voicesense

Speech Profiling Validation Study

BIG 5 Personality

voicesense

Background: Voicesense specializes in real-time, prosodic speech analysis and specifically on personality and state of mind profiling. Our patented personality speech analysis introduces a new bio-metric concept, based on the fact that people tend to have typical, repeating speech patterns over time. VoiceSense offers speech based analytic solutions for Human resources, Health, Fin-Tech, and Enterprise markets.

Purpose: The purpose of this in-house study was to attain generic speech profiling structures that would measure the Big 5 personality scales accurately, based on a combined sample of English and Hebrew subjects.

The study was designed to assist verifying that the Voicesense personality speech profiling system is (a) valid; (b) language independent; and (c) accurate.

We are aware that there are some differences in speech patterns across cultures, languages, gender and age. However, our claim is that the basic speech patterns that represent personality are generic for all people, and that the cultural differences would be just small biases on top of these general patterns.

Sample: The sample included 199 subjects. 95 were American, English speaking subjects and 104 were Israeli, Hebrew speaking subjects.

The 95 English speaking subjects consisted of three ethnic groups: Black (28 subjects); Hispanic (22 subjects); White (45 subjects).

92 subjects were males and 107 were females. Out of the English speaking subjects, 40 were males and 55 females; Out of the Hebrew speaking subjects 52 were males and 52 were females.

All subjects were between 20 and 70 years old. The English speaking subjects were divided into three age groups: 20-34 years old (42 subjects); 35-49 years old (30 subjects); over 50 years old (23 subjects). There was no age data for the Hebrew speaking subjects.

The English-speaking subjects were sampled from different US regions: Northeast (19 subjects); South (26 subjects); Midwest (23 subjects); West (27 subjects). There was no residence region data for the Israeli subjects.

The subjects were reached and approached by a market research and survey company. The subjects were compensated for their participation in the study.

Method: All the subjects downloaded Voicesense's mobile app (HR dialer) and completed 10-12 valid calls (at least 45 seconds of user voice per call). The calls were routed to Voicesense cloud servers and the subject's voices were analyzed in order to calculate the speech parameters for each subject.

After having completed the calls, each subject filled standard Big 5 online questionnaires (44 items).

Reference criteria (questionnaires):
The questionnaires were scored according to the five Big 5 scales. The scores were then normalized in order to reach equal variances and averages for all scales.

12 subjects with suspicious response patterns were removed from the sample. Suspicious patterns included responses with almost no variance (e.g. all 5: strongly agree) or responses with consistently strong positive self-presentation bias (e.g. only high responses in "positive" scales and only low responses in "negative" scales).

Response norms were calculated for the two languages, for the four combinations of language and gender, and for the three age groups. The questionnaire scores were corrected according to these norms in order to reach unbiased reference criteria. The questionnaire scores were not corrected for ethnicity or residence data.

For each scale, subjects were divided into 5 equal sized groups according to their score percentiles (using percentiles 20, 40, 60 and 80)—these groups represented the reference criteria categories for each scale—Low, Moderate Low, Moderate, Moderate High, High.

Speech parameters:
41 different raw speech parameters were calculated by the Voicesense speech analysis per subject per call.

The parameters were calibrated according to their amplitude and frequency values in order to correct speech differences caused by amplitude and frequency.

Various combinations of the raw speech parameters were used in order to calculate 142 additional,
high level speech parameters.

All 183 raw and high level speech parameters were normalized in order to reach equal variances and averages for all parameters.

For each subject, the same parameters in all the calls were averaged in order to reach 183 final characteristic speech parameters per subject.

Speech patterns: We define a speech pattern as a structure or combination of several speech parameters that represent a certain personality tendency. In this study five speech patterns were defined, one per each Big 5 scale.

The study focused on determining the speech parameters that should be included in each of the five speech patterns. In order to do so Pearson correlations were calculated between each of the 183 speech parameters and between the five reference criteria Big 5 scores for all subjects.

After initial selection of the participating parameters in each of the five speech patterns, regression analysis was run in order to understand the relative contribution of each parameter. Based on the regression another selection was made in order to reach the final speech parameter participation.

At the end of the process, each Big 5 speech pattern consisted of between 13 to 18 speech parameters.

Language, gender and age speech normalization: The final five Big 5 speech pattern scores were first normalized in order to bring them into equal variances and averages.

Then speech pattern norms were calculated for the two languages, for the four combinations of language and gender and for the three age groups. The speech pattern scores were corrected according to these norms in order to reach unbiased speech scores. The scores were not corrected for ethnicity or residence data.

For each Big 5 speech pattern the subjects were divided into 5 equal sized groups according to their score percentiles (percentiles 20, 40, 60 and 80)—these groups represented the final speech profiling categories for each scale—Low, Moderate Low, Moderate, Moderate High, High.

Matching the reference criteria scores with the speech profiling scores: As mentioned, both Pearson correlations and regression analysis were used in order to determine the statistical match and significance between the speech patterns and the reference criteria scores.

However, the interesting operational question relates to the actual percentage of subjects whose speech pattern scores match their reference criteria (the questionnaire responses).

Measuring this percentage is not trivial, since the definition of a good match is somewhat vague. For example, if someone replies that she's high on a certain scale and according to the speech analysis she's only moderately high—what would be the match score?

Given the subjective nature of questionnaire self responses: self awareness differences, positive self presentation bias differences, self esteem differences, response pattern differences (tendency to respond with extreme or moderate rankings), differences in the perception of personality concepts, and so on—it is clear that some degrees of freedom are required when trying to match personal and external personality evaluations.

We decided to use the following approach: a match is considered positive when the speech profiling category (Low, Moderate Low, Moderate, Moderate High, High) matches the reference criteria with the exact same score category or the adjacent category, otherwise the match is considered negative.

This means that when someone marked herself as High on a certain scale (after normalization), both High or Moderate-high categories by the speech profiling would be considered a positive match, whereas Moderate or lower categories would not be considered a match.

This approach has a 52% probability for random match. It is close to the 50% random match probability of a simple high/low score approach, but it was preferred as it maintains higher category differentiation.

Results

Overall study results (Table 1):

All five speech profiling scores were highly correlated to the Big 5 reference criteria scores for the entire subject sample with high significance. The average match percentage between the reference criteria scores and the speech profiling scores was 76%.

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	74%	78%	79%	74%	73%	76%
Correlation	0.43	0.56	0.49	0.47	0.42	
Significance	P< 0.00001	P< 0.00001	P< 0.00001	P< 0.00001	P< 0.00001	

Table 1. Big 5: Reference and speech profiling scores match – Overall sample

Language comparison results (Tables 2, 3):

All five speech profiling scores were highly correlated to the Big 5 reference criteria scores for both English speaking subjects and Hebrew speaking subjects, with high significance. The average match percentages between the reference criteria scores and the speech profiling scores were 77% for English speakers and 74% for Hebrew speakers.

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	77%	80%	78%	77%	74%	77%
Correlation	0.51	0.63	0.49	0.56	0.53	
Significance	P< 0.00001	P< 0.00001	P< 0.00001	P< 0.00001	P< 0.00001	

Table 2. Big 5: Reference and speech profiling scores match – English speakers

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	72%	75%	79%	72%	72%	74%
Correlation	0.34	0.49	0.50	0.35	0.32	
Significance	P< 0.0005	P< 0.00001	P< 0.00001	P< 0.0005	P< 0.0005	

Table 3. Big 5: Reference and speech profiling scores match – Hebrew speakers

Ethnicity comparison results (Tables 4, 5, 6):
All five speech profiling scores were significantly correlated to the Big 5 reference criteria scores for the White subjects group and the Hispanic subjects group. Four speech profiling scores were significantly correlated to the Big 5 reference criteria scores for the Black subjects. The openness speech profiling score was not significantly correlated for the Black subjects group, but it was on the verge of significance (P<0.09). With a larger sample this profiling score would probably reach significance as well.

The average match percentages between the reference criteria scores and the speech profiling scores were 78% for the Black subjects group, 81% for the Hispanic subjects group and 75% for the White subjects group.

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	68%	84%	80%	84%	72%	78%
Correlation	0.52	0.68	0.44	0.69	0.28	
Significance	P< 0.005	P< 0.0001	P< 0.05	P< 0.0001	P< 0.09	

Table 4. Big 5: Reference and speech profiling scores match – Black subjects

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	82%	86%	73%	86%	77%	81%
Correlation	0.56	0.56	0.40	0.47	0.72	
Significance	P< 0.005	P< 0.005	P< 0.05	P< 0.05	P< 0.0001	

Table 5. Big 5: Reference and speech profiling scores match – Hispanic subjects

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	80%	75%	80%	68%	73%	75%
Correlation	0.48	0.63	0.57	0.47	0.54	
Significance	P< 0.001	P< 0.00001	P< 0.0001	P< 0.005	P< 0.001	

Table 6. Big 5: Reference and speech profiling scores match – White subjects

Gender comparison results (Tables 7, 8):

All five speech profiling scores were significantly correlated to the Big 5 reference criteria scores for both female and male subjects. The average match percentages between the reference criteria scores and the speech profiling scores were 74% for women and 77% for men.

Big 5 Scale N=187	Extraversion	Agreeability	Conscientiousness	Emotion stability	Openness	Average
Match %	68%	79%	80%	74%	72%	74%
Correlation	0.28	0.55	0.53	0.54	0.34	
Significance	P< 0.01	P< 0.00001	P< 0.00001	P< 0.0001	P< 0.001	

Table 7. Big 5: Reference and speech profiling scores match – Female subjects

Big 5 Scale N=187	Extraversion	Agreeability	Conscientiousness	Emotion stability	Openness	Average
Match %	82%	76%	77%	75%	74%	77%
Correlation	0.61	0.56	0.46	0.38	0.52	
Significance	P< 0.00001	P< 0.00001	P< 0.00001	P< 0.001	P< 0.00001	

Table 8. Big 5: Reference and speech profiling scores match – Male subjects

Age comparison results (Tables 9, 10, 11):

All five speech profiling scores were significantly correlated to the Big 5 reference criteria scores for the 20-34 years old age group and the 35-49 years old age group. Four speech profiling scores were significantly correlated to the Big 5 reference criteria scores for the over 50 age group. The Conscientiousness speech profiling score was not significantly correlated for this age group (P<0.13). With a larger sample this profiling score would probably reach significance as well. The average match percentages between the reference criteria scores and the speech profiling scores were 76% for both the 20-34 years old age group and the 35-49 years old age group and 81% for the over 50 group.

Big 5 Scale N=187	Extraversion	Agreeability	Conscientiousness	Emotion stability	Openness	Average
Match %	73%	85%	80%	78%	65%	76%
Correlation	0.47	0.70	0.53	0.59	0.29	
Significance	P< 0.01	P< 0.00001	P< 0.001	P< 0.0001	P< 0.05	

Table 9. Big 5: Reference and speech profiling scores match – 20-34 years old subjects

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	75%	75%	82%	75%	75%	76%
Correlation	0.43	0.62	0.57	0.57	0.70	
Significance	P< 0.05	P< 0.001	P< 0.001	P< 0.001	P< 0.00001	

Table 10. Big 5: Reference and speech profiling scores match – 35-49 years old subjects

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	89%	79%	68%	79%	89%	81%
Correlation	0.72	0.56	0.28	0.47	0.69	
Significance	P< 0.001	P< 0.01	P< 0.13	P< 0.05	P< 0.00001	

Table 11. Big 5: Reference and speech profiling scores match – over 50 years old subjects

Residence region comparison results (Tables 12, 13, 14, 15):
For the most part, the five speech profiling scores were significantly correlated to the Big 5 reference criteria scores for the different residence regions. All Northeast speech profiling scores were significantly correlated to the reference criteria scores. Four speech profiling scores were significantly correlated to the reference criteria scores for the South, Midwest and West regions. The Conscientiousness speech score for the South region, the Emotion stability speech score for the West region and the Agreeability speech score for the Midwest region were not significantly correlated or were on the verge of significance ($p<0.07$; $p<0.07$, ;$P<0.17$ in accordance), probably due to the small sample sizes. The average match percentages between the reference criteria scores and the speech scores were 86% for the Northeast region, 76% for the South region, 70% for the Midwest region and 75% for the West region.

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	83%	89%	89%	89%	78%	86%
Correlation	0.52	0.74	0.73	0.82	0.41	
Significance	P< 0.05	P< 0.001	P< 0.001	P< 0.0001	P< 0.00001	

Table 12. Big 5: Reference and speech profiling scores match – Northeast region subjects

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	65%	100%	70%	83%	83%	76%
Correlation	0.43	0.82	0.33	0.60	0.77	
Significance	P< 0.05	P< 0.00001	P< 0.07	P< 0.01	P< 0.00001	

Table 13. Big 5: Reference and speech profiling scores match – South region subjects

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	76%	48%	81%	76%	67%	70%
Correlation	0.41	0.22	0.59	0.42	0.56	
Significance	P< 0.05	P< 0.17	P< 0.01	P< 0.05	P< 0.01	

Table 14. Big 5: Reference and speech profiling scores match – Midwest region subjects

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	84%	84%	76%	64%	68%	75%
Correlation	0.56	0.68	0.34	0.31	0.35	
Significance	P< 0.01	P< 0.0001	P< 0.05	P< 0.07	P< 0.001	

Table 15. Big 5: Reference and speech profiling scores match – West region subjects

Overall sample results without language, gender and age corrections (Table 16):
All five speech profiling scores were highly correlated to the Big 5 reference criteria scores for the entire subject sample with high significance, even when not correcting the speech profiling scores for language, gender and age norms.

The average match between the reference criteria scores and the speech profiling scores without the language, gender and age corrections was 72%. The average match when including the corrections was 76% (Table 1), so the combined contribution of these corrections to the overall match was 4%.

When examining separately the contribution of each norm correction to the match level, the language correction added 2%, the gender correction added 3% and the age correction added 2% to the match percentage. Part of these corrections overlap, as the combined contribution of the three was only 4%.

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	72%	78%	73%	71%	65%	72%
Correlation	0.42	0.54	0.49	0.47	0.39	
Significance	P< 0.00001	P< 0.00001	P< 0.00001	P< 0.00001	P< 0.00001	

Table 16. Big 5: Reference and speech profiling scores match – without language, gender and age corrections

Overall sample results without removing subjects with suspicious response patterns (Table 17):

All five speech profiling scores were highly correlated to the Big 5 reference criteria scores for the entire subject sample with high significance, even when not removing the subjects that had suspicious response patterns from the sample.

The average match between the reference criteria scores and the speech profiling scores when not removing the subjects who had suspicious response patterns from the sample was 74%. The average match when removing these subjects was 76% (Table 1), so the effect of removing these subjects to the overall match was 2%. The match difference was higher when examining only the English speaking subjects (77% vs 73%), while there was no match difference for the Hebrew speaking subjects (74% vs 74%).

Big 5 Scale N=187	Extraversion	Agreeability	Conscien-tiousness	Emotion stability	Openness	Average
Match %	72%	75%	77%	72%	72%	74%
Correlation	0.36	0.51	0.45	0.38	0.38	
Significance	P< 0.00001	P< 0.00001	P< 0.00001	P< 0.00001	P< 0.00001	

Table 17. Big 5: Reference and speech profiling scores match – without deleting suspicious subjects

Conclusions and discussion

Validity: The main study target was successfully achieved—generic speech profiling patterns for measuring the Big 5 personality scales were developed, while all five speech profiling patterns reached strong correlations to the five Big 5 reference criteria scores, with high statistical significance.

The speech profiling patterns were found to correlate significantly with the reference criteria for both English and Hebrew speakers and across age, gender, ethnicity and residence region differences.
The strong statistical correlations along with the generality of the speech patterns beyond language, gender, age and region strengthen the confidence that these speech patterns are indeed valid. This supports the concept that speech analysis can offer objective personality measurement.

Nevertheless, further assurance for such validation would require a replication study, which would test the speech patterns that were developed in the current study on a separate, independent sample of subjects. Such a replication study is currently in process and results would be reported in the coming weeks.

Generality: The results of the study indicate clearly that the same speech patterns are significantly correlated with personality tendencies for both English and Hebrew speakers, as well as for different ethnic groups and across gender, age and regional differences.

We did calculate separate norms for language, gender and age speech patterns and we corrected the speech scores accordingly. However, these were minor biases corrections. The meaningful point is that the same speech parameters and the same speech patterns were found to be related to the personality tendencies for both languages.

Moreover, the correlations between the speech patterns and the personality scales remained highly and significantly correlated even when the corrections for language, gender and age were not applied. The accumulated contribution of all corrections to the match percentage between the speech scores and the reference scores was relatively small—4%, and only 2-3% for each correction separately. The match percentage without any correction remained high: 72%.
These finding support our claim that speech patterns, as measured by Voicesense speech analysis, are language independent, and may be common to humans beyond cultural, gender and age differences.
The clear advantage is that it places all people on one, generic and uniform scale of measurement.
Naturally, replication studies with more languages are required to further establish these conclusions.

Accuracy: The practical accuracy of the speech profiling is best perceived by the match percentage between the speech scores and the Big 5 reference criteria scores. Simply, this number represents the percentage of subjects whose speech personality scores are similar to their questionnaire personality scores.
According to the study results, this match percentage was around 75% in most aspects that were examined.

The overall match percentage for the entire sample was 76%; For English speakers: 77%; For Hebrew speakers: 74%; For black subjects: 78%; Hispanic: 81%; White: 72%; For Men: 77%; For Women: 74%. Similar accuracies were consistent across age and regional groups geographically.

The consistency of these percentages in all dimensions suggests that this is the match accuracy that should be expected when applying speech profiling analysis for personality evaluation.

However, an important aspect to keep in mind is the well known subjective and inaccurate nature of questionnaire self responses in general. There are many biasing factors in self reported questionnaires, especially when applied to domains that are subjective in their nature, such as personality evaluation.

Biasing factors include:
The strong tendency of people to present themselves positively rather than accurately; the limited awareness of people regarding their real nature; the effect of high or low personal self-esteem on the way people perceive themselves; the difficulty of people to see themselves objectively; personal differences in response patterns (some people tend to reply with extreme rankings while others tend to provide moderate rankings); the ambiguity in the way people perceive personality concepts (is there one clear and common definition for openness?).

Another biasing factor, relevant especially to the current study, is the incentive people have when filling the personality questionnaire. When someone is applying for a job, we may expect that the positive self presentation would play a strong role. In this study, people were compensated for participation. This probably caused some people to fill the questionnaire as quickly as possible, without paying too much attention to answer accurately. As mentioned, we removed from the overall sample 12 subjects whose response patterns where clearly suspicious. Considering all the biasing effects described above regarding self questionnaires, we can assume that at least part of the 25% of speech profiling scores that did not match to the reference criteria scores, can be explained by the questionnaires' lack of accuracy rather than the speech profiling inaccuracy.

We can therefore assume that the overall accuracy of the speech profiling is even higher than the 75% that was found in the study. How much higher? Hard to say, but above 80% is certainly reasonable.

To summarize, this study suggests that using speech profiling offers an objective, valid, accurate and language independent method for performing Big 5 personality evaluation in English and Hebrew. Further studies are required to replicate the results and to widen the scope for more languages and other personality constructs.

www.voicesense.com