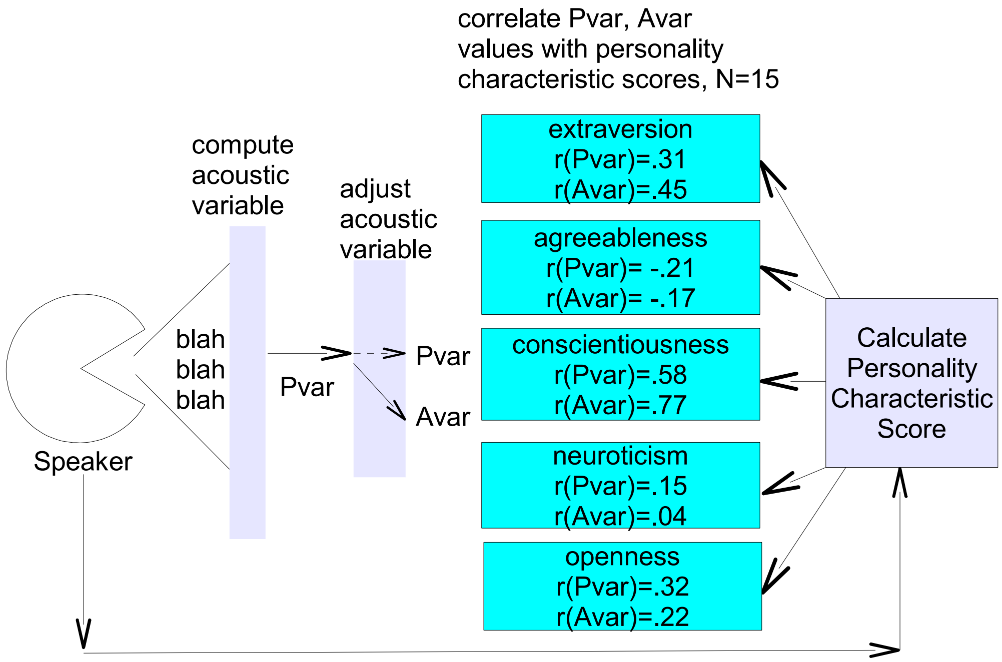
DRAFT: NOT FOR PUBLICATION OR DISTRIBUTION. AVAILABLE ONLY WITH SIGNED NON-DISCLOSURE AGREEMENT

Discriminating the "Conscientiousness" Personality Characteristic of the Big Five Inventory Using Speech Acoustics.

AUTHOR:
Dan Begel, M.D.
AP Systems
1642 19th Street
Santa Monica, California 90404
tel: 310.453.1531 FAX: 310.453.1561 email: Dan@danbegelmd.com

ABSTRACT

In this study, correlation coefficients were computed between the scores for personality characteristics of the Big Five Inventory (BFI) and acoustic variables derived from the natural speech of fifteen subjects. Acoustic variables were derived by computing the simultaneous transformations of pitch and intensity values at 10 millisecond intervals and separating prevailing transformations from infrequent ones. The difference between the intensity means of prevailing and infrequent transformations was the initial variable of interest. This variable was adjusted to account for a relative preference for pitch or intensity modulation in each speaker to derive a second variable of interest. The correlation of acoustic variables with BFI indices was high for the quality of conscientiousness, mild for the quality of extraversion, and minimal or non-existent for the qualities of agreeableness, neuroticism, and openness. Coefficients of determination were also computed.



Discriminating the "Conscientious" Personality Characteristic of the Big Five Inventory Using Speech Acoustics.

1. INTRODUCTION

Competent dialogue between humans and machines requires that machines are able to discern human characteristics within a speech signal. The characteristic of emotion is usually considered to be the most important. Various authors have pointed out that there are significant differences among individuals in the ways that they express similar emotions [1], and that the experience of emotion itself may vary with personality style [2]. As a result, the development of effective methods for the automated detection of emotion within a speech signal may depend upon methods that discern human personality characteristics as well.

The Big Five Inventory (BFI) is a well-studied questionnaire that discriminates between five personality characteristics-- extraversion, agreeableness, conscientiousness, neuroticism, and openness--as these are defined by the creators of the test [3]. To the extent that these five characteristics are distinct aspects of personality it is of interest to attempt to discriminate them using an acoustic analysis of natural speech. In the study reported here, correlation coefficients between each of the BFI indices and two acoustic variables derived from specimens of the natural speech of fifteen subjects are computed.

2. MATERIALS AND METHOD

Adult subjects, male and female, were recruited from the group of patients visiting the outpatient psychiatry practice of the investigator during the course of a given week. The subjects were relatively high functioning. All but three were employed or in school. None had been hospitalized within the previous six months. No attempt was made to distinguish subjects by diagnosis, severity, medication regimen, or other medical illness. Patients for whom an invitation to participate in research might have been disruptive to their therapy, such as those in immediate crisis, were not invited to participate. Every other patient agreed to participate in the research, perhaps because of their relationship with the investigator and the relative simplicity of the protocol. Fifteen subjects were enrolled in the study. None were removed in the course of it.

Upon giving informed consent, each subject filled out the 44-item version of the Big Five Inventory and provided a sample of speech of roughly two minutes duration. Speech was recorded in a clinical office using a head-mounted hypercardioid microphone placed approximately two centimeters from the mouth.The sampling rate was 44.1KHz. The subject was asked to speak for approximately two minutes without response or interruption by the investigator on the following topic: "If you were to design your own elevator, what would it be like?"

3. THEORY AND COMPUTATION

The BFI's were scored according to the protocol of the test publisher, with each subject receiving a score for each of the five personality characteristics measured by the test-- extraversion, agreeableness, conscientiousness, neuroticism, and openness.

The acoustic analysis and the computation of two acoustic variables of interest were carried out in Praat [4], the widely used speech analysis tool, and in Matlab [5], the multi-purpose scientific computing environment.

3.1 Theory

The method of analysis used here has not, to the writer's knowledge, been reported elsewhere, and so it may be useful to briefly explain the underlying assumptions and logic of the computational steps described in section 3.2. The acoustic variables of interest for which correlation coefficients with BFI indices were computed were drawn from an inventory of variables calculated according to a specific method of acoustic psychometry®, the measurement of mental activity using voice acoustics. By describing the logic of this method, the computational steps described in 3.2 that are based on it may become less arbitrary for the reader. The computations themselves stand on their own, however, regardless of the strength or weakness of the theoretical support for their execution, and the correlations of BFI indices with acoustic variables are as reported.

The acoustic analysis is guided by five ideas of the behavioral sciences. The first idea, articulated by many authors, among them Edelson [6], is that language is a system of representation, including the representation of inner states. As a result, the patterns or structures generated in the act of speaking are likely to provide fertile clues to the hallmarks of personality, including those personality characteristics defined by the BFI. The second idea is Piaget's notion that the essence of a structure is found not in its products but in the actions that structure performs simultaneously on multiple elements [7]. In the computations performed here, pitch and intensity values were extracted every 10 ms, and the variables of interest are based on the simultaneous changes of pitch and intensity in that interval. The third idea is that of Chomsky, who showed that intelligible speech is dependent on the syntactic structures of its generation [8]. In the method described here phonated speech is processed independently of semantic content.

The fourth idea, also one articulated by many authors, among them Mahl [9], is the clinical idea that representation has a dynamic aspect consisting of an interplay of conflict and defense. The therapist in clinical practice attends to the transitory shifts of tone and intensity, expressive of impulses, fears, and lapses of attention, as well as to the habitual measures of defense taken by the patient to preserve a sense of intactness. On this principle, the computations performed in section 3.2 distinguish between those prevailing acoustic transformations that occur most often and the subordinate, less frequently occurring transformations to yield two distinct acoustic structures. These two acoustic structures are assumed to bear a relationship to one another.

In a hypothesis-generating exploration of the inventory of statistics describing the prevailing and infrequent acoustic transformations, it appeared that "conscientious" persons may be best able to maintain a consistent intensity during infrequent activity. As an initial acoustic variable of interest, then, the mean intensity transformation of the prevailing acoustic structure was subtracted from the mean intensity transformation of the acoustic structure representing infrequent activity.

The fifth idea, articulated by Shapiro [ibid.], among others, is that an individual's cognitive style pervades every representational activity of that person, from expressing oneself in speech to comprehending the world. Thus, if there exists an acoustic variable that correlates with a specific personality characteristic, persons who possess that characteristic equally may display that variable to unequal degrees if they are of divergent cognitive styles. The acoustic variabile of interest, in other words, may become more meaningful if a correction for cognitive style is made. In this study, and index of cognitive style consisting of the relative degree of pitch versus intensity modulation was selected on intuitive grounds.

3.2 Computation

In Praat, one specimen of 25 seconds was excised from the audio file of each subject and saved for analysis. In excising these specimens, no attention was given to content, other than attempting to avoid long silences or extraneous sounds such as coughing.

In Praat, pitch and intensity values were computed at 10 ms intervals throughout the phonated portions of each 25-second specimen. In Matlab, differences were computed between successive values of pitch and intensity at each 10ms interval. Differences in pitch were computed relative to the initial pitch value in Hz of any given pair of values, similar to the method of Nilsonne[6]. Differences in intensity were computed by simple subtraction of succeeding values given in dB.

Next, data points representing simultaneous pitch/intensity transformations were located within equivalent-sized neighborhoods on a two-dimensional plane. This is shown graphically for an exemplary subject in figure 1.

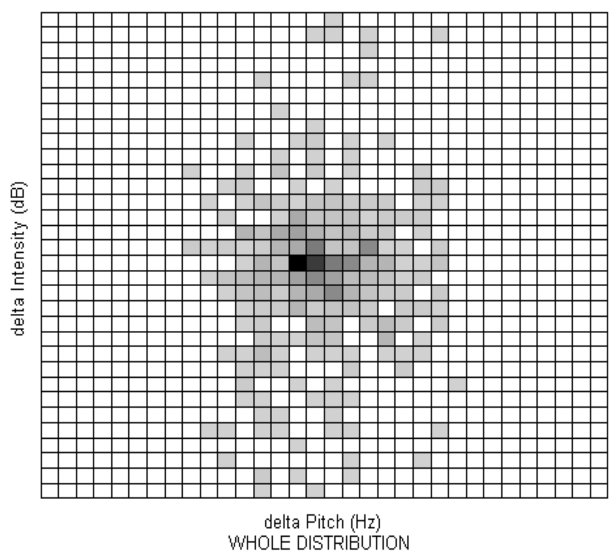


Figure 1. Relative Density of Neighborhoods of Pitch/Intensity Transformations

In this figure, the horizontal axis represents the magnitude of pitch differences, while the vertical axis represents the magnitude of simultaneous intensity differences. The shading within each square of the grid represents the relative density of data points within the neighborhood defined by the boundaries of the square, with the darker neighborhoods being more dense.

Locating pitch/intensity pairs within neighborhoods makes it is possible to discern patterns of activity and to separate patterns representing those transformations executed most often by the speaker from those transformations executed less often. The result of separating prevailing from infrequent pitch/ intensity tranformations may be depicted graphically. In figures 2 and 3, the separating of prevailing, or "core" transformations, from infrequent, or "border" transformations is shown for the exemplary subject of figure 1. In these figures as well as in the analysis of all other experimental specimens, an arbitrary threshold value of neighborhood densities was chosen to separate the core from border transformations. This arbitrary value provided a sufficient number of data points for analysis within both prevailing and infrequent sets as well as an easy visualization of core and border contours.

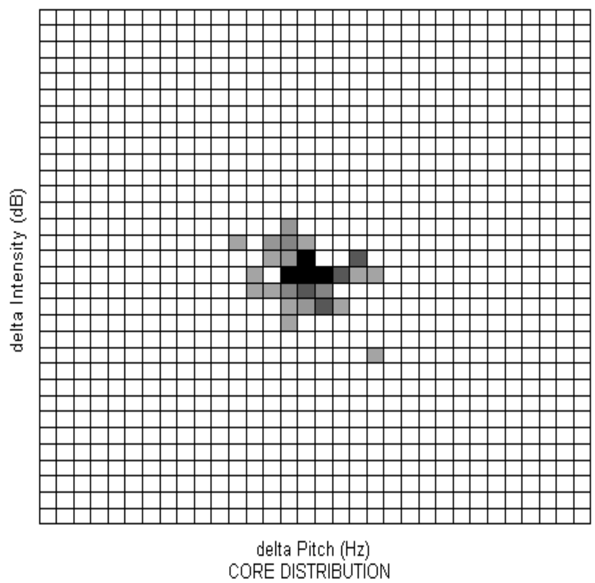


Figure 2. Relative Densities of Neighborhoods of Prevailing (Core) Pitch/Intensity Transformations

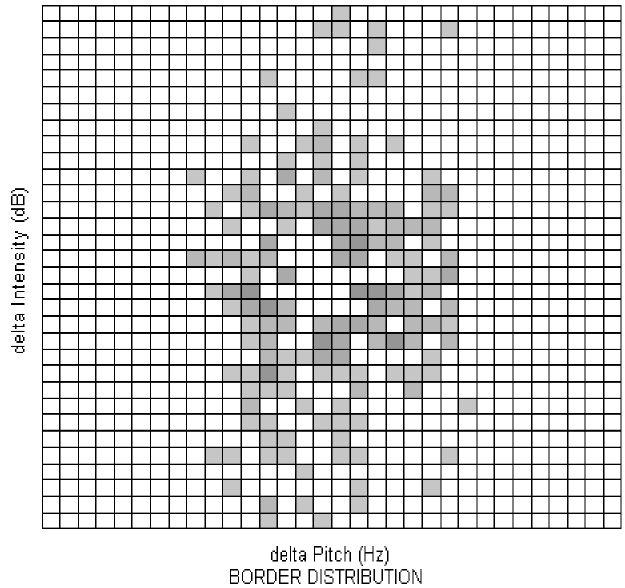


Figure 3. Relative Densities of Neighborhoods of Infrequent (Border) Pitch/Itensity Transformations

Mean intensity values were then computed for the core and border subsets of each specimen, and the core mean intensity difference was subtracted from the border mean intensity difference to provide an initial acoustic variable of interest. This pre-adjusted acoustic variable of interest may be expressed as:

Pvar=mnIborder-mnIcore

where
Pvar=the pre-adjusted variable of interest
mnIborder=mean of infrequent (border) intensity differences
mnIcore=mean of prevailing (core) intensity differences

This initial variable was than adjusted for cognitive style as referred to in section 3.1. The adjustment factor, "sdPcore/sdIcore" is defined below. The adjusted acoustic variable of interest is then:

Avar=(mnIborder-mnIcore)/(sdPcore/sdIcore), where

Avar=the adjusted acoustic variable of interest

sdPcore=standard deviation of prevailing (core) pitch differences
sdIcore=standard deviation of prevailing (core) intensity differences

A correlation coefficient, r, and its associated p-value between each BFI variable and each of the two acoustic variables of interest, Pvar and Avar, were then calculated for the 15 subjects. A coefficient of determination, r^2 , between each BFI variable and each of the two acoustic variables of interest was also calculated.

4. RESULTS

Table 1 depicts the BFI indices, the pre-adjusted variable of interest score, and the adjusted variable of interest score for each subject. In this table the personality indices are abbreviated as follows: E=extraversion, A=agreeableness, C=conscientiousness, N=neuroticism, and O=openness. The pre-adjusted variable of interest is abbreviated "Pvar," and the adjusted variable of interest is abbreviated "Avar."

Subject #	E-score	A-score	C-score	N-score	O-score	Pvar	Avar
1	3.5	4.22	2.78	2.63	4.5	-18.9	-10.07
2	4	4.78	4.67	1.5	3.5	-8.56	-4.57
3	2.63	3.89	3.78	2.75	2.5	-16.51	-5.38
4	4	3	3.44	2.38	3.8	0.74	0.66
5	2	3.56	1.89	4.5	3.7	-11.49	-11.69
6	3.38	2.44	4.56	3.5	4.9	4.16	1.56
7	3.13	3.44	3.33	3	4.8	-13.69	-5.65
8	3.13	4.11	3.67	3.75	4.5	-4.33	-2.32
9	2.75	4.89	4.11	2.75	3.9	1.99	0.62
10	4.38	4	4.11	4.88	3.6	-2.91	-1.37
11	3.63	3	2.56	4.13	3.1	-11.41	-6.12
12	2.88	4.33	3.11	2.88	3.4	-9	-5.08
13	4.13	3.78	3.89	3.25	3.9	-1.43	-0.5
14	3.38	3.44	4	4.25	4.5	-1.34	-0.83
15	4.38	3.44	3.89	3.38	3.8	-4.9	-2.66

Table 1. BFI scores, preadjusted(Pvar) and adjusted (Avar) acoustic variable scores per subject. E=extraversion, A=agreeableness, C=conscientiousness, N=neuroticism, O=openness.

Table 2 depicts the Pearson correlation coefficient, r, of the BFI scores with the Pvar and Avar scores along each BFI axis. For each correlation the associated p-value is also given.

BFI factor	r (Pvar)	p-value	r (Avar)	p-value
E	0.31	0.25	0.45	0.09
A	-0.21	0.45	-0.17	0.54
C	0.58	0.02	0.77	0.0008
N	0.15	0.6	0.04	0.88
O	0.32	0.25	0.22	0.42

Table 2. Correlation coefficients (r) and p-values of pre-adjusted(Pvar) and adjusted (Avar) acoustic variables with BFI indices. E=extraversion, A=agreeableness, C=conscientiousness, N=neuroticism, O=openness

One can see from table 2. that the adjusted acoustic variable, Avar, correlates quite highly with conscientiousness scores on the BFI ($r = .77$, $p=.0008$). There is a mild correlation of Avar with the extraversion scores, although the significance is weak ($r=.45$, $p=.09$). There is little or no correlation of Avar with the agreeableness, neuroticism, or openness scores.

It can also be seen from table 2 that adjusting the difference between mean border and core intensity differences to take account of the relative degrees pitch and intensity variability improved the correlation between conscientiousness scores and acoustic measures considerably. Without adjustment $r=.58$ with $p= .02$, in comparison to an adjusted $r= .77$ and $p= .0008$. This adjustment also strengthened the correlation between extraversion scores and acoustic measures, with a preadjustment $r= .31$, an adjusted $r=.45$, and p-values of .25 and .09, respectively.

Another way to look at the effect of adjusting acoustic measures with respect to pitch and intensity activity is to calculate a coefficient of determination, r^2, for acoustic measures with each BFI parameter. For two variables, this statistic is an estimate of the percent of variablility of one measure accounted for by variability in the other. Squaring the correlation coefficients of Pvar and Nvar with BFI scores yields the coefficients of determination depicted in table 3.

BFI factor	r^2 Pvar	r^2 Avar
E	0.1	0.2
A	0.04	0.03
C	0.34	0.59
N	0.02	0
O	0.1	0.05

Table 3. Coefficients of determination, r^2 , of BFI indices and acoustic measures.

One can see from table 3. that the adjusted acoustic measure, Avar, accounts for 59% of the variability of the conscientiousness scores. The converse is also implied, that 59% of the variability of Avar may be predicted from the BFI conscientiousness score. For the BFI extraversion score the coefficient of determination is 20%. Prior to adjustment, these figures are 34% for conscientiousness and 10% for extraversion. As with correlation coefficients, there is an improved separation of coefficients of determination for conscientiousness and extraversion when acoustic measures are adjusted for relative pitch/intensity variability.

5. DISCUSSION

The results of this study demonstrate, with qualifications, one possible way of discerning the personality characteristic of conscientiousness within a signal of natural speech and discriminating it from the four other personality characteristics composing the Big Five Inventory. The degree of discrimination may be considered moderately high between conscientiousness and extraversion, and very high between conscientiousness and the other three personality factors of agreeableness, neuroticism, and openness. The results also indicate, with qualification, that the discriminating of conscientiousness is strengthened when acoustic analysis takes into account the relative preference of the speaker for pitch or intensity modulation.

A number of qualifications pertain. The subjects of the study, for one, are drawn from the population of persons defined as "ill," and it would be interesting to repeat this protocol with persons defined as "normal." The subjects also have a prior relationship with the investigator that could have influenced the measurements, especially the BFI's, in some unknown way. Blind protocols will be preferable in follow-up studies. The validity of the BFI itself is taken for granted in this study with no account taken of the potential for overlapping characteristics. The 60-item BFI may represent an improvement in subsequent studies. Test-retest reliability of the acoustic feature extraction system was not assessed, nor was the system itself fully assessed for engineering adequacy. It is likely that improvements could be made in the method of computing the acoustic variables of interest if the present method were refined.

Finally, it is unclear how the variables of interest may relate to other acoustic variables that could also have been extracted, or how the personality characteristic of conscientiousness may relate acoustically to the remaining four characteristics, if at all. These issues, it seems, would be important to address in the process of designing a machine that conveys a coherent sense of realism.

If the potential errors of the analysis described in this study can be identified and reduced, and if modified research protocols turn out to validate the approach taken in this study, it would, in principle, be possible to embed the signal-processing algorithms used here within a machine for automated performance. Such a machine could recognize the personality characteristic of conscientiousnss in a speaker and be taught to appropriately modify its response in dialogue. Similarly, such a machine could, in principle, modify its synthetic speech to convey the quality of conscientiousness, though this would perhaps be more complicated to achieve.

REFERENCES

1. K.R.Scherer, Vocal affect expression: a review and a model for future research, Psychological Bulletin 99(2)(1986) 143-165.

2. D. Shapiro, Neurotic Styles, Basic Books, New York, 1965.

3. O.P. John, S. Srivastava, The big five trait taxonomy: history, measurement, and theoretical perspectives, in: L. A. Pervin, O. P. John (Eds.), Handbook of Personality: Theory and Research, Guilford Press, New York, 1999, 102-138.

4. P.Boersma, D. Weenink, Praat version 4.6.06, Amsterdam, The Netherlands, copyright 1992-2007.

5. The Mathworks, Inc., Matlab version 7, R14, Natick, MA, copyright 1984-2004.

6. M. Edelson, The Idea of a Mental Illness, Yale University Press, New Haven, CT, 1971.

7. J. Piaget, Structuralism, Harper & Row, New York, 1970.

8. N. Chomsky, Syntactic Structures, Mouton & Co., The Hague, The Netherlands, 1957.

9. G.F. Mahl, Psychological Conflict and Defense, Harcourt Brace Jovanovich, New York, 1969

10. A. Nilsonne, J. Sundberg, S.Ternstrom, A. Askenfelt, Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression, J. Acoust. Soc. Am. 83(2) (1986) 716-728.