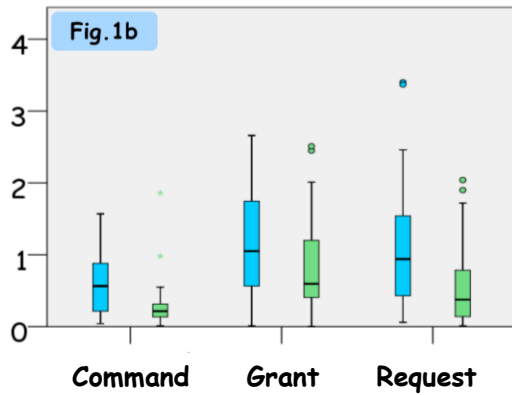
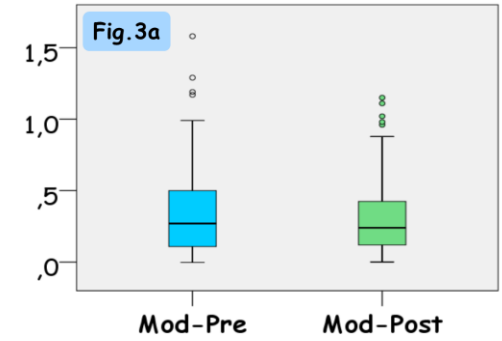
F0 mean



F0 max



syll. duration



## What about perception?

Does the training improve the communicative effectiveness of L2 speakers?

Does the degree of correct match between intended and perceived pragmatic functions increases after self-imitation training?

# The perception test










- ▶ **17 native Italian listeners**
  - ▶ Aged 23-30, students at the University of Naples L'Orientale
  - ▶ no prior knowledge of Japanese language
  - ▶ familiar with different foreign accents
- ▶ **60 utterances in L2 Italian**
  - ▶ 30 pre-training productions (Japanese speakers only)
  - ▶ 30 post-training productions
- ▶ **Administration on line through the software**
  - ▶ Individual listening with headphones
- ▶ **Assessments**
  - ▶ to identify the conveyed pragmatic functions (requests, commands, grant, statements, other)

# Hypotheses

- ▶ **If self-imitation is effective**
- ▶ **Production**
  - ▶ L2 speakers converge prosodically towards the native model after self-imitation training
    - ▶ differences between L1 and L2 in duration and F0 decrease after the training
- ▶ **Perception**
  - ▶ the degree of correct match between intended and perceived pragmatic functions increases after self-imitation training



# MEAN PERCENTAGE OF CORRECT MATCH BY SPEECH ACT AND TRAINING PHASE

	Pre-training (A)	Post-training (B)	Difference (B – A)
 Requests	 52,52%	 75,21%	+22,69
 Orders	 39,92%	 57,98%	+18,06
 Grantings	 8,40%	 47,06%	+38,66

# Conclusions

## ► Production

The results of acoustic analysis show that after self imitation training:

- L2 utterances converge more with the native model in terms of F0 mean, F0 max and syllable duration
- F0 mean is the acoustic feature for which there is a consistent convergence to the native model after training over all speech acts
- L2 speakers converge more to the native model in terms of F0 max for Requests and in terms of syllable duration for Grant

# Conclusions

## Perception

- ▶ The percentage of **correct match** between intended and perceived communicative intentions **increases** significantly after the training session.
- ▶ The improvement regards all the three speech acts, especially **grantings**.
- ▶ Self-imitation prosodic training helps learners memorize and reproduce intonation and duration patterns corresponding to the native listeners' expectation
- ▶ **Production + Perception = Self Imitation is effective**



▶ Next steps

- ▶ Are the learners able to generalize the patterns to sentences that they have heard before?
- ▶ How long is the effect of training?

THANKS

The background features abstract, overlapping geometric shapes in various shades of green, ranging from light lime to dark forest green. These shapes are primarily located on the right side of the frame, creating a modern, layered effect against the white background.

# Results

## PRE-TRAINING PHASE

Perceived pragmatic functions						
Intended pragmatic functions		O	G	R	S	Other
	O	39,92%	10,92%	32,35%	13,87%	2,94%
	G	47,68%	8,44%	20,25%	18,57%	5,06%
	R	16,88%	5,06%	52,74%	11,81%	13,50%

O= Order

R= Request

G= Granting

S= Statement

# Results

## CONFUSION MATRIX POST-TRAINING PHASE

		Perceived pragmatic functions				
Intended pragmatic functions		O	G	R	S	Other
	O	57,98%	11,34%	14,29%	14,71%	1,68%
	G	11,34%	47,06%	17,23%	17,23%	7,14%
	R	12,61%	4,20%	75,21%	5,88%	2,10%

Frase 1

RICHIESTA Accendi la radio?

質問 ラジオつけてくれない?

Shitsumon rajio tsukete kurenai?<sup>1</sup>

COMANDO Accendi la radio!

命令 ラジオつけて!

Meirei rajio tsukete!<sup>2</sup>

CONCESSIONE Accendi la radio.

譲歩 ラジオつけていいよ

Jōho rajio tsukete ii yo.<sup>3</sup>

Frase 2

RICHIESTA Chiudi la finestra?

質問 窓閉めてくれない?

Shitsumon mado shimete kurenai?

COMANDO Chiudi la finestra!

命令 窓閉めて!

Meirei mado shimenasai!

CONCESSIONE Chiudi la finestra.

譲歩 窓閉めていいよ

Jōho mado shimete ii yo



# Results

## MEAN PERCENTAGE OF CORRECT ANSWERS BY TRAINING PHASE

	<b>Pre-training (A)</b>	<b>Post-training (B)</b>	<b>Difference (B – A)</b>
<b>Average</b>	33.61%	60.04%	+ 26.43

# First data on Italian

## De Meo et al. (2013) **Imitation/self-imitation in computer-assisted prosody training for Chinese learners of L2 Italian**

- ▶ Experimental test on **self-imitation**
- ▶ Chinese learners with an upper-intermediate level of competence
- ▶ Suprasegmental level
- ▶ **Rhythmic-prosodic transplantation**
  - ▶ **Modified elements:**  $F_0$ , segment duration
  - ▶ **Unchanged elements:** segmental level and speaker's identity

# First data on Italian

- ▶ Both **imitation** and **self-imitation improved** the non-native speakers' rhythmic-prosodic performance
  - ▶ Increase in the degree of match between the intended and perceived communicative functions and the L1 listener comprehension.
  - ▶ Reduction in foreign accentedness
  
- ▶ **Self-imitation** generally achieved more satisfactory results

The background features abstract, overlapping green geometric shapes in various shades, creating a modern and dynamic feel. The shapes are primarily triangles and polygons, some semi-transparent, layered against a white background.

# Thank you!

[epellegrino@unior.it](mailto:epellegrino@unior.it)

[dvigliano@yahoo.it](mailto:dvigliano@yahoo.it)