Analyzing the Impact of Gender on the Automation of Feedback for Public Speaking

Astha Singhal[1], Mohammad Rafayet Ali[2], Raiyan Abdul Baten[2]
[1]University of Maryland, College Park
[2]University of Rochester
United States

Chigusa Kurumada[2], Elizabeth West Marvin[2], and Mohammed (Ehsan) Hoque[2]
[2]University of Rochester
United States

***Abstract*— This paper explores gender differences in the evaluation of male and female speakers' affective features in public speaking. We analyzed 260 two-minute behavioral videos (200 of females and 60 of males), collected from an online public speaking practice tool. We adopted a linear regression model that utilized facial and prosodic features, including facial action units (AU), word count, pitch, and volume, to automatically assess speaker performance. The model was evaluated against ratings from 2 expert speakers from Toastmasters, an international public speaking club, on speaker performance. Our feature analysis suggests that certain combinations of features are correlated with higher ratings only in males, such as the combined increase of speech rate and vocal pitch variation. Moreover, our clustering analysis suggests that exhibiting certain negative emotions correlates with higher ratings for males but not for females, illustrating the impact of gender in generating effective feedback on public speaking.**

I. INTRODUCTION

Previous research has established that the fear of public speaking is common among Americans [19,29]. It has been proposed that automated evaluation tools can help speakers improve their public speaking skills and eventually overcome the fear itself [6]. However, providing effective computer-generated recommendations on a multi-dimensional social behavior such as public speaking is a relatively uncharted territory. For instance, Big Interview [20] does not provide feedback for improvement at all, and Interview Coach relies on human instructors to provide personalized feedback [27]. MACH gives its users automatic feedback on visual and prosodic features in the context of a job interview but does not provide concrete suggestions for improvement [18]. In the broader realm of public speaking recommendations, little is known about how to adjust feedback strategies for differences in the users' socio-cultural parameters, such as gender or age [22,33,34]. Given common gender stereotypes (e.g., women smile more than men), the same physical gesture can have different interpretations depending on the speaker's gender [15]. It is therefore worthwhile to include a speaker's gender as a factor towards building an effective automated feedback system.

In this paper, we explore gender-dependent effects of facial and prosodic features on performance ratings in public speaking. We analyze our observations against gender stereotypes found in past literature and discuss how automatic feedback systems might utilize these insights. We present the analyses of affective features in 260 behavioral videos, gathered through a publicly available semi-automated public speaking practicing platform, ROCSpeak [17]. Participants were asked to upload two-minute videos answering five of the most common behavioral questions, as suggested by the Career Center of a private research university in the United States. Two female experts from the local Toastmasters club then independently evaluated the videos. Though the experts' gender can be a factor in affecting their ratings, it is controlled for all the participants and hence is not analyzed further in the current paper. The evaluation metrics included one composite overall performance score and seven other categories (Table 1), designed to assess participants' speaking performance on a 7 point Likert scale.

To assess their performance automatically, we extracted several prosodic and facial features (facial action units) from the videos and applied a linear regression model. Our model was able to predict the human-assigned rating scores with a root mean squared error (RMSE) of 1.03. The RMSE score is negatively affected by a rather poor performance on predicting ratings in one category (Speaker's explanation of the concept.) This, however, is to be expected because the features examined were closely related to speaker's prosodic features, and not directly relevant to the content of the speech. Excluding this category improves the classifier's RMSE to 0.529.

Our analysis identified features that are strongly predictive of higher ratings on speakers' performance. Interestingly, these features differed significantly between male and female speakers. Overall, females had a greater average smile intensity compared with males and males spoke at a faster rate. Our results from multivariate correlation analysis indicated that male speakers who spoke at a faster rate and smiled more were rated higher ($r = 0.334$, $p < 0.001$). Again, high speech rate and average pitch variation also resulted in higher ratings for male speakers ($r = 0.424$, $p < 0.001$). This correlation was weak in female speakers ($r < 0.3$).

We further attempted to identify the facial expressions associated with higher ratings. We found features that were predictive of higher ratings regardless of the participants' gender (e.g., smiling more, having happier facial expressions, and showing less fearful expressions). In addition, we found some gender-specific effects, such as the fact that male speakers received higher performance ratings when they demonstrated frequent instances of Action Unit (AU) 04 (Brow Lowerer) and AU20 (Lip

978-1-5386-2335-0/18/$31.00 ©2018 IEEE

Stretcher), which are primarily associated with sadness when combined. In contrast, female speakers who demonstrated more of these AUs associated with sadness received lower performance scores. This suggests that though it was better for both genders to smile more, demonstrating negative emotions through facial features is more advantageous for male speakers than for female speakers. This motivates further probing into customized recommendations on speaking performances depending on gender.

II. Background and Related Work

A. Automated Feedback Tools

The field of Social Signal Processing (SSP) [30,38] explores the nonverbal cues (e.g., gestures and postures [32], eye movements, facial expressions [39,26,40], and vocal characteristics [2]) that communicate social meanings. Harnessing the progress in the SSP field, many attempts have been made to create automated systems to help people develop effective speaking skills. For instance, "Logue" [7] facilitates the awareness of users' non-verbal behaviors. "AwareMe" [5] utilizes a detachable wristband to give feedback on pitch, use of filler words, and words per minute. "Rhema" [35] gives real-time feedback on speech rate and volume using Google Glass. "AutoManner" [36] makes speakers cognizant of the idiosyncrasies of their body language. "MACH" [18] utilizes a reactive 3D avatar to allow anyone to practice for job interviews, with visual feedback provided at the end. Though these systems vary in the devices and algorithms they employ and the feedback they generate, none of these customizes its feedback based on gender differences, opening up a potential scope for improvement in automated feedback systems.

B. Gender Differences in Job Interviews

Differences in speech styles between men and women have been widely studied and documented [4,9,10,14,22,34]. Some differences have been attributed to social inequalities between and stereotypes of the two genders [23]. Typically, women are expected to be helpful, supportive, and concerned for the well-being of others, while men are expected to be assertive, competitive, and goal-oriented [8]. These stereotypes shape the social norms and expectations in speaking, with low-pitched voices being considered as more masculine, and high-pitched voices more feminine [16].

In a behavioral analysis of job interviews [28], interview outcomes were better predicted for male participants than for female participants. This was primarily because men's outcomes had a higher correlation with standard numerical measurements such as speaking time, turn duration, speech energy, and silence. In fact, psychologists often refer to these characteristics as powerful speech [13]. Male speakers who used more of these cues were more likely to be perceived as powerful and persuasive, which predicted their success in job interviews [24]. In contrast, the same set of features were not as straightforward in predicting outcomes for female speakers.

This power dynamic plays a large role in public speaking stereotypes on the whole. Because men are typically regarded as being more powerful and assertive [33], many of their speaking stereotypes involve displaying it. They are often described to be louder and more blunt, and subdue in emotive and affective states [4,22]. Females, on the other hand, are considered to be less direct and conservative with more expressions to convey their emotions. Thus, females are generally considered to have faster and more enthusiastic speech, smile more, and be more emotionally expressive [4,22]. In this paper, we take an exploratory approach towards understanding the effects of these gender stereotypes on ratings of public-speaking performances, using spontaneous interview data collected from online workers.

III. Data Collection

260 videos from 52 independent speakers (12 males and 40 females) were collected via an online public speaking practicing platform, ROCSpeak (available at www.rocspeak.com). This data was amassed from a randomized control study conducted by the researchers. The participants were recruited from Amazon Mechanical Turk for this 10-day study. Every other day, the participants were given a prompt. The five prompts were — *(1) Tell me about yourself, (2) Describe your biggest weakness, (3) Tell me about your greatest achievement, (4) Describe a conflict or challenge you faced, and (5) Tell me about yourself.* The repeated prompt (i.e., (1) and (5)) was used to assess the improvement of speaking skills over time. Using the ROCSpeak system, the participants recorded 2-minute videos in response to each of the prompts and subsequently received subjective feedback from other participants on how to improve their performance. In addition to the subjective feedback from

TABLE 1. THE MEANS, STANDARD DEVIATIONS, AND KRIPPENDORFF'S ALPHA OF EACH RATING CATEGORY

Rating Categories	Mean	SD	α
Speaker's overall performance.	5.14	0.80	0.36
I'd like to see this person speak again	4.20	0.83	0.23
Speaker's friendliness.	5.17	0.99	0.43
Speaker's eye contact.	5.55	0.89	0.51
Speaker's body gestures.	3.95	1.26	0.42
Speaker's vocal variety.	5.23	0.89	0.33
Speaker's articulation.	5.57	0.75	0.14
Speaker's explanation of the concept.	5.47	0.76	0.12

TABLE 2. THE RMSE OF THE LINEAR REGRESSION FOR EACH RATING CATEGORY

Rating Categories	RMSE
Speaker's overall performance.	0.01
I'd like to see this person speak again	0.67
Speaker's friendliness.	0.87
Speaker's eye contact.	0.12
Speaker's body gestures.	0.89
Speaker's vocal variety.	0.78
Speaker's articulation.	0.36
Speaker's explanation of the concept.	4.52

Figure. 1. Five of the highest and lowest weights of features from the linear regression model in the "See Again" category. These features held the largest influence in determining the model's predicted score in this category.

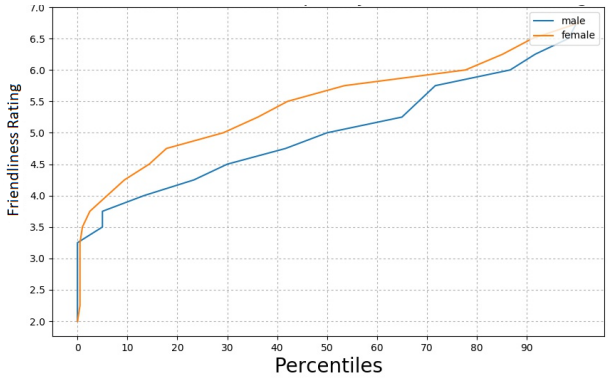

Figure. 2. The Cumulative Relative Frequency Graph of Friendliness Rating Category. In this category, women had a significantly higher score than men.

peers, the automatically extracted facial and prosodic features were shown to the treatment group. The control and treatment conditions were assigned randomly with equal probability.

We recruited two experts from Toastmasters to assess the performance of the speakers. They examined all of the videos and rated them in eight categories using a 7-point Likert scale. Table 1 summarizes the categories with the mean and standard deviation of the ratings, independent of gender.

IV. Analysis

In our analysis, we first attempted to automatically predict the performance scores provided by the human raters using a machine-learning model. Because this model predicts scores without taking gender into account, we investigated the gender differences between speakers to further refine the classifier and lower the RMSE values. In doing so, we identified styles of effective speaking by pairing AUs, a method of modifying the activation of facial muscles [12] and performing a cluster analysis.

A. Feature Collection

To extract the features, we used *Praat* [3] for audio analysis, *Openface* [1] for facial feature extraction, and Google's speech recognizer [31] for transcript generation. With *Praat*, we extracted statistics regarding volume, pitch, intensity, and pauses. We utilized Ekman and Friesen's Facial Action Coding System (FACS) to examine facial expressions in the videos [11,12,38]. An open source framework called *OpenFace* was used to extract the Action Units (AUs) from each frame of the videos. *OpenFace* uses a 0/1 classification scheme to indicate whether an AU is present in a frame, and also gives an indicator of AU intensity in the range of 1-5. We additionally computed word counts and speech rates from the automatically generated transcript. For each of the videos we took the average and standard deviation of each features.

B. Rating Prediction

We employed several machine-learning techniques including linear regression, support vector regression, AdaBoost classifier, and k-nearest neighbors to automatically predict the rating scores. After performing a 10-fold cross validation, we found that linear regression was performing the best (lowest RMSE on the test sets). Table 2 lists the RMSE values obtained from our linear regression model. This illustrates the differences between the scores predicted by the model and the scores given by the experts. One category (Speakers' explanation of concept) resulted in a noticeably high RMSE value. This is presumably due to the fact that the features examined in the model were audio-visual, and not directly tied to the content of the speech.

We examined the weights that the classifier assigned to identify the features with the greatest influence on the model's predictions. The classifier found different visual features (in particular, AUs) to be the most salient among all features and rating categories. Some of these features are shown in Figure 1. This precipitated a more careful and detailed analysis of the features and ratings themselves, focusing on the gender of the speaker to build a better feedback system.

C. Gender Differences

We analyzed the facial movement of participants in relation to their gender, to see which movements were more helpful for men or women exclusively. To explore the gender-dependent effects of audio-visual features, we modeled the expert ratings with the gender of the speakers. Since the majority of the videos were from female speakers, in each analysis, we sampled 60 videos randomly multiple times from the female pool to match the number of male videos.

In examining the rating categories, we found a statistically significant difference between males and females in 4 of the 8 rating categories: *I'd like to see this person speak again*, *friendliness, eye contact*, and *body gestures*. Figure 2 depicts the cumulative relative frequency of friendliness ratings. We know that at the 50th percentile, half of the women have a score of approximately 5.6 or below, whereas half of the men have a score of 5 or below, implying that on the whole, women have higher scores than men. This conclusion holds under statistical analysis as well. The significance was measured using a two-sample unpaired t-test for $\alpha = 0.05$, adjusted for the number of categories. Interestingly, while women

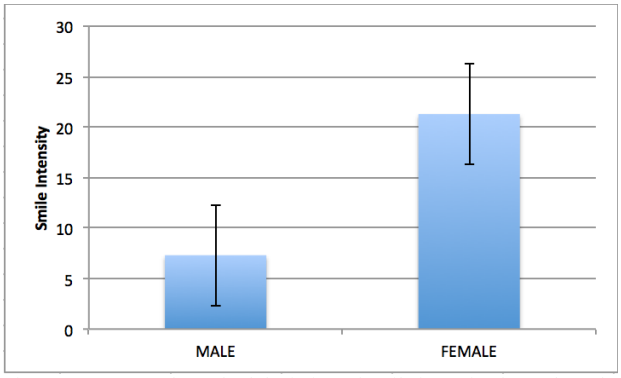

Figure. 3. Mean and standard deviation of smile intensity for male and female speakers.

were outperforming men in 4 of 8 categories, there was no statistically significant difference in the overall performance category, hinting that men and women employ different strategies to enhance their public speaking ability.

To better isolate the differences between male and female speaking styles, we compared different audio-visual characteristics by gender. Since females had higher friendliness scores than males, we looked more closely at the smile intensity feature to see if there were any significant differences. Indeed, as Figure 3 indicates, smile intensity was significantly higher ($t(118) = -5.42$, $p < 0.05$) for female speakers. Additionally, the correlation between the smile intensity and the friendliness score for females supports the finding that smiling more improves friendliness scores ($r = 0.42$, $p < 0.001$). On the other hand, men had a statistically higher speech rate of 110.439 WPM versus the female average of 98.065 WPM ($t(118) = 1.58$, $p < 0.001$).

We then evaluated the possible multivariate correlations between the features and the ratings. None of the features individually showed a strong correlation with the performance ratings. However, when speakers' gender was included in the model, we found that speech rate and smile intensity had a moderate correlation with overall performance ($r = 0.33$, $p < 0.001$) in males, along with high speech rate and average voiced pitch with overall performance ($r = 0.42$, $p < 0.001$). This means that when males both spoke faster and smiled more, their overall performance scores tended to be higher. Similarly, when males' pitch variation increased in conjunction with their speech rate, their overall performance scores tended to be higher as well. As speaking faster, smiling more, and speaking with a higher pitch are stereotypically female public speaking traits [22], it was interesting to note that men who adopted these more "female" speaking characteristics received higher performance scores.

D. Identifying Speaking Styles

From our analyses of the individual features and the correlations, it was evident that providing identical feedback to males and females may not be effective. Research has been done to define the different facial expressions, such as happy, sad, fearful, or angry, using AUs [21,25]. In particular, Kohler *et al.* found distinct clustering of pairs of AUs in different facial expressions,

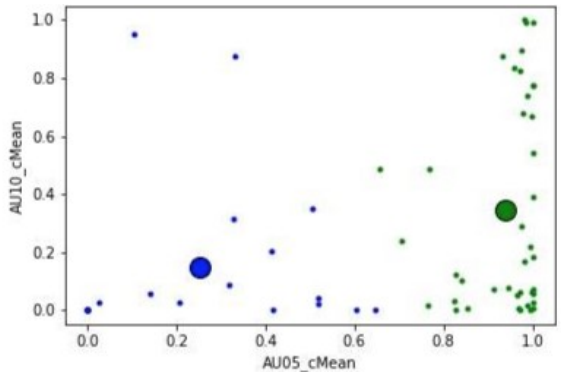

Figure. 4. Two styles of male speakers. Blue cluster has higher overall rating, but only minimally. There is no significant difference between either cluster. The larger circles show the cluster center and smaller circles represent individual speakers.

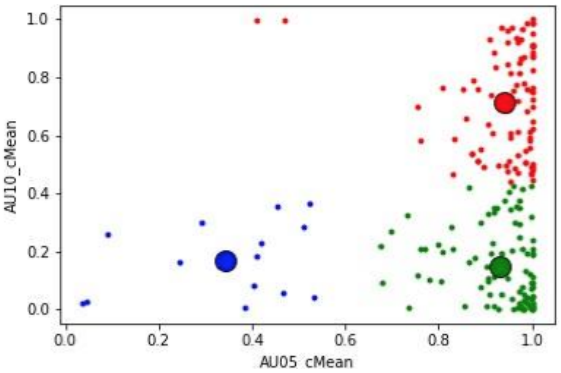

Figure. 5. Three styles of female speakers. The red cluster has higher overall rating, whereas the green cluster has the lowest ratings. The red and green clusters differ by 0.7695. The larger circles show the cluster center and smaller circles represent individual speakers.

indicating that even across different people, the same pairs of AUs are involved to convey happiness, fear, anger, and sadness [21].

We examined AU pairs and grouped the speakers who expressed similar amount of activation by calculating the average counts of an AU's presence per video. The grouping was performed using the k-means clustering algorithm and the number of groups (clusters) was identified by maximizing the Silhouette score [35]. We only analyzed those results when the identified clusters showed a significant difference in overall ratings.

An interesting pattern emerged when splitting the clustering analysis by gender. Though the distributions for male and female videos in terms of the presence of AU05 (Upper Lid Raiser) and AU10 (Upper Lip Raiser) are fairly similar, the clustering analysis reveals the futility of giving the same feedback to both males and females. A higher activity of this two AUs combined can be related to a 'fear' facial expression. As Figure 4 indicates, though there were 2 groups of men who exhibited distinct amounts of AU05 and AU10, their overall performance ratings were indistinguishable. Thus, for males, one could say that the amount of AU05 and AU10 exhibited does not matter in terms of their public speaking performance. However, this was not the case for females, where there was a large difference of 0.77 between exhibiting more AU10 versus less AU10 while keeping AU05 roughly constant. It would be impractical to give males the advice

of increasing their expression of AU10, whereas for women it could potentially yield better performance scores. These observations formed the basis for our further analysis.

However, just giving recommendations on individual action units is not practical because it would be non-intuitive for a user to simply stretch their lips more or to furrow their brows less while speaking. Therefore, we attempted to provide feedback at the level of the speaker's emotions and their correlates of their facial expressions. By examining the AUs specifically associated with certain expressions, we found that males and females both received higher ratings when exhibiting happier facial expressions and less fearful ones. Happy expressions often contain clusters of different AU pairs, including AU06 (Cheek Raiser), AU12 (Lip Corner Puller), and AU07 (Lid Tightener) [21]. For both men and women, displaying increased amounts of AU06 and AU12, as well as AU07 and AU12, resulted in higher ratings, suggesting that both men and women can improve their public speaking by smiling more or displaying a happier face. Similarly, we found that for both men and women, it was better to express less of AU05 (Upper Lid Raiser) and AU26 (Jaw Drop), associated with fearful expressions.

Demonstrating expressions related to sadness differentially affected performance scores for males and females. We found that for women, it was effective to express less AU04 (Brow Lowerer) and AU07 (Lid Tightener) together, which are associated with sadness. The same cannot be said for men. Men were perceived to be better speakers when more frequently displaying sadness than females who did the same. Thus, we found that expressing similar facial expressions can be either effective or detrimental depending on one's gender.

V. Discussion and Limitations

Our clustering analysis was more or less consistent with past literature discussing gender stereotypes. It has been argued that it is advantageous for females to have a happier demeanor, as women are generally expected to be more cheerful and expressive. Our results support this: women received better performance ratings when they expressed AUs associated with happiness more frequently. Similarly, being fearful does not align itself with the typical alpha male stereotype, and our results supported that breaking away from this stereotype would lead to lower scores. Expressing the AUs associated with fear was found to lead to worse performance ratings in males. Additionally, perhaps because sad facial expressions can be construed as a sign of seriousness, male conformity to gender roles also increased their public speaking scores in that case as well. Thus, overall, aligning with gender stereotypes led to better public speaking performance. However, this was not always the case. Although women were considered to be less confident public speakers in the past literature, exhibiting more fearful expressions or shyness would not be to their benefit. In this sense, there is a potential advantage associated with breaking away from gender stereotypes. Nevertheless, it bears noting that Kohler *et al.* only focused upon happiness, sadness, fear, and anger in their analysis, and not seriousness nor shyness [21].

Some caveats and limitations of our analyses are in order. Each of the videos are 2-minute-long, which does not fully represent features determining successful performances of public speaking. Perhaps the genre examined here is closer to so-called elevator pitch. Also, we need more male speakers in our data set to perform the analyses more accurately. In this paper, we tried to avoid this skewness by sampling videos from female speakers. Collection and analysis of longer, gender-balanced, video data will be necessary in future work. Additionally, the two expert raters were female, which might have some biases in the ratings. In the future, we plan to recruit both male and female experts to evaluate the videos. Finally, in our current analysis, we did not consider the verbal content of the speech. It is critical to analyze verbal contents using advanced natural language processing tools.

VI. Conclusions

In this paper, we identified prosodic and facial features that predicted human ratings of public speaking in a job interview context. We focused on gender-based effects of such features. Importantly, conforming to traditional gender norms and stereotypes is often, but not always, key to effective performances. Overall, it was better for both males and females to smile more, and not be visibly scared. However, men were rated more successful when exhibiting a higher speech rate and smile intensity, as well as a higher speech rate and average pitch variance, both of which are stereotypically considered to be more female. Similarly, the demonstration of negative emotions, associated with sadness, was found to be beneficial only for male speakers. Since the sample included fewer male speakers than female speakers, much work needs to be done to conclusively determine how feedback mechanisms can be tailored for male and female speakers. Nevertheless, our findings strongly suggest that providing generic feedback for male and female speakers does not necessarily result in a comparable improvement. Our observation holds promise to help researchers improve their feedback and intervention strategies in the future.

Acknowledgement

This work was supported by National Science Foundation Award #IIS-1464162 and NSF REU Site Award #1659250. We thank Ethan Cole for his contributions to creating the linear regression classifier.

References

[1] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. 2016. *OpenFace: A general-purpose face recognition library with mobile applications.* Technical Report. CMU-CS-16-118, CMU School of Computer Science.

[2] Sankaranarayanan Ananthakrishnan and Shrikanth S Narayanan. 2008. Automatic prosodic event detection using acoustic, lexical, and syntactic

evidence. *IEEE Transactions on Audio, Speech, and Language Processing* 16, 1 (2008), 216–228.

[3] Paul Boersma and David Weenink. 2017. Praat, a system for doing phonetics by computer [Computer program], Version 6.0.32. (2017). Retrieved September 17, 2017 from http://www.praat.org/.

[4] Nancy J Briton and Judith A Hall. 1995. Beliefs about female and male nonverbal communication. *Sex Roles* 32, 1 (1995), 79–90.

[5] Mark Bubel, Ruiwen Jiang, Christine H Lee, Wen Shi, and Audrey Tse. 2016. AwareMe: addressing fear of public speech through awareness. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 68–73.

[6] Lei Chen, Gary Feng, Jilliam Joe, Chee Wee Leong, Christopher Kitchen, and Chong Min Lee. 2014. Towards automated assessment of public speaking skills using multimodal cues. In *Proceedings of the 16th International Conference on Multimodal Interaction*. ACM, 200–203.

[7] Ionut Damian, Chiew Seng Sean Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 565–574.

[8] Alice H Eagly and Steven J Karau. 2002. Role congruity theory of prejudice toward female leaders. *Psychological Review* 109, 3 (2002), 573.

[9] Penelope Eckert and Sally McConnell-Ginet. 1999. New generalizations and explanations in language and gender research. *Language in Society* 28, 2 (1999), 185–201.

[10] Penelope Eckert and Sally McConnell-Ginet. 2003. *Language and Gender.* Cambridge University Press.

[11] Paul Ekman and Wallace V Friesen. 1976. Measuring facial movement. *Environmental Psychology and Nonverbal Behavior* 1, 1 (1976), 56–75.

[12] Paul Ekman and Wallace V Friesen. 1978. FACS - Facial Action Coding System. (1978). Retrieved September 17, 2017 from https://www.cs.cmu.edu/~face/facs.htm.

[13] Bonnie Erickson, E Allan Lind, Bruce C Johnson, and William M O'Barr. 1978. Speech style and impression formation in a court setting: The effects of "powerful" and "powerless" speech. *Journal of Experimental Social Psychology* 14, 3 (1978), 266–279.

[14] Anita Esposito. 1979. Sex differences in children's conversation. *Language and Speech* 22, 3 (1979), 213–220.

[15] Coreen Farris, Teresa A Treat, and Richard J Viken. 2010. Alcohol alters men's perceptual and decisional processing of women's sexual interest. *Journal of Abnormal Psychology* 119, 2 (2010), 427.

[16] David R Feinberg, Lisa M DeBruine, Benedict C Jones, and Anthony C Little. 2008. Correlated preferences for men's facial and vocal masculinity. *Evolution and Human Behavior* 29, 4 (2008), 233–241.

[17] Michelle Fung, Yina Jin, RuJie Zhao, and Mohammed Ehsan Hoque. 2015. ROC speak: semi-automated personalized feedback on nonverbal behavior from recorded videos. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1167–1178.

[18] Mohammed Ehsan Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W Picard. 2013. Mach: My automated conversation coach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 697–706.

[19] Christopher Ingraham. 2014. America's top fears: Public speaking, heights and bugs. (30 October 2014). Retrieved September 17, 2017 from https://www.washingtonpost.com/news/wonk/wp/2014/10/30/clowns-are-twice-as-scary-to-democrats-as-they-are-to-republicans/?utm_term=.e1a1cf6c186a.

[20] Big Interview. Simple Software for Better Interview Skills. (n.d.). Retrieved September 17, 2017 from https://biginterview.com/.

[21] Christian G Kohler, Travis Turner, Neal M Stolar, Warren B Bilker, Colleen M Brensinger, Raquel E Gur, and Ruben C Gur. 2004. Differences in facial expressions of four universal emotions. *Psychiatry Research* 128, 3 (2004), 235–244.

[22] Cheris Kramer. 1977. Perceptions of female and male speech. *Language and Speech* 20, 2 (1977), 151–161.

[23] Robin Lakoff. 1973. Language and woman's place. *Language in Society* 2, 1 (1973), 45–79.

[24] Campbell Leaper and Rachael D Robnett. 2011. Women are more likely than men to use tentative language, aren't they? A meta-analysis testing for gender differences and moderators. *Psychology of Women Quarterly* 35, 1 (2011), 129–142.

[25] James Jenn-Jier Lien, Takeo Kanade, Jeffrey F Cohn, and Ching-Chung Li. 2000. Detection, tracking, and classification of action units in facial expression. *Robotics and Autonomous Systems* 31, 3 (2000), 131–146.

[26] Gwen Littlewort, Jacob Whitehill, Tingfan Wu, Ian Fasel, Mark Frank, Javier Movellan, and Marian Bartlett. 2011. The computer expression recognition toolbox (CERT). In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 298–305.

[27] Carole Martin. Interview Coach. (n.d.). Retrieved September 17, 2017 from http://interviewcoach.com/.

[28] Skanda Muralidhar, Laurent Son Nguyen, Denise Frauendorfer, Jean-Marc Odobez, Marianne

Schmid Mast, and Daniel Gatica-Perez. 2016. Training on the job: Behavioral analysis of job interviews in hospitality. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 84–91.

[29] Chapman University Survey of American Fears. 2016. America's Top Fears 2016. (11 October 2016). Retrieved September 17, 2017 from https://blogs.chapman.edu/wilkinson/2016/10/11/americas-top-fears-2016/.

[30] Alex Pentland. 2007. Social signal processing [exploratory DSP]. *IEEE Signal Processing Magazine* 24, 4 (2007), 108–111.

[31] Google Cloud Platform. 2017. Cloud Speech API. (2017). Retrieved September 17, 2017 from https://cloud.google.com/speech/.

[32] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. 2013. Real-time human pose recognition in parts from single depth images. *Commun. ACM* 56, 1 (2013), 116–124.

[33] David M Siegler and Robert S Siegler. 1976. Stereotypes of males' and females' speech. *Psychological Reports* 39, 1 (1976), 167–170.

[34] Elizabeth A Strand. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18, 1 (1999), 86–100.

[35] Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. (2002, April 01). Retrieved September 19, 2017, from http://www.sciencedirect.com/science/article/pii/0377042787901257?via%3Dihub

[36] M Iftekhar Tanveer, Emy Lin, and Mohammed Ehsan Hoque. 2015. Rhema: A real-time in-situ intelligent interface to help people with public speaking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 286–295.

[37] M Iftekhar Tanveer, Ru Zhao, Kezhen Chen, Zoe Tiet, and Mohammed Ehsan Hoque. 2016. Automanner: An automated interface for making public speakers aware of their mannerisms. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 385–396.

[38] Y-I Tian, Takeo Kanade, and Jeffrey F Cohn. 2001. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 2 (2001), 97–115.

[39] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image and Vision Computing* 27, 12 (2009), 1743–1759.

[40] Jiabei Zeng, Wen-Sheng Chu, Fernando De la Torre, Jeffrey F Cohn, and Zhang Xiong. 2015. Confidence preserving machine for facial action unit detection. In *Proceedings of the IEEE International Conference on Computer Vision*. 3622–3630.

[41] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1 (2009), 39–58.