Examining Linguistic Content and Skill Impression Structure for Job Interview Analytics in Hospitality

Skanda Muralidhar
Idiap & EPFL, Switzerland
smuralidhar@idiap.ch

Daniel Gatica-Perez
Idiap & EPFL,Switzerland
gatica@idiap.ch

ABSTRACT
First impressions are critical to professional interactions especially in the context of employment interviews. This work investigates connections between linguistic content and first impressions in job interviews and the structure of ten soft skills and overall impressions. Towards this, we transcribe 169 role-played job interviews conducted at a hospitality school and analyze the linguistic content using off-the-shelf software. To understand the structure of the soft skill impressions, we conduct a principal component analysis. We then develop methods to automatically infer impressions using verbal and nonverbal features and their combination. Results indicate low predictive power of verbal cues for overall impression ($R^2 = 0.11$). Combined verbal and nonverbal cues explain up to 34% of variance, a marginal improvement over $R^2 = 0.32$ using only nonverbal cues. The use of principal components reveals a major component associated to overall positive and negative impressions that when used as labels for supervised learning results in a regression performance of $R^2 = 0.41$.

ACM Classification Keywords
I.7.m. Document and Text Processing: Miscellaneous; J.4 Social and Behavioral Sciences

Author Keywords
Verbal content; job interviews; first impressions; hospitality

INTRODUCTION
Job interviews are ubiquitous and the impact of nonverbal behavior (NVB) on job interview outcomes has been studied in psychology [1] and social computing [16, 13]. Nonverbal communication is known to play an important role in the outcome of negotiations [6], leadership [20], hirability [16], cohesion [9] up to certain levels. Verbal communication is also an important aspect of communication. This paper studies connection between spoken words and first impressions in the context of job interviews for hospitality students in a multi-sensor environment equipped with Kinect devices and microphone arrays.

MUM 2017, November 26–29, 2017, Stuttgart, Germany

ISBN 978-1-4503-5378-6/17/11…$15.00

DOI: https://doi.org/10.1145/3152832.3152866

First impressions are critical in the context of hospitality. First impressions are the mental image one forms about someone after a first encounter [11]. In the context of job interviews, these initial judgments can influence whether a candidate is hired or not. While NVB is an established channel of communication on which first impressions are formed [11], research in social psychology also shows that choice of words while speaking and writing reveal aspects of a person's identity [5] and thus potentially influence first impressions. Words people use provide cues to their thought processes, emotional states, intentions, and motivations [24]. For example, analysis of verbal content has shown that people tend to use positive emotion words (e.g., love, nice, sweet) when describing or writing about a positive event, and more negative emotion words (e.g., hurt, ugly, nasty) when writing about a negative event.

Research in fields of psychology and hospitality has relied on time-consuming manual labeling of behavior by trained experts. The ubiquitous availability of sensors and audio visual analytic techniques facilitate the automatic analysis of interactions. Previous literature investigates the role of words in a number of settings like video blogs [3, 21] and group meetings [22]. Using a dataset of YouTube video blogs, the work in [3] showed that verbal cues outperformed nonverbal cues in predicting three of the Big-Five personality traits. Similarly, it was shown that use of verbal cues and keyword analysis could predict perception of leadership during group meeting [22].

In the context of workplaces, some recent studies have suggested the feasibility of using linguistic content for predicting hirability [14, 4]. The work in [14] reported that their framework, trained on college students, detected the use of *We* instead of *I*, and the use of less fillers and more unique words as having links to positive impression. Similarly, the work in [4] reported improved prediction of expert scores using verbal content from LIWC in addition to a Doc2Vec method.

In this work, we investigate the role of verbal content in the inference of soft skills impressions in job interviews for young hospitality students. To this end, we extract verbal cues from 169 job interview transcripts. We analyze the relationship between extracted verbal cues and impressions of professional, communication and social skills through correlation analysis. We also perform a Principal Component Analysis (PCA) [25] on the annotated ratings of these social variables. The motivation of using PCA is to capture (in a few components) the variance of impressions that are mutually correlated. We then define a regression task to infer the overall impression scores and other skill variables by utilizing the extracted verbal and

Skill Type	Professional			Social			Communication			Overall Impression
Social Variable	Motivated (motiv)	Competent (compe)	Hardworking (hardw)	Sociable (socia)	Enthusiastic (enthu)	Positive (posit)	Communicative (commu)	Concise (consi)	Persuasive (persu)	ovImp
{ICC(2,k)}	0.52	0.56	0.54	0.57	0.68	0.60	0.60	0.55	0.69	0.73

Table 1: **Intra class correlation (*ICC*) of the annotated variables for the job interview data corpus.**

nonverbal cues. We find that the projection of the skill impressions onto the first principal component, which corresponds to the positive or negative general assessment of the candidates, can be inferred with $R^2 = 0.41$.

DATA CORPUS

Data Collection

In this study, we use a dataset of 169 dyadic interactions in a job interview setting, previously collected by our group [13]. This corpus was then transcribed manually for analysis. The interviews were structured (each interview followed the same sequence of questions), thus making it possible to compare results across subjects. Structured interviews are one of the most valid tools for personnel selection [8]. The interviews were conducted by research assistants, and were recorded synchronously using two Kinect cameras (one for each protagonist) at 30 frames per second. Audio was recorded at 48kHz using a microphone array placed at the center of the table. A detailed description of the data collection can be found in [13].

Annotations of Impression

The dataset was annotated for various social variables such as overall impressions, personality, as well as professional, communication and social skills [13]. To obtain impressions, the first two minutes (thin slices) of the interview were rated using a 7-point scale by five raters, who were masters student in psychology. The use of thin slices is common in psychology [17]. The agreement between the raters was assessed using Intraclass Correlation Coefficient (ICC) [23]. As a sample of raters annotated each video $ICC(2,k)$ is utilized as concrete reliability measure. Table 1 summarizes the annotated variables and their ($ICC(2,k)$ values taken from [13]). We observe that $ICC(2,k) > 0.5$ for all impressions, indicating moderate to high agreement between the raters.

Transcripts

As our first step towards the analysis of linguistic content, we manually transcribed all the interviews in the data corpus using a pool of five master's students in organizational psychology, who were native French speakers and fluent in English. Each question by the interviewer and answer by the applicant was transcribed, but only the applicant answers were utilized for linguistic analysis. The average number of transcribed words for an interview (applicant answers only) was 813, with a minimum of 358 and a maximum of 2587 words. In aggregate, the corpus comprised of 1690 minutes of interviews (mean duration: 10 minutes).

Figure 1: A snapshot of the job interview setup with the applicant on the left and the interviewer on the right.

FEATURE EXTRACTION

Verbal Cues

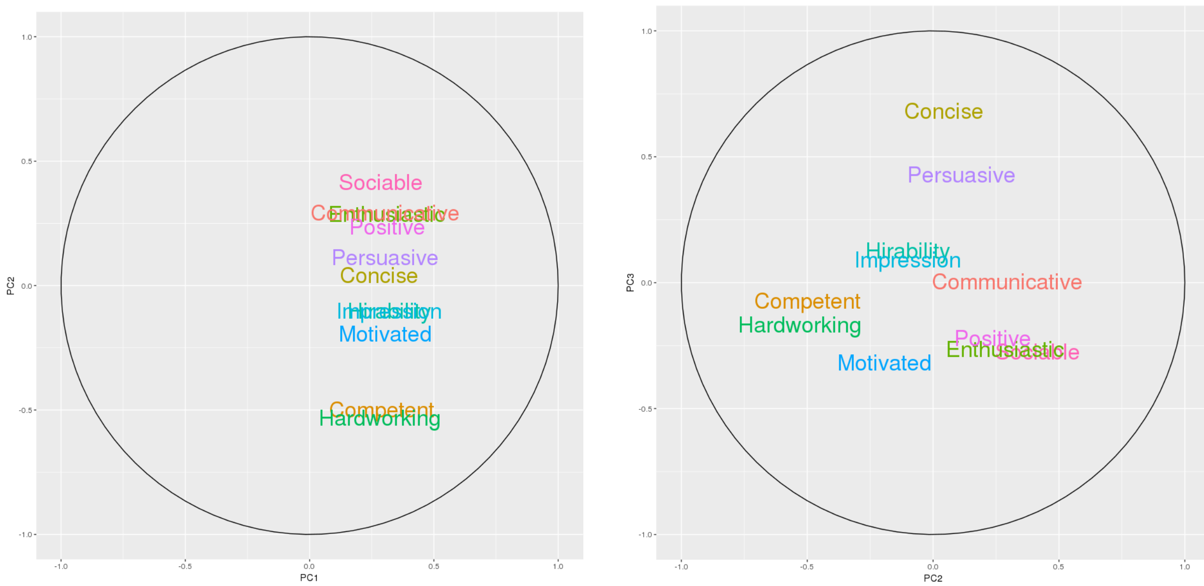
The job interview transcripts were processed to extract lexical features with the Linguistic Inquiry and Word Count (LIWC) [19], a software package widely used in social psychology [5] and social computing [3, 22] to extract verbal content. This tool was developed based on research in social psychology with an aim to link linguistic and para-linguistic categories to various psychological constructs. The English dictionary of LIWC is composed of $4,500$ words and word stems, while the French dictionary contains $39,164$ words. Each word in the interview transcript is looked up in the dictionary, and in case of a match the appropriate word category (out of 71 categories) is incremented. It must be noted that in LIWC, words can be assigned to more than one category at a time. After an interview transcription is processed, LIWC divides the count of categories by the total number of words in the document. Since LIWC is designed to process raw text, transcripts were not pre-processed.

Nonverbal Cues

We extracted various nonverbal cues from both visual and audio modality for behavioral representation of the dyadic interaction. These nonverbal cues include prosodic cues (pitch, energy, voice loudness modulation, spectral entropy), speaking activity features (speaking time, speaking turns, pauses, short utterances), visual motion (WMEI [2]) and head nods. A detailed description of the extracted features can be found in [13]. The choice of nonverbal cues extracted was based literature in social psychology [7, 10] and social computing [16, 17]. The nonverbal cues were extracted for the full interview and for both applicant and interviewer.

PRINCIPAL COMPONENT ANALYSIS OF SKILLS

As shown in [13], the manually annotated impression variables are correlated. Pairwise Pearson correlations was found to be in the range $r \in [0.60, 0.96]$ (median$= 0.81$). This suggests that a lower dimensionality representation of impressions could be found through principal component analysis (PCA) of the annotated variables. As a first step towards PCA, the annotated values were first pre-processed so that each variable had zero mean and unity variance. Then, PCA was conducted on the skill variables using the inbuilt *prcomp* function in R. The first three principal components (PC) are visualized in Figure 2, which displays the original variables projected onto the coordinate system. We observe that these three PCs account for 94.3% of the variance. Further, we observe that the first component, accounting for 82.6% of the variance, essentially corresponds to overall positive and negative impressions. Note



(a) (b)

Figure 2: **Clustering of perceived skills after principal component analysis (PCA). The first three principal components, accounting for** 94.3% **of the variance, are displayed here by projecting the original axes onto the PCA space** ($N = 169$)**.**

that all the variables in Table 1 are positively phrased. We also observe that *Communicative* and *Enthusiastic* overlap, and *Hirability* and *Overall Impressions* overlap. The second principal component, accounting for 6.1% of the variance, seemed to distinguish professional skills (Motivated, Competent) from communication (Communicative, Concise) and social (Positive, Sociable) skills (Figure 2b). The third component accounted for 5.6% of the variance. We can thus see that this lower dimensional representation of the impressions is appealing. The projection onto the first PCs were used as the labels in an inference task (Section 6.1).

CORRELATION ANALYSIS

As a first step in our analysis, pairwise correlation (using Pearson's correlation) between all the impression variables and verbal cues was calculated. For this analysis, the mean rating of all variables by the five raters and the features extracted from LIWC were utilized. Table 2 shows that a few LIWC features significantly correlate with overall impression scores.

We observe that there is a low correlation between verbal content and *Overall Impression* (ovImp). The use of personal pronouns (*pproun*), 1st person singular (*i, me*), 3rd person singular (*she, him*) are negatively correlated to overall impressions while use of 3rd person plural (*they, their*) are positively correlated. These observations are somewhat in line with those of reported by authors in [14]. Other weak effects observed are that participants who used negation (*negate*), non fluency (*hmm, er*) and asked questions (*QMark*) had lower overall impression scores.

INFERENCE

The inference of impressions was defined as a regression task and was evaluated using two machine learning methods, random forest (RF) and support vector machines (SVM). Towards this, the CARET R package [12] was utilized. Hyper parameters (i.e., number of trees, cost, polynomial degree) were automatically tuned by using an inner 10-fold cross-validation (CV) on the training set. The final machine-inferred scores was obtained by repeating the leave-one-video-out CV process 100 times.

SocVar	Personal Pronoun	1st pers singular	3rd pers singular	3rd pers plural	Negation	Non fluency	Question
motiv	**-0.17***	-0.12	**-0.21****	0.12	-0.10	-0.15	-0.09
compe	-0.15	-0.10	**-0.25****	**0.22****	**-0.21****	**-0.25****	-0.08
hardw	-0.14	-0.07	**-0.27****	**0.24****	-0.11	**-0.17***	-0.08
socia	-0.12	-0.12	-0.14	0.11	-0.09	-0.11	-0.15
enthu	-0.13	**-0.16***	**-0.16***	**0.17***	-0.13	-0.12	**-0.17***
posit	-0.14	-0.13	**-0.19***	**0.17***	-0.14	-0.12	-0.13
commu	-0.13	-0.13	-0.11	0.11	-0.13	-0.08	**-0.18***
conci	**-0.22****	**-0.16***	**-0.15***	0.12	**-0.17***	**-0.25****	**-0.18***
persu	**-0.18***	**-0.20***	-0.13	0.14	**-0.16***	**-0.20***	**-0.19***
ovImp	**-0.18***	**-0.17***	**-0.17***	**0.17***	**-0.17***	**-0.17***	**-0.20****

Table 2: **Correlation between linguistic cues and impressions** ($N = 169$) $^{**}p < 0.001$**;**$^*p < 0.05$. **Others not significant.**

The performance of automatic inference models was evaluated by two metrics: the root-mean-square error (RMSE) and the coefficient of determination (R^2). Both metrics are widely used measures. As the baseline regression model, the average impression score was utilized as the predicted value. RMSE is the difference between a model's predicted values and the values actually observed. The coefficient of determination (R^2) is based on the ratio between the mean squared errors of the predicted values, obtained using a regression model and the baseline-average model. Due to space constraints only the best performing model (RF) is presented.

Table 3 shows the performance of the RF models for the inference of impressions. We observe that performance of verbal

Variables	Baseline	Nonverbal		Verbal		Nonverbal + Verbal	
	R^2	R^2	RMSE	R^2	RMSE	R^2	RMSE
motiv	0.0	0.26	0.46	0.13	0.54	0.27	0.46
compe	0.0	0.21	0.36	0.17	0.37	0.26	0.33
hardw	0.0	0.21	0.41	0.16	0.43	0.26	0.38
socia	0.0	0.27	0.45	0.03	0.59	0.22	0.48
enthu	0.0	0.33	0.70	0.04	0.98	0.32	0.69
posit	0.0	0.35	0.55	0.06	0.80	0.33	0.57
commu	0.0	0.26	0.52	0.04	0.67	0.26	0.52
conci	0.0	0.21	0.52	0.09	0.59	0.26	0.48
persu	0.0	0.33	0.67	0.08	0.92	0.32	0.67
ovImp	0.0	0.32	0.82	0.11	1.07	0.34	0.79

Table 3: **Regression results ($N = 169$) for verbal, nonverbal and combining both cues using RF ($p < 0.05$ for all).**

cues in inference of social variables is low ($R^2 \in [0.03, 0.17]$), with the best performance achieved for *Competent* (compe).

This results are slightly better then the values reported in [15], and are similar to those reported in other settings like inference of leadership [22], mood [21] and personality [3]. In the job interview setting, the work in [4] used a corpus consisting of 36 participants, extracted verbal cues used LIWC and a Doc2Vec method. Using Pearson's correlation r as their evaluation measure, they reported $r = 0.39$ with Ridge regression. For comparison, converting r to R^2 (by computing the square of correlation coefficient r) indicates $R^2 = 0.15$ which is higher than our results.

In comparison, the model trained on nonverbal cues performs better ($R^2 \in [0.21, 0.35]$) for the same dataset. Combining the nonverbal and verbal cues leads to an marginal increase in performance of inference for some social variables like *Overall Impression* ($R^2 = 0.34$), *Concise* ($R^2 = 0.26$) and all the professional skills ($R^2 \in [0.26, 0.27]$) indicating that verbal components adds some information. The work in [14] investigated words used and nonverbal behavior displayed in a job interview setup using college students. They examined a different set of social variables and used correlation coefficient r as their evaluation measure. They too used LIWC for extracting lexical features but then applied LDA to learn common topics in the data corpus. By combining lexical and nonverbal features, the authors reported a prediction accuracy of $r = 0.70$ for *Overall Performance*, which indicates a $R^2 = 0.49$ compared to our $R^2 = 0.34$. This dataset is not publicly available (to the best of our knowledge) and thus, there is no direct way to assess the performance difference.

To understand the contributions of each of the feature sets, we determine the top 20 features used by RF model, presented in Table 4. We observe that while most of the features are nonverbal cues (*Speaking Energy, Turn Duration, Silence Events etc*), some verbal cues like *Question, Word Count, Proper pronouns* also contribute to the inference of *Overall Impression*, indicating that verbal cues add albeit marginally to improved inference.

Inference of Principal Components

We define a second regression task with the aim of predicting the first three PCs which account for 94.3% of the variance in the annotation data. The best inference performance was achieved by using RF (Table 5). We observe that predicting the first PC using nonverbal cues achieves $R^2 = 0.41$ which is better than the performance for all the individual social variables. Essentially, this predicts positive or negative impression.

Applicant Features	
Speaking Energy	Max Energy Derivative
Energy Derivative Lower Quartile	Min Energy Derivative
Avg Speaking Energy	Speaking Energy Lower Quartile
Avg Speaking Turn Duration	Speaking Energy Upper Quartile
Silence Ratio	Energy Derivative Upper Quartile
Number of Silence Events	Max Turn Duration
Max Silence Duration	Questions
Number of Speaking Turns	Word Count
Number of Head Nods	3rd Person Singular
Max Speaking Energy	Proper Pronouns

Table 4: **List of top 20 features from RF regression model ($N = 169$) for *Overall Impression* (left:$1-10$; right:$11-20$)**

Cues	PC1	PC2	PC3
Nonverbal	0.41	0.04	0.01
Verbal	0.12	0.02	0.02
Nonverbal + Verbal	0.34	0.02	0.02

Table 5: **Inference of PC using RF and verbal, nonverbal and combining both cues as predictors with $p < 0.05$.**

Similarly, using verbal components we can infer only up to 0.12. This suggests that use of PCs removes some of the noise in the annotations data leading to slightly improved inference. We also observe that the second and third PC are not recognizable, likely due to the little variance (6.1% and 5.6% respectively) they account for.

CONCLUSION

This work studied the possible links between linguistic style and impressions in the context of employment interviews for hospitality students. Towards this, we utilized a data corpus consisting of 169 interviews. To understand the connection between linguistic content and impressions, the interactions were first manually transcribed. Then, verbal cues were extracted using LIWC, which links linguistic and paralinguistic categories to psychological constructs. A correlation analysis between use of words and impression scores provided interesting insights into the weak effect of linguistic content on impressions, a result in line with some existing literature. An inference of impressions scores defined as a regression task showed that verbal features had low performance as compared to nonverbal cues, indicating the importance of the latter in a structured job interview context. We then assessed the underlying structure of the annotations using principal component analysis. The first PC accounted for more that 82% of the variance and was found to distinguish the overall impression. Using this PC as labels in a regression task showed performance of $R^2 = 0.41$.

ACKNOWLEDGMENTS

This work was funded by the UBImpressed project of the Sinergia program of the Swiss National Science Foundation (SNSF). We thank Marianne Schmid Mast (UNIL) for discussions, and Denise Frauendorfer (UNIL) and the research assistants for their support with speech transcriptions.

REFERENCES

1. Murray R Barrick, Jonathan A Shaffer, and Sandra W DeGrassi. 2009. What you see may not be what you get: relationships among self-presentation tactics and ratings of interview and job performance. *J. Applied Psychology* 94, 6 (2009), 1394.

2. Joan-Isaac Biel and Daniel Gatica-Perez. 2013. The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Trans. on Multimedia* 15, 1 (2013).

3. Joan-Isaac Biel, Vagia Tsiminaki, John Dines, and Daniel Gatica-Perez. 2013. Hi youtube!: Personality impressions and verbal content in social video. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 119–126.

4. Lei Chen, Gary Feng, Chee Wee Leong, Blair Lehman, Michelle Martin-Raugh, Harrison Kell, Chong Min Lee, and Su-Youn Yoon. 2016. Automated scoring of interview videos using Doc2Vec multimodal feature extraction paradigm. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 161–168.

5. Cindy Chung and James W Pennebaker. 2007. The psychological functions of function words. *Social communication* (2007), 343–359.

6. Jared R Curhan and Alex Pentland. 2007. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *J. Applied Psychology* 92, 3 (2007).

7. Timothy DeGroot and Janaki Gooty. 2009. Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *J. Business and Psychology* 24, 2 (2009).

8. Allen I Huffcutt, James M Conway, Philip L Roth, and Nancy J Stone. 2001. Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *J. Applied Psychology* 86, 5 (2001).

9. Hayley Hung and Daniel Gatica-Perez. 2010. Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia* 12, 6 (2010), 563–575.

10. Andrew S Imada and Milton D Hakel. 1977. Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *J. Applied Psychology* 62, 3 (1977).

11. Mark Knapp, Judith Hall, and Terrence Horgan. 2013. *Nonverbal communication in human interaction*. Cengage Learning.

12. Max Kuhn. 2016. A Short Introduction to the caret Package. (2016).

13. Skanda Muralidhar, Laurent Son Nguyen, Denise Frauendorfer, Jean-Marc Odobez, Marianne Schimd-Mast, and Daniel Gatica-Perez. 2016. Training on the Job: Behavioral Analysis of Job Interviews in Hospitality. In *Proc. 18th ACM Int. Conf. Multimodal Interact.* 84–91.

14. Iftekhar Naim, M Iftekhar Tanveer, Daniel Gildea, and Mohammed Ehsan Hoque. 2015. Automated prediction and analysis of job interview performance: The role of what you say and how you say it. *Proc. IEEE FG* (2015).

15. Laurent Son Nguyen. 2015. *Computational Analysis Of Behavior In Employment Interviews And Video Resumes*. Ph.D. Dissertation. École Polytechnique Fédérale de Lausanne.

16. Laurent Son Nguyen, Denise Frauendorfer, Marianne Schmid Mast, and Daniel Gatica-Perez. 2014. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Trans. on Multimedia* 16, 4 (2014).

17. Laurent Son Nguyen and Daniel Gatica-Perez. 2015. I Would Hire You in a Minute: Thin Slices of Nonverbal Behavior in Job Interviews. In *Proc. ACM ICMI*.

18. Richard E Nisbett and Timothy D Wilson. 1977. The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology* 35, 4 (1977), 250.

19. James W Pennebaker and Laura A King. 1999. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology* 77, 6 (1999), 1296.

20. Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2010. Identifying emergent leadership in small groups using nonverbal communicative cues. In *Int Conf. on Multimodal interfaces and the workshop on machine learning for Multimodal Interaction*. ACM, 39.

21. Dairazalia Sanchez-Cortes, Joan-Isaac Biel, Shiro Kumano, Junji Yamato, Kazuhiro Otsuka, and Daniel Gatica-Perez. 2013. Inferring mood in ubiquitous conversational video. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 22.

22. Dairazalia Sanchez-Cortes, Petr Motlicek, and Daniel Gatica-Perez. 2012. Assessing the impact of language style on emergent leadership perception from ubiquitous audio. In *Proceedings of the 11th international conference on mobile and ubiquitous multimedia*. ACM, 33.

23. Patrick E Shrout and Joseph L Fleiss. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86, 2 (1979).

24. Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

25. Svante Wold, Kim Esbensen, and Paul Geladi. 1987. Principal component analysis. *Chemometrics and intelligent laboratory systems* 2, 1-3 (1987), 37–52.