AUTOMATIC CHARACTERIZATION OF SPEAKING STYLES IN EDUCATIONAL VIDEOS

Soroosh Mariooryad[1*], *Anitha Kannan*[2], *Dilek Hakkani-Tür*[2], *Elizabeth Shriberg*[3,4†]

[1]The University of Texas at Dallas, [2]Microsoft Research,
[3]SRI International, [4]International Computer Science Institute, Berkeley

ABSTRACT

Recent studies have shown the importance of using online videos along with textual material in educational instruction, especially for better content retention and improved concept understanding. A key question is how to select videos to maximize student engagement, particularly when there are multiple possible videos on the same topic. While there are many aspects that drive student engagement, in this paper we focus on presenter speaking styles in the video. We use crowd-sourcing to explore speaking style dimensions in online educational videos, and identify six broad dimensions: liveliness, speaking rate, pleasantness, clarity, formality and confidence. We then propose techniques based solely on acoustic features for automatically identifying a subset of the dimensions. Finally, we perform video re-ranking experiments to learn how users apply their speaking style preferences to augment textbook material. Our findings also indicate how certain dimensions are correlated with perceptions of general pleasantness of the voice.

Index Terms— speaking style, speaking rate, liveliness

1. INTRODUCTION

In the context of education, the proliferation of low-cost and increasingly accessible tablet devices [1] present important new opportunities. Replacing traditional textbooks with tablet-based digital content can enable "anytime, anywhere learning" across the globe [2]. While many current electronic textbooks tend to be digital versions of their printed counterparts, recent research has shown the value of augmenting content with relevant supplementary materials including text, images and videos mined from the web [3, 4, 5, 6, 7, 8]. Web-based supplementary materials obviously need to be content appropriate, but all else being equal, they should also be as engaging as possible to the student using them. As multiple different online videos are often available to teach a particular concept, a question arises: how can they best be sorted based on a particular user's preferences?

In this study we assume that the set of relevant videos for a textbook is already identified (a research question in and of itself) and focus on the question of a user's preference for the speaking style in the video. We are interested in the presenter's speaking style because it is a dimension that cross-cuts domains, and one that has both a preference aspect as well as an impact on learning. For example some students may prefer a more serious tone, while others prefer an upbeat presentation. And for a topic that a student is having difficulty with, a presenter with a fast speaking rate may make it difficult to learn the material. The end goal is to allow users to sort retrieved videos according to their preference along the speaking style dimensions that we provide. As a first step to this end, we look at large sets of single-speaker short educational videos matched for various topics. We also restrict our features to those that can be extracted without using lexical information. While words are clearly an important component of style and should also be studied, acoustic features obviate the need for a speech recognizer and provide a useful starting point for style characterization.

A number of recent studies have focused on analyzing speaking styles. Using subjective evaluations, Rosenberg and Hirschberg [9] identified personal attributes that were highly correlated with perceived charisma. They also identified acoustic, prosodic and lexical features relevant for this task. In similar studies, Strangert [10]; Strangert and Gustafson [11] presented analyses to differentiate between good and bad speakers. The recent Interspeech paralinguistic challenge included a task on speaker likability as an attempt to provide predictive models for identifying good speakers [12, 13].

The question of preference for speaking style in educational videos is a new and open area. To explore the dimensions of user preference, we conduct a crowd-sourcing study to identify what types of information users report when listening to different presenters on the same topic (§ 2). In further studies we focus on two dimensions that show good agreement for crowd-based annotation: liveliness and speaking rate. We then look at how well a set of simple acoustic features can predict these dimensions (§ 3). We present extensive experiments using educational videos from the wild (*e.g.,* YouTube), relevant to sections in a textbook, to showcase the efficacy of our approach (§ 4). We also study how these two dimensions correlate with the pleasantness of the voice (§ 4).

2. DIMENSIONS OF SPEAKING STYLES

We first sought to identify speaking style dimensions that listeners perceive when viewing an educational video. To this end we performed a large-scale user study using *Amazon mechanical Turk* (AMT). Each *human intelligence task* (HIT) corresponded to a judge watching an educational video and reflecting on the speaker characteristics using three phrases of his/her own choice. Each HIT was judged by seven evaluators, and we obtained judgments for 100 unique educational videos. This experiment yielded a list of over 200 unique descriptors. Many of the descriptors appeared to reflect similar underlying characteristics. We manually binned these descriptors into six dimensions. Table 1 depicts these dimensions, examples from the corresponding descriptors and the proportions in which were referred by the evaluators. Speech liveliness (or lack thereof) plays the most significant role in these videos and judges described it using varied descriptors including boring, dull, dynamic, flat, monotonous, lively and toneless. The second most frequent dimension is pleasantness of the voice, which seems to be a function of other characteristics of the voice such as liveliness. Other frequently reported characteristics were voice clarity, formality, confidence, and speaking rate. Note that we will use the term "speaking rate" as a dimension (speaker is slow or fast) because it was used this way

[*]Work done when the author was an intern at Microsoft Research
[†]Work done when the author was at Microsoft Research

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

Table 1. Speaking style dimensions identified using large scale human judgments

Inferred Dimensions	Example descriptors used by the judges	Relative proportions
liveliness (dull vs. lively)	boring, dull, dynamic, flat, lively, monotonous, toneless	.26
pleasantness (pleasant vs. unpleasant)	awesome, good, ineffective, interesting, unappealing, unpleasant	.25
speaking rate (slow vs. fast)	fast, quick, slow, speedy	.14
clarity (unclear vs. clear)	blurry, clear, unclear	.14
formality (formal vs. casual)	bookish, casual, conversational, narrative	.12
confidence (hesitant vs. confident)	bold, confident, hesitant, insecure, unsure, wobbly	.10

by annotators–even though it is also used to refer to specific feature metrics (§ 3).

3. AUTOMATIC DETECTION OF SPEAKING STYLES

We explored methods to learn functions that score videos for two of the dimensions frequently identified by annotators: liveliness and speaking rate. In particular, we build two independent models that take an audio segment as input and score it for a speaker's liveliness and speaking rate. The ground truth Likert-scale ratings of each dimension is collected through subjective evaluations (§ 4.1). We used *least absolute shrinkage and selection operator* (LASSO) regression [14] as the mapping model. In particular, the regressor captures the relationship between features, $\phi(\mathbf{x})$, extracted from the segment $\mathbf{x}$ and speaking rate (or liveliness) scores y_i through the functional form $y = \mathbf{w}^T\phi(\mathbf{x})$, given some annotated training data $\mathbf{x}_i \in \mathcal{X}$. The parameters $\mathbf{w}$ are learned by solving the optimization function:

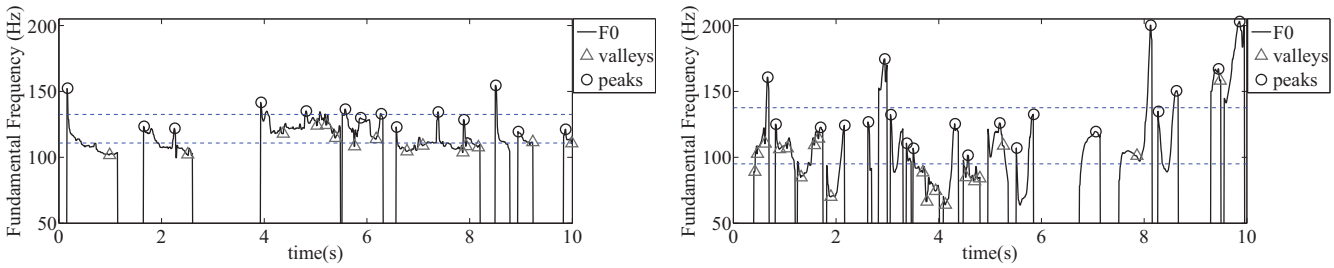
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \lambda\|\mathbf{w}\|_1 + \sum_i (\mathbf{w}^T\phi(\mathbf{x}_i) - y_i)^2 \qquad (1)$$

where λ is a regularization parameter that trades off model sparsity and squared error, which is set using 5-fold cross-validation in the training set. The unique optimum of Eq. 1 can be found via least angle regression. $\phi(\mathbf{x})$ corresponds to acoustic and prosodic features derived from the audio segment; these are summarized in Table 2 and described next. Since we have a large feature set, LASSO regression is used to avoid overfitting and also to perform feature analysis (§ 4.3).

Features: The first column in Table 2 summarizes the features used. To capture voice liveliness, we extracted a set of statistics from fundamental frequency (F0) and intensity contours representing their dynamic ranges. A peak finding algorithm [16] is used to extract the peaks and valleys in the contours. Figure 1 illustrates the detected valleys and peaks for a dull and a lively voice. The features include F0 mean peak to mean valley distance (i.e., the distance between dashed lines in Figure 1), mean/max jump from a peak/valley to the following peak, standard deviation of peaks in F0 contour, F0 distribution interquartile distances (i.e., $Q_2 - Q_1$ and $Q_3 - Q_2$) and mean/max jumps from valley to peak in intensity contour. Jump is defined as the distance along the vertical axis in the contours. F0 features representing dynamic range are normalized with respect to gender. Mean short time *discrete cosine transformation* (DCT) of the intensity contour is used to capture rhythmicity of the voice [17]. The amplitude of the first five DCT coefficients are extracted over 300-ms windows with 150-ms shift. Since voice liveliness can be attributed to higher vocal effort, we have used openSMILE [18] to extract the 50% spectral roll-off point. This is the frequency beyond which the accumulative signal energy exceeds 50% of total signal energy, representing the spectral tilt. For each audio segment, $1^{st} - centile$ (∼min), $99^{th} - centile$ (∼max) and mean value of 50% spectral roll-off point over the voiced segments are included in the feature set. Mean F0 segment length, number of F0 segments and mean distance between consecutive peaks in the intensity contour are potential features for capturing speaking rate. In addition to these low level features, we included estimations of speaking rate based on syllable detection by peak-counting in the energy contour, proposed by Jong and Wempe [15]. They defined speaking rate and articulation rate as the number of syllables in energy contour per time unit for entire utterance duration and entire duration excluding the silence, respectively. Our preliminary analysis showed that mrate [19] as a measure of speaking rate does not yield high correlation with the manual annotations in our corpus. This might be due to the noisy samples in the videos. Given the high correlation between the collected rates of liveliness and speaking rate (see Figure 2), we decided to include all the features in the experiments for both tasks.

Features	liveliness		speaking rate	
	%	+/-	%	+/-
Fundamental Frequency (F0)				
mean peak to mean valley distance	**95**	+	4	
interquartile distance (Q_2-Q_1)	**100**	+	33	
interquartile distance (Q_3-Q_2)	**99**	+	66	+
max peak to peak jump	0		28	
mean peak to peak jump	5		**100**	+
max valley to peak jump	0		1	
mean valley to peak jump	0		29	
standard deviation of peaks	**100**	+	**100**	+
mean segment length	71	-	76	+
number of segments	0		**99**	+
rising time ratio	**100**	+	27	
Energy				
mean short-term DCT_0	**100**	+	**100**	+
mean short-term DCT_1	0		29	
mean short-term DCT_2	0		19	
mean short-term DCT_3	0		16	
mean short-term DCT_4	0		63	+
mean of peak distance	3		**100**	+
standard deviation of peak distance	**100**	-	**100**	-
mean peak to mean valley distance	0		0	
mean peak to peak jump	1		**100**	+
max peak to peak jump	54	+	86	+
mean valley to peak jump	87	+	1	
max valley to peak jump	**100**	+	**95**	+
Spectral				
1^{st}-centile of 50% spect roll-off freq	**100**	+	**92**	+
99^{th}-centile of 50% spect roll-off freq	**100**	+	**100**	+
mean of 50% spectral roll-off point	**100**	+	**100**	+
Estimations of Speaking Rate[15]				
estimated articulation rate	**100**	+	**100**	+
estimated speaking rate	**0**		**100**	+

Table 2. Acoustic and prosodic features used. The percentage of time in which the feature was selected by the LASSO regression model across cross-validation folds and sign of the corresponding coefficients to predict liveliness and speaking rate.



(a) a dull speaker (b) a lively speaker

Fig. 1. Detected peaks and valleys in F0 contour for (a) a speaker rated as dull, and (b) a speaker rated as lively. Dashed lines show mean valley and peak values.

4. EXPERIMENTS

4.1. Setup

Data set: The training set consisted of educational videos relevant to ten sections corresponding to four chemistry chapters from a high school science textbook. The videos are mined from the web using a variant of the COMITY algorithm [4]. This resulted in a training set of 100 videos, 10 from each section. We similarly created a test set corresponding to three sections from a biology chapter in the same book. Since these videos are user-uploaded educational content created by self-appointed "teachers", they exhibit wide variability in terms of various dimensions of speaking styles, and have wide-ranging recording conditions, speaker accent, age and culture. Both males and females are represented. For both training and test sets, the only manual pre-processing we did is to remove videos that have background music or multiple speakers, resulting in 79 and 27 videos, respectively.

To regularize the rating task over videos differing in length, and to speed the annotation process, we created fixed-duration segments. From a listening study we determined that a length of about 20 seconds was sufficient for judging speaking style in this task, as long as segments did not contain long silent portions (while the speaker was writing, for example). We segmented each into 20-second segments such that each segment had no contiguous silent regions longer than two seconds. For each video, we randomly chose three non-boundary segments as its representative segments. In the rest of the paper, the entire audio is referred to as session and the 20-second snippets as segments. The models are trained on the segment-level data.

Human judgments: We designed a HIT on AMT to label each segment for liveliness and speaking rate on a discrete scale between zero and four, where zero was slow (or dull) and four was fast (or lively). Each HIT contained three audio segments from three randomly selected speakers. Our experimentation showed that presenting samples from multiple speakers helps the evaluators to calibrate the rates by comparison. Each HIT is evaluated by seven judges. We used similar approach to obtain judgments for the test set, but at the video and not the segment level. We asked evaluators to ignore the content of the videos in their ratings.

In order to capture the overall pleasantness of voice, we had each video judged by nine evaluators as to whether they like the speaker's voice in the video as a teacher. Then, majority of the votes is used to set the pleasantness label for the corresponding audio segment.

Figure 2 shows the average speaking rate and liveliness assigned to each audio segment in the training set. For both dimensions, most samples lie in the middle range indicating that a typical educational video has average speaking rate and liveliness. We can see high positive correlation between speaking rate and liveliness ($\rho = 0.66$), which is consistent with findings of Traunmuller and Eriksson [20]. However, the two dimensions capture different aspects of a speaker: While average speaking rate can be accompanied by high liveliness, average liveliness is less likely to be accompanied by high speaking rate.

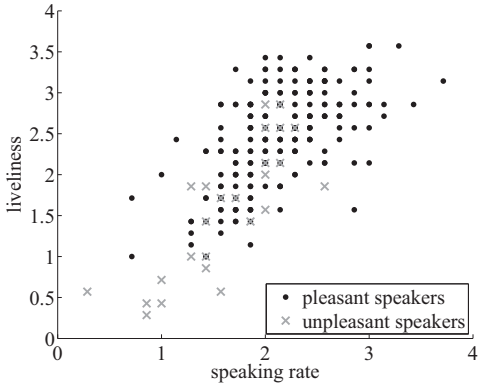


Fig. 2. The average speaking speed and liveliness collected with subjective evaluations for each audio segment. The audio segments which are identified as bad speakers are marked with a cross sign.

In the same figure, the samples with unpleasant voice are marked with a cross sign. We can see that dull voices (i.e., liveliness$<$1) are generally perceived as unpleasant, and speakers with fast speaking rate or high liveliness are perceived as pleasant. Assigning the votes for unpleasant and pleasant voice to -1 and 1 and taking the average to have a single rating for pleasantness yields positive correlations with ratings of liveliness ($\rho = 0.49$) and speaking rate ($\rho = 0.33$). A similar connection has been made between charismatic speech and ratings of being enthusiastic [9]. Further analysis shows that unpleasant speakers lying in the mid-range of speaking rate and liveliness correspond to videos with accented speech and/or bad recording conditions (*e.g.,* distant microphone).

Evaluator agreement: We had 43 unique evaluators annotate all of the data, with each evaluator rating a different number of segments. To quantify the inter-evaluator agreement, we computed the average cross-correlation between ratings of each evaluator and average rating of the rest of the evaluators. We found this to be 0.5 and 0.48 for liveliness and speaking rate, respectively. This shows agreements higher than the ones reported in [12] for speaker likability. It has been previously shown that subjectivity of the ratings in similar tasks often results in low inter-evaluator agreements [9, 10]. This is partly because most speakers fall in the middle ranges and

Dimension	Cross-validation		Test set	
	ρ	MSE	ρ	MSE
liveliness	0.67	0.24	0.56	0.40
speaking rate	0.64	0.16	0.36	0.29

Table 3. Performance of our algorithm measured using Pearson correlation (ρ) and mean squared error (MSE)

agreement is higher for the extremes, which we are most interested in for this application.

Within video segment consistency: We also analyzed if short segments can be used to obtain consistent human judgments for the entire video. We compared the standard deviation of average ratings of the three segments of a single video lecture (i.e., intra-speaker variability) to the standard deviation of the average ratings in three randomly selected segments from different speakers (i.e., inter-speaker variability). According to population mean z-test, the mean inter-speaker variability is significantly higher than the mean intra-speaker variability (p-value $< 1e-20$) and thereby confirming that short audio segments can be used for annotation and also for predicting the speaking style characteristics of the speakers. This is consistent with the thin-slicing notion [21].

Metric: We evaluate our approach on three different metrics: Pearson's correlation coefficient [22], *mean squared error* and information retrieval metric, *normalized discounted cumulative gain* (NDCG) [23]. Evaluation is performed using (a) leave-one-speaker-out cross validation and (b) a separate test set.

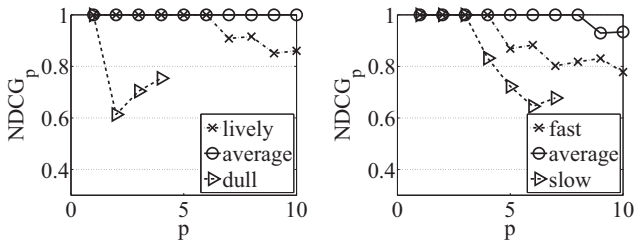
The use of NDCG as a metric stems from the application: Given the set of relevant videos from a topic of interest (in our setting, a chapter from a textbook), the goal is to retrieve top p re-ranked subset in accordance with preference for a particular dimension (*e.g.,* lively videos) elicited by the user. Therefore, we would like to capture the relevance of the video based on its position in the re-ranked list such that the relevance in the higher ranks are valued more than those with lower ranks:

$$NDCG_p = \frac{DCG_p}{IDCG_p} = \frac{\sum_{i=1}^{p} \frac{2^{rel_i}-1}{log_2(i+1)}}{IDCG_p}, \quad (2)$$

DCG is the gain associated to a ranking, which penalizes the relevant samples appearing in lower ranks with the log term; rel_i is the relevance of retrieved result at position i in the ranking. For instance, to retrieve lively (or fast) speakers, the relevance of samples with liveliness (or speaking rate) score in [0, 2) is set to 0 and the relevance of samples with liveliness (or speaking rate) in [2, 4] is set to 1. Likewise, the samples with scores in [0, 2] and [1, 3] are defined relevant for retrieving low-range and mid-range of the dimensions, respectively. $IDCG_p$ is the ideal DCG_p and is obtained by optimal ordering of the videos based on the ground truth ratings. Thus, $NDCG_p$ varies between 0 to 1 with 1 being most consistent with ground truth judgments across all p retrieved videos.

4.2. Results

Table 3 reports the regression results in terms of Pearson's correlation coefficient (ρ) and *mean square error* (MSE) between ground truth ratings and predicted ratings. We observe a reasonable correlation between the ground truth and the prediction showcasing the efficacy of our approach. Figure 3 show the NDCG evaluations on the test set for retrieving up to the top 10 videos. Since judgments for the test set were obtained at the video-level, we set the predicted value to be the median of the predictions for the multiple segments corresponding to the video. As shown, our approach has high NDCG for retrieving videos with average speaking rate or liveliness. We see



(a) Liveliness (b) Speaking rate

Fig. 3. Test set evaluation: $NDCG$ values by varying p. NDCG computed for retrieving only up to k videos when the algorithm found only $k \leq p$ videos satisfying the query.

a drop in performance for retrieving the dull and slow videos due to a smaller number of samples in this category for our data. In the training dataset only 25% and 39% of the samples are relevant for retrieving dull and slow videos, respectively. In the test set this ratio is 30% for both dimensions. Overall, the experiments show good separation of the different descriptors within each dimension.

4.3. Feature Analysis

The LASSO regression penalizes the absolute value of the regression coefficients and shrinks a subset of them to zero to yield a sparse model (Eq. 1). Thus, it can provide insights for feature selection. Table 2 reports the selection percentage across cross-validation folds. The features selected in more than 90% of time are highlighted in bold for each task. These features are used to build the test models. The features which yielded coefficients with consistent sign across folds are also identified with the corresponding sign.

speaking rate: The table indicates that speaking rate is positively correlated with Jong and Wamp's estimation of speaking and articulation rates [15], number of F0 segments and total signal energy captured by zeroth intensity DCT coefficient.

liveliness: The selected features relevant for liveliness capturing the dynamic range include mean peak to mean valley distance in F0 contour, F0 interquartile distances, standard deviation of F0 peaks and max valley to peak jump in intensity contour. Likewise, liveliness is positively correlated with spectral roll-off features capturing the vocal effort. According to this analysis, lively voices have higher F0 rising time ratio compared to dull voices. Figure 1 also depicts this effect. Interestingly, from estimations of speaking/articulation rates [15] only articulation rate is always selected for liveliness. This indicates that the subjects have ignored the pauses to rate the liveliness, which is expected.

5. CONCLUSION AND FUTURE DIRECTIONS

Automatic estimation of speaking style can allow students to select preferred online videos for lesson augmentation. Using crowd sourcing, we identified six broad style dimensions (liveliness, speaking rate, pleasantness, clarity, formality and confidence) that users perceive. We proposed techniques for automatically detecting liveliness and speaking rate, and found these to be correlated with general pleasantness of the voice. We showed the benefit of the approach in a re-ranking experiment in which users selected among multiple videos with similar content, based on their preferred speaking style. Automatic estimation of speaking styles could also help teachers or speakers improve their presentation skills. Future directions include exploring additional dimensions of speaking style, understanding the relationship between features and higher order percepts, and extending the work to include visual and other information in the video.

6. REFERENCES

[1] J. Ubrani, *IDC Worldwide Quarterly Tablet Tracker*. International Development Corp. Framingham, MA, February 2013.

[2] *Policy guidelines for mobile learning*. UNESCO, 2013.

[3] R. Agrawal, S. Gollapudi, K. Kenthapadi, N. Srivastava, and R. Velu, "Enriching textbooks through data mining," in *ACM DEV*, 2010.

[4] R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi, "Enriching textbooks with images," in *CIKM*, 2011.

[5] ——, "Data mining for improving textbooks," *ACM SIGKDD Explorations Newsletter*, vol. 13, no. 2, 2011.

[6] ——, "Studying from electronic textbooks," in *CIKM*, 2013.

[7] P. Tantrarungroj, "Effect of embedded streaming video strategy in an online learning environment on the learning of neuroscience," Ph.D. dissertation, Indiana State University, 2008.

[8] M. Miller, "Integrating online multimedia into college course and classroom: With application to the social sciences," *MERLOT Journal of Online Learning and Teaching*, vol. 5, no. 2, 2009.

[9] A. Rosenberg and J. Hirschberg, "Acoustic/prosodic and lexical correlates of charismatic speech," in *Interspeech 2005*. Proceedings of Eurospeech'05, September 2005, pp. 513–516.

[10] E. Strangert, "What makes a good speaker? subjective ratings and acoustic measurements," in *Proceedings from Fonetik 2007: speech, music and hearing, quarterly progress and status report, TMH-QPSR, Vol 50, 2007*, 2007, pp. 29–32.

[11] E. Strangert and J. Gustafson, "What makes a good speaker? subject ratings, acoustic measurements and perceptual evaluations." in *Interspeech 2008*, vol. 8, 2008, pp. 1688–1691.

[12] F. Burkhardt, B. Schuller, B. Weiss, and F. Weninger, ""would you buy a car from me?" - on the likability of telephone voices." in *Interspeech 2011*. Florence, Italy: ISCA, August 2011, pp. 1557–1560.

[13] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. van Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, and B. Weiss, "The interspeech 2012 speaker trait challenge," in *Interspeech 2012*, Portland, OR, USA, September 2012, pp. 254–257.

[14] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[15] N. Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior research methods*, vol. 41, no. 2, pp. 385–390, 2009.

[16] N. C. Yoder, "Peakfinder (matlab program)," http://www.mathworks.com/matlabcentral/fileexchange/25500-peakfinder, 2011.

[17] E. Shriberg, A. Stolcke, D. Hakkani-Tür, and L. Heck, "Learning when to listen: Detecting system-addressed speech in human-human-computer dialog." in *Interspeech 2012*, Portland, OR, USA, September 2012, p. 334337.

[18] F. Eyben, M. Wöllmer, and B. Schuller, "OpenSMILE: the Munich versatile and fast open-source audio feature extractor," in *ACM International conference on Multimedia (MM 2010)*, Florence, Italy, October 2010, pp. 1459–1462.

[19] N. Morgan and E. Fosler-Lussier, "Combining multiple estimators of speaking rate," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, May 1998, pp. 729–732 vol.2.

[20] H. Traunmüller and A. Eriksson, "The perceptual evaluation of f excursions in speech as evidenced in liveliness estimations," *The Journal of the Acoustical Society of America*, vol. 97, pp. 1905–1915, 1995.

[21] N. Ambady and R. Rosenthal, "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis." *Psychological bulletin*, vol. 111, no. 2, p. 256, March 1992.

[22] K. Pearson, "Notes on the history of correlation," *Biometrika*, vol. 13, no. 1, pp. 25–45, October 1920.

[23] K. Järvelin and J. Kekäläinen, "IR evaluation methods for retrieving highly relevant documents," in *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. Athens, Greece: ACM, July 2000, pp. 41–48.