voicesense

Speech Profiling

Hogan HPI Validation Study

voicesense

Background: Voicesense specializes in real-time, prosodic speech analysis and provides personality and behavioral tendencies profiling. Our patented personal profiling speech analysis introduces a new biometric concept, which correlates individual speech patterns to personal behavioral tendencies. Based on the speech profiling, Voicesense offers solutions for enterprise analytics, personal health tracking and personal assessment for human resources.

Purpose: The purpose of the current study was to characterize the generic speech patterns that reflect the seven basic Hogan personality scales (HPI). The study used a combined sample of English, Hebrew and Hungarian speaking subjects.

The study was designed to assist verifying the validity, the language independence and the accuracy of Voicesense vocal analysis for the HPI instrument.

Sample: The sample included 151 subjects. 67 were English speaking subjects—65 American and 2 from New Zealand. 50 subjects were Israeli, Hebrew speaking subjects. 34 were Hungarian speaking subjects.

70 subjects of the overall sample were males and 81 were females. Out of the English-speaking subjects, 28 were males and 39 females; Out of the Hebrew speaking subjects 25 were males and 25 were females; and out of the Hungarian speaking subjects 17 were males and 17 were females.

All subjects were between 20 and 70 years old.

The English and Hebrew speaking subjects were reached and approached by a market research and survey company. The subjects were compensated for their participation in the study. The Hungarian and New Zealand subjects were part of the local Hogan personnel, who participated voluntarily.

Overall, the sample consisted of 2252 audio recordings (smart phone calls of regular interactions) for all the subjects (an average of 14.9 call recordings per subject).

Method: Reference criteria (questionnaires):
All subjects filled the Hogan HPI questionnaires (short form), which were scored according to Hogan standards, using four categories per each scale (High / Above average / Below average / Low). We assume that the scores in the US, NZ and Hungary were calculated according to known Hogan norms. Israeli norms were not available, so it is assumed that the US norms were used.

Speech parameters:
All the subjects downloaded Voicesense's mobile app (HR dialer) and made 10-20 valid calls using the app. When using the app, the user makes regular calls from the contact list and the calls are routed through Voicesense cloud servers for analysis. A valid call required at least 45 seconds of the user's voice. Voicesense cloud servers analyze the recorded calls and calculate the speech parameters per each call.

211 raw speech parameters were calculated by Voicesense speech analysis per each call. The calculation includes calibration for amplitude and frequency values according to Voicesense generic calibration norms.

Each of the 211 parameters (per every call) receives a high/low score by comparing it to Voicesense generic speech norms.

The 211 speech parameters consist of 22 different parameter categories. The speech parameters within each category are related and dependent; therefore, no more than one parameter from each category is allowed in the parameter selection process.

Training and test samples:
The overall sample was divided into a training sample and a test sample using a random routine, while verifying that both samples will retain the same proportion of English, Hebrew and Hungarian subjects. The training sample included 104 subjects and consisted of 1542 calls (about 69% of the overall sample). The test sample included 47 subjects and consisted of 710 calls (about 31% of the overall sample).

Training sample folds and iterations (Pearson correlation):
Based on cross validation practices, the training sample calls were divided into seven separate folds of about 220 calls each. Three folds represented English calls, three folds represented Hebrew calls and one fold represented Hungarian calls.

Pearson correlations were calculated between each of the 211 speech parameters in each fold and between all the reference criteria (the seven Hogan questionnaire scores).

Still in accordance with cross validation practices, this process was repeated five times (five iterations) while the calls in the training sample folds were distributed randomly for each iteration and the correlations were recalculated (we did retain the structure of separate English, Hebrew and Hungarian folds in the different iterations).

Training sample T-test: In addition to the Pearson correlation, we sorted the training sample for each Hogan scale according to the Hogan score, divided the sample into two groups of high/low scores and ran a t-test analysis for each of the 211 speech parameters per each Hogan score.

Parameter selection for the Hogan speech patterns in the training sample: We define a speech pattern as a combination of speech parameters that are found related to each personality scale. Seven speech patterns were calculated to represent the seven HPI scales, consisting on the statistical findings of the 211 parameters.

Regression Analysis: As part of the training stage, regression analysis was performed on the training sample to assess the validity and predictive ability of the speech scores.

Test sample: After reaching the final speech scores for the HPI scales, based on the training model per each scale as described above, these HPI speech scores were calculated for all the calls in the test sample. Naturally, no tuning was performed on the test sample, only calculation of the final speech scores.

Results

In order to overcome the difficulties of analyzing a relatively small sample (151 subjects), we started by treating the calls of each subject separately. There is a total of 2252 calls and analysis was made as if each call was a new subject (all calls from the same subject had the same Hogan score). The training sample therefor consisted on 1542 subjects (calls) and the test sample consisted on 710 subjects (calls).

t-test: In each Hogan scale, the sample was divided into two groups: “High” and “Low” according to the Hogan scores (High= 3-4, Low=1-2). T-test was performed on the calculated speech scores of each Hogan scale to examine the significance of the difference between the two groups (High and Low). The results of both the training and test samples are significant in all Hogan scales.

Hogan HPI		Adjustment		Ambition		Sociability		Interpersonal Sensitivity		Prudence		Inquisitive		Learning Approach	
		High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Training Sample N= 1542	Average speech score	0.49	0.43	0.51	0.45	0.54	0.46	0.52	0.47	0.55	0.46	0.49	0.40	0.52	0.44
	T- test	1.11E-12		6.71E-17		4.85E-30		2.09E-07		3.06E-21		7.81E-17		1.36E-16	
Test Sample N= 710	Average speech score	0.45	0.40	0.51	0.46	0.52	0.46	0.52	0.48	0.58	0.53	0.43	0.38	0.49	0.42
	T- test	5.09E-04		6.63E-06		4.55E-09		4.60E-03		9.40E-04		6.66E-04		1.92E-08	

Table 1. HPI: T-test analysis results

Pearson Correlation:

Correlations were calculated between Hogan scores and the speech scores for each Hogan scale in both samples (training/test). All seven speech profiling scores were highly correlated to the Hogan reference criteria scores in the training and the test samples (Average correlations of 0.20 and 0.18) with high significance (P-value =<0.0002).

Hogan HPI	Adjustment	Ambition	Sociability	Interpersonal Sensitivity	Prudence	Inquisitive	Learning Approach	Average
Training sample Correlation N=1542	.17	.17	.22	.17	.22	.20	.24	.20
Corr. Significance	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Test sample Correlation N=710	.13	.14	.22	.16	.26	.16	.22	.18
Corr. Significance	0.0002	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

Table 2. HPI: Reference and speech profiling scores correlations

After assembling successful speech scores using a large sample (treating each call as a subject, N=2252), we grouped the calls by subject and tested the results on an averaged speech data for each subject. The results have improved tremendously and the advantage of multiple speech samples for a subject is clearly demonstrated - the correlations for all Hogan scales (Average correlations of 0.32 in the training sample and 0.35 in the test sample) had increased drastically with high significance (P-value =<0.01)

Hogan HPI	Adjustment	Ambition	Sociability	Interpersonal Sensitivity	Prudence	Inquisitive	Learning Approach	Average
Training sample Correlation N=1542	.32	.32	.36	.28	.30	.34	.35	.32
Corr. Significance	0.0004	0.0004	<.0001	0.001	0.0009	0.0002	0.0001	0.0004
Test sample Correlation N=710	.32	.39	.37	.35	.41	.28	.36	.35
Corr. Significance	0.01	0.003	0.005	0.007	0.002	0.02	0.006	0.007

Table 3. HPI: Reference and speech profiling scores correlations – Averaged speech data

Regression Analysis:

Regression analysis was performed in order to examine the ability of the speech scores to predict the Hogan scores. The analysis was made on the training sample on which the speech score model was based. The ANOVA demonstrates significant results for all seven Hogan scales (P<0.0001) and a relatively low Adjusted R2 value. The speech score successfully predicts 3.1%-6.6% of the variance.

Hogan HPI N=1542	Adjustment	Ambition	Sociability	Interpersonal Sensitivity	Prudence	Inquisitive	Learning Approach	Average
Adj. R2	3.4%	3.1%	6.6%	3.6%	5.3%	3.9%	5.7%	.32
ANOVA p-value	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	0.0004

Table 4. Regression analysis results

Analysis of the ESM and Narrative samples

The new sample consisted on two samples (ESM and Narrative) joined into one. The origin of this new speech data sample is in recorded interviews taken with students who filled in among others the Hogan HPI, as part of two research projects. The number of subjects who had speech data was a total of 118 (ESM: N=37, Narrative: N=81) and for each subject there was one speech sample.
First, we intended to use the new sample as an extension of the sample presented in this report. We combined the old and the new, creating one big database of 269 subjects (2370 calls).
Our speech analysis revealed significant differences between the two samples in manner of the speech parameters behavior, which means that the data in the new sample is essentially different. We tried to apply the thresholds of the previous sample on the new one but still the different behavior of parameters was present. This essential difference in speech data between the old and the new sample is assumed to be related to the different communicational context in which the speech samples were taken – in the old sample the context was of natural conversations and in the new sample, it was of interviews. It is highly probable that these different contexts had generated essential differences in the speech patterns.
Considering these differences, we could not continue with the analysis of the two samples joined into one.
Secondly, we intended to use the new sample as an additional independent test sample, to examine once again the results we have presented in the previous test sample. Due to the significant differences in speech parameters behavior, this examination was not possible.

Conclusions and discussion

1. Validation- The main study target was successfully achieved – generic speech profiling patterns for measuring the Hogan- HPI were developed, while all seven speech profiling patterns reached strong correlations to the seven HPI reference criteria scores, with high statistical significance.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.
2. As shown in the grouped/averaged speech data per subject method, multiple speech samples for each subject contributes significantly to the level of accuracy that is possible within each subject. Still, a large sample, as we generated by treating each call as a subject, is critical for training and creating speech profiling scores.
3. The attempts of using the new sample (ESM + Narrative), revealed differences in speech data that are assumed to be related to the communicational contexts differences. In further research, we aim to reach two sets of analysis calibration – conversation and self-talking. Additional speech samples of this communicational context are of course necessary.
4. In addition to the statistical analysis performed and presented in this report, a calculation of the matching percent between the Hogan reference scores and the speech profiling scores will provide us with more information about the accuracy of our method. This kind on analysis will be possible with a given Hogan scores distribution in the population for each scale, so that we can set the percentiles for a proper comparison with the speech scores. If such information can be given we will gladly perform the analysis.
5. In further research, we can relate to the subjects by age, gender, language and ethnic groups (US), and examine the consistency of the results along the different groups.

voicesense

www.voicesense.com