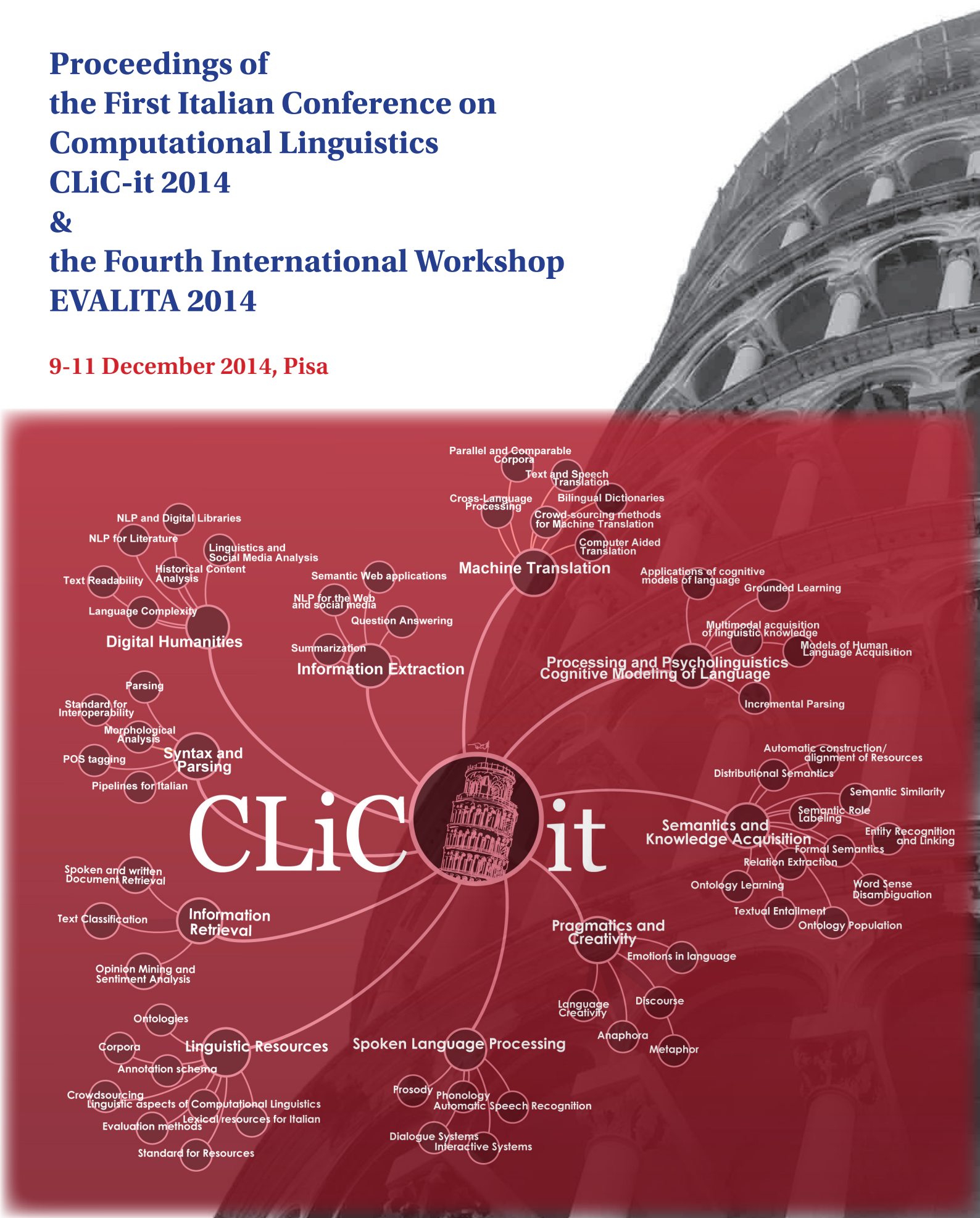
Human and Machine Language / Dialect Identification from Natural Speech and Artificial Stimuli: a Pilot Study with Italian Listener

UNIVERSITÀ DEGLI STUDI DI TORINO

Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & the Fourth International Workshop EVALITA 2014

9-11 December 2014, Pisa



PISA UNIVERSITY PRESS

10.12871/clicit2014225

Human and Machine Language / Dialect Identification from Natural Speech and Artificial Stimuli: a Pilot Study with Italian Listeners

Antonio Romano **Claudio Russo**
Università degli Studi di Torino, Dip. Lingue e Lett. Str. e Cult. Mod.
Laboratorio di Fonetica Sperimentale "Arturo Genre"
via Sant'Ottavio, 24 I-10124 Torino, Italia
antonio.romano@unito.it clrusso@unito.it

Abstract

English. After a short review of the state of the art, this paper illustrates a selection of the most important Automatic Language Identification and Accent Identification approaches. A series of tasks is presented, providing some evaluation measures about the overall human performance on the basis of language/dialect identification by Italian listeners. Results confirm that humans are able to easily detect linguistic features of languages they have been directly exposed to, thus being able to perform a swift identification when listening even to short samples. Identification rates rise in familiar dialect id. tasks, and a sharp separation is usually established between unknown foreign languages, guessed languages and local varieties of one's own country.

Italian. *Dopo una breve introduzione sullo stato dell'arte, quest'articolo riassume una selezione dei più diffusi approcci all'Identificazione Automatica delle Lingue e degli Accenti (LID/AID). Alcune misure sono offerte riguardo a una serie di test che sono stati svolti per valutare le modalità con cui è avvenuta l'identificazione di una selezione di lingue e dialetti da parte di alcuni uditori italiani. I risultati confermano che gli esseri umani hanno una certa abilità nell'individuare i principali tratti linguistici ai quali sono esposti più spesso e sono, anche per questo, in grado d'identificare agevolmente le lingue conosciute sulla base di campioni di parlato anche piuttosto brevi. Le prestazioni migliorano, infatti, nell'identifica-zione di dialetti con i quali si abbia una certa familiarità. Una separazione netta si può infine stabilire tra lingue straniere sconosciute, lingue indovinate in base a supposizioni e varietà del proprio Paese.*

1 Introduction

Since its origins, the challenge of Automatic Language Identification (*LID*) encountered the problems raised by the presence of dialectal variation and the difficult task of accent identification (*AID*): "the absolute acoustic differences of the native accents is very subtle and sensitive so that they might be an order magnitude smaller than the differences between speech sounds, and be secondary to the individual speaker differences" (Wu *et alii* 2004).

These problems have been tackled by different research teams with a wide set of phone- or acoustic-based techniques (*n-grams*, *phone-lattice* and so on). The state of the art provided by Muthusamy *et alii* (1994) and Geoffrois (2004) during the *MIDL* event of 2004 "Identification des langues et des variétés dialectales par les humains et par les machines" (Paris, France, 29-30 nov. 2004, see Adda Decker *et alii* 2004) needs an update since relevant milestones have been achieved after the *NIST LID* contest of 2003 and the following *NIST LRE* 2005 and 2009. Discriminative *LID* based on *Support Vector Machines* or on *Multi-corpus* and *out-of-set LID* received positive attention since then, and training datasets have been purposefully created and expanded in various *LRE* tasks (following the model of the *Callfriend* corpus, based on labelled speech stuff, and other *LDC* corpora).

Even though the most successful *LID* systems implement more than one component modeling different information types at various levels, several *LID* systems are still nowadays mostly phone-based (cp. Kirchhoff *et alii* 2002, Singer *et alii* 2003, Timoshenko & Bauer 2006; for a review, see, Schultz & Kirchhoff 2006, Wang 2008). Nevertheless, 'acoustic' *LID* systems tend to rely on spectral features in order to extract language-discriminating information encoded within speech productions, whereas language-specific sequences of speech units are traced by 'phonotactic' *LID* systems.

The linguistic information is then usually extracted from the test speech sample with phone recognition modules that rely on either language-

dependent or cross-linguistic acoustic phone models (cp. Yan & Bernard 1995).

According to the scientific literature on human language/dialect identification (Ohala & Gilbert 1981, Romano 1997, Ramus & Mehler 1999), we expect that prosodic level of organisation, such as intonation and rhythm, provides a reliable cue for this purpose (Vaissière & Boula de Mareüil 2004). However, prosodic cues are still less explored in *LID* systems (Navrátil 2006, Leena & Yegnanarayana 2008, Timoshenko 2012) and results of listening tasks aiming to assess the role of the related variables have not yet been achieved for the present study.

After a short review of *LID/AID* models, this paper proposes a discussion about the results of two listening tasks performed by Italian listeners; 54 students were exposed to speech stimuli of 18 foreign languages whereas a selection of 32 of them was asked to identify 20 dialectal varieties.

2 Motivation

Besides the perspective of shedding light on the reasons why automatic speech recognition systems succeed (or fail) when dealing with speech samples encoded in an unknown language, research on human and machine performances in language identification are *per se* interesting.

The challenge for *IT* developers (and for institutions investing on it) is to implement automatic procedures aimed at achieving human performances in language and dialect identification.

On the one hand, that means looking at the inherent language variation in the world (thanks to well documented *DB* and archives, see references) and, on the other hand, trying to emulate human skills in this kind of task.

By the way, also humans do face a challenge when they experience multi-lingual spoken or written communication and are intrigued by language diversity. Whatever their success in dealing with languages which are used in these situations, human beings are amazed by this surprising diversity and are usually challenged to guess the unknown languages they listen to. That explains the large public success of amateur websites such as the "Great language game" (http://greatlanguagegame.com/).

While language variation in specific areas have been captured by various speech/accent archives, significant knowledge about world's languages comes from well-known projects such as *Ethnologue* (Lewis *et alii* 2014) or the *Rosetta* project (rosettaproject.org/). Academic research recently yielded a relevant progress thanks to authoritative sources such as *WALS*, but has also benefited by recent contributions such as *Langscape* or *Phoible*. These projects gathered questionable but useful speech samples as well as phonetic/phonological and bibliographic data on sound structure (this aspect founds a consolidated reference in the *UCLA Phonetic Segment Inventory Database* and the more recent *Lyon-Albuquerque Phonological Systems Database*).

As the individual sensitivity is generally very poor when facing dialectal variation outside the area of origin or residence, so is the knowledge gathered about such variation in large repository sites. Furthermore, dialectal variation is heterogeneous within the different countries. In some areas, a monolingual situation is attested, with potential accent variation throughout the whole territory, but some other regions may be characterised by a jumble of different languages and each of them strongly affected by dialectal variation (cp. Tsai & Chang 2002). This is the situation of Italy and its surrounding countries.

Languages and dialects spoken in Italy are surveyed and discussed in several dialectological studies (among others, Maiden & Parry 1997, Loporcaro 2009) and a remarkable quantity of lexical and phonetic data is provided by linguistic atlases such as the *ALI* (Massobrio *et alii* 1996) who helped in the definition of the dataset (§3.2). Nevertheless, the available information is hardly exploitable for testing since no speech samples are included and data is not intended for IT purposes or language identification tasks. Experiments on the perception of foreign accent in Italian are carried out by some research teams (De Meo *et alii* 2011), but native accented speech is less studied and the general knowledge of Italian speakers about regional varieties/dialects is almost completely ignored.

2.1 Automatic *LID*/*AID* methods

Within the last twenty years, universities from all over the world jointly worked with *IT* companies to produce effective automated speech recognition systems. Thanks to this striking cooperative effort, the research community witnessed a wide range of different techniques, which can be roughly classified as:

- techniques based on parallel phone recognition for phone lattice classification (*PPLRM*; cp. Gauvain *et alii* 2004). These approaches relied mostly on language-dependent *n-gram* models and context-independent phone models to classify the salient features of phonotac-

tic traits. Both context-dependent Hidden Markov Models (*CD-HMM*) and null-grammar *HMM* have been exploited by this particular approach (Damashek 2005, Suo *et alii* 2008);

- techniques focused on spectral change representation (*SCR*) and extraction of prosodic features. These approaches usually look at utterances as collections of independent spectral vectors. For accent identification (*AID*) purposes, such vectors are combined in a supervector that is assigned to each speaker; to achieve *LID*, the vector collection is usually modeled by Gaussian Mixture Models (*GMM*s) or similar (Kirchhoff *et alii* 2002). Within these approaches, an unusual solution has been explored with the Bag-of-sounds (*BOS*) technique, which exploits a universal sound recogniser to create a sound sequence that is converted into a count vector at a second stage. The classifier being trained, the *BOS* technique does not need any acoustic modelling to add new language capabilities;
- hybrid techniques have been refined thanks to different technologies (such as Deep Neural Networks, *DNN*s, used as state probability estimators; Lopez Moreno *et alii* 2014). Recently, further attempts towards *GMM*-free approaches have been made, aiming at improving segmentations through online interaction with a parameter server and graph-based semi-supervised algorithms for speech processing (Liu & Kirchhoff 2013).

3 Tasks for human listeners

Since human perception of identification cues are unconscious, listening experiments are needed in order to empirically assess in which way human language identification occurs.

In this research, three listening tasks have been proposed to test human abilities in language and dialect identification.

Testing scripts and soundwave files were freely distributed at the following website: http://www.lfsag.unito.it/evalita2014/index.html. The execution of the listening tasks required the installation of the *PRAAT* software and the creation of a *HMDI* folder on the PC. Instructions on how to carry out each experiment were illustrated by a *.pps* slideshow.

HMDI (see §3.1 and 3.4) was a task aiming at testing human abilities to identify languages from short speech samples.

The two following tasks *HMDI_DIA* and *HMDI_TON* were intended to test dialect identification by natural and synthetic speech samples. *HMDI_DIA* (see §3.2 and 3.5) was a task mainly intended for listeners living in Italy and it aimed at testing their abilities to identify dialectal varieties whereas *HMDI_TON* was conceived to test the possibility to identify dialect just relying on prosodic values extracted from real sentences. Results of the latter are not reported here.

3.1 First Dataset (*HMDI*)

The *HMDI* task was based on a sample of 18 languages represented by natural stimuli recorded in a soundproof booth. Two samples based on passages from a local version of the IPA narrative "The North Wind and the Sun" were submitted to the listeners' judgment. All the recordings are original and belong to a larger ongoing speech archive available at the *LFSAG*.

All the speakers were women aged between 20 and 28. Stimuli are coded with a number corresponding to each language as it follows:

1. Albanian (Durrësi-Duras accent)
2. Arabic (Tunisian accented *SMA*)
3. Baoulé (from Bouaké, Ivory Coast)
4. Chinese (from the Jiangsu region)
5. Farsi (from Tehran)
6. Bavarian German (Südtyrolian dialect)
7. Hebrew (from Jerusalem)
8. Hungarian (from Eger)
9. I.-Veneto (from Vodnjan-Dignano, Istria)
10. Latvian (from Riga)
11. Macedonian (from Bitola)
12. Polish (from Krakow)
13. Portuguese (Capeverdean accent)
14. Romanian (from Braşov)
15. Serbian (from Beograd)
16. Spanish (from Buenos Aires, Argentina)
17. Sardinian (from Orosei)
18. Vietnamese (Hanoi accent).

Speech samples have a variable length (between 7.2 and 13.3 s) and more or less the same number of syllables belonging to a text which corresponds to the narrative's last passages: "And so the North Wind was obliged to confess that the Sun was the stronger of the two. Did you like the story? Do you want to hear it again?".

Listeners sat before a PC monitor wearing a headset and decided when to run the PRAAT script. Speech stimuli for this experiment were played twice in random order and listeners were asked to select the corresponding language label in an interactive window as quickly as possible.

The overall duration of the each test session was about 6-10 min.

3.2 Second Dataset (*HMDI_DIA*)

The *HMDI_DIA* task relied on a sample of 20 dialects. Even in this case, stimuli were extracted from a local version of "The North Wind and the Sun".

All the speakers were female aged between 20 and 28 except for one who was in her 40s.

The task was intended for Italian listeners and is mainly based on samples selected from dialects which are spoken in Italy or nearby but includes several dialects of foreign languages as distractors/control languages.

The test was administered by means of a PRAAT script (see above) and through an interactive window allowing the listener to choose a language label on the screen after listening to each of the 20 stimuli (randomly played once). Since the task was intended for Italian listeners, languages were labelled in Italian.

The stimuli were taken from recordings collected for the following languages: *Arabo M.* (Moroccan Arabic), *Arabo T.* (Tunisian accented S.M. Arabic), *Napoletano* (Neapolitan), *Occitano P.* (Piedmont Occitan), *Pugliese* (Apulian), *Polacco K.* (Polish from Krakow), *Polacco W.* (Polish from Wrocław), *Piemontese* (Piedmontese from Saluzzo), *Portoghese C.V.* (Capeverdean Portuguese), *Portoghese T.E* (Portuguese from East Timor), *Romeno V.* (Romanian from Braşov), *Romeno M.* (Moldavian from Chişinau), *Siciliano Or.* (East Sicilian from Catania), *Siciliano Occ.* (West Sicilian from Erice), *Siciliano Mer.* (Southern Sicilian from Pachino), *Salentino* (Sallentinian from Mesagne), *Spagnolo A.* (Argentinian Spanish), *Spagnolo V.* (Venezuelan Spanish), *Sardo* (Sardinian), *I.-Veneto* (Veneto-Istrian dialect from Vodnjan-Dignano).

Even in this dataset, the length of the stimuli was well below the usual *LID* values and it was variable between 5.5 and 13.2 s.

3.3 Listeners' samples

Listeners were 54 students, or visiting students at the Uni.TO, aged between 18 and 35 (34 women and 20 men; 93% were students of foreign languages). 37% were first-degree students and the remaining 63% was almost equally represented by MA and PhD students. 17% of the sample was constituted by students of foreign origins (2 Spanish, 2 Romanian, 2 Macedonian, 1 Moroccan, 1 Iranian and 1 Albanian).

For the *HMDI_DIA* task the sample was reduced to 34 listeners (mainly of Italian origins or living since various years in Italy and very proficient in Italian). Many of them had Piedmontese origins (24, that is 71%) and declared a passive knowledge of a local dialect (6 of them of another dialect spoken in Italy: 2 Sicilian, 2 Apulian and 2 Sardinian). Furthermore, 14 listeners (41%) reported an active competence of a foreign language (1 Spanish, 1 Romanian) or another dialect spoken in Italy (3 Calabrian, 3 Sicilian, 3 Apulian, 2 Sallentinian and 1 Sardinian).

3.4 Evaluation measures for *HMDI*

Generally speaking, for the first task (*HMDI*) listeners answered correctly 713 times, which means that 36.7% languages of the tested sample have been correctly identified.

A negligible learning effect has been observed from the first to the second passage of the same stimulus: 350 correct responses were collected for the first repetition vs. 363 for the second one.

Individual responses were displayed in confusion plots such the one showed in Fig. 1, whereas overall results are summarised in Fig. 2.

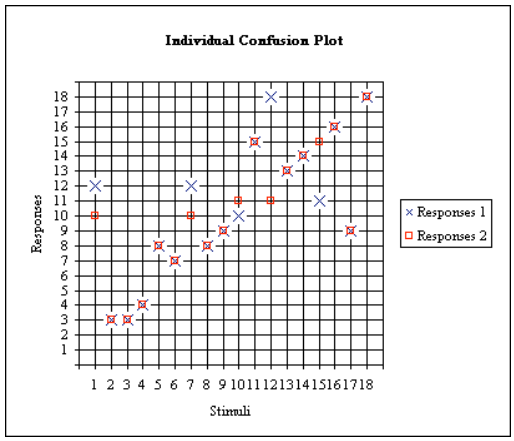


Fig. 1 – Individual plot of responses given to each pair of language stimuli.

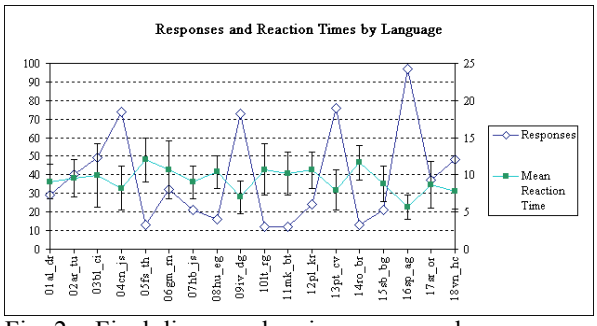


Fig. 2 – Final diagram showing scores and mean reaction times for each test language.

All the responses were statistically analysed by using *R* functions and scripts. Of course, re-

sults have not been assessed in *DET* curves diagrams, as for automatic systems, since only one sample per language was tested. Even though Miss probabilities and False Alarm rates could be extensively discussed for human listener too (cp. Swets 1964), the sample was reduced (and responses were highly non-linear). Therefore, general results (plotted in Fig. 3 and summarized in table I) are discussed in a more adapted way.

As shown in Fig. 3, the listeners responded variously. The top-four, most-identified languages were Spanish (row 16), Portuguese (r. 13), Chinese (r. 4) and Veneto-Istrian (r. 9).

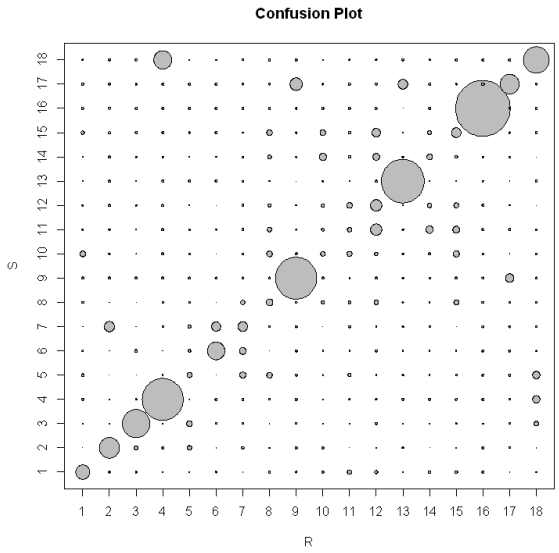


Fig. 3 – Confusion plot for the 18 stimuli (*S* axis) and responses (*R* axis) for the first task. See the text for language codes (§3.1).

The four least-identified languages were Latvian (r. 10), Macedonian (r. 11), Romanian (r. 14) and Farsi (r. 5). The error rate (*ER*) for Spanish, Portuguese, Chinese and Veneto-Istrian is 6%, 26%, 29% and 29% respectively, whereas it rises to 87-89% for the less identified languages. It is worth noticing how Latvian has been uniformly confused among Arabic, Hungarian, Portuguese and Serbian. Macedonian has been confused mostly with Polish, Serbian and Romanian and the latter with Latvian, Polish and Hungarian. Finally, it is interesting to notice how the listeners identified Vietnamese (r. 18) despite their lack of any kind of knowledge about it. A similar score was achieved for Baoulé (r. 3).

When guessing the right answer, the listeners expressed their preference for some languages in particular: Polish, Portuguese and Chinese above others. Conversely, Sardinian, Arabic and Südtyrolian German scored preference values below their actual presence in the task. This may signal a sort of prototypical reference role of the former languages for listeners of this almost homogeneous sample.

Finally, the dispersion plot in Fig. 4 allows establishing an inverse proportionality between the number of correct answers and the reaction times (RT) as a general trend for all the listeners. RT were significantly lower for the declared known languages (5,4 s) than for unknown or guessed languages (10,7 s; a two-sample Welch t-test gave t = -9.36, df = 65.98, p-value = 1.009e-13).

Table I. Confusion matrix (Task HMDI, see §3.1)

	Responses																	
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
01al_dr	29	1	1	2	3	1	3	8	1	8	12	10	2	8	8	0	5	2
02ar_tu	4	40	11	0	11	3	8	2	0	7	3	2	1	4	2	0	1	5
03bl_ci	2	2	49	3	14	1	1	3	1	3	2	0	5	2	1	2	1	12
04cn_js	0	2	1	74	5	0	0	1	0	2	1	0	0	0	0	0	1	17
05fs_th	8	4	4	6	13	4	14	14	0	6	10	2	1	2	1	0	0	15
06gm_rn	1	3	9	3	9	32	16	4	0	2	2	8	2	2	5	0	0	6
07hb_js	2	22	3	1	9	18	21	7	0	4	8	2	1	1	4	0	0	1
08hu_eg	8	3	4	0	7	3	12	16	0	10	8	11	1	2	12	0	1	6
09iv_dg	0	0	0	0	0	0	0	0	73	0	2	0	1	1	0	8	19	0
10lt_rg	14	1	3	0	1	3	1	13	0	12	13	10	7	8	15	1	1	1
11mk_bt	6	1	3	0	1	2	0	12	2	8	12	24	1	15	16	0	0	1
12pl_kr	7	0	1	0	2	4	0	7	2	9	14	24	3	12	13	0	2	4
13pt_cv	2	0	2	0	0	0	1	2	4	1	2	0	76	5	3	1	5	0
14ro_br	5	0	1	1	2	2	6	11	1	17	8	16	7	13	8	1	1	4
15sb_bg	9	0	0	0	1	0	2	14	1	13	8	19	5	11	21	0	0	0
16sp_ag	0	0	0	0	0	0	0	0	1	0	0	0	4	0	1	97	1	0
17sr_or	0	0	1	0	0	0	1	1	23	1	0	0	21	8	1	9	37	1
18vn_hc	1	0	7	35	5	2	1	0	0	2	1	1	0	1	0	0	0	48

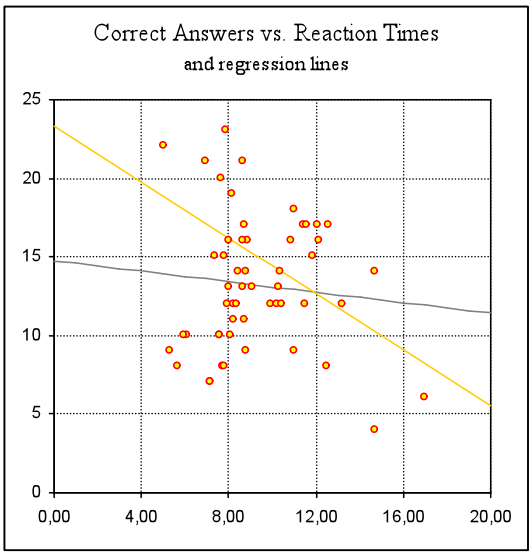


Fig. 4 – Dispersion plot of the number of correct answers vs. Reaction time for all the listeners.

3.5 Evaluation measures for *HMDI_DIA*

As for the second task (*HMDI_DIA*), listeners answered correctly 289 times out of 680 stimuli, which means a 42.5% score of language/dialect identification. Dialects within the Italo-Romance space were correctly identified at 57.3% (184 judgments out of 321).

We did not expect the Italian listeners to identify the dialects of those foreign languages which had not been identified in the first task (see §3.4); these stimuli were intended for foreign listeners and acted as distractors/reference noise for native Italian listeners. Conversely, the possibility of discrimination among Eastern, Western and Southern Sicilian was too ambitious for the current composition of the listener sample and served for comparisons. Partial scores are then collapsed into a total score (01-05 for the foreign languages and 10 for Sicilian, see *Table II*).

Fig. 5 shows the overall sample's responses in the second task. The plot clearly highlights that local dialects are perceived as such, in contrast with foreign languages. Appropriate responses to stimuli in languages other than Italian dialects are classified in the small, top-left square of Table II: while it is true that some listeners failed to positively identify some foreign languages (i.e. Polish and Romanian), they straightforwardly perceived such languages as unrelated to Italian dialects. The bigger, bottom-right square summarises the responses to dialect stimuli: again, the listeners generally identified the language they had listen to, Sardinian being the only exception. Sardinian has been correctly identified 8 times and confused 5 times with Veneto-Istrian, Sicilian and Portuguese, and 4 times with Spanish (minor confusion with other languages and dialects aside), with an extraordinary *ER* of 76%.

It is worth noticing that Sardinian has been perceived as a foreign language in 32% of cases whereas Veneto-Istrian has been confused with a foreign language in only one case (with Spanish).

Foreign languages have been identified as such with a 96% accuracy (325 correct answers), but listeners' also scored a 94% accuracy ratio in recognising dialect data as such. Of course, specific dialects scored 100% from listeners who previously declared a competence of them. Generally speaking, we may say instead that Sicilian (and Neapolitan), as well as Veneto-Istrian, provided good references for southern and northern broad dialectal areas for listeners who were not trained to detect subtler differences.

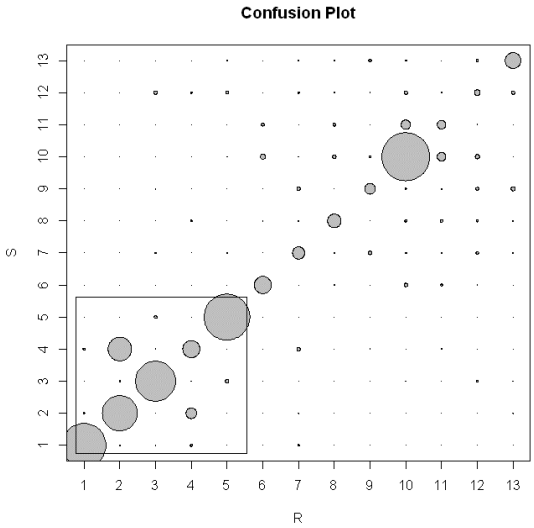


Fig. 5 – Confusion plot for the 13 stimuli (*S* axis) and responses (*R* axis) for the second task. See table II for language codes.

Table II. Confusion matrix (HMDI_DIA, §3.2)

	Responses												
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
01AR	61	1	1	3	0	0	2	0	0	0	0	0	0
02PL	2	49	0	15	0	0	1	0	0	0	0	0	1
03PT	0	2	56	1	5	0	0	0	0	1	0	2	1
04RO	3	33	1	24	0	0	5	0	1	0	1	0	0
05SP	0	0	4	0	64	0	0	0	0	0	0	0	0
06NA	0	0	0	0	0	24	0	1	0	5	3	1	0
07OC	0	0	2	0	1	0	17	1	5	2	1	4	1
08PG	0	0	0	2	0	1	1	19	0	3	4	3	1
09PM	0	0	1	0	0	0	5	0	15	2	1	4	6
10SC	0	0	0	0	1	7	1	5	3	67	12	6	0
11SL	0	0	0	0	0	4	0	4	0	13	12	1	0
12SR	0	0	5	2	4	0	2	1	1	5	1	8	5
13IV	0	0	0	0	1	0	1	1	4	2	0	3	22

4 Task for *LID/AID* systems

The speech samples presented in §2 were also designed for testing machine performances after a training of the *LID/AID* systems of each participant on longer and multispeaker samples downloadable in a *HMDI_TRAINING* folder. Candidates in testing their *LID/AID* systems were also invited to run it on telephonic or noisy samples available in the *HMDI_NOISY* folder.

4.1 Participation-results

Unfortunately, no participant chose to fully complete the proposed task procedure. Only three research teams previously showed their interest in it, but no documentation has been produced.

As a first attempt to compare human performances and the possibilities for automatic procedure to approximate them, we tested a few variables in our data that may prompt a more extensive pilot study on Italian dialects identification.

We particularly took into account listeners' comments pointing out the relevance for them of intonation cues. By the way, some listeners easily distinguished Polish and Portuguese, as well as Sardinian and Apulian, from the other languages or dialects, and reported that they relied on the overwhelming presence of fricative sounds in the stimuli for these varieties.

In facts, the stimuli used for Polish and Portuguese are characterised by the presence of 26 and 16 sharp fricative segments, respectively, vs. e.g. the number of fricatives affecting the passages in other languages (e.g. in the stimuli for Vietnamese, Baoulé or even Spanish and Veneto-Istrian, fricatives were limited to a selection of 6-9 fricatives with generally flat spectrum).

Overall variables accounting for general spectral properties, such as *CoG*, standard deviation (*st.dev*) or spectral tilt, are well taken into account for speech recognition and *LID* purposes (Wu *et alii* 2004). In our case, *CoG* and *st.dev* alone account for the discrimination of the two language groups (*st.dev* ranged over 1000 Hz for the former, whereas it was particularly low, < 700 Hz, for the latter). Even the zero-crossing scores discriminated the two groups, with higher values for 'sharp fricative languages' (> 2000 zc/s) vs. 'flat fricative languages' (< 1300 zc/s). Nevertheless, familiarity as well as areal, lexical or phonotactic features must have played a discriminating role within the same group, so allowing these listeners to distinguish e.g. Portuguese from Polish or Sallentinian from Occitan (all mostly ignored by the listeners). In particular, local prosodic signals and phonotactic regularities (whose importance is highlighted since Arai 1995; cp. Tong *et alii* 2006, 2009) are supposed to provide cues for human dialect identification.

5 Conclusion

Since no report about automatic *LID* on the proposed language/dialect datasets was delivered, this paper aimed at provisionally surveying only the main results of a series of experiments on language/dialect identification carried out with the help of a sample of 54 Italian listeners.

In particular, after a short review of the most widespread techniques in automatic *LID*, a pilot study has been proposed, which explores responses and reaction times and try to match individual scores with linguistic biographies.

An areal sensitivity has been confirmed and a clear-cut separation emerged between known, guessed and unknown dialects in terms of scores and reaction times.

The next step will consist in testing how a training may improve listeners' performances.

6 References

Adda Decker M. *et alii* (eds.) (2004). *Identification des langues et des variétés dialectales par les humains et par les machines - Proc. of MIDL* (Paris, France, Nov. 2004), Paris, ENST.

ALI – Massobrio L. *et alii* (1996-). *Atlante Linguistico Italiano* (http://www.atlantelinguistico.it/, last accessed July 2014).

Arai T. (1995). "Automatic language identification using sequential information of phonemes". *IEICE Trans.*, E78-D/6, 705-711.

Damashek M. (2005). "Gauging Similarity with n-Grams: Language Independent Categorization of Text". *Science*, 267/10, 843-848.

De Meo A., Vitale M., Pettorino M. & Martin Ph. (2011). "Acoustic-perceptual credibility correlates of news reading by native and non-native speakers of Italian". *Proc. of ICPhS2011* (Hong Kong, August 2011), 1366-1369.

Gauvain J.L., Messaoudi A. & Schwenk H. (2004). "Language Recognition Using Phone Lattices". *Proc. of ICSLP '04* (Jeju Island, South Korea, October 2004), 1283–1286.

Geoffrois E. (2004). « Identification automatique des langues : techniques, ressources, et évaluations ». In Adda Decker *et alii* (eds.), 43-44.

Suo H., Li M., Liu T., Lu P. & Yan Y. (2008). "The Design of Backend Classifiers in PPRLM System for Language Identification". *EURASIP Journal on Audio, Speech and Music Processing*, 6 p. (doi: 10.1155/2008/ 674859).

Kirchhoff K., Parandekar S. & Bilmes J. (2002). "Mixed-memory Markov Models for Automatic Language Identification". *Proc. of ICASSP2002* (Orlando, USA, May 2002), 2841-2844.

Ladefoged P. & Maddieson I. (1996). *The sounds of the world's languages*. Oxford, Blackwell.

Langscape – Maryland Language Science Center, University of Maryland - *Language Identification Tool* and *Language Familiarization Game* (http://langscape.umd.edu/, last accessed 27 Oct. 2014).

LDC – Linguistic Data Consortium - University of Pennsylvania (https://www.ldc.upenn.edu/, last accessed 27 Oct. 2014).

Leena M. & Yegnanarayana B. (2008). "Extraction and representation of prosodic features for language and speaker recognition". *Speech Communication*, 50, 782–796.

Lewis M.P., Simons G.F. & Fennig Ch.D. (eds.) (2014). *Ethnologue: Languages of the World*. Dallas, SIL International (17th ed.; http://www.ethnologue.com, last accessed 14 Oct. 2014).

Liu Y. & Kirchhoff K. (2013). "Graph-Based Semi-Supervised Learning for Phone and Segment Classification". *Proc. of Interspeech 2013* (Lyon, France, August 2013), 1839-1842.

Lopez-Moreno I., Gonzalez-Dominguez J., Plchot O. *et alii* (2014). "Automatic Language Identification Using Deep Neural Networks". *Proc. of ICASSP 2014* (Florence, Italy, May 2014), 5374-5378.

Loporcaro M. (2009). *Profilo linguistico dei dialetti italiani*. Roma-Bari, Laterza.

Maiden M. & Parry M. (eds.) (1997). *The Dialects of Italy*. London-New York, Routledge.

Muthusamy Y.K., Barnard E. & Cole R.A. (1994). "Reviewing automatic language identification". *IEEE Signal Processing Magazine*, 11/4,. 33-41.

Navrátil J. (2006). "Automatic Language Identification". In Schultz & Kirchhoff (eds.), 233-268.

Ohala J.J. & Gilbert J.B. (1981). "Listeners' ability to identify languages by their prosody". In: P. Leon & M. Rossi (eds.), *Problèmes de Prosodie: vol. 2*, Paris, Didier, 123-131.

Phoible – Moran S. & McCloy D. & Wright R. (eds.) 2014. *PHOIBLE Online*. Leipzig, Max Planck Institute for Evolutionary Anthropology (http://phoible.org/, last accessed 24 Oct. 2013).

PRAAT – Boersma P. & Weenink D. (1995-2013). *Praat: doing phonetics by computer* (http://www.fon.hum.uva.nl/praat/ v. 5.3.03, 2011).

Ramus F. & Mehler J. (1999). "Language identification with suprasegmental cues: A study based on speech resynthesis". *J. A. S. A.*, 105/1, 512-521.

R-language – The R Project for Statistical Computing (http://www.r-project.org/, last acc. 14 Oct. 2014).

Romano A. (1997). "Persistence of prosodic features between dialectal and standard Italian utterances in six sub-varieties of a region of Southern Italy (Salento): first assessments of the results of a recognition test". *Proc. of EuroSpeech97* (Rhodes, Greece, September 1997), 175-178.

Schultz T. & Kirchhoff K. (eds.) (2006). *Multilingual Speech Processing*. Amsterdam, Elsevier Academic Press.

Singer E., Torres-Carrasquillo P.A., Gleason T.P., Campbell W.M. & Reynolds D.A. (2003). "Acoustic, phonetic, and discriminative approaches to automatic language identification". *Proc. of Eurospeech 2003 - Interspeech 2003* (Geneva, Switzerland, September 2003), 1345-1348.

Swets J.A. (1964). *Signal detection and recognition by human observers: contemporary readings*. New York, Wiley & sons.

Timoshenko E. & Bauer J.G. (2006). "Unsupervised adaptation for acoustic language identification". *Proc. of ICSLP2006* (Pittsburgh, USA, September 2006), 409-412.

Timoshenko E. (2012). "Rhythm Information for Automated Spoken Language Identification". *PhD Thesis*, Technischen Universität München (https://mediatum.ub.tum.de/doc/1063301/106330.pdf, last accessed 28 Oct. 2014).

Tong R., Ma B., Li H. & Chng E.S. (2009). "A target-oriented phonotactic front-end for spoken language recognition". *IEEE Transactions on Audio, Speech and Language Processing*, 17/7, 1335-1347.

Tong R., Ma B., Zhu D., Li H. & Chng E.S. (2006). "Integrating Acoustic, Prosodic and Phonotactic Features for Spoken Language Identification". *Proc. of ICASSP2006* (Toulouse, France, May 2006), 205-208.

Tsai W.H. & Chang W.W. (2002). "Discriminative training of Gaussian mixture bigram models with application to Chinese dialect identification". *Speech Communication*, 36, 317-326.

Vaissière J. & Boula de Mareüil Ph. (2004). "Identifying a language or an accent: from segments to prosody". In Adda Decker *et alii* (eds.), 1-4.

WALS – B. Comrie *et alii* (eds.), *World Atlas of Linguistic Structures* (http://wals.info/, last accessed 14 Dec. 2013).

Wang L. (2008). "Automatic Spoken Language Identification". *PhD Thesis*, The Univ. of New South Wales (http://www.nicta.com.au/pub?doc=1784, last accessed 28 Oct. 2014).

Wu T., Van Compernolle D., Duchateau J., Yang Q. & Martens J.P. (2004). "Spectral Change Representation and Feature Selection for Accent Identification Tasks". In Adda Decker *et alii* (eds.), 57-61.

Yan Y. & Bernard E. (1995). "An approach to automatic language identification based on language-dependent phone recognition". *Proc. ICASSP '95* (Detroit, USA, May 1995), 3511-3514.