A CORRELATION STUDY BETWEEN SPEECH-RELATED FEATURES AND PERSONALITY TRAITS

A.Guidi[1,2], C. Gentili [3], E.P. Scilingo[1,2] and N.Vanello [1,2]

[1] Research Center "E. Piaggio", University of Pisa, Pisa,
[2] Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy
[3]Department of General Psychology, University of Padua, Padua, Italy
guidi.andrea84@gmail.com, c.gentili@unipd.it, e.scilingo@ing.unipi.it, nicola.vanello@unipi.it

Abstract: **Voice signal has been widely investigated to characterize mood and emotional states. A further interesting dimension could regard the personality traits. In fact, speech production can be related to personality traits evaluated by others. The relationship between personality traits and specific speech features is not yet fully understood and requires further investigation. In this study, a correlational analysis between some speech-related features and the personality traits, as described by the Zuckerman-Kuhlman model, is performed. An experimental protocol was administered to eighteen healthy subjects to investigate both fundamental frequency and voice quality related features. Results showed that a skewness-like measure of the fundamental frequency is negatively correlated with the Sociability dimension. The impact of personality traits and speech production studies on the characterization of mental disorders and the estimation of emotional/mood state of the speaker are discussed.**

Keywords : **personality traits, Zuckerman-Kuhlman model, Fundamental frequency, spectral slope, jitter**

I. INTRODUCTION

The analysis of speech signal allows to explore several psychological dimensions: emotion [1], mood [2], and stress [3] were widely studied in relation to the speakers' speech production. A further interesting dimension could be related to the personality traits, whose effects might overlap to the ones related to emotion and/or mood. Speech intonation parameters might be related to a set of individual and sociocultural means that can allow reaching different communication goals [4]. Probably, such a relation might be stronger in people showing some particular personality trait. According to the trait theory [5], traits can be defined as "stable internal characteristics that people display consistently over time and across situations". Different studies attempted to investigate the relationship between voice and personality. Sapir [6] proposed the hypothesis of "speech as a personality trait". Addington [7] reported that a higher pitch variation in males was perceived as more dynamic, feminine and aesthetically inclined, while in female was rated as more dynamic and extravert. Again, some prosodic features, such as mean pitch, pitch variation and speaking rate, were found to be related to the perception of competence, benevolence, extraversion, dominance and political charisma in [8,9]. These studies showed that a voice characterized by a high (low) pitch variation and a high (low) speaking rate was perceived as index of high (low) competence, while a voice showing a low pitch variation and a high speaking rate was judged with low benevolence ratings, and vice versa. Similarly, a negative (positive) correlation between mean pitch and both extraversion and dominance was detected in American female (male) speakers [8]. Political charisma and leadership were reported to be positively correlated with higher pitch in [9]. Spectra and voice quality features were also investigated in relation to personality perception in [10,11]. A correlation between speech fluency and both extroversion and neuroticism was observed in [12]. The INTERSPEECH 2012 Speaker Trait Challenge showed that the classification of personality traits, as defined by the OCEAN five personality dimensions [13], is feasible. Within this challenge, hundreds of short clips, on average lasting 10 s, were evaluated by a pool of judges to assess the personality traits by using the Big Five Inventory questionnaire [14]. Acoustic features outperformed linguistic and psycholinguistic features to achieve an automatic recognition of speaker personality trait [15].

Although some useful indications about a significant relationship between personality traits and voice production can be drawn from the literature review, the work on this topic is far from being concluded. For instance, the currently available studies mostly rely on the estimation of the perceived personality traits, without exploring the possibility of using dedicated personality tests. Moreover, the relationship of personality traits and specific speech features have still to be clarified.

In this study, a correlational analysis between some speech-related features and the personality traits, as

Claudia Manfredi (edited by), *Models and analysis of vocal emissions for biomedical applications : 10 th international workshop : December 13-15, 2017*, ISBN 978-88-6453-606-4 (print) ISBN 978-88-6453-607-1 (online)

described by the Zuckerman-Kuhlman model [16], is performed. Specifically, a correlation analysis between features related to speech fundamental frequency (F_0) and voice quality, and the six factors of personality traits defined in the above cited models, will be conducted. This study will be performed on healthy subjects using a structured speech task.

II. METHODS

Experimental Protocol: Eighteen healthy subjects (12 females, 23.66 ± 2.28 year) without any history of psychiatric disorder were enrolled. Subjects were asked to fill out the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ) at home, about 4 days before performing the experimental protocol. The ZKPQ is a self-report questionnaire that provides information about personality in terms of five dimensions: Impulsive Sensation Seeking (ImpSS), Aggression-Hostility (Agg-Host), Sociability (Sy), Neuroticism-Anxiety (N-Anx) and Activity (Act). Subjects were asked to read a neutral text ("The universal declaration of Human rights", lasting 3 minutes) twice, at the beginning and at the end of the experimental protocol, after about 30 minutes. In the following, the first and the second reading task will be indicated as Rd1 and Rd2, respectively. In addition, they were asked to comment a set of Thematic Apperception Test (TAT) images [17], between the two neutral text reading tasks. Subjects were driven to comment all of them or were stopped after 3 minutes of speaking. One task was chosen to provide a neutral baseline of the vocal production (reading of neutral text), while the other was customized to emphasize some particular phenomena related to personality traits. In fact, TAT is a traditional projective test used to assess personality disorders. Furthermore, since anxiety can play a role as a confounding factor in speech-related features dynamics [19,20] we asked subjects to fill out the short form of the State-Trait Anxiety Inventory (STAI) for state anxiety, i.e. the STAI-X2 test [20]. This form has shown comparable psychometric properties to the original one and therefore is preferred in case of multiple administrations [21]. Subjects were asked to compile the scale at the beginning and at the end of each single task. Audio signals were acquired by means of a high quality system (AKG P220 Condenser Microphone, M-Audio Fast-Track), with a sampling frequency equal to 48 KHz and a resolution of 32 bits.

Speech Feature Extraction. The estimated speech features took into account the overall F_0 dynamics and the voice quality of the speakers. More in detail, skewness-like measurement of F_0 (*Median/Mean*), a frame-to-frame Jitter Factor (*LPJit*) estimate, and the Glottal Flow Spectral Slope (*Slope*) were investigated. The *Median/Mean* provides a global information about the tone of the speaker, while the *LPJit* and the *Slope* carry information about the quality of the speakers' voice. Specifically, *LPJit* describes the short-term variability of the voice, while *Slope* can be used to describe different phonation types (e.g creaky, tense or breathy). As a first step, voiced sounds are extracted from speech signals by means of a Voice Activity Detection algorithm that exploits signal energy and Zero Crossing Rate as described in [22]. Then, the proposed features are estimated within each segment from the F_0 contour, obtained according to the double iteration method as described in [18], based on Camacho's SWIPE' [23] algorithm. This latter algorithm estimates speech fundamental frequency using a spectral matching approach. The *Median/Mean* is computed as the ratio of median over mean of F_0. This ratio also acts as a normalizing procedure across subjects to face individual differences in tone. The *LPJit* is estimated in each segment using 4 glottal cycles-long time windows according to the following formula

$$LPJit = \frac{1}{N-1}\sum_{i=1}^{N-1}|F_{i+1} - F_i| \Big/ \frac{1}{N}\sum_{i=1}^{N} F_i \qquad (1)$$

where F_i is the fundamental frequency at the i-th window. *LPJit* represents a low-pass version of the classical jitter measure. *Slope* is obtained according to the procedure described in [24]. According to this approach, the glottal flow spectrum is estimated after the removal of the vocal tract effects. This result is obtained by averaging all the energy-normalized frames, obtained from voiced speech spectra using sliding windows. At the end, the glottal flow spectral slope is estimated by fitting a straight line over 300-3000Hz frequency band of the glottal flow spectrum. Both *LPJit* and *Slope* are already normalized measures and can be directly used in a correlation study at group level.

Statistical Analysis. A non-parametric Sign Test [25] is used to compare short-form STAI scores acquired before and after each task to evaluate possible effects on subject anxiety due to task execution. Moreover, a correlation analysis between monitored anxiety levels and speech features is also performed at group level by means of the non-parametric Spearman method. Similarly, the Spearman method is used to estimate the correlation coefficient between the personality trait dimensions and the corresponding speech-related features ($\alpha<0.05$). The Benjamini-Hochberg procedure is used to correct p-values for the false discovery rate.

III. RESULTS

No statistically significant differences were observed by investigating short-form STAI scores acquired before and after the reading tasks, thus confirming that these tasks did not induce anxiety level

changes. Interestingly, no statistically significant differences between short-form STAI scores related to the beginning and the end of the whole experiment were found, while no significant correlation coefficients, between STAI scores and speech-related features, were reported. In Table 1, the Spearman's correlation coefficients between ZKPQ scores and speech features are reported.

Table 1. Spearman's correlation coefficients between ZKPQ scores and speech features.

#task	Feat	Imp-SS	Agg-Host	Sy	N-Anx	Act
Rd 1	Median/Mean	0.01	0.29	**-0.66**	0.16	0.05
Rd 1	LPJit	-0.51	-0.08	-0.07	0.09	-0.27
Rd 1	Slope	0.09	0.55	0.10	-0.39	0.11
TAT	Median/Mean	-0.22	0.21	-0.41	-0.07	0.43
TAT	LPJit	**-0.60**	-0.01	-0.10	0.19	0.02
TAT	Slope	0.00	**0.66**	-0.02	-0.29	0.08
Rd 2	Median/Mean	-0.02	0.02	**-0.82**	0.35	-0.06
Rd 2	LPJit	-0.24	-0.29	0.23	0.00	0.00
Rd 2	Slope	0.00	0.55	0.32	-0.48	0.10

The values that report a significant p-value according to the Benjamini-Hochberg procedure are highlighted in bold.

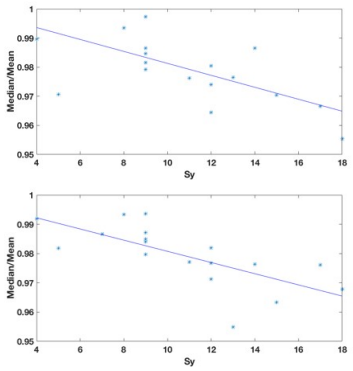

Figure 1. Reading task: scatter plot of *Median/Mean* vs *Sy*. Upper: Rd1 (ρ=-0.66). Lower: Rd2 (ρ=-0.82).

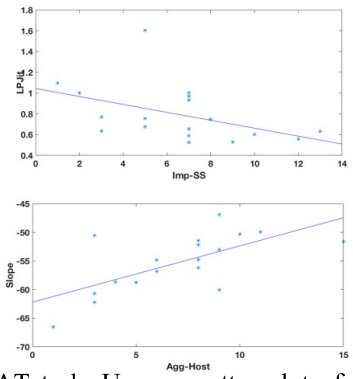

Figure 2. TAT task. Upper: scatter plot of *LPJit* and *Imp-SS*. (ρ=-0.60). Lower: Slope and Agg-Host (ρ=0.66).

Interestingly, *Median/Mean* reports a negative Spearman's correlation coefficient with the Sociability trait dimension in both reading tasks. Scatter plots of *Median/Mean* vs *Sy* values obtained in both reading tasks are shown in Fig. 1. In addition, the analysis of the commenting of TAT image task shows that *LPJit* correlates negatively with Impulsive Sensation Seeking trait dimension and *Slope* positively with Aggression-Hostility trait. In Fig. 2 while the scatter plots related to the TAT task are shown.

IV. Discussion and Conclusion

The results obtained on this study revealed some significant correlations. Interestingly, different results were obtained with the two tasks. Neutral text reading and TAT image commenting might in fact emphasize specific phenomena related to personality traits. As regards neutral text reading, a negative Spearman's correlation between *Median/Mean* feature and the Sociability trait dimension score is reported. This is verified in both repetitions of this task. Such a result could indicate that the more the speaker shows a sociable personality, the more the F_0 distribution shows a negative-skewed behaviour. A negative-skewed F_0 distribution is usually reported in relaxed and calm voices. The results on TAT images showed a negative correlation between *LPJit* and *Imp-SS*. This result might indicate a possible less hoarse voice in persons with a marked Impulsive Sensation Seeking trait. In this task, a positive correlation between *Slope* and *Agg-Host* was found. According to the fact that Slope is always negative and a steeper value is usually associated with a breathier voice, while a flat spectrum to a tenser or creakier voice [26], this result might indicate that a more aggressive trait is associated to a tenser or creakier voice. The coherent results, obtained between the two repetitions of the neutral reading task, seem to indicate a robust behavior of *Median/Mean*. Since personality traits have long-temporal dynamics, an analysis performed on longer time intervals might further elucidate the relevance and the robustness of this feature. Interestingly, no significant correlations were found between speech-related features and anxiety levels. We have to stress that anxiety levels were not significantly different before and after the recordings. Future studies could explore the use of stressful tasks to further investigate possible interactions of anxiety, personality traits and speech features.

The results of this study could have an impact on the comprehension of mental disorders. In fact, according to Zuckerman [16] severe personality disorders such as psychopathy, antisocial behaviour and forms of paranoid hostility would be a

combination of research of impulsive sensation seeking and low sociability.

REFERENCES

[1] P. Gangamohan, S. R. Kadiri, and B. Yegnanarayana, "Analysis of Emotional Speech---A Review," in Toward Robotic Socially Believable Behaving Systems-Volume I, Springer, 2016, pp. 205–238.

[2] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," Speech Commun., vol. 71, pp. 10–49, 2015.

[3] C. L. Giddens, K. W. Barron, J. Byrd-Craven, K. F. Clark, and A. S. Winter, "Vocal indices of stress: a review," J. Voice, vol. 27, no. 3, pp. 390--e21, 2013.

[4] A. S. Silnitskaya and A. N. Gusev, "character and temperamental determinants of prosodic parameters in natural speech," Psychol. Russ. State art, vol. 6, no. 3, 2013.

[5] D. Bernstein, Essentials of psychology. Cengage Learning, 2013.

[6] E. Sapir, "Speech as a personality trait," Am. J. Sociol., pp. 892–905, 1927.

[7] G. B. Ray, "Vocally cued personality prototypes: An implicit personality theory approach," Commun. Monogr., vol. 53, no. 3, pp. 266–276, 1986.

[8] K. R. Scherer and U. Scherer, "Speech behavior and personality," Speech Eval. psychiatry, pp. 115–135, 1981.

[9] F. Weninger, J. Krajewski, A. Batliner, and B. Schuller, "The voice of leadership: Models and performances of automatic analysis in online speeches," Affect. Comput. IEEE Trans., vol. 3, no. 4, pp. 496–508, 2012.

[10] G. Mohammadi, A. Origlia, M. Filippone, and A. Vinciarelli, "From speech to personality: Mapping voice quality and intonation into personality differences," in Proceedings of the 20th ACM international conference on Multimedia, 2012, pp. 789–792.

[11] C. Hu, Q. Wang, L. A. Short, and G. Fu, "Speech spectrum's correlation with speakers' Eysenck Personality Traits," PLoS One, vol. 7, no. 3, p. e33906, 2012.

[12] B. Gawda, "Neuroticism, extraversion, and paralinguistic expression," Psychol. Rep., vol. 100, no. 3, pp. 721–726, 2007.

[13] P. H. Lodhi, S. Deo, and V. M. Belhekar, "The Five-Factor Model of Personality," in The five-factor model of personality across cultures, Springer, 2002, pp. 227–248.

[14] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," J. Res. Pers., vol. 41, no. 1, pp. 203–212, 2007.

[15] F. Alam and G. Riccardi, "Fusion of acoustic, linguistic and psycholinguistic features for speaker personality traits recognition," in 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 955–959.

[16] M. Zuckerman, "Zuckerman-Kuhlman Personality Questionnaire (ZKPQ): an alternative five-factorial model," Big five Assess., pp. 377–396, 2002.

[17] H. A. Murray, "Uses of the thematic apperception test," Am. J. Psychiatry, vol. 107, no. 8, pp. 577–581, 1951.

[18] N. Vanello, A. Guidi, C. Gentili, S. Werner, G. Bertschy, G. Valenza, A. Lanata, and E. P. Scilingo, "Speech analysis for mood state characterization in bipolar patients," in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2012.

[19] E. Moore, M. A. Clements, J. W. Peifer, L. Weisser, and others, "Critical analysis of the impact of glottal features in the classification of clinical depression in speech," Biomed. Eng. IEEE Trans., vol. 55, no. 1, pp. 96–107, 2008.

[20] C. D. Spielberger, "Manual for the State-Trait Anxiety Inventory STAI (form Y)(' self-evaluation questionnaire')," 1983.

[21] T. M. Marteau and H. Bekker, "The development of a six-item short-form of the state scale of the Spielberger State---Trait Anxiety Inventory (STAI)," Br. J. Clin. Psychol., vol. 31, no. 3, pp. 301–306, 1992.

[22] B. Atal and L. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition," Acoust. Speech Signal Process. IEEE Trans., vol. 24, no. 3, pp. 201–212, 1976.

[23] A. Camacho and J. G. Harris, "A sawtooth waveform inspired pitch estimator for speech and music," J. Acoust. Soc. Am., vol. 124, no. 3, pp. 1638–1652, 2008.

[24] A. Ozdas, R. G. Shiavi, S. E. Silverman, M. K. Silverman, and D. M. Wilkes, "Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk," Biomed. Eng. IEEE Trans., vol. 51, no. 9, pp. 1530–1540, 2004.

[25] J. D. Gibbons and S. Chakraborti, Nonparametric statistical inference. Springer, 2011.

[26] J. Kuang and M. Libermann, "The effect of spectral slope on pitch perception," in Sixteenth Annual Conference of the International Speech Communication Association, 2015.