Fluency Articulation and Speech Rate as new parameters in the Speaker Recognition

Luciano Romito, Rosita Lio, Vincenzo Galatà

(Laboratorio di Fonetica, Dipartimento di Linguistica, Università della Calabria)

Abstract

The subject of this study is to compare, in the Speaker Recognition area, a new objective method, dealing with the domain of time, with the perceptual method.

After a brief review of confused definitions in linguistics literature about fluency indexes, we will focus on the Articulation Rate and Speech Rate as parameters to be used for the identification of a speaker in a closed-set of known voices. We calculated the Articulation and Speech Rate of samples of read speech using five female speakers with same diatopic and diastratic features having also parental relationships.

Introduction

Over the years, many and various solutions have been proposed to solve the problem of the identification of an unknown speaker.

The human ability to recognize, in daily life, the voice of a known person[1], has led to the belief that the identification by means of voice can have a possible application in judicial enquiries.

On account of its variable nature, the voice would not appear to be suitable for such a complex field. However, scientific knowledge of phonetic parameters depending on the speaker (such as the fundamental frequency F_0, the vowel formants' frequencies[2], the resonance of nasal consonants FN1, FN2, FN3, the Articulation Rate and the Speech Rate[3]), allows the use of the voice for speaker recognition in forensics.

The ascription of a voice to a speaker requires analysis, observation, evaluation and comparison of parameters.

Since this study focuses on the search for new parameters to be used in Speaker Recognition, the typical procedures of the semi-automatic method have been adopted:

- choice of acoustic parameters to be analysed;

[1] The identification procedures are however different in forensic speaker recognition because the voices are always unknown.

[2] These parameters are approved and used in judicial enquiries:

[3] These parameters are for the moment only used in laboratory for experimentation.

- evaluation of the statistical relevance of the considered parameters;
- estimation and comparison of samples by means of a statistical test (using the results of previous measurements).

The goal of this study is to verify the reliability of the fluency indexes of Articulation Rate and Speech Rate, through the verification of the identity of a speaker in a closed-set of known voices.

The audio-verbal communication is realized through the production and reception of phonic blocks of varying size (sentence, word groups, single words or fragments of a word). These blocks are internally modeled by a certain pitch, by speed variation and by stress position. The features of the fluency indexes are prosodic phenomena or, more frequently, over-segmental ones[4]. The over-segmental features of phones differ from the segmental ones because of their intrinsic nature in relation to the surrounding phones: a phone is considered in terms of more or less duration, intensity, quality considered not in absolute terms but always in reference to the context in which it finds itself.

The Articulation Rate depends on the intrinsic duration of the various phones, on the rate of articulation movements and on the rules of co-articulation.

The Speech Rate depends on speaker-specific features, on the communicative situation (diatopic and diastratic variables), but it varies also in a sentence with accelerations and contractions indicating, among other things, the degree of attention that the speaker asks the listener (speed for less attention, slowness for more attention). Furthermore, the rhythm of speech in a similar situation depends on the phonetics alphabet used[5].

The study that follows is divided into three steps:

1. Study of reliability of the fluency index of Articulation Rate;
2. Study of reliability of the fluency index of Speech Rate;
3. re-evaluation of the perceptual method using a group of *naïf* listeners.

Articulation Rate and Speech Rate: some definitions

According to the definition of the indexes, it is important to establish what parameters it is necessary to consider and how to calculate them.

[4] This means above the segments because they refer to the entire sequence.

[5] Studies on three different varieties (Bari, Napoli and Pisa) demonstrate that the production of syllables in less time is not a voluntary strategy, but it depends on the different phonetics alphabets used (Pettorino 2003).

By means of a brief review of some relevant definitions in linguistics literature, it is possible to observe how, over the years, there have been various attempts to define the fluency indexes of Articulation Rate and Speech Rate.

For example, for the Articulation Rate, there is:

- *"La vitesse d'articulation. Véritable vitesse de phonation, quisqu'on lui retrace le temps de pauses, elle est exprimée en nombre de syllabe/sec et s'obtient en divisant le nombre de syllabes émises par le temps d'articulation du locuteur"* (Grosjaen and Deschamps 1975: 148) "Articulation Rate. Real speech rate, since we count the time of pauses, it is expressed in number of syllables per second and can be calculated by dividing the number of syllables uttered by the speaker's articulation time";
- "The articulation rate of each group ('utterance') was computed by dividing the total number of syllables in the group by the cumulative duration of the runs comprising the group (excluding any pause time)", where run is "the stretch of speech that contains no pauses, with a pause defined as a silent interval of 250 ms or greater" (the calculus of Articulation Rate is made on a group of runs with a minimum of 30 syllables, (Miller *et al.* 1984: 218-219);
- "*il rapporto tra il numero delle sillabe e la durata della catena fonica*[6]" (Sorianello 1996:95) "The ratio between the number of syllables and the duration of the speech chain*"*;
- "*il numero delle sillabe diviso per la durata della catena fonica*[7]" (Magno Caldognetto and Vagges 1993: 101, recalling Duez 1982) "The number of syllables divided by the duration of the speech chain";
- "*espresso come il numero di sillabe al secondo della sequenza articolata*[8]*, risultante dalla formula: numero delle sillabe della sequenza articolata/durata della sequenza articolata*" (Zmarich *et al.* 1996:120) "Expressed as the number of syllables per second of the speech chain resulting from the formula: number of syllables of the uttered sequence/duration of the uttered sequence";
- "average number of (phonetic) syllables per second of the articulation phase of speech[9] (number of syllable/[duration- combined duration of all pauses])" (Künzel 1997: 1358);

[6] Here the writer uses the definition of speech chain by Pettorino and Giannini 1994: "*La catena fonica è la porzione di un enunciato compresa tra due pause vuote*".

[7] The speech chain is for the authors: "*Total articulation time: durata globale della produzione verbale del soggetto, costituita da catene foniche e dai silenzi*".

[8] That is the speech chain without the non silent pauses.

[9] In this case it is the duration of each speaker's task, equivalent to half a minute.

- "*la velocità di articolazione è data dal rapporto tra il numero delle sillabe e la durata delle catene foniche. Nel computo delle sillabe vengono di norma inclusi tutti quei fenomeni udibili quali pause piene e prolungamenti vocalici.*" (Giannini 2000: 253) "The articulation rate is given by dividing the number of syllables by the duration of the speech chains. Any audible phenomenon, such as full pauses and lengthened vowels, are normally included in the number of syllables";
- "*equivale al rapporto tra il numero delle sillabe realmente pronunciate e il tempo impiegato per realizzarle*" (Pettorino 2003: 228) "equals the ratio between the number of syllables really pronounced and the time used to realize them".

The first difficulties that we run into after this brief but important review of definitions are those related to the different ways to consider what parts of the signal to examine: "*temps d'articulation du locuteur*" (articulation time of the speaker), run, speech chain, uttered sequence, the articulation phase of speech, the time used to realize the syllable and the different definitions of uttered sequence and speech chain. Moreover, the different definitions of silent pauses[10] and the various limits given when silence occurs followed by stop consonants [11].

Some definitions of Speech Rate are:

- "*il rapporto tra il numero delle sillabe e la durata dell'enunciato*" (Sorianello 1996: 95) "The ratio between the number of syllables and the duration of the enunciation";
- *"Speech Rate: espresso come numero di sillabe al secondo della sequenza articolata, in relazione alla durata della catena fonica comprensiva di esitazioni e disfluenze, risultante dal rapporto tra il numero delle sillabe della sequenza articolata e la durata dell'intera catena fonica"* (Zmarich *et al.* 1996: 120) "Speech rate: expressed as the number of syllables per second of the uttered sequence in relation to the duration of the phonic chain including hesitations and disfluencies resulting from the ratio between the number of syllables in the uttered sequence and the duration of the entire phonic chain";
- "*rapporto tra il numero di sillabe e il tempo totale dell'enunciato*" (Pettorino 2003: 228) "Ratio between the number of syllables and the total time of enunciation".

In general, the result is a certain degree of confusion.

[10]Miller *et al* 1984: "a silent interval of 250 ms or greater"; Sorianello 1996 "*una momentanea sospensione dell'attività fonatoria e conseguente assenza di rumore spettrografico, di almeno 100 ms nel parlato spontaneo*"; Künzel 1997 gives the limit of 100 ms.

[11] Magno Caldognetto and Vagges 1993 give 100 ms to the phase of consonantic occlusion, whilst Zmarich *et al* 1997, give 50 ms.

In this study, we will refer to Zmarich *et al.* (1996: 120) to calculate the Articulation Rate while to calculate the Speech Rate we used the definition of these writers considering the formula: number of syllables of the uttered sequence/duration of the phonic chain, expressed in syll/sec.

Materials and Methods

The materials used in this study are recordings of read speech taken from four female speakers having the same diastratic and diatopic features (same peer-group, same culture degree etc). The four speakers, indicated as D1, D2, D3, D4, have also parental relationships: D1 (age 25) is sister of D4 (age 24); D2 and D3 are twins (age 23) and are aunts of D1 and D4. The recordings consist of a list of ten sentences and a list of ten repetitions of a same sentence. The sentences have been constructed trying to avoid the beginning of the words by stop consonants: in this way it was possible to detect the presence of possible silent pauses the speaker used within a sentence. The segmentation[12] of the sentences was hand made by means of the visualization of three windows: waveform, spectrogram, energy curve and with the help of auditory feed-back. The syllable computation was carried out by identifying the phonetic syllable[13] according to the rules of syllabification[14].

We considered only uttered sequences and phonetic chains with a number of syllables superior or equal to eight, to remove the possibility of misleading data. We did not establish limits to the duration of pauses. The material as described was used for the calculation of articulation rate and speech rate.

Considering the parental situation of our speakers, we decided to apply the perceptual method to verify if the voices would be recognized in the right way although their similarity and although the different recording channels. For this task we made two types of recording for each one of the speakers: one in a silent room and the other one through the telephone line. To the four speakers used for the preceding task we added here another female speaker, D5 (age 45) mother of D1 and D4 and sister of D2 and D3 to complicate the listeners' recognition task. The speakers were asked to read two passages from a newspaper of no less than 120-130

[12] A phonetic transcription with informations on the temporal setting of unit boundaries (see Salza 1990: 24-25).

[13] The notion of syllable has been object of study for a long time because of its phonetic and phonological "double nature"; in this study we refer respectively to: "*la struttura elementare che sta alla base d'ogni raggruppamento di fonemi*" (see Jakobson 1974: 94) and "*l'unità prosodica costituita da uno o più foni, agglomerati intorno ad un picco d'intensità*"(see Leoni and Maturi 1998: 74-5).

[14] Differences between phonetic and phonological syllabification are not so great, only in the case of group /s/ and /z/ + consonant (for example "*festa*" is phonetically divided into ['fɛs-ta] and phonologically in ['fɛ-sta]), and some connections, not autochthonous, such as /tm/ and /tl/ (for example '*atmosfera*, *atleta'*), see Muljačič 1969: 471.

seconds. The reading of the two passages was repeated more times to achieve more spontaneity in the reading. For the experiment we extracted only 6 sentences for each speaker (3 ones for the silent room recording and 3 ones for the telephone line) of minimum 2,7 seconds of length[15]. For the test we used 60 *naïf* listeners (university students of both sexes) who were asked to answer "yes" if the voice compared belonged to the same speaker and "no" in the case it did not. In doubtful cases the listeners could answer "perhaps" thus expressing that it could belong to the same speaker. To verify the validity of the listeners' answers we used 26 control tests excluding those listeners who did not answer them correctly.

Data Analysis

General Consideration

The first thing to keep in mind is that we are dealing here read speech.

In the case of repetitions of a same sentence it is possible to see how each speaker plans approximately the same duration for the production and how it differs among the four speakers.

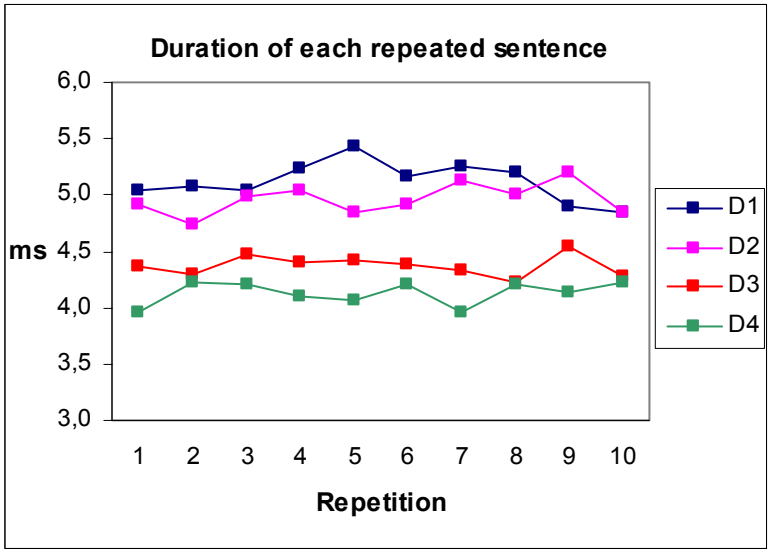

Figure 1 Duration of each repeated sentence considering all kind of pauses.

Articulation Rate

We remember that for the Articulation Rate we used here the formula of Zmarich *et al.* (1996: 120) resulting from: number of syllables of the uttered sequence/duration of the uttered sequence.

[15] According to Romito L., 2000, pp.199-203.

The indexes of Articulation Rate calculated for the four speakers are presented in Table 1 and Table 2 for the two sentences' lists.

	10 repetitions of the same sentence																												μ	σ	n
speaker	**1**			**2**			**3**			**4**			**5**			**6**			**7**			**8**			**9**			**10**					
D1	6,2	-		6,0	-		6,0	-	-	6,2	6,0	-	6,3	5,7	-	6,4	5,8	-	6,6	5,5	-	6,7	6,0	-	6,6	6,1	7,1	6,3	-	6,2	0,4	17
D2	6,9	6,1		6,9	6,6		7,1	6,4	6,3	6,8	5,9	-	7,2	6,4	6,0	7,3	6,0	-	6,8	3,6	5,7	6,8	6,4	5,9	6,8	6,1	7,1	6,4	6,2	6,4	0,7	25
D3	6,9	-		7,0	-		6,7	-	-	6,8	-	-	6,8	-	-	6,9	-	-	6,9	-	-	7,1	-	-	6,5	7,0	7,0	-	-	6,9	0,2	11
D4	7,6	-		7,5	-		7,8	-	-	7,6	7,9	-	7,9	7,9	-	7,9	-	-	7,9	7,7	-	7,4	-	-	7,5	7,7	7,5	-	-	7,7	0,2	14
μ: Mean Articulation Rate; **σ:** Articulation Rate's standard deviation; **n:** number of values																															

Table 1 Articulation Rate index calculated for each uttered sequence of the 10 repetitions of the same sentence.

	single sentences																												μ	σ	n
speaker	**1**			**2**			**3**			**4**			**5**			**6**			**7**			**8**			**9**			**10**					
D1	5,1	-		5,1	-		5,0	5,0	-	5,7	4,9	-	5,3	5,3	-	5,3	5,0	-	5,3	6,0	5,1	5,3	4,8	-	6,0	-	5,5	5,7	-	5,3	0,4	18
D2	5,4	-		7,0	-		6,7	6,4	-	7,3	5,8	-	7,1	5,7	6,1	6,6	5,5	6,8	6,8	7,4	6,2	6,6	5,8	-	6,9	-	7,2	7,1	-	6,5	0,6	20
D3	6,0	-		6,6	-		6,9	-	-	7,4	6,1	-	6,6	-	-	6,1	-	-	7,0	6,0	-	6,2	-	-	6,8	-	6,8	-	-	6,5	0,4	12
D4	5,6	-		6,2	-		6,2	-	-	6,4	6,9	6,3	6,6	-	-	6,7	6,0	6,8	7,2	5,6	-	6,8	-	-	8,1	6,7	7,8	-	-	6,6	0,7	16
μ: Mean Articulation Rate; **σ:** Articulation Rate's standard deviation; **n:** number of values																															

Table 2 Articulation Rate index calculated for each uttered sequence of the 10 single sentences.

At a first look we can see that within the same speaker there is already great difference if we consider the mean articulation rate and the standard deviation for the two sentences' lists.

10 repetitions of the same sentence					single sentences			
D1	**D2**	**D3**	**D4**		**D1**	**D2**	**D3**	**D4**
6,2	6,4	6,9	7,7	**μ**	5,3	6,5	6,5	6,6
0,4	0,7	0,2	0,2	**σ**	0,4	0,6	0,4	0,7
17	25	11	14	**n**	18	20	12	16
μ: Mean Articulation Rate; **σ:** Articulation Rate's standard deviation; **n:** number of values								

We used here the T-test to calculate the dissimilarity percentages, which provide the results in Table 3 for each comparison.

Intra-speaker repetitions vs. single	p	% dissimilarity
D1 - D1	0,0000000598	99,9999940152
D2 - D2	0,4748176757	52,5182324259
D3 - D3	0,0311056454	96,8894354560
D4 - D4	0,0000150707	99,9984929281

Table 3 T-test results for the Articulation Rate.

The difference between repetitions of a same sentence and single sentences within the same speaker is probably due to the increasing confidence with the repeated sentence and according to us it does not represent a problem in this case considering it quite normal.

Inter-speaker repetitions	p	% dissimilarity
D1 - D2	0,86573179349	**67,1623751786**
D1 - D3	0,00000342724	99,9996572758
D1 - D4	0,00000000001	99,9999999998
D2 - D3	0,00006998279	99,6105945181
D2 - D4	0,00000000009	99,9999996662
D3 - D4	0,00000000073	99,9999999922

Table 4 T-test results for the Articulation Rate.

According to what we said before for the repetitions the test seems to work well for all the speakers giving results higher than 99% except for the comparison between D1 and D2 with 67,2%.

Inter-speaker single sentences	p	% dissimilarity
D1 - D2	0,00000002956	99,9999982412
D1 - D3	0,00000009297	99,9999873364
D1 - D4	0,00000113624	99,9999240993
D2 - D3	0,78793137618	**5,0165460178**
D2 - D4	0,56213540878	**32,0611884799**
D3 - D4	0,71477394077	**28,7261784387**

Table 5 T-test results for the Articulation Rate.

On the other hand, considering single sentences, all the speakers seem to plan differently the single productions with a larger range of variability. This situation has its repercussions on the statistical results where we find only D1 well differentiated from all the other speakers.

Speech Rate

If the goal of this study is to find out a temporal parameter that characterizes the speaker and that describes his production task using the syllable number and the temporal aspect, it is necessary to consider all of the phonic production, including the syllables of non-silent pauses.

This choice is also suggested by results in Duez (1982)[16]: for this author, the phenomenon of hesitation is a speaker-specific feature, above all in the tendency to use or not to use full pauses.

To calculate the Speech Rate we used here the formula: number of syllables of the uttered sequence/duration of the phonic chain, expressed in syll/sec.

The indexes of Speech Rate calculated for the four speakers are presented in Table 3 and Table 4 for the two sentences' lists.

[16] Duez D. (1982): in a study on three styles of speech in French.

	10 repetitions of the same sentence																												μ	σ	n
speaker	**1**		**2**		**3**			**4**			**5**			**6**			**7**			**8**			**9**			**10**					
D1	6,2	-	6,0	-	6,0	-	-	6,2	6,0	-	6,3	5,7	-	6,4	5,8	-	6,6	5,5	-	6,7	6,0	-	6,6	6,1	7,1	6,3	-	6,2	0,4	18
D2	6,6	6,0	6,5	6,4	6,8	6,1	6,3	6,5	5,9	-	7,0	6,4	6,0	7,0	6,0	-	6,5	3,6	5,7	6,5	6,4	5,9	6,5	6,1	6,6	6,4	6,2	6,2	0,6	25
D3	6,9	-	7,0	-	6,7	-	-	6,8	-	-	6,8	-	-	6,9	-	-	6,9	-	-	7,1	-	-	6,5	7,0	7,0	-	-	6,9	0,2	11
D4	7,6	-	7,5	-	7,6	-	-	7,4	7,9	-	7,7	7,8	-	7,5	-	-	7,7	7,7	-	7,4	-	-	7,5	7,7	7,3	-		7,6	0,2	14
μ: Mean Speech Rate; **σ:** Speech Rate's standard deviation; **n:** number of values																															

Table 6 Speech Rate index calculated for each uttered sequence of the 10 repetitions of the same sentence.

	single sentences																												μ	σ	n
speaker	**1**		**2**		**3**			**4**			**5**			**6**			**7**			**8**			**9**			**10**					
D1	5,1	-	5,1	-	5,0	5,0	-	5,7	4,9	-	5,3	5,3	-	5,3	5,0	-	5,3	6,0	5,1	5,3	4,8	-	6,0	-	5,5	5,7	-	5,3	0,4	18
D2	5,3	-	7,0	-	6,7	6,4	-	7,3	5,8	-	7,1	5,7	6,1	6,6	5,5	6,8	6,8	7,0	5,9	6,6	5,8	-	6,9	-	6,9	7,1	-	6,5	0,6	20
D3	6,0	-	6,6	-	6,9	-	-	7,3	6,1	-	6,6	-	-	6,1	-	-	7,0	6,0	-	6,2	-	-	6,8	-	6,8	-	-	6,5	0,4	12
D4	5,6	-	6,2	-	6,2	-	-	6,1	6,9	6,3	6,6	-	-	6,7	6,0	6,8	7,2	5,6	-	6,8	-	-	8,1	6,7	7,8	-	-	6,6	0,7	16
μ: Mean Speech Rate; **σ:** Speech Rate's standard deviation; **n:** number of values																															

Table 7 Speech Rate index calculated for each uttered sequence of the of the 10 single sentences.

10 repetitions of the same sentence					single sentences			
D1	**D2**	**D3**	**D4**		**D1**	**D2**	**D3**	**D4**
6,2	6,2	6,9	7,6	**μ**	5,3	6,5	6,5	6,6
0,4	0,6	0,2	0,2	**σ**	0,4	0,6	0,4	0,7
17	25	11	14	**n**	18	20	12	16
μ: Mean Speech Rate; **σ:** Speech Rate's standard deviation; **n:** number of values								

Also in this case we have the same results as those presented for the Articulation Rate. This is due to the fact that we are here considering read speech.

Inter-speaker repetitions	p	% dissimilarity
D1 - D2	0,86573179	**13,4268206514**
D1 - D3	0,00000343	99,9996572758
D1 - D4	0,00000000	99,9999999988
D2 - D3	0,00006998	99,9930017206
D2 - D4	0,00000000	99,9999999910
D3 - D4	0,00000000	99,9999999273

Table 8 T-test results for the Speech Rate.

Inter-speaker single sentences	p	% dissimilarity
D1 - D2	0,00000003	99,9999970441
D1 - D3	0,00000009	99,9999907030
D1 - D4	0,00000114	99,9998863761
D2 - D3	0,78793138	**21,2068623821**
D2 - D4	0,56213541	**43,7864591221**
D3 - D4	0,71477394	**28,5226059233**

Table 9 T-test results for the Speech Rate.

Perceptual method

According to the control test we considered only 40 listeners as reliable. From the results presented in Table 10 and Table 11 it is possible to see how the greatest part of the identification percentages are over the 90 %. The tables do not need further comments.

silent room recordings								
Comparison			Yes	No	Perhaps	Yes	No	Perhaps
1	D4	D5	9	105	6	7.5%	87.5%	5%
2	D4	D1	16	101	3	13.33%	84.16%	5%
3	D4	D2	4	115	1	3.33%	95.83%	0.83%
4	D4	D3	9	111	0	7.5%	92.5%	0%
5	D5	D1	2	115	3	1.66%	95.83%	2.5%
6	D5	D2	5	112	3	4.16%	93.33%	2.5%
7	D5	D3	7	105	8	5.83%	87.5%	6.66%
8	D1	D2	12	104	4	10%	86.66%	3.33%
9	D1	D3	11	103	6	9.16%	85.83%	5%
10	D2	D3	11	104	5	9.16%	86.66%	4.16%

Table 10 Results for the perceptual test.

telephone line recordings								
Comparison			Yes	No	Perhaps	Yes	No	Perhaps
1	D4	D5	7	112	1	5.83%	93.33%	0.83%
2	D4	D1	16	97	7	13.33%	80.83%	5.83%
3	D4	D2	6	112	2	5%	93.33%	1.66%
4	D4	D3	9	106	5	7.5%	88.33%	4.16%
5	D5	D1	8	109	3	6.66%	90.83%	2.5%
6	D5	D2	2	117	1	1.66%	97.5%	0.83%
7	D5	D3	9	104	7	7.5%	86.66%	5.83%
8	D1	D2	14	103	3	11.66%	85.83%	2.5%
9	D1	D3	17	91	12	14.16%	75.83%	10%
10	D2	D3	13	105	2	10.83%	87.5%	1.66%

Table 11 Results for the perceptual test.

CONCLUSIONS

On the basis of a careful analysis of the used methods and of the obtained results, we can conclude as follows:

- We have statistically significant results for Articulation Rate and Speech Rate through the objective method when the temporal planning of the speaker is characteristic. On the other hand, considering speakers with a similar fluency rate we reach low recognition percentages;
- For Speech Rate we indicate here the mean production of syllables for each phonetic chain, where for phonetic chain we mean the sequence of phonetic segments (including the non-silent pauses) between two silent pauses, and/or portions of signal included between one inspiration and another.

- To analyse the importance of the fluency index of Speech Rate, as a parameter in Speaker Recognition, it is necessary to consider the whole phonic realization of a speaker and making statistical comparisons with samples of voice of the same, and/or of different speakers, to support the used method considering a common segmentation protocol. The need to consider the speaker's entire phonic realization means in fact identifying the phonetic syllables that build the uttered sequence (including the non-silent pauses).
- Considering the perceptual method as we considered it, the test seems to give good results even with similar sounding voices (even if we have percentages that never reach the 99%).

Our purpose for the future is to calculate the indexes of Articulation Rate and Speech Rate using the same speakers (for their parental situation) analysing read passages to verify if it is possible to obtain more homogeneous data. On the other hand spontaneous speech will be taken into account because of the characteristic fragmentation of on-line planning[17] considering those features strictly bound to the realisation of uttered sequences and/or phonic chains of differing length caused by the syllables of the so-called non-silent pauses[18], the lengthening of vowels[19], the reduction and the diphthongization[20] and the fall of syllables or parts of them.

Once obtained and evaluated the results we will compare the voices of the above said speakers using the parametric approach used in Speaker Recognition (weighting in statistical terms the Articulation and Speech Rate indexes as information to be used in SR).

Moreover it is very important for this experiment to create a corpus containing spontaneous speech (with pragmatic variables) to be used as training and test set.

Last but not least, we will try to put down some guidelines for the tasks of segmentation, calculus and analysis of the various indexes defining some rules and finding homogeneous and non-ambiguous definitions to be used to achieve the same results on the same material in different laboratories.

---

[17] Spontaneous speech is characterized by a tendency towards the reduction of articulatory effort, causing phenomena of assimilations, centralizations, de-accentuations, elisions and, from the acoustic point of view, medium shortening of all the phones, fall of phones and/or syllable tendency to omit segments of the sequence, see Kohler, 1995, op. cit. in Zmarich *et al.* 1997.

[18] "*Pause piene: esitazioni, interiezioni, allungamenti di vocale, disfluenze, ecc.*": Zmarich *et al.* 1997.

[19] The spectro-acoustical analysis of some vowels' "length" at the end of a word or in hesitations, suggests the production of two or more phonetic syllables.

[20] Because of the combination and/or meeting of vowels in adjacent words.

Bibliography

Duez D. (1982), Silent and non-silent pauses in three speech styles, *Language and Speech*, 3: 179-192.

Grosjaen F., Deschamps A. (1975), Analyse contrastée des variables temporelles des l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation, in *Phonetica*, 31: 144-184.

Jakobson R. (1974), *Saggi di linguistica generale*, a cura di L.Helmann, Milano, Feltrinelli.

Künzel H. J. (1997), Some general phonetic and forensic aspect of speaking tempo, in *Forensic Linguistics*, 4(1): 48-83.

Leoni F.A., Maturi P. (1998)*, Manuale di fonetica*, Roma, Carocci.

Magno Caldognetto E., Vagges K. (1993), Le pause quali indici diagnostici per lo stile del parlato spontaneo, in *Atti delle 2^{e} Giornate di Studio del G.F.S*., Calabria, 28-29 novembre 1991, 19: 97-106.

Miller J.L., Grosjaen F., Lomanto C. (1984), Articulation Rate and its variability in spontaneous speech: a reanalysis and some implication, in *Phonetica*, 41: 215-225.

Mori L., Paoloni A., (2004), Sulla sociolinguistica forense:la costituzione di corpora vocali per l'analisi della velocità di articolazione in italiano, in *Atti delle XIV Giornate del GFS, Viterbo, 4-6 Dicembre 2003.*

Muljačič Z. (1969), *Fonologia generale e fonologia della lingua italiana*, Bologna, il Mulino.

Pettorino M. (2003), Caratteristiche prosodiche dell'italiano dialogico in *Voce Canto Parlato, studi in onore di F.Ferrero*, a cura di P. Cosi, E. Magno Caldognetto, A. Zamboni, Unipress, Padova, pp. 227-230.

Romito L. (2000), *Manuale di fonetica articolatoria, acustica e forense*, Centro Editoriale e Librario Università della Calabria, Cosenza.

Romito L. Blefari M. (2003), *Verso un nuovo parametro nel riconoscimento del parlatore* unpublished thesis, University of Calabria.

Salza P. L. (1990), La problematica della segmentazione del segnale vocale, in *Atti delle 1^{a} Giornata di Studio del G.F.S*, Padova, 3-6-novembre.

Sorianello P. (1996), Dal parlato letto al parlato spontaneo: indici prosodici a confronto, in *Atti delle 7^{e} Giornate di Studio del G.F.S,* Napoli, 14-15 novembre, 7: 89-110.

Zmarich C., Magno Caldognettto E., Ferrero F. (1997), Analisi confrontativa di parlato spontaneo e letto: fenomeni macroprosodici e indici di fluenza, in *Quaderni del CNR*, 16: 266-290.