Analysis of speech features and personality traits

A. Guidi [a,b], C. Gentili [c], E.P. Scilingo [a,b], N. Vanello [a,b,*]

[a] University of Pisa, Dipartimento di Ingegneria dell'Informazione, Via G. Caruso 16, 56122, Pisa, Italy
[b] Research Center "E.Piaggio", University of Pisa, Largo Lucio Lazzarino 1, 56122, Pisa, Italy
[c] Department of General Psychology, University of Padua, via 8 febbraio 2, 35122, Padova, Italy

ARTICLE INFO

ABSTRACT

Voice signal has been widely investigated to characterize mood and emotional states. A further interesting dimension could regard the personality traits. The relationship between personality traits and specific speech features is known, however this topic requires further investigation. Specifically, most studies are focused on perceived personality traits, without adopting dedicated personality tests. Moreover, the relationship among speaker 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 described by the Zuckerman-Kuhlman model and the Toronto Alexithymia Scale, is performed. An experimental protocol, consisting of two structured speech tasks, was administered to eighteen healthy subjects. Speech features were estimated to describe fundamental frequency (F_0) and voice quality related features from whole speech recording and tilt-related features, describing F_0 dynamics at voiced segment level. Significant correlations among personality traits and speech features were observed using both feature sets. Interestingly, the adopted speech task was found to influence the obtained results. Specifically, no feature reports the same significant correlation in both adopted tasks. 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.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Voice signal is modulated by the autonomic and somatic nervous systems. Almost the whole central nervous system (CNS) is involved in voice production. Voice patterns are strictly dependent on the interactive process of phonation, articulation and audition since auditory feedback play a primary role in acquiring and emitting one's own voice [1]. Both the autonomic and somatic nervous systems influence the speaker's prosody by modulating the action of the respiratory system. Moreover, the vagus nerve, that supplies motor parasympathetic fibers and it is responsible for the control of heart rate and sweating, controls the activation of some mouth and laryngeal muscles [1]. Noticeably, different psychological and/or mental conditions modulate speech production. Hence, speech analysis represents an interesting, non-invasive and economic approach to study speaker's mental state [2].The analysis of speech signal allows to explore several psychological dimensions: emotion [3–6], mood [7], and stress [8–11] 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 [12]. Probably, such a relation might be stronger in people showing some particular personality trait [12]. According to the trait theory [13], 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 [14] proposed the hypothesis of "speech as a personality trait". Addington [15] 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. The relationship between attractive voices and acoustic characteristics, especially voice quality, were investigated in the study of Zuckerman and Miyake [16]. 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 [17–19]. These studies

* Corresponding author at: University of Pisa, Dipartimento di Ingegneria dell'Informazione, Via G. Caruso 16, 56122, Pisa, Italy.
E-mail addresses: andrea.guidi@for.unipi.it (A. Guidi), c.gentili@unipd.it (C. Gentili), e.scilingo@centropiaggio.unipi.it (E.P. Scilingo), nicola.vanello@iet.unipi.it (N. Vanello).

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 [17,18]. Political charisma and leadership were reported to be positively correlated with higher pitch in [19]. Spectra and voice quality features were also investigated in relation to personality perception in [20,21]. A correlation between speech fluency and both extroversion and neuroticism was observed in [22]. Silnitskaya et al. reported that demonstrativity trait was affected by mean speech fundamental frequency variability, tone span and speech rate in the presence of an interlocutor [22]. The INTERSPEECH 2012 Speaker Trait Challenge [23] showed that the classification of personality traits, as defined by the OCEAN five personality dimensions [24], 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 [25]. Acoustic features outperformed linguistic and psycholinguistic features to achieve an automatic recognition of speaker personality trait [26]. One issue in the analysis of the voice and personality traits interaction is related to the fact that many personality models define personality traits in different ways. The personological traits that are more commonly described are neuroticism and sociability/extraversion [27,28]. These dimensions are consistent across different models and scales. Moreover, the long temporal dynamics of the personality traits, and therefore the difficulty in observing trait dynamics, may complicate such an analysis [29].

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 among speaker 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 described by the Zuckerman-Kuhlman model [30] and the Toronto Alexithymia Scale [31], is performed. Specifically, a correlation analysis between features related to speech fundamental frequency (F0) 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. Anxiety levels will be also monitored to evaluate their effects on the observed features.

2. Material and methods

2.1. Experimental protocol

Eighteen healthy subjects (23.66 ± 2.28 year, 12 females) without any history of psychiatric disorder were enrolled. Subjects were asked to fill out the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ) and the Toronto Alexithymia Scale (TAS-20) at home [32], 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). The TAS-20 is an instrument characterized by high objectivity, good internal consistency, reliability and validity [33]. Such a scale is a self-report questionnaire based on 20 items and provides information about alexithymia trait in terms of three factors: difficulty identi-

fying feelings, difficulty describing feelings, and external-oriented thinking. In this study, the TAS-20 sum score was investigated. Alexithymia is a personality trait that is related to the incapability of recognizing and describing the emotions in the self [34]. More specifically this trait is characterized by the difficulty in identifying feelings and distinguishing between feelings and the bodily sensations of emotional arousal and by the difficulty in describing feelings to other people [32].

Subjects were asked to read a neutral text ("The universal declaration of Human rights", lasting 3 min) twice, at the beginning and at the end of the experimental protocol, after about 30 min. 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 [35], between the two neutral text reading tasks. In this study five TAT images constituted the image set. Subjects were driven to comment all of them or were stopped after 3 min 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 [36,37] 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 [38]. While the N-Anx dimension of the ZKPQ comprehends, together with emotional instability and depression also proneness to anxiety (also known as state anxiety) [39,40] the STAI-X2 measures the actual anxiety symptoms during the test. This short form of STAI has shown comparable psychometric properties to the original one and therefore is preferred in case of multiple administrations [41]. 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. The experiment timeline is shown in Fig. 1.

2.2. Speech feature extraction

In this study, two different sets of speech-related features were studied to focus on two different description levels of the investigated phenomena. The first set aimed at providing a general description of the recorded speeches and took into account the overall F_0 dynamics and the voice quality of the speakers. Differently, the second set of features investigated F_0 dynamics within every voiced segments.

More in detail, in the first set a skewness-like measurement of F_0 (Median/Mean), a frame-to-frame Jitter Factor (LPJit) estimate, and the Glottal Flow Spectral Slope (Slope) [42] 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 [43]. Then, the proposed features are estimated within each segment from the F_0 contour, obtained according to the double iteration method as described in [36], based on Camacho's SWIPE' [44] 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. This feature gives information about the median use of F_0 with respect to the whole observed F_0 interval, indicating when the central
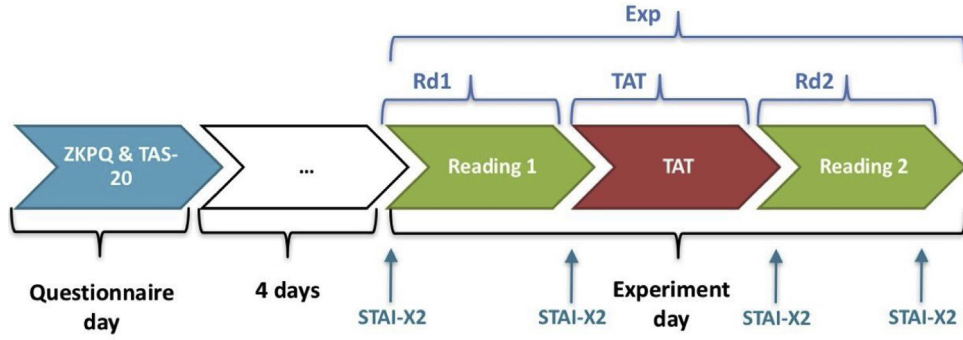
Fig. 1. Experiment timeline. The ZKPQ and TAS-20 questionnaires were administered 4 days before the speech task. The three speech tasks were recorded in one session. The STAI-X2 test was administered before and after each speech task.

moment of the F0 distribution is at the higher end or at the lower end of the distribution range. The LPJit is estimated in each segment using 4 glottal cycle-long time windows according to the following formula

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

where F_i is the fundamental frequency at the i-th window [36]. LPJit represents a low-pass version of the classical jitter measure. Slope is obtained according to the procedure described in [42]. According to this approach, the glottal flow spectrum is estimated after the reduction 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 300–3000 Hz 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.

The second set of speech-related features used the VAD outcomes and the F_0 estimates, previously described, to estimate a set of geometrical features, i.e. tilt features, related to the F_0 dynamics within every voiced segment, namely Amplitude* (Amp), Duration* (Dur) and Tilt* (Tilt). These features, describing the prosodic behaviour in the vocalized sounds, aimed at characterizing rising and falling F_0 events in terms of amplitudes and durations. Originally proposed by Taylor [45] to specifically characterize intonational events, they are here applied to each voiced segment. These features were already used to characterize the speech in patients affected by bipolar disorder in [46]. They are related to the amplitude and duration relative sizes of rises and falls for an event and can be defined as in Eqs. (2)–(4):

$$Amplitude^* = (|A_{rise}| - |A_{fall}|) / (|A_{rise}| + |A_{fall}|) \qquad (2)$$

$$Duration^* = (D_{rise} - D_{fall}) / (D_{rise} + D_{fall}) \qquad (3)$$

$$Tilt^* = (Amplitude^* + Duration^*) / 2 \qquad (4)$$

where A_{rise} and A_{fall} are the F0 changes during the rising and falling intervals, D_{rise} and D_{fall} are the duration of the rising and falling intervals (Fig. 2). The value of these features range from -1 to 1 in each voiced segment. The median value across all voiced segments was estimated for each subject and each speech task.

2.3. Statistical analysis

A non-parametric Sign Test [47] is used to compare short-form STAI scores acquired before and after each task (i.e. first and second reading tasks, Rd1 and Rd2 respectively, and image commenting, TAT). Moreover, the same test was applied to the scores obtained
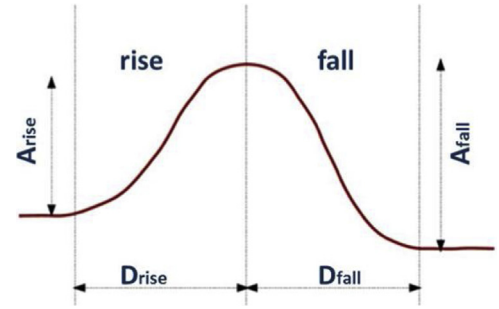

Fig. 2. Parameters of the Tilt Model.

before and after the overall experimental session, that will be indicated by Exp. The test on STAI scores aims at evaluating 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 nonparametric 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 number of commented TAT images was also related to the clinician scores by means of the Spearman's correlation.

The Benjamini-Hochberg procedure is used to control the false discovery rate for multiple comparisons [48]. Such a correction is performed distinctly within the two personality scales and the two features sets under the hypothesis that in different scales and at different description levels might act different processes. The adjusted p-values according to this procedure are reported.

3. 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. Specifically, the nonparametric sign test performed to explore possible differences between the STAI scores before and after each task (e.g. between the scores obtained after and before the Rd1, TAT and Rd2 tasks) as well as between the STAI scores before and after the overall experiment, gave nonsignificant results (see Table 1). No significant correlation coefficients, between STAI scores and speech-related features, were reported (data not shown).

In Table 2, the Spearman's correlation coefficients between ZKPQ scores, TAS-20 score and the first set of speech features are reported. The values that report a significant p-value according to the Benjamini-Hochberg procedure are highlighted in bold. Inter-
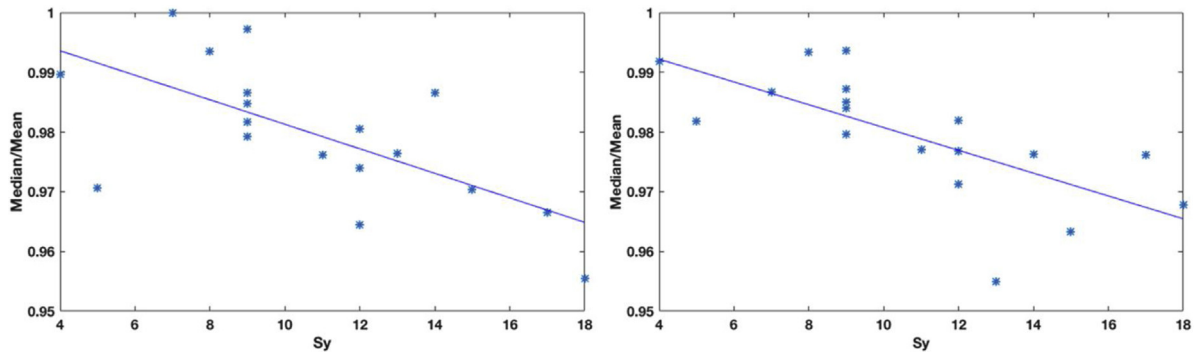
Fig. 3. Reading task: scatter plot of *Median/Mean* vs *Sy*. Left: Rd1 (ρ=-0.66). Right: Rd2 (ρ=-0.82). ρ refers to the Spearman's correlation coefficient between the corresponding speech feature and personality score across 18 subjects.

Fig. 4. TAT task. Left: scatter plot of *LPJit* and *Imp-SS*. (ρ=-0.60). Right: Slope and Agg-Host (ρ=0.66). ρ refers to the Spearman's correlation coefficient between the corresponding speech feature and personality score across 18 subjects.

Table 1
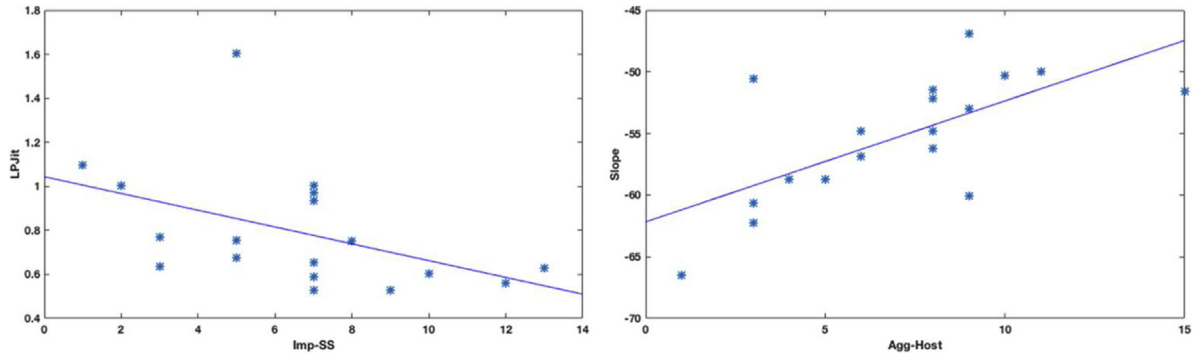Statistical analysis of short-form STAI scores. The p-values about the statistical test on the STAI values before and after each task (Rd1, TAT and Rd2) and the overall experimental session (Exp.) are shown.

Rd1	TAT	Rd2	Exp
0.5169	0.8036	0.1530	0.4828

estingly, 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. 3. 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. 4, the scatter plots related to the TAT task are shown.

In Table 3 the Spearman correlation coefficients related to the second set of features, i.e. the tilt features investigating F_0 dynamics within every voiced segment, are reported. Interestingly, this

Table 3
Within-voiced segments investigation - Spearman correlation coefficients between ZKPQ, TAS-20 scores and speech features belonging to the second feature set. Significant results are marked in bold (α<0.05, Benjamini-Hochberg corrected).

#task	Feat	Imp-SS	Agg-Host	Sy	N-Anx	Act	TAS-20
Rd 1	Amp	0.01	0.04	−0.07	0.20	0.45	0.25
	Dur	0.02	0.01	−0.04	0.18	0.51	0.26
	Tilt	0.03	0.02	−0.06	0.22	0.49	0.25
TAT	Amp	0.07	0.07	−0.11	0.15	**0.67**	**0.56**
	Dur	0.00	0.13	−0.13	0.16	**0.56**	**0.48**
	Tilt	0.06	0.10	−0.12	0.13	**0.63**	**0.54**
Rd 2	Amp	−0.17	0.07	0.25	0.07	0.33	0.08
	Dur	−0.04	−0.05	0.15	0.04	0.53	0.26
	Tilt	−0.16	0.10	0.20	0.15	0.33	0.11

kind of features reported significant correlation only by taking into account commenting of TAT image recordings. Specifically, *Amp* (Fig. 5), *Dur* (Fig. 6), *Tilt* (Fig. 7) reported significant correlation with Activity (Act) trait and with the TAS-20 score.

Table 2
Overall investigation - Spearman correlation coefficients between ZKPQ, TAS-20 scores and speech features belonging to the first feature set. Significant results are marked in bold (α<0.05, Benjamini-Hochberg corrected).

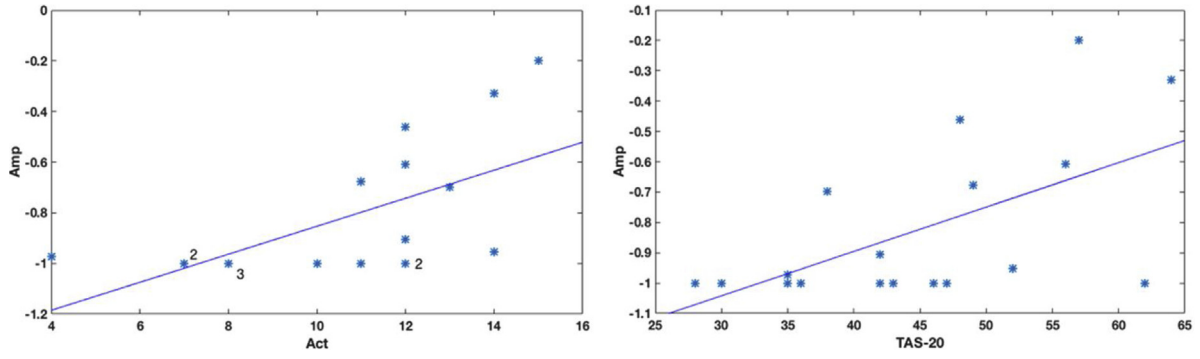| #task	Feat	Imp-SS	Agg-Host	Sy	N-Anx	Act	TAS-20
Rd 1	Median/Mean	0.01	0.29	**−0.66**	0.16	0.05	0.00
	LPJit	−0.51	−0.08	−0.07	0.09	−0.27	−0.51
	Slope	0.09	0.55	0.10	−0.39	0.11	−0.17
TAT	Median/Mean	−0.22	0.21	−0.41	−0.07	0.43	0.22
	LPJit	**−0.60**	−0.01	−0.10	0.19	0.02	−0.05
	Slope	0.00	**0.66**	−0.02	−0.29	0.08	−0.012
Rd 2	Median/Mean	−0.02	0.02	**−0.82**	0.35	−0.06	0.02
	LPJit	−0.24	−0.29	0.23	0.00	0.00	−0.18
	Slope	0.00	0.55	0.32	−0.48	0.10	−0.17

Fig. 5. TAT task. Left: scatter plot of *Amp* vs *Act* (ρ = 0.67). Right: scatter plot of *Amp* vs *TAS-20* (ρ = 0.56). ρ refers to the Spearman's correlation coefficient between the corresponding speech feature and personality score across 18 subjects. The multiplicity of coinciding measurements is shown.
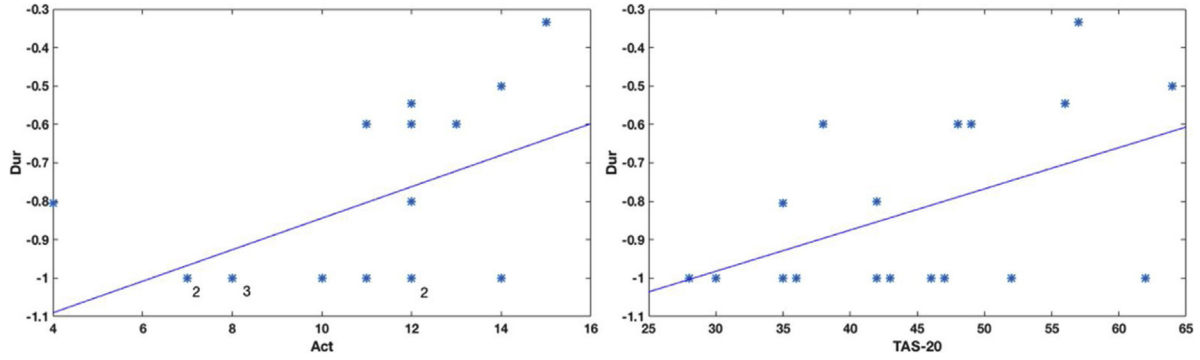

Fig. 6. TAT task. Left: scatter plot of *Dur* vs *Act* (ρ = 0.56). Right: scatter plot of *Dur* vs *TAS-20* (ρ = 0.48). ρ refers to the Spearman's correlation coefficient between the corresponding speech feature and personality score across 18 subjects. The multiplicity of coinciding measurements is shown.

Fig. 7. TAT task. Upper: scatter plot of *Tilt* vs *Act* (ρ = 0.63). Lower: scatter plot of *Tilt* vs *TAS-20* (ρ = 0.54). ρ refers to the Spearman's correlation coefficient between the corresponding speech feature and personality score across 18 subjects. The multiplicity of coinciding measurements is shown.
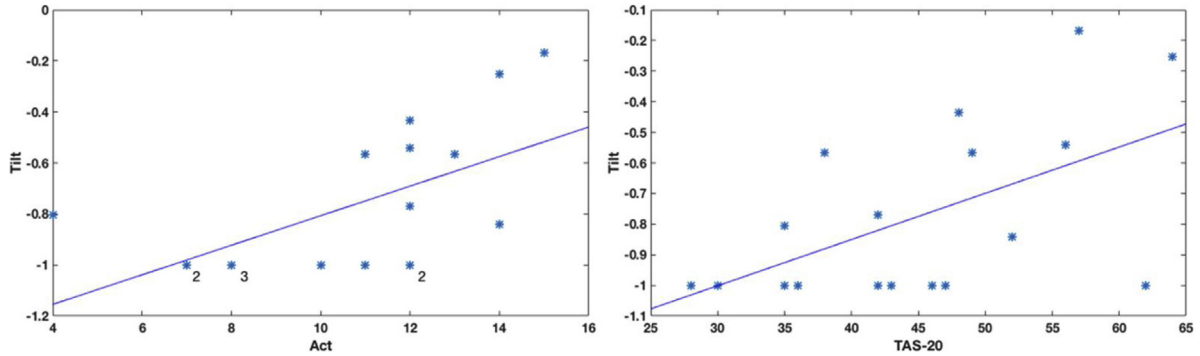
Finally, the correlation analysis between number of commented TAT images and clinician scores revealed a negative correlation (ρ = -0.63) only with TAS-20 score. In fact, no significant correlation was observed between number of commented images and the five fundamental personality traits defined by the ZKPQ model.

4. Discussion and conclusions

The results obtained in this study revealed some significant correlations between personality traits and speech features in both feature sets. 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. In fact, according to Scherer [49], relaxed voice is characterized by a F_0 close to the lower end of the range. Sociability/extraversion regards not only to feel at ease in social context, but also to like being in social context and interacting with others. Speech is one of the most important channel in the social interaction process, therefore is not surprising that a relation can be yield between voice features and this ability.

The results on TAT images showed a negative correlation between LPJit and Imp-SS. On one hand high levels of jitter were often correlated with hoarse voice [50], on the other hand low jitter was observed in stressed people [9].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 also 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 [51], this result might indicate that a more aggressive trait is associated to a tenser or creakier voice.

Concerning the second feature set, i.e. tilt features, significant results were observed only when considering commenting of TAT image task. In fact, *Amp*, *Dur* and *Tilt* reported significant correlation with Activity trait (Act) and the TAS-20 score. This features aim at describing the F_0 contour within every voiced segment and therefore the prosodic dynamics.

Hence, the results let us suppose that information about speakers' prosodic production might be related to both alexithymia scores and Activity trait. Specifically, alexithymia seems to be associated with a generally weakened neural response to speech prosody. This reduced prosodic processing might be partially responsible for an impaired social communication in people with this personality trait [52]. We might speculate that the impaired prosody processing might also affect the prosodic speech production.

We did not find any significant correlation between speech features and Neuroticism Anxiety (N-Anx). Neuroticism-Anxiety defines a personality trait, which includes different aspect of emotion regulation and disregulation. Although in the ZKPQ the name of this personality trait comprehends the term anxiety, it corresponds to the Neuroticism personality dimension, which in turn is one of the fundamental traits in personality theories. Its core characteristics include emotional instability, along with depression and anxiety symptoms [39][40]. Therefore, part of this personality trait includes proneness to anxiety. Anxiety in a specific situation or event that is perceived as unpleasant or threatening, can be defined as state anxiety. Of note we did not find any significant correlation between STAI scores and voice features, as well.

A further consideration to be drawn regards the fact that STAI scores did not significantly change across the tasks. The use of this scale was meant to measure whether changes in speech features may be related to change in anxiety levels. If one hand, the lack of correlation suggests that, in our experiment, state anxiety did not interfere with speech features, on the other the lack of changes suggest that our task cannot be considered a psychological stress task. Future studies could explore the use of stressful tasks to investigate possible interactions of anxiety, personality traits and speech features.

A significant negative correlation was observed between the number of commented TAT images and the TAS-20 score. This might mean that a higher TAS-20 score could be observed in people who spent more time in attempting to provide a comment to the presented images, and therefore in people who commented a lower number of images.

Moreover, if the first set of speech-related features reports significant correlations in both tasks, tilt features seem to be more significant in commenting of TAT task. Again, no feature reports the same significant correlation in both kinds of task. These phenomena are related to an already-observed task-dependent behaviour of speech-related features. In fact, in voice analysis, the choice of the vocal task can be crucial. Generally, tasks such as reading, free speech, interviews, commenting of images, dialogues, and sustained vowels are investigated. Anyway, it is important to highlight that a task-dependent behaviour of F_0 was observed in some speech-related features [46,53]. Moreover, Laver et al. [54] stated that the kind of vocal task, i.e. sustained vowels or continuous speech, can be considered as a further confounding factor when investigating some features as jitter, shimmer and HNR. In fact, the extraction of these cues is easier in sustained vowels thanks to the intrinsic stationarity notwithstanding the differences in sound pressure levels both intra- and inter-subjects [55].

The coherent results, obtained between the two repetitions of the neutral reading task, seem to indicate a robust behaviour 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. Overall, these results indicate a possible interaction between personality and speech task. Indeed, TAT task is conceived to induce an emotional stimulation to subjects. Given this we could hypothesize that people with different personality profiles could behave differently and that this could be highlighted analysing speech. A more complete analysis of this interaction among mood, personality and speech task would require more samples, and this could not be accomplished at this time.

The results of this study could have an impact on the comprehension of mental disorders. In fact, according to Zuckerman [30] 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. Interestingly, a strong connection between personality and mental disorders was reported by Kotov et al. [56,57]. More in detail, neuroticism was seen to be strongly correlated with anxiety, depression, and substance use disorders. Moreover, since high alexithymia scores were observed in people with eating disorders [58], this study might be relevant for clinical practice.

Finally, the correlation of personality traits and speech production could be very useful in the attempt to characterize both mood and emotional states. Specifically, the speaker personality profile should be taken into account for building a model relating speech production and subject emotional and mood state. We believe that, given the large number of psychological dimensions when mood and/or emotion state are under investigation, further studies may succeed in characterizing other psychological traits and try to build a psychological-voice map.

References

[1] J. Kreiman, D. Sidtis, Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception, John Wiley & Sons, 2011.

[2] C.S. Hopkins, R.J. Ratley, D.S. Benincasa, J.J. Grieco, Evaluation of voice stress analysis technology, in: System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on, 2005, 20b-20b.

[3] D. Ververidis, C. Kotropoulos, Emotional speech recognition: resources, features, and methods, Speech Commun. 48 (9) (2006) 1162–1181.

[4] S.G. Koolagudi, K.S. Rao, Emotion recognition from speech: a review, Int. J. Speech Technol. 15 (2) (2012) 99–117.

[5] K.S. Rao, S.G. Koolagudi, Emotion Recognition Using Speech Features, Springer Science & Business Media, 2012.

[6] P. Gangamohan, S.R. Kadiri, B. Yegnanarayana, Analysis of emotional speech—a review, in: Toward Robotic Socially Believable Behaving Systems-Volume I, Springer, 2016, pp. 205–238.

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

[8] R. Fernandez, R.W. Picard, Modeling drivers' speech under stress, Speech Commun. 40 (1) (2003) 145–159.

[9] N. Sharma, T. Gedeon, Objective measures, sensors and computational techniques for stress recognition and classification: a survey, Comput. Methods Programs Biomed. 108 (3) (2012) 1287–1301.

[10] J. Luig, A. Sontacchi, A speech database for stress monitoring in the cockpit, Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng. (2013), p.954410012467944.

[11] C.L. Giddens, K.W. Barron, J. Byrd-Craven, K.F. Clark, A.S. Winter, Vocal indices of stress: a review, J. Voice 27 (3) (2013) 390–e21.

[12] A.S. Silnitskaya, A.N. Gusev, Character and temperamental determinants of prosodic parameters in natural speech, Psychol. Russ. State Art 6 (3) (2013).

[13] D. Bernstein, Essentials of Psychology, Cengage Learning, 2013.

[14] E. Sapir, Speech as a personality trait, Am. J. Sociol. (1927) 892–905.

[15] D.W. Addington, The relationship of selected vocal characteristics to personality perception, J. Speech Hear. Disord. Monogr. Suppl. 35 (4) (1968).

[16] M. Zuckerman, K. Miyake, The attractive voice: What makes it so? J. Nonverb. Behav. 17 (2) (1993) 119–135.

[17] K.R. Scherer, Effect of stress on fundamental frequency of the voice, J. Acoust. Soc. Am. 62 (S1) (1977) S25–S26.

[18] K.R. Scherer, U. Scherer, Speech behavior and personality, Speech Eval. Psychiatry (1981) 115–135.

[19] F. Weninger, J. Krajewski, A. Batliner, B. Schuller, The voice of leadership: models and performances of automatic analysis in online speeches, Affect. Comput. IEEE Trans. 3 (4) (2012) 496–508.

[20] G. Mohammadi, A. Vinciarelli, Automatic personality perception: prediction of trait attribution based on prosodic features, Affect. Comput. IEEE Trans. 3 (3) (2012) 273–284.

[21] G. Mohammadi, A. Origlia, M. Filippone, 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.

[22] B. Gawda, Neuroticism, extraversion, and paralinguistic expression, Psychol. Rep. 100 (3) (2007) 721–726.

[23] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. Van Son, F. Weninger, F. Eyben, T. Bocklet, et al., The INTERSPEECH 2012 speaker trait challenge INTERSPEECH, 2012, 2012, pp. 254–257.

[24] P.H. Lodhi, S. Deo, V.M. Belhekar, The Five-factor model of personality, in: The Five-factor Model of Personality Across Cultures, Springer, 2002, pp. 227–248.

[25] B. Rammstedt, 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. 41 (1) (2007) 203–212.

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

[27] M. Zuckerman, S.B. Eysenck, H.J. Eysenck, Sensation seeking in England and America: cross-cultural, age, and sex comparisons, J. Consult. Clin. Psychol. 46 (1) (1978) 139.

[28] P.T. Costa, R.R. McCrae, Normal personality assessment in clinical practice: the NEO personality inventory, Psychol. Assess. 4 (1) (1992) 5.

[29] B. Schuller, A. Batliner, Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing, John Wiley & Sons, 2013.

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

[31] R.M. Bagby, J.D.A. Parker, G.J. Taylor, The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure, J. Psychosom. Res. 38 (1) (1994) 23–32.

[32] G.J. Taylor, R.M. Bagby, O. Luminet, et al., Assessment of alexithymia: Self-report and observer-rated measures, Handb. Emot. Intell. (2000) 301–319.

[33] M. Franz, K. Popp, R. Schaefer, W. Sitte, C. Schneider, J. Hardt, O. Decker, E. Braehler, Alexithymia in the German general population, Soc. Psychiatry Psychiatr. Epidemiol. 43 (1) (2008) 54–62.

[34] P.E. Sifneos, The prevalence of 'alexithymic' characteristics in psychosomatic patients, Psychother. Psychosom. 22 (2–6) (1973) 255–262.

[35] H.A. Murray, Uses of the thematic apperception test, Am. J. Psychiatry 107 (8) (1951) 577–581.

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

[37] E. Moore, M.A. Clements, J.W. Peifer, L. Weisser, et al., Critical analysis of the impact of glottal features in the classification of clinical depression in speech, Biomed. Eng. IEEE Trans. 55 (1) (2008) 96–107.

[38] C.D. Spielberger, Manual for the State-trait Anxiety Inventory STAI (form Y) ('Self-evaluation Questionnaire'), 1983.

[39] M. Zuckerman, Psychobiology of Personality, 2nd ed., Cambridge University Press, 2005.

[40] M. Zuckerman, D.M. Kuhlman, J. Joireman, P. Teta, M. Kraft, A comparison of three structural models for personality: the Big three, the Big Five, and the Alternative five, J. Pers. Soc. Psychol. 65 (4) (1993) 757–768.

[41] T.M. Marteau, 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. 31 (3) (1992) 301–306.

[42] A. Ozdas, R.G. Shiavi, S.E. Silverman, M.K. Silverman, 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. 51 (9) (2004) 1530–1540.

[43] B. Atal, L. Rabiner, A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition, Acoust. Speech Signal Process. IEEE Trans. 24 (3) (1976) 201–212.

[44] A. Camacho, J.G. Harris, A sawtooth waveform inspired pitch estimator for speech and music, J. Acoust. Soc. Am. 124 (3) (2008) 1638–1652.

[45] P. Taylor, Analysis and synthesis of intonation using the tilt model, J. Acoust. Soc. Am. 107 (3) (2000) 1697–1714.

[46] A. Guidi, N. Vanello, G. Bertschy, C. Gentili, L. Landini, E.P. Scilingo, Automatic analysis of speech F0 contour for the characterization of mood changes in bipolar patients, Biomed. Signal Process. Control 17 (March) (2015) 29–37.

[47] J.D. Gibbons, S. Chakraborti, Nonparametric Statistical Inference, Springer, 2011.

[48] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B (1995) 289–300.

[49] K.R. Scherer, Vocal affect expression: a review and a model for future research, Psychol. Bull. 99 (2) (1986) 143.

[50] W.J. Hardcastle, J. Laver, F.E. Gibbon, The Handbook of Phonetic Sciences, John Wiley & Sons, 2010.

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

[52] K.S. Goerlich-Dobre, J. Witteman, N.O. Schiller, V.J.P. van Heuven, A. Aleman, S. Martens, Blunted feelings: alexithymia is associated with a diminished neural response to speech prosody, Soc. Cogn. Affect. Neurosci. 9 (8) (2014) 1108–1117.

[53] R. Horwitz, T.F. Quatieri, B.S. Helfer, B. Yu, J.R. Williamson, J. Mundt, On the relative importance of vocal source, system, and prosody in human depression, in: Body Sensor Networks (BSN), 2013 IEEE International Conference on, 2013, pp. 1–6.

[54] J. Laver, S. Hiller, J.M. Beck, Acoustic waveform perturbations and voice disorders, J. Voice 6 (2) (1992) 115–126.

[55] R.F. Orlikoff, J.C. Kahane, Influence of mean sound pressure level on jitter and shimmer measures, J. Voice 5 (2) (1991) 113–119.

[56] R. Kotov, W. Gamez, F. Schmidt, D. Watson, Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis, Psychol. Bull. s136 (5) (2010) 768.

[57] D.N. Klein, R. Kotov, S.J. Bufferd, Personality and depression: explanatory models and review of the evidence, Annu. Rev. Clin. Psychol. 7 (2011) 269.

[58] K. Kucharska-Pietura, V. Nikolaou, M. Masiak, J. Treasure, The recognition of emotion in the faces and voice of anorexia nervosa, Int. J. Eat. Disord. 35 (1) (2004) 42–47.