



gipsa-lab

Grenoble | images | parole | signal | automatique | laboratoire

WACA 2010
Lille - 25, 26 novembre

A la recherche d'indices de culture et/ou
de langue dans les micro-événements
audio-visuels de l'interaction face à face

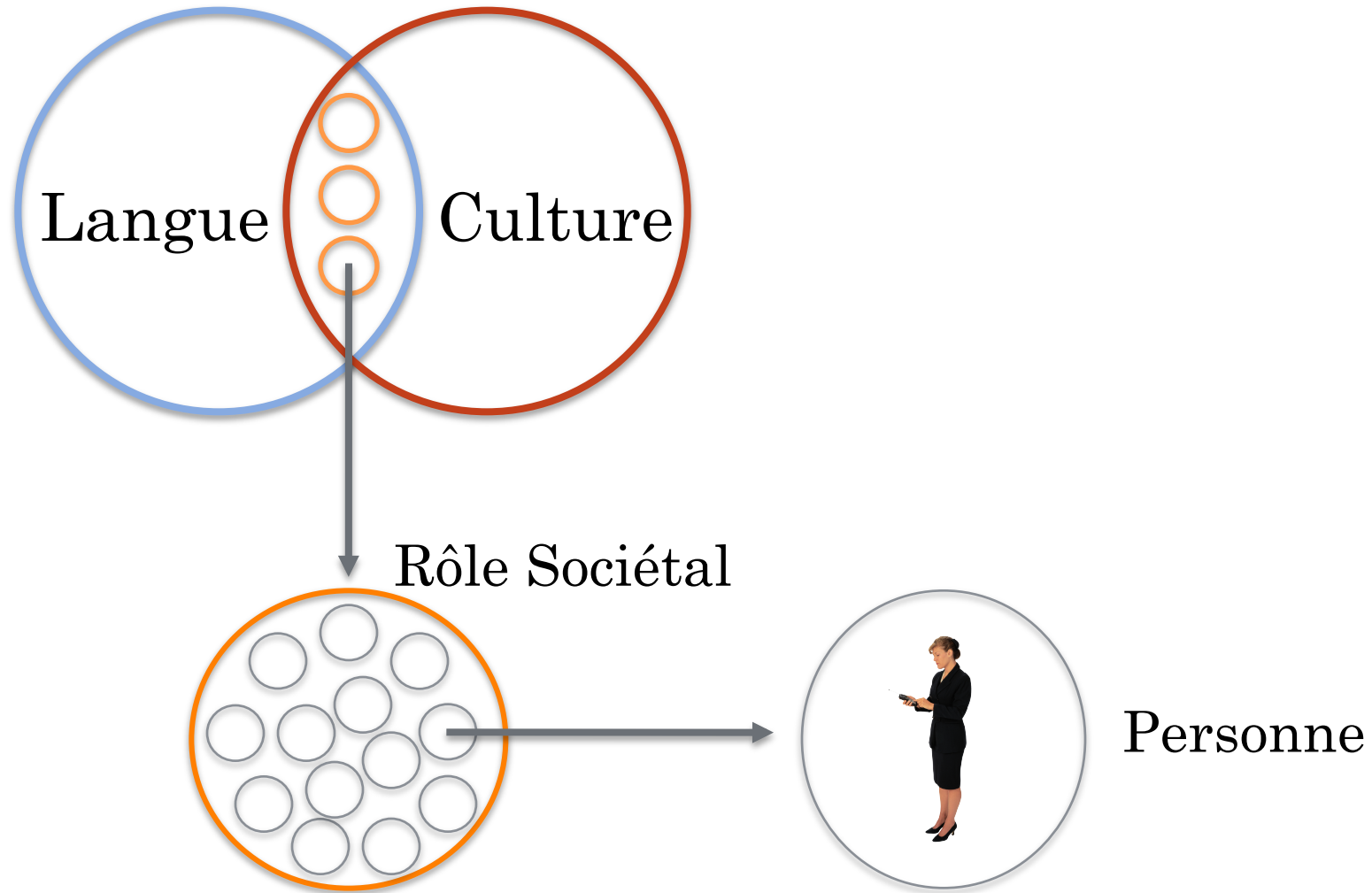
Rosario Signorello¹, Véronique Aubergé^{1,2},
Anne Vanpé^{1,2}, Lionel Granjon¹ et Nicolas Audibert³

¹ GIPSA-Lab, Département Parole et Cognition, UMR CNRS 5216, Université Stendhal, Grenoble, France

² Laboratoire d'Informatique de Grenoble, équipe GETALP, UMR CNRS 5217, Grenoble, France

³ Laboratoire Informatique d'Avignon, Université d'Avignon, Avignon, France

LANGUE ET CULTURE



INTERACTION : *Quand, où, pourquoi, à qui, qui parle ?*

**Pertinence de l'expression dans sa multimodalité
(*voix, parole, langage, face, gestualité*)**

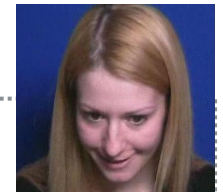
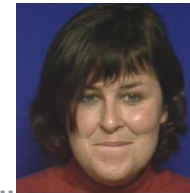
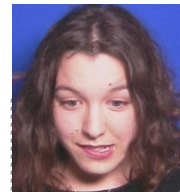
**instanciation dans le
contexte**

interaction face à face
continuité communicative
locuteur ↔ auditeur

tour de parole ↔ *backchannel*

Communication

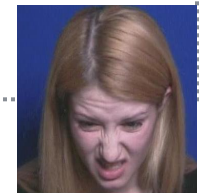
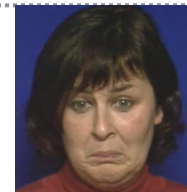
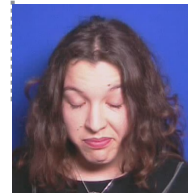
flux permanent d'informations
structuré par les tours de parole



sujet

comportement → personnalité
intentions/buts - motivation
expression des états mentaux
attitudes, émotions, humeurs

→ *Feeling of Thinking* [Aubergé, 2007]



PROBLÉMATIQUE

- Peut-on discriminer deux langues/cultures à travers des micro-événements non langagiers du *Feeling of Thinking* ?
- Existe-t-il une interaction perceptive entre des faits de langue/culture et des indices de personnalité ?

OBJECTIFS DE L'ÉTUDE

- Identification perceptive de l'appartenance culturelle de sujets locuteurs « francophones » par des juges FRA et ITA à travers la présentation de micro-événements non langagiers

Langues typologiquement proches
variantes familières de culture euro-méditerranéenne

- Hypothèse d'apprentissage sur une suite de stimuli
 - « quantitatif » : exposition perceptive répétée
 - « qualitatif » : stimuli de qualité informative supposée croissante
- Évaluation perceptive du degré d'extraversion
 - effet croisé extraversion perçue du sujet et langue/culture ?
 - interaction choix ita *vs.* fra et degré d'extraversion perçu

LE PROTOCOLE EXPÉRIMENTAL

- 6 sujets locuteurs « francophones »
 - corpus Ewiz/SoundTeacher [Aubergé *et al.*, 2006]
 - annotation fine [Vanpé *et al.*, 2008 à 2010]
 - 3 femmes et 3 hommes – variation du degré d’extraversion

- 90 juges participants : 45 français et 45 italiens

- Test de perception sans contact personnel interculturel des juges
 - 3 modalités de passation du test
 - Audio seul – Visuel seul – Audio Visuel
 - 1 modalité = 1 groupe de 15 juges

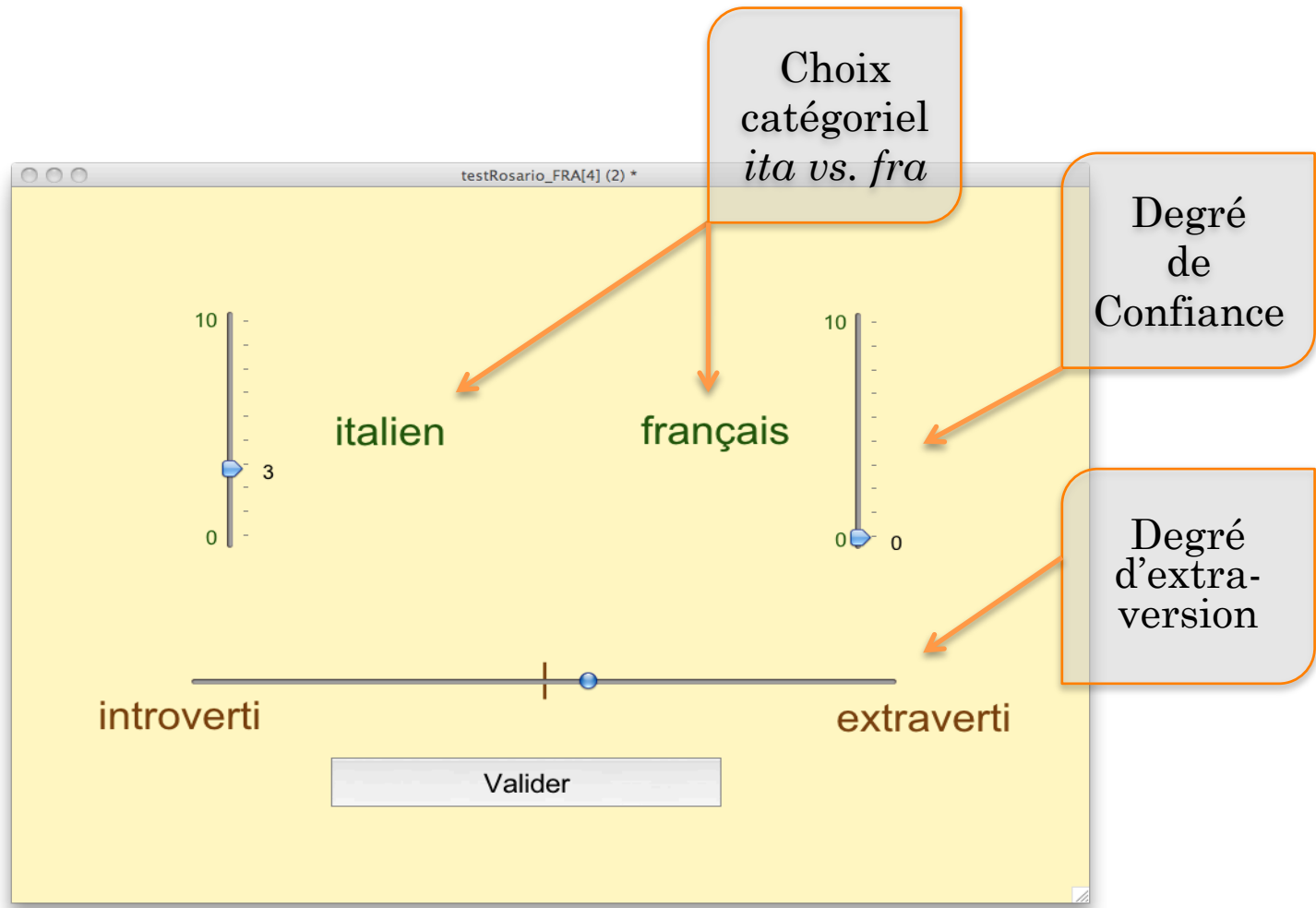
STIMULI: MICRO-ÉVÉNEMENTS AUDIBLES ET VISIBLES NON LANGAGIERS

- Bruits de bouche, *grunts*, *bursts*, *fillers*, interjections
- Hypothèses
 - Les premiers faits de langage apparaissent bien **avant** des sons doublement articulés
 - Nature langagière **est** construite par le contrôle prosodique
 - Qualité informative **variable** (e.g. indices langue/culture)
- Buts
 - Mettre en évidence l'identité langagière possible
 - Mesures perceptives pour vérifier cette identité

STIMULI: STRUCTURE DE PRÉSENTATION

- Stimuli présentés séquentiellement selon leur qualité informative supposée d'indices sur la langue/culture :
 - 24 stimuli ordonnés sur un axe
bruit de bouche (e.g. « *burst* ») → item pré-lexical (e.g. interjection)
- Qualité informative des stimuli :
 - de 1 à 10/14 : sans prosodie, non phonétiques, non phonologiques
 - de 10/14 à 13/20 : prosodie contrôlée, pré-phonétiques, non phonologiques
 - de de 13/20 à 22/24 : prosodie contrôlée, sons phonétiques, pré-phonologiques
 - de 22/24 à 24 : prosodie contrôlée, sons phonétiques et phonologiques

INTERFACE DU TEST (en français)



EXEMPLES DE STIMULI : AUDIO SEUL



friction ingressive
mouillée bilabiale

fra (56%) *ita* (44%)



« *ouh* »

fra *ita*
70% 30%



« *ok ben* »

fra *ita*
90% 10%



raclement
de gorge

fra *ita*
63% 37%



inspiration
expiration

fra *ita*
53% 47%



« *euh* »

fra *ita*
57% 43%



« *pff lala* »

fra *ita*
90% 10%

1

15

24

sons non phonétiques
sans contrôle prosodique
(bruit de bouche, *grunt*, *burst*)

sons non phonétiques
avec contrôle prosodique
(expiration, qualité de voix)

sons phonologiques
prosodie contrôlée
(onomatopée, *filler*, interjection)

EXEMPLES DE STIMULI : VISUEL SEUL

friction inspirée
mouillée



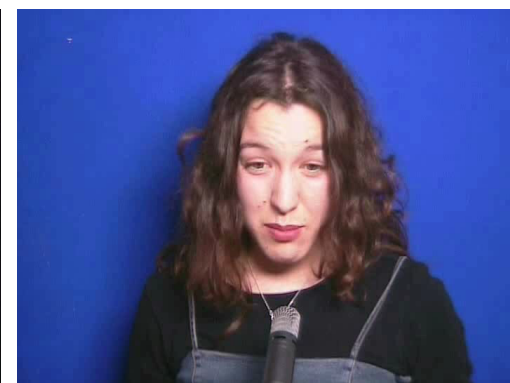
fra ita
30% 70%

« euhmmh »
cluster VC



fra ita
63% 37%

« ben mmh »
creaky voice



fra ita
46% 54%

1

sons non phonétiques
sans contrôle prosodique
(bruit de bouche, *grunt*, *burst*)

15

sons non phonétiques
avec contrôle prosodique
(expiration, qualité de voix)

24

sons phonologiques
prosodie contrôlée
(onomatopée, *filler*, interjection)

EXEMPLES DE STIMULI : AUDIO-VISUEL

inspiration rapide



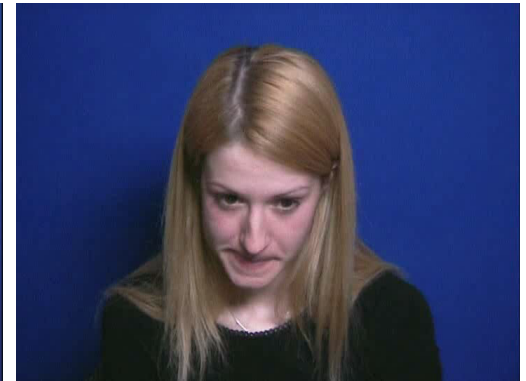
fra ita
46% 54%

avale salive
raclement de gorge
« *mmmh* » inspiré



fra ita
60% 40%

« *ouhla* »
interjection



fra ita
96% 4%

1

sons non phonétiques
sans contrôle prosodique
(bruit de bouche, *grunt*, *burst*)

15

sons non phonétiques
avec contrôle prosodique
(expiration, qualité de voix)

24

sons phonologiques
prosodie contrôlée
(onomatopée, *filler*, interjection)

CHOIX LANGUE/CULTURE : *français* vs. *italien*

60,2%

Juges FRA

	A		AV		V	
	<i>fra</i>	<i>ita</i>	<i>fra</i>	<i>ita</i>	<i>fra</i>	<i>ita</i>
F-M	77,2 %	22,8 %	73,3%	26,7%	27,8 %	72,2 %
F-S	75 %	25 %	63,1%	36,9%	42,8 %	57,2 %
F-T	66,1 %	33,9 %	51,4%**	48,6%**	66,7 %	33,3 %
M-J	66,4 %	33,6 %	49,7%**	50,3%**	62,2 %	37,8 %
M-N	68,1 %	31,9 %	70,8%	29,2%	73,1 %	26,9 %
M-R	48,6 %**	51,4 %**	44,7%*	55,3%*	55,8 %*	44,2 %*
TOTAL	66,9 %	33,9 %	58,8%	41,2%	54,7%	45,3%

Juges ITA

56,7%

	A		AV		V	
	<i>fra</i>	<i>ita</i>	<i>fra</i>	<i>ita</i>	<i>fra</i>	<i>ita</i>
F-M	53,3%**	46,7%**	73,3%	26,7%	39,2%	60,8 %
F-S	44,2%*	55,8%*	68,1%	31,9%	44,4%*	55,6%*
F-T	48,9%**	51,1%**	66,4%	33,6%	55%**	45%**
M-J	26,9%	73,1%	59,2%	40,8%	55%**	45%**
M-N	40,3%	59,7%	66,1%	33,9%	60,6%	39,4%
M-R	40,8%	59,2%	87,2%	12,8%	91,9%	8,1%
TOTAL	42,4%	57,6%	70,0%	30,0%	57,7%	42,3%

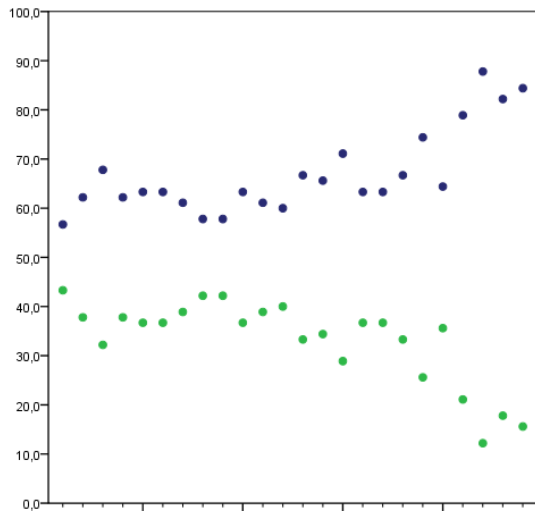
A posteriori

	A		AV		V	
	<i>fra</i>	<i>ita</i>	<i>fra</i>	<i>ita</i>	<i>fra</i>	<i>ita</i>
	93,3%	6,7%	100,0%	0,0%	26,7%	73,3%
	93,3%	6,7%	93,3%	6,7%	46,7%	53,3%
	100,0%	0,0%	100,0%	0,0%	73,3%	26,7%
	53,3%	46,7%	86,7%	13,3%	73,3%	26,7%
	93,3%	6,7%	93,3%	6,7%	73,3%	26,7%
	66,7%	33,3%	93,3%	6,7%	60,0%	40,0%

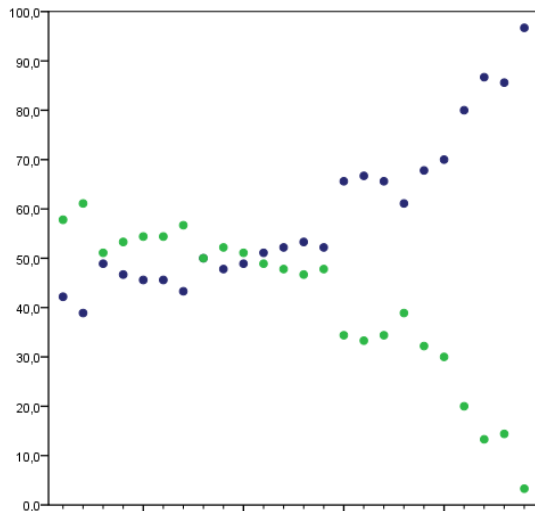
	A		AV		V	
	<i>fra</i>	<i>ita</i>	<i>fra</i>	<i>ita</i>	<i>fra</i>	<i>ita</i>
	73,3%	26,7%	93,3%	6,7%	33,3%	66,7%
	80,0%	20,0%	73,3%	26,7%	53,3%	46,7%
	80,0%	20,0%	80,0%	20,0%	73,3%	26,7%
	26,7%	73,3%	66,7%	33,3%	60,0%	40,0%
	73,3%	26,7%	93,3%	6,7%	66,7%	33,3%
	86,7%	13,3%	93,3%	6,7%	100,0%	0,0%

PROGRESSION DES RÉPONSES : CHOIX *français* vs. *italien*

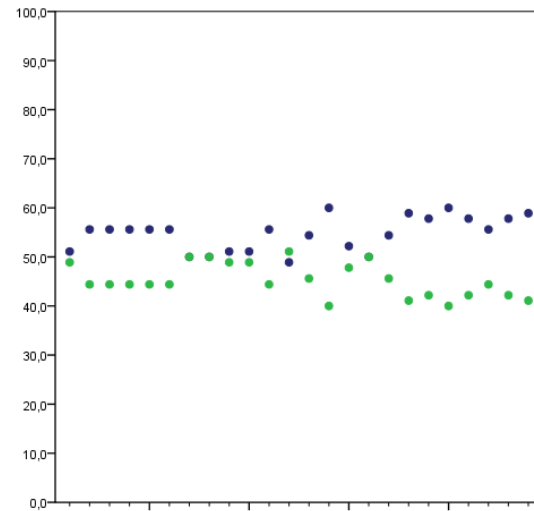
Juges FRA – A



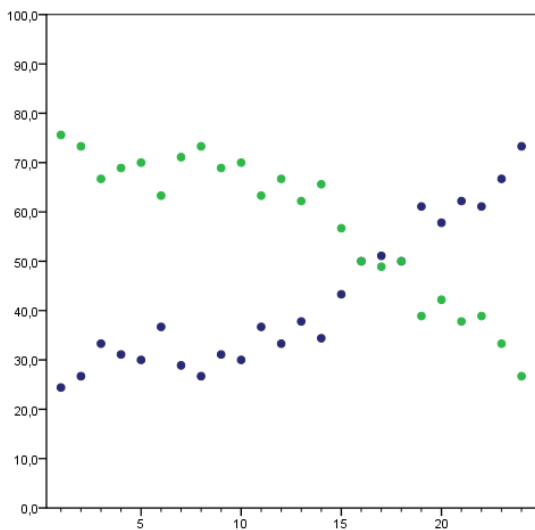
Juges FRA – AV



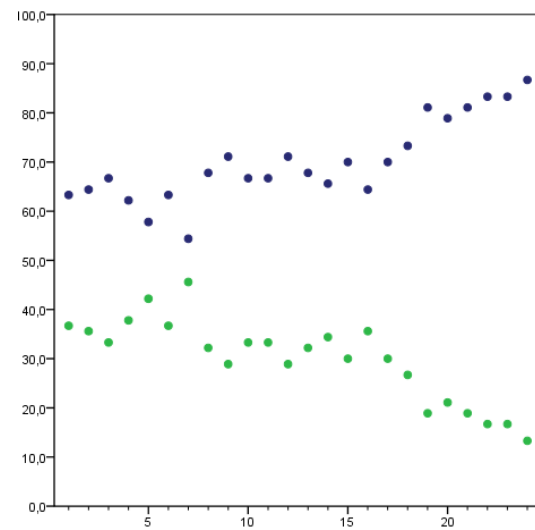
Juges FRA – V



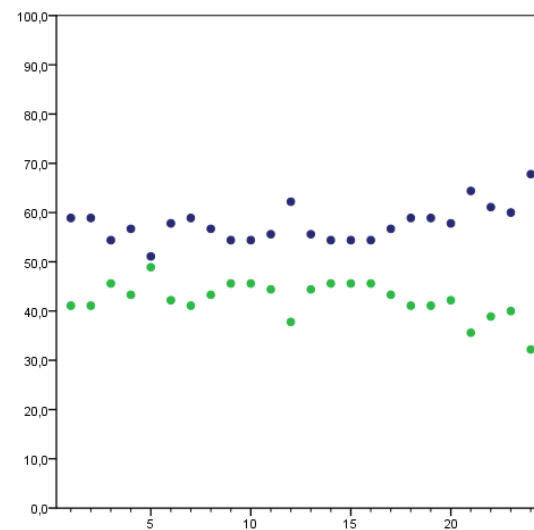
Juges ITA – A



Juges ITA – AV

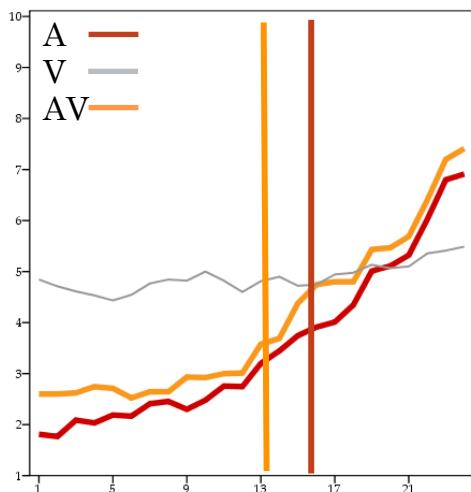


Juges ITA – V

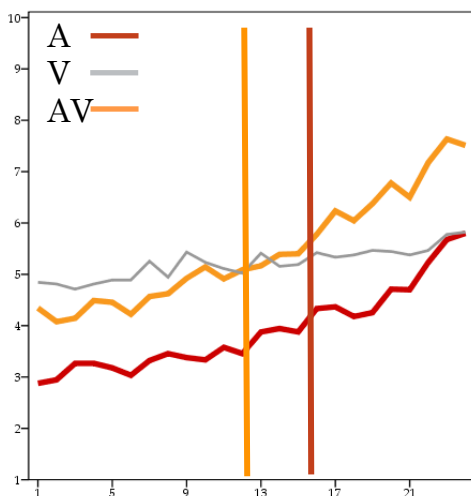


POINT DE STABILITÉ: CHOIX LANGUE/CULTURE ET DEGRÉ DE CONFIANCE

Juges FRA



Juges ITA



- En Audio et Audio-Visuel « point de stabilité »
 - juge ne modifie plus son choix
 - augmentation du degré de confiance
 - à partir du 15^{ème} stimulus en A Seul
 - à partir du 12^{ème}/14^{ème} en AV



à l'apparition des stimuli
pré-phonétiques ou phonétiques
contrôlés prosodiquement

- En Vidéo : point de stabilité non significatif
 - degré de confiance élevé dès le début
 - stable pendant la présentation de la suite des stimuli

DEGRÉ D'EXTRAVERSION

- Patrons de réponse « différents » pour chaque sujet selon :
 - Nationalité du juge ($p < .001$)
 - Modalité de passation du test (A, V, AV)
- L'ordre d'affichage des stimuli influence le choix ($p < .001$)
Ex. *Juges français* perçoivent
→ sujets plus « introverti » pendant les premiers stimuli pour progresser vers « extraverti »
- Effet du sujet
- Pas d'interaction entre choix culture/langue et degré d'extraversion

CONCLUSIONS

○ Visuel seul

- degré de confiance plus élevé du début
- **aucun** de « point de stabilité » :
→ l'aspect est décisif, peu d'information dynamique, ni de renforcement du choix

○ Audio seul

- pas d'information discriminante suffisante pour les juges italiens
- information suffisante pour les français
- **mais** « point de stabilité » pour les stimuli prosodiquement contrôlés
→ quand le choix ne change plus, le degré de confiance augmente

○ Audio-Visuel

- condition de référence qui permet aux italiens de faire le « bon » choix
- **mais** l'ajout de la vidéo à l'audio « dérange » les juges français
- comme en A « point de stabilité » pour les stimuli « prosodiques »

○ Degré d'extraversion

- Influence de la nationalité du juge et de la modalité de passation

PERSPECTIVES DE L'ÉTUDE

- Autres expérimentations similaires
 - avec des sujets locuteurs/stimuli italiens
 - suite de stimuli qui commence au « point de stabilité »
 - distribution équilibrée a priori des sujets par rapport à leur variation de personnalité (extraversion)
 - présentation par paires des sujets locuteurs
→ « lequel des deux est français » ?
 - variation de la distance langue/culture
→ Ex. français *vs.* mandarin ou au contraire espagnol *vs.* argentin

APPLICATIONS : ACA « CULTUREL » ET « PERSONNALISÉ »

- Pertinence cognitive des « détails »
- Présence/absence des facteurs de langue/culture modifie et/ou active des stratégies dans l'ACA [Poggi *et al.*, 2005]
- Contribuer à donner à un ACA
 - un comportement « culturellement » pertinent en situation réaliste et écologique
 - une capacité cognitive de haut niveau
 - une capacité de pré-planification de l'interaction
- « *en fonction de comme il/elle est, je lui dis ça ou pas* » ?

**MERCI
POUR VOTRE
ATTENTION**

BIBLIOGRAPHIE

- Aubergé, V., Audibert, N., et Rilliard, A. (2006). « De E-Wiz à C-Clone. Recueil, modélisation et synthèse d'expressions authentiques ». In *Revue d'Intelligence Artificielle « Interactions émotionnelles »*, 20(4-5), pages 499-528.
- Mancini, M., Björn, H., Pelachaud, C. (2008). « L'expressivité nella comunicazione non-verbale degli agenti conversazionali ». In Poggi, I., éditeur : *La mente del cuore. Le emozioni nel lavoro, nella scuola, nella vita*, chapitre 6, pages 163-177. Armando Editore, Rome, Italie.
- Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V. et De Carolis, B. (2005). GRETA. A Believable Embodied Conversational Agent. In Stock, O. et Zancarano, M., éditeurs : *Multimodal Intelligent Information Presentation*, pages 1–23. Kluwer Academic Publishers, Belgique. (Cité aux pages 43 et 44)
- Poggi, I. (2008). *La mente del cuore. Le emozioni nel lavoro, nella scuola, nella vita*. Armando Editore, Rome, Italie.
- Vanpé, A. et Aubergé, V. (2008). « Prosodie visuelle expressive du Feeling of Thinking perception statique ou dynamique ? ». In *WACA 2008*.
- Vanpé, A. et Aubergé, V. (2010b). « Prosodie expressive audio-visuelle de l'interaction personne-machine ». In *Technique et Science Informatiques, numéro spécial Agents Conversationels Animés*, 29.