Awe the Audience: How the Narrative Trajectories Affect Audience Perception in Public Speaking

M. Iftekhar Tanveer[1], **Samiha Samrose**[2], **Raiyan Abdul Baten**[3], **M. Ehsan Hoque**[4]
ROC-HCI Lab
University of Rochester
{[1]itanveer,[2]ssamrose,[4]mehoque}@cs.rochester.edu, [3]rbaten@ur.rochester.edu

Figure 1. In his 2012 TED Talk titled "404, the story of a page not found", Renny Gleeson discussed the evolution of the 404 error page. His enchanting delivery undoubtedly strengthened the message. However, computational analysis of just the words he spoke is still able to capture the seesaw in 'joyful' emotion in the speech, as plotted in the figure. The lowest point is observed when he explained how the 404 pages used to give the viewers a feeling of failure in the past. In contrast, the peak came when he explained how one brilliant use of a funny video later inspired others to positively utilize the page as well. This shape corresponds to one of the three major patterns we have found in TED Talks, as shown in figure 6.

ABSTRACT

Telling a great story often involves a deliberate alteration of emotions. In this paper, we objectively measure and analyze the narrative trajectories of stories in public speaking and their impact on subjective ratings. We conduct the analysis using the transcripts of over 2000 TED talks and estimate potential audience response using over 5 million spontaneous annotations from the viewers. We use IBM Watson Tone Analyzer to extract sentence-wise emotion, language, and social scores. Our study indicates that it is possible to predict (with AUC as high as 0.88) the subjective ratings of the audience by analyzing the narrative trajectories. Additionally, we find that some trajectories (for example, a flat trajectory of joy) correlate well with some specific ratings (e.g. "Longwinded") assigned by the viewers. Such an association could be useful in forecasting audience responses using objective analysis.

ACM Classification Keywords

H.5. INFORMATION INTERFACES AND PRESENTATION (e.g., HCI); I.5.3 PATTERN RECOGNITION: Clustering

CHI 2018, April 21–26, 2018, Montreal, QC, Canada
 ISBN 978-1-4503-5620-6/18/04…$15.00
DOI: https://doi.org/10.1145/3173574.3173598

Author Keywords

Affective Computing; Styles of Storytelling; Narrative Trajectories; TED Talks; Clustering; Pattern Recognition

INTRODUCTION

Great stories progress through conflict, suspense, tension, rises, and falls in the plots [1]. Satirist Kurt Vonnegut anecdotally claimed that many stories could be plotted on a plane of "Great Fortune–Ill Fortune" vs. "Beginning–End" axes[1], and several popular stories follow similar trajectories [49]. Reagan et al. [38] showed that it is possible to trace emotional trajectories through statistical analysis of large volumes of texts from English story books. Inspired by these works, we seek to answer if it is possible to computationally capture these trajectories in the setting of *public speaking*. In addition to emotions, we analyze how the linguistic styles of the speech vary over time—which we collectively define as the "narrative trajectory". We also measure how the audience responds to various narrative trajectories. These analyses have significant implications in gaining deeper understanding of the nature of human behavior. In addition, it would allow the use of computer algorithms to predict potential audience response of public speaking, creating the possibility of building automated training tools.

There is, however, a challenge that must be addressed for effectively answering these questions. A narrative trajectory

[1] http://www.youtube.com/watch?v=oP3c1h8v2ZQ

must be observed with a reasonably high statistical confidence in order to claim its existence. To ensure the statistical confidence, it is important to analyze the phenomenon over a large set of data. However, manual analysis of a large dataset is prohibitively expensive in terms of required time and effort. In this paper, we present an algorithmic approach to resolve the problem with current off-the-shelf technologies and freely available resources. Such an approach also ensures the reproducibility of the experiments, makes verifiable claims, and provides specific steps to test if the claims generalize across different domains.

To gain access to a large dataset of public speaking, we resort to the largest open repository of high quality public speeches that we know of—the TED (Technology, Entertainment, Design) conference talks. In TED.com, videos and transcripts of over 2000 public speeches are freely available. Many of them are quite popular and influential. In addition, there are more than five million responses from the spontaneous viewers of the videos, where the speeches are annotated in 14 different categories: Beautiful, Courageous, Confusing, Fascinating, Funny, Ingenious, Informative, Inspiring, Jaw-dropping, Longwinded, Ok, Obnoxious, Persuasive, and Unconvincing. We use IBM Watson Tone Analyzer [24] to analyze the affective components expressed in each sentence of the transcripts. The tone analyzer can evaluate each sentence to calculate objective scores for Emotion (anger, disgust, fear, joy, and sadness), Language (analytical, confident, and tentative), and Social Personality (openness, extraversion, emotional range, conscientiousness, and agreeableness). We calculate how these scores vary over time and then use clustering techniques [2] to group similar trajectories. Finally, we statistically analyze how the audience rate the TED talks pertaining to various clusters of narrative trajectories.

We conduct classification and regression experiments to test if it is possible to predict the audience ratings just from the narrative trajectories. We find that it is indeed possible with much higher accuracy than random chance. We can classify the ratings in two classes with an average AUC 0.76 (maximum 0.88), where random chance is just 0.5. We conduct various hypothesis tests to analyze *how* the audience ratings get affected by the narrative trajectories. The most prominent result we observe from these tests is as follows—when the trajectories of the scores do not vary much over the time (i.e. showing a "flat" trajectory), it is likely that the audience would rate those talks as "Longwinded". There are other interesting results which are enlisted in Table 4. We obtain these results with significantly high statistical confidence. In summary, we have made the following contributions in this paper:

1. We propose a technique to objectively compute the most prominent patterns of narrative trajectories in TED talks.

2. We use a large dataset ($N > 2000$), thus ensuring high confidence for the validity of the claims. The dataset contains both the public speeches and the audience responses.

3. We report our observations that narrative trajectories result in predictable changes to the audience ratings in the public speaking setting.

4. Narrative trajectories were previously shown to exist in English stories. To the best of our knowledge, this is the first work showing that such trajectories also exist in the public speech domain.

5. We release our source code and the experimental data for the scientific community[2].

LITERATURE REVIEW

In this section we contextualize our work with respect to similar research.

Patterns of Storytelling

Quantitative analyses of literary styles have a rich history, from Augustus de Morgan's comment on the attributions of Pauline epistles [11] to Wincentry Lutosławski's statistical analysis of Plato's word usage [28]. In the realm of written fictions, there has been some progress in extracting storyline trajectories from sentimental analysis of texts, in line with Kurt Vonnegut's proposition mentioned earlier [49]. The idea being, sentence-wise sentiment scores (are meant to) correspond to the instantaneous sentiments a reader experiences; but when the scores are smoothed out (filtered) over a large amount of text, the remaining sentiment variation corresponds to the narrative development of the novel [19].

Samothrakis et al. [39] extracted the trajectories of 6 basic emotions [15] from the fictions of Project Gutenberg, using the WordNet Affect Lexicon [41]. With this information, they were able to predict the fiction genre significantly better than random chance. Mohammed created visualizations of emotional trajectories in famous novels [30] using the NRC Word-Emotion Association Lexicon [31]. Jockers created Syuzhet [26] that extracts sentiment and plot arcs from texts. Using hierarchical clustering, he proposed the existence of six, or possibly seven, archetypal plot shapes in a corpus of 41,383 books [25].

More recently, Reagan et al. [38] conducted a research to find all such trajectories in English stories. They performed sentiment analysis on 1327 stories in the Project Gutenberg fiction collection. They claimed that there are six dominant "emotional trajectories" of stories as obtained through Principal Component Analysis (PCA) [2]. Our work is different from previous literature on several aspects. We narrow our scope of interest down to public speeches (TED Talks) only, rather than storytelling in a broader sense. We use density based clustering in contrast to hierarchical or PCA approaches. PCA imposes a constraint of orthogonality over principal components, which would unnecessarily bias the clusters from their natural shape. More importantly, unlike previous work, we are rather keen to understand the *audience perception* of various narrative trajectories, in addition to just identifying the trajectories.

Works Related to Public Speaking

In recent years, there have been significant research efforts to build automated systems for helping in public speaking. "ROC Speak" [17] is an open online system that provides semi-automatic feedback to a user on his/her public speaking skills,

[2]Code and data available in: https://github.com/ROC-HCI/TEDTalk_Analytics

in terms of smile intensity, voice modulation and body gestures. Curtis et al. [9] emphasized on both presentation ratings and audience engagement by analyzing the data of conference scientific presentations having both audience and speaker views. Damian et al. proposed a system named "Logue" [10] that increases public speakers' awareness of their nonverbal behaviors. Bubel et al. created "AwareMe" [3] to give feedback on pitch, use of filler words, and words per minute. AwareMe uses a detachable wristband to provide feedback to the speakers while they are practicing. "Rhema" [43], a Google Glass application, provides real-time feedback to public speakers on prosodic attributes like speech rate and volume. Nguyen et al. [33] implemented an online system to provide feedback on a speaker's body language. Tanveer et al. [44] developed "AutoManner", a system that helps public speakers become aware of the idiosyncrasies in their body language. Chollet et al. [7] applied thin slice technique to analyze the behaviors of public speakers. They collected the audio, video, and kinect data for public speakers from the Cicero public speaking virtual agent tool.

In the book "Talk Like TED" [18], Carmine Gallo describes various tactics for great public speeches. It is focused more on practical skills and discusses techniques like dispersing passion among audience, the art of storytelling, verbal and nonverbal cues, creating striking or jaw-dropping moments etc. Bull et al. [4] proposes theoretical implication of speaker-audience interaction in the concept of dialogue between speaker and audience through applause, laughter, cheering, chanting, booing, delivery, speech content, and uninvited applause.

TED Talks have been analyzed in some of the previous works as well. Tsou et al. [47] conducted research on the comments left by the viewers of the talks in the TED.com website and in YouTube. Their analysis shows that viewers in the YouTube platform are more likely to discuss the characteristics of the presenters, whereas discussions in the TED website focus on the contents of the talks. Sugimoto et al. [42] analyzed the presenters' backgrounds in correspondence to the impact of the videos. Their results reveal that giving a TED presentation has no impact on the number of citations received by the academics. Drasovean et al. [12] conducted an experiment to evaluate the users' reactions from a linguistic point of view. They use the Appraisal framework to emphasize the social meanings of various linguistic patterns. Chen et al. [6] analyzed the TED Talk and Pun of the Day corpus for humor detection. Even though their accuracy improved significantly for Pun of the Day data, the accuracy for TED Talks was not promising. Liu et al. [27] analyzed the transcripts of TED talks to predict audience engagement. They came up with 24 rhetorical devices as triggers for audience applause.

IBM Watson Tone Analyzer

We use the IBM Watson Tone Analyzer [24, 23] to obtain the sentence-wise *scores* (a number in the rage of 0 to 1) representing various aspects of the transcripts. We construct the narrative trajectories by pre-processing these scores, which is detailed in the later sections. The tone analyzer can analyze three different aspects of each sentence in a given text document—the Emotion, Language, and Social aspects. IBM provides technical details on how these scores are computed in their online manual [23]. In general, the techniques involve human annotation of large volumes of text and using machine learning algorithms to predict the annotations from a number of *features*. IBM uses n-gram features, lexical features (e.g. LIWC [45]) from various dictionaries, person-based features, dialogue-specific features, and several higher-level features such as the existence of consecutive question marks or exclamation marks etc. Additionally, they use Support Vector Machine (SVM) [48, 2] as the learning algorithm [23]. We describe the scores in the sequel.

The *Emotion scores* represent the likelihood that a sentence portrays one of the following emotions: anger, disgust, fear, joy, or sadness. This is a subset of the 6 basic emotions proposed by Ekman [15] and the 8 basic emotions proposed by Plutchik [36]. The scores are computed by training the system with a large amount of text using a constrained optimization approach [50]. The optimization problem is designed to handle co-occurrences of multiple emotions and noisy training data, which are relevant for our purposes.

The *Social scores* [5, 24] indicate the likelihood of a sentence portraying the characteristics of the Big Five personality model [34, 20]. These characteristics are: *openness*, *conscientiousness*, *extraversion*, *emotional range*, and *agreeableness*. High openness indicates the property of being open to try out new ideas. Agreeableness represents the tendency to be compassionate, caring, cooperative, compromising, and trustworthy. Conscientiousness is the characteristic of being methodical, disciplined, and organized. Extraversion represents the tendency to seek stimulation in the company of others. Finally, emotional range, which is also referred to as "neuroticism", represents the extent to which a person's emotion is sensitive to the environment. IBM uses various psycholinguistic features to predict the social personality scores [5, 21].

The *Language scores* evaluate three qualities in the words of a sentence: *analytical*, *confidence*, and *tentative*. Analytical score represents the amount of reasoning and technical substance in the language used. Confidence represents the degree of certainty: a highly confident expression denotes an assured and optimistic attitude. A tentative use of language is perceived to be questionable, doubtful, and debatable.

RESEARCH QUESTIONS

In this work, we strive to answer the following research questions:

1. Do narrative trajectories exist in a public speaking setting?
2. If they do, how might they impact the audience perception of a speech?

Notably, we narrow our scope down to the domain of public speaking only. We choose TED Talks as our testing ground, with the assumption that the audience perception is reflected in the spontaneous ratings that the videos receive. Towards gaining insight to our primary queries, we need to answer several secondary questions: 1. What is the statistical distributions and characteristics of the TED dataset? 2. Can we predict the ratings using the scores from the Tone Analyzer?

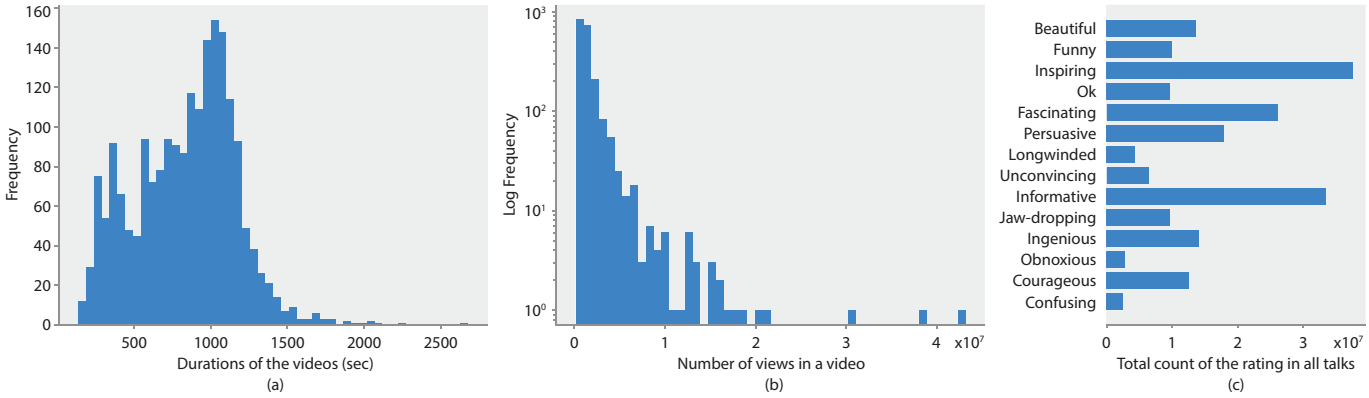

Figure 2. Statistical properties of the dataset

Property	Quantity
Total Number of talks	2007
Total length of all talks	465.24 hours
Average duration of a talk	13.9 minutes
Total number of ratings	5,069,897
Total word count	5,081,321
Total sentence count	271,567

Table 1. Dataset Properties

In order to answer all these questions, we formulate several experiments: 1. We extract several statistics regarding the dataset and calculate how various user ratings correlate with the total view counts in the dataset; and see whether that matches with our intuition or not. 2. We analyze if it is possible to predict the user ratings from the objective measurements of the talks. 3. Finally, we perform cluster analysis to identify the major patterns of narrative trajectories in storytelling and perform various hypothesis tests to identify how the trajectories affect audience ratings. We describe all these experiments and their results in the subsequent sections.

DATASET

We collect the transcripts and meta information about the talks from TED.com. The dataset is described below.

Amount of Data

As of February 7, 2017, we found a total of 2212 TED talks in TED.com. However, not all of those were public speeches; many were videos of music, dance and other performances. Some of the talks were recently-published and we suspected they might not have had enough time to be watched and be rated in a quantity typical of TED talks. Therefore, we filtered any talk that 1. was published less than 6 months prior to February 2017, 2. contained any of the following keywords: live music, dance, music, performance, entertainment, or, 3. contained less than 450 words in the transcript. Our heuristic analysis indicates that transcripts containing less than 450 words are rarely public speeches. After filtering, we had a total of 2007 TED talks in our dataset which constitute approximately 465 hours of videos. A summary of the dataset is shown in Table 1.

Contents of the Dataset

The dataset contains several meta information for each talk from TED.com. It contains the number of viewers who rated a talk with a specific label (e.g. Beautiful, Inspiring). These labels are binary; and aggregated using check-boxes in the web page. The dataset contains the total number of viewers providing a certain label, which we refer to as the "ratings". We calculate a scaled version of these ratings as well, which can be defined mathematically as follows:

$$r_{i,\text{scaled}} = \frac{r_i}{\sum_i r_i} \tag{1}$$

where r_i represents the rating for the i^{th} label in that talk. Besides the ratings, we also collect other information such as the title, presenter name, keywords, total number of views, publishing date and the crawling date for each TED talk.

The website also offers a human-generated transcript for each talk. The transcripts contain timestamps and tags describing additional tags (e.g. applause, laughter, music). In this work, we use only the textual transcripts, and not the descriptive tags. The IBM Tone Analyzer gives us objective scores for each of the sentences in the transcripts which are also included in the dataset.

Statistical Properties

A histogram of the video durations is shown in Figure 2(a). This histogram has two peaks at around 5 and 18 minutes. This is reflective of the TED talk formats of short talks (3 to 5 minutes) and long talks (18 minutes).

We show the histogram of the total-views in Figure 2(b). The vertical axis of this histogram is drawn in the logarithmic scale. Noticeably, the total-views show a heavy-tailed distribution similar to *power-law* [14]. In other words, although there are progressively fewer talks with higher total-views, the number of talks having very high total-views is not very low. We know this type of "*rich gets richer*" phenomenon [14] to be abundant in *preferential attachment* models of networks. We think the same phenomenon plays a role in the total-views of TED talks as well. When many people watch a particular talk, it is likely to be viewed by even more people because of the recommendations made by the early viewers in their network.

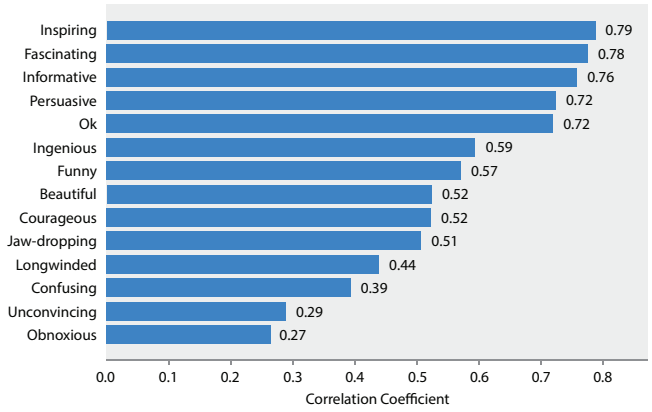

Figure 3. Correlation coefficient of various ratings and total-views

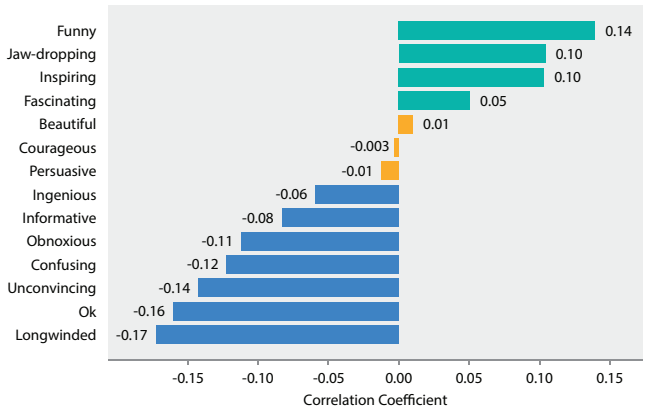

Figure 4. Correlation coefficient of scaled ratings and total-views

In Figure 2(c), we illustrate the total counts of various ratings in the dataset. The figure shows that "Inspiring" and "Informative" are the two most used ratings. There are only a few occurrences of strongly negative ratings, like obnoxious, confusing, long-winded, unconvincing, etc.

Sanity Check

In order to ensure the quality of the dataset, we perform a few experiments checking its consistency. We calculate the correlation coefficients between the ratings and the total-views to analyze their relations. The result of the analysis is illustrated in Figure 3. The figure shows that all the ratings are positively correlated with total-views. This is reasonable because the more people watch a particular TED Talk, the more ratings it gets. Some of the ratings (inspiring, fascinating, informative etc.) are more correlated with total-views than others (obnoxious, unconvincing, confusing etc.). We notice that the negative ratings demonstrate lower correlation with the total-views. This result is reasonable because a negative rating suggests that the talk is less appealing, and therefore less likely to attain a high view count.

This analysis captures an inherent bias of the ratings towards videos with higher total-views. We use the *scaled ratings* as calculated in equation (1) to remove this bias. This metric effectively cancels the effect of total-views, because, for a single TED talk, both the numerator and the denominator terms are equally affected. To test our rationale, we again calculate the correlation coefficients of these scaled ratings with the total-views. The results are shown in Figure 4. The figure illustrates that after scaling, the ratings are weakly correlated with the total-views. Interestingly, the intuitive *positive* ratings (e.g. Funny, Jaw-Dropping, Inspiring) are positively correlated with the total-views and the *negative* ratings (e.g. Longwinded, Unconvincing, Confusing) are negatively correlated. Although there are three exceptions ("Ingenious", "Informative", and "Ok") to this general pattern, the overall trend of correlation gives us a general idea about the consistency of the dataset.

Constructing the Narrative Trajectories

As mentioned previously, we use the IBM Watson Tone Analyzer to extract the sentence-wise *scores* from the transcripts of the TED talks. We use the extracted scores to construct the narrative trajectories according to the procedure outline in Algorithm 1. We consider the sequence of sentence-wise scores as the raw signal S. We use a 5-point averaging kernel to filter S into a smoothed version S_{smooth}. We crop the signal in order to remove the boundary effects caused by filtering. This process removes either ends of S_{smooth} and keeps only the $\max(M,N) - \min(M,N) + 1$ elements from the middle of the signal. Here, N and M refers to the signal length and the filter length respectively. Finally, the smoothed and cropped signal S_{crop} is interpolated using a piece-wise linear technique to have a canonical length of $N = 100$ samples. We refer to this final signal as the narrative trajectory, $T[n]$ for a specific TED talk, where n is the discreet time index in $0 \leq n < N$.

Algorithm 1 Constructing the Narrative Trajectories

Input: Array of sentence-wise scores, S for a TED Talk
Output: Narrative Trajectory $T[n]$ for that TED Talk
  procedure BUILD_TRAJECTORY(S)
    Filter: Apply averaging filter on S to get S_{smooth}
    Crop: Remove the boundary effects to form S_{crop}
    Interpolate: Interpolate to make length = 100, S_{intp}
    $T[n] = S_{\text{intp}}$
  return $T[n]$

EXPERIMENTS

In this section, we describe the experiments conducted to answer the research questions. In all of these experiments, we use the scaled ratings from Equation (1). We conduct classification and regression analyses to test the predictability of the audience ratings from the narrative trajectories. Then we proceed to extract the global and local patterns of the narrative trajectories.

Classification

In this experiment, we strive to classify the TED talks into *Highly Rated* and *Poorly Rated* classes for each audience rating. We use the 33^{rd} and 66^{th} percentiles of each rating to divide the dataset into three chunks. We denote any talk having a rating greater than 66^{th} percentile as the positive class, and talks having a rating less than 33^{rd} percentile as the negative

Ratings	Lin. SVM	RBF SVM	Logistic Reg.
Beautiful	**0.88** (78.8)	0.80 (71.6)	0.80 (71.6)
Confusing	0.57 (65.8)	0.55 (68.6)	0.62 (67.4)
Courageous	0.83 (76.3)	0.82 (73.5)	0.77 (72.3)
Fascinating	0.85 (77.8)	0.83 (**77.1**)	**0.84 (77.1)**
Funny	0.70 (69)	0.7 (66.4)	0.68 (69.3)
Informative	0.84 (74.6)	**0.84** (77)	**0.84** (77)
Ingenious	0.78 (68.6)	0.78 (69.6)	0.76 (70)
Inspiring	0.79 (74.6)	0.81 (73.3)	0.79 (74.3)
Jaw-dropping	0.75 (69.1)	0.7 (63.8)	0.68 (65.1)
Longwinded	0.69 (70.1)	0.69 (68.4)	0.66 (67.8)
Obnoxious	0.57 (64.7)	0.54 (70)	0.60 (64.6)
Ok	0.70 (63.6)	0.74 (67.6)	0.68 (61.6)
Persuasive	**0.88 (80)**	0.83 (75.8)	**0.84** (75.1)
Total-views	0.65 (69.8)	0.6 (57.6)	0.60 (57.4)
Unconvincing	0.68 (63.6)	0.67 (61)	0.71 (64.8)
Average	**0.74 (71.1)**	0.73 (69.4)	0.72 (69)

Table 2. Area under the ROC Curve (AUC) and classification accuracy in percent (numbers within the parenthesis) for predicting the ratings and total views using three different classifiers

Ratings	SVR	Ridge	GP	LASSO
Beautiful	0.51	0.43	**0.28**	0.47
Confusing	0.28	0.19	0.07	0.09
Courageous	0.52	0.49	**0.28**	0.47
Fascinating	0.53	0.54	0.27	**0.56**
Funny	0.32	0.42	0.18	0.39
Informative	**0.61**	**0.56**	0.26	0.46
Ingenious	0.40	0.39	0.16	0.39
Inspiring	0.42	0.46	0.22	0.41
Jaw-dropping	0.29	0.31	0.16	0.25
Longwinded	0.26	0.30	0.11	0.26
Obnoxious	0.12	0.15	0.04	0.00
Ok	0.38	0.42	0.17	0.28
Persuasive	0.46	0.50	0.22	0.50
Total-views	0.22	0.22	0.12	0.15
Unconvincing	0.19	0.31	0.02	0.21
Average	0.37	**0.38**	0.17	0.33

Table 3. Correlation Coefficient for predicting the ratings and total views using several regression techniques

class. We do not use the talks with mediocre ratings in the classification analyses. We divide the dataset into three splits, instead of two, in order to train the classifiers with extreme ratings. In a real-life scenario, the mediocre talks could be identified from the class association probabilities of the trained classifiers.

Features

We extract various time-wise statistics (i.e. minimums, maximums, averages, and standard deviations) from the narrative trajectories and feed them as the input features to the classifier. The time-wise statistics are calculated over the time axis. For example, if $T_i[n]$ is the narrative trajectory of the i^{th} TED talk, the time-wise average is defined as in Equation (2).

$$\bar{T}_i = \frac{1}{N}\sum_{n=0}^{N-1} T_i[n] \tag{2}$$

We calculate these statistics for all the trajectories of the scores in a single talk and use them as features for classification.

Classifiers

We use three different classifiers in this task: a logistic regression classifier [32], a Linear Support Vector Machine (SVM) [48, 32], a Kernel SVM with Radial Basis Function (RBF) [32] as the kernel. We use the logistic regression as the simplest form of the classifier, the linear SVM to test if there is any improvement due to its max-margin constraint, and the RBF kernel to test if there is any gain from projecting the data to higher dimensions [32]. We divide the available data-points into a 7:3 split to train over the larger part and test over the smaller part. We tune the hyper-parameters (e.g. the slack parameter C in SVM or the bandwidth parameter σ in RBF) by performing randomized search with 3-fold cross-validation and 100 iterations. We use a python library named Scikit-learn [35] for all these operations.

Results of Classification

We report the area under the ROC curve [37, 22] and the classification accuracy (in percentage) for evaluating the classifier performances on the test split of the dataset. The results of classification is shown in Table 2. It is evident that the Linear SVM can classify the ratings with the highest average AUC (0.74) and accuracy (71.1%). "Beautiful" and "Persuasive" can be classified with the highest test AUC (0.88). "Fascinating", "Informative", and "Courageous" are also among the top few highly predictable ratings. "Obnoxious" is the least predictable rating. Note that the ratings with lowest AUC are the ones having fewer counts in the dataset (compare with Figure 2(c)). Lack of the representative samples might be a reason for low predictability of these ratings.

Regression

We perform regression over the narrative trajectories to predict the audience ratings. In this experiment we analyze how accurately it is possible to predict the *continuous* ratings. We use the same set of features for regression as we used in the classification task. We use four different regression models. The Ridge Regression [32] and LASSO [46] are simple linear regressions with $\ell 2$-norm and $\ell 1$-norm regularizations respectively. We chose these two regressors because Ridge regression is robust to handle noisy datasets while LASSO suppresses correlated features through enforcing sparsity. We also use Support Vector Regression (SVR) [40] which is similar to SVM and thus maximize the margin. Finally, we use a Gaussian Process Regressor [32] which assumes a Gaussian process prior for the regression. We divide the dataset into 7:3 training-test splits and use randomized search to tune the parameters in a similar fashion to the classification experiments. Scikit-learn is used for the regression tasks as well.

We calculate the correlation coefficient of the predicted user ratings and the actual audience ratings which are shown in

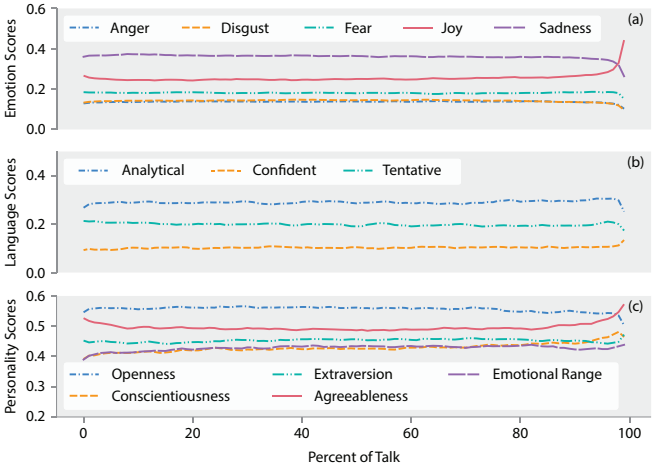

Figure 5. Progression of Average Scores

Table 3. The results are obtained from the "test" split of the dataset while the regressors were trained on the "train" split. It is evident that both SVR and Ridge regression provides similar prediction performance while Ridge has a slightly higher average correlation coefficient. "Informative" has the highest test correlation coefficient (0.61). "Obnoxious" shows the lowest performance (0.15) in regression as well. The overall trend of regression performance is similar to the one seen in classification performance—the ratings with lesser samples are more difficult to predict.

Global Trends in the Narrative Trajectories

We were curious to see if there are any global patterns in the narrative trajectories. Therefore, we compute the *ensemble average* of the trajectories for all the TED talks in the dataset. If $T_1[n], T_2[n], T_3[n] \ldots T_K[n]$ be the narrative trajectories for K TED talks, the ensemble average is defined as follows.

$$\tilde{T}[n] = \frac{1}{K}\sum_{i=0}^{K} T_i[n] \quad \forall n \in \{0,1,2,\ldots N-1\} \tag{3}$$

Here K is the total number of TED talks in the dataset ($K = 2007$). The ensemble averages for all the different scores are shown in Figure 5.

Note that the average of the trajectories are almost flat in the middle section, implying that there is no single major pattern in this area for the aggregated dataset. However, there is a distinct rise towards the end of the averages. The joy score increases significantly in the last ten percent of the average trajectory. Sadness decreases during this time. Fear, anger, and disgust show a slight decrease in their values. Language and personality scores also show change. Overall, the positive scores (joy, confidence, agreeableness etc.) increase towards the end of the talks and the negative scores (sadness, tentative, emotional range) decrease. We think this is a prominent characteristic of TED talks that they tend to finish with a positive note on average.

Clusters of Narrative Trajectories

To identify the major patterns in the narrative trajectories, we cluster similar trajectories together. We describe this process in the following subsections.

Processing the Narrative Trajectories

While clustering, it is important to consider only the relative rises and falls of each narrative trajectory with respect to its own peaks and troughs. To achieve this, we standardize the trajectories by subtracting the time-average and dividing by the time-wise standard deviation. Mathematically the standardized narrative trajectory for the i^{th} TED talk, $\hat{T}_i[n]$ can be defined as follows:

$$\hat{T}_i[n] = \frac{1}{S}\left(T_i[n] - \bar{T}_i\right) \tag{4}$$

where, $\bar{T}_i$ is defined in Equation (2) and S is the time-wise standard deviation which is defined in the following equation.

$$S = \sqrt{\frac{\left(\sum_{n=0}^{N-1} T_i[n] - \bar{T}_i\right)^2}{N-1}} \tag{5}$$

Clustering

We use density based clustering (DBSCAN) [16] to find patterns among the standardized narrative trajectories in the dataset. Unlike several other common clustering algorithms (e.g. k-means), DBSCAN does not enforce any model-based constraint over the shape of clusters. It works by grouping together the data-points distributed at a higher density than the neighboring regions. Consequently, it groups together the most similar trajectories in the dataset. We use the Euclidean distance as the distance metric. We heuristically set the DBSCAN parameter (ε) to be 6.25. This value resulted in two to seven clusters for all the scores which is similar to Reagan et al. [38] and Jockers' [25] observations. Too high or low values of ε result in too few or too many clusters. We use the Scikit-learn implementation of the algorithm. The shapes of the cluster means (ensemble average) are given in the supplementary materials.

Hypothesis Testing for Audience Responses

Once we obtain the clusters of the TED talks, we proceed to analyze how the audience ratings correspond to these clusters. Several hypothesis tests are conducted to discover this relationship. We perform ANOVA [29] to analyze if there is any significant difference in the audience ratings over all the clusters. As we conduct repeated ANOVA tests for multiple ratings, we perform Bonferroni Adjustment [13] by multiplying the p-values with the number of repeated experiments. This accounts for the statistical likelihood of obtaining significant results by chance. We use a significance level of 0.05 in all the tests.

However, ANOVA can not provide information on which two clusters are different in terms of audience ratings. For this purpose, we perform pair-wise t-tests [29] for each pair of the clusters. We perform Bonferroni Correction by multiplying the obtained p-values with the total number of repeated tests (i.e. $\binom{k}{2}$, where k is the number of clusters) and the number of audience ratings. Once we find any significant difference in the audience ratings within a pair of clusters, we compute

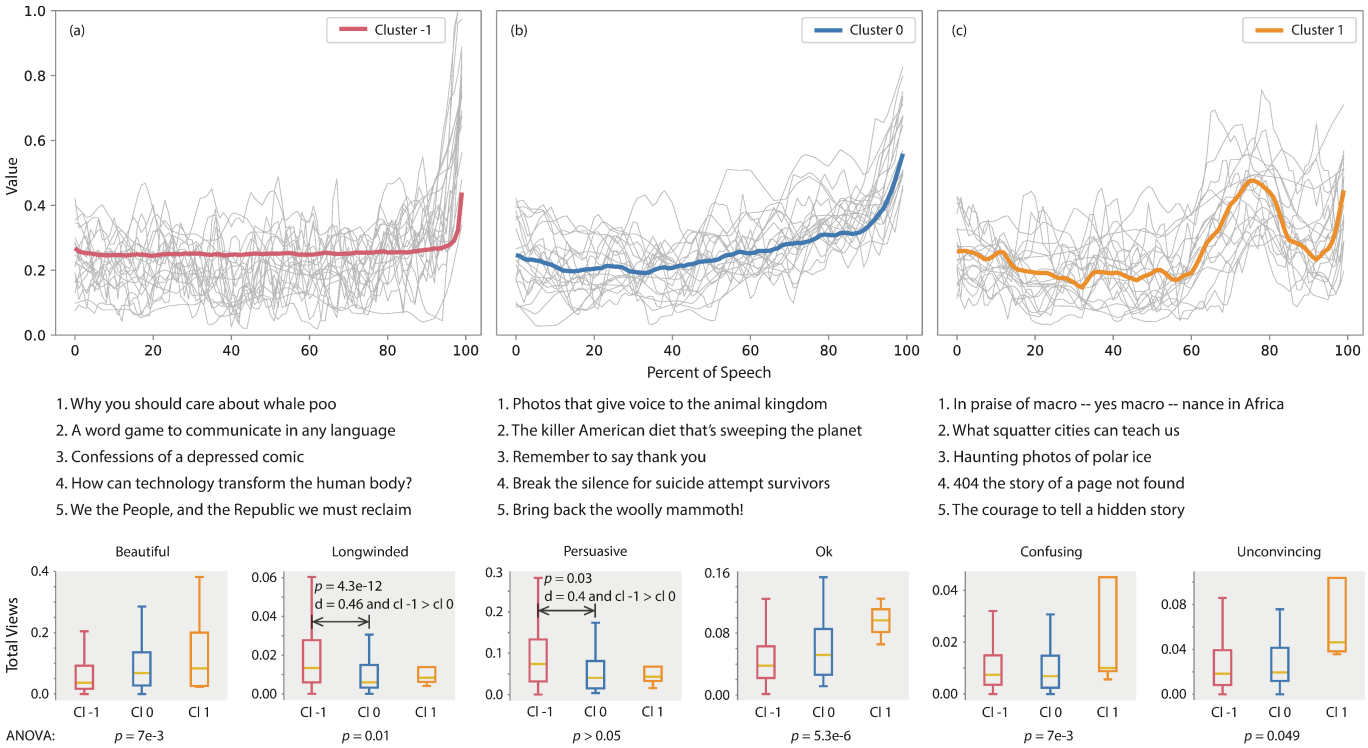

Figure 6. Three clusters as obtained by the DBSCAN algorithm for Joy score.

the effect size using Cohen's d [8]. Along with the effect size, it also provides information about which cluster in a pair has higher audience ratings than the other.

Results of Cluster Analysis

In Figure 6, we show the three clusters found using the DBSCAN algorithm for trajectories of the "Joy" score. The upper row in the figure (a, b and c) shows the clusters with ID -1, 0 and 1. The ensemble average of the trajectories for each cluster is shown in a bold line. The thinner background lines show the trajectories of the top twenty individual TED talks similar to the cluster means. The middle row shows the titles of the top five TED talks. Notice that the average of cluster -1 looks mostly like a flat line. The average tends to increase gradually in cluster 0 and it shows a peak towards the end in cluster 1. Notably, all the clusters show an increase in the mean joy score towards the end of the talks—reinforcing the phenomenon we described in the section Global Trends in the Narrative Trajectories.

The bottom row illustrates the box plots of the subjective ratings. The numbers below the box plots represent the p-values in ANOVA. If there is a significant difference between any two clusters in the pair-wise t-tests, the annotations within the boxplots represent the Bonferroni corrected p-values. Those annotations also show the effect size (Cohen's d) and the directionality of the audience ratings. We observe from the results of the ANOVA test that the ratings "Beautiful", "Longwinded", "OK", "Confusing", and "Unconvincing" show statistically significant difference over the three clusters of the trajectories. The pairwise t-tests show that the flat trajectory (cluster -1) receives a significantly ($p = 4.3e-12 << 0.05$) more "Longwinded" rating than the gradually increasing trajectory (cluster 0). We could not find any significant difference in the "Persuasive" rating from the ANOVA test. However, the pair-wise t-test shows that there is significant difference in Persuasiveness between cluster -1 and cluster 0.

Summary for All Other Scores

The clustering of the narrative trajectories provide us with insights regarding audience perception. Unfortunately, it is impossible to show the complete set of results here due to space constraints. We summarize the cluster analysis and the most prominent observations in Table 4. In the rows of Table 4, we discuss the trajectories for each objective score. In the "Total Cluster" column, we provide the total number of clusters revealed by the DBSCAN [16] algorithm. The next column represents the names and the p-values of the subjective ratings which are found significantly different in the clusters of ANOVA test. The rightmost column describes the effect of the clusters on the audience ratings. We also describe the shapes of the cluster-means in this column. The actual shapes of the cluster-means are included in the supplementary material. Our purpose here is to shed light on the concurrences of the temporal patterns with subjective ratings from a data scientific approach. We do not imply any causation or condition for successful public speaking—which are subject to manual analysis of the data from a psycholinguistic point of view.

	Scores	Total Clust.	Results of ANOVA (with Bonferroni)	Results of pair-wise t-tests (with Bonferroni corrections) and the effect sizes (Cohen's d)
Language Scores	Analytical	4	OK (p=4.7e-6)	Nothing Significant
	Confident	6	Beautiful (p=0.02) OK (p=1e-8)	A flat trajectory of confidence (cluster -1) is rated more "longwinded" than a trajectory with a peak in the beginning (cluster 0, p=1.4e-8, d=0.55) or peaks towards the end (cluster 3, p=0.003, d=0.53).
	Tentative	7	Beautiful (p=9e-6) OK (p=1e-4)	A flat trajectory of tentative score (cluster -1) is rated more "funny" than starting off highly tentative (cluster 4, p=0.001, d=0.4). However, it (cluster -1) is rated more "longwinded" than trajectories with high values in the beginning (cluster 4, p=0.029, d=0.51), in the middle (cluster 0, p=7.8e-11, d=0.63) or in the 2nd quarter (cluster 1, p=0.01, d=0.54)
Emotion Scores	Anger	5	Beautiful (p=1.24e-4)	A flat trajectory of anger (cluster -1) is rated more "longwinded" (p=3.4e-6, d=0.61) than the trajectory with relatively higher value in the first half (cluster 1).
	Joy	3	Beautiful (p=7e-3) Confus. (p=7e-3) Longwin. (p=0.01) OK (p=5.3e-6) Unconv. (p=0.049)	A flat trajectory of joy (cluster -1) is rated more "persuasive" (p=0.03, d=0.4) and "longwinded" (p=4.3e-12, d=0.46) than an increasing trajectory (cluster 0).
	Sadness	3	Unconv. (p=0.006) OK (p=0.002)	A flat trajectory of sadness (cluster -1) is rated more "obnoxious" (p=0.02, d=0.42), "confusing" (p=0.046, d=0.54), "funny" (p=0.003, d=0.43), "longwinded" (p=1.7e-4, d=0.62), and "unconvincing" (p=2.5e-4, d=0.46) than the trajectory with a peak in the middle (cluster 1). Additionally, a flat trajectory is rated more "longwinded" than a gradually decreasing trajectory (cluster 0, p=1.7e-7, d=0.55).
	Fear	2	OK (p=0.014)	A flat trajectory of fear (cluster -1) is rated more "courageous" (p=0.02, d=0.47) and "longwinded" (p=6e-4, d=0.5) than one with a peak in the middle (cluster 0).
	Disgust	7	OK (p=9.2e-8) Unconvincing (p=0.007)	Disgust score with a flat trajectory (cluster -1) is rated more "longwinded" than trajectories with high values in the beginning (cluster 1, p=0.014, d=0.55), or in 2nd quarter (cluster 2, p=0.01, d=0.54) or in 3rd quarter (cluster 3, p=0.009, d=0.65).
Social Scores	Agree-ableness	5	Unconv. (p=0.016) OK (p=1e-8)	Flat agreeableness trajectory (cluster -1) is rated more "funny" than a trajectory ending with high values but low values in the middle (cluster 3, p=0.01, d=0.46). Flat trajectory is more "longwinded" than a hat-shaped (cluster 1, p=0.029, d= 0.55) or a bowl-shaped trajectory (cluster 0, p=9.1e-12, d=0.49)
	Consci-entious-ness	3	Fascina.(p=4.2e-8) Longwin.(p=0.01)	Flat trajectory (cluster -1) of conscientiousness score is rated more "fascinating" than a gradually increasing trajectory (cluster 0, p=4.2e-8, d=0.54). Flat trajectory is rated more "longwinded" than the gradually increasing (p=1.5e-10, d=0.52) or a largely alternating trajectory (cluster 1, p=1.2e-4, d=0.6).
	Emotion Range	4	Beautiful (p=1e-4)	Flat trajectory of emotional range (cluster -1) is more "funny" (p=1.7e-8, d=0.49) but "longwinded" (p=0.005, d=0.6) than the one with a trough in the middle (cluster 2). Flat trajectory is even more "longwinded" (p=6.6e-7, d=0.52) than the one ending with high values (cluster 1) of emotion range towards the end.
	Extra-version	5	Beautiful (p=5e-4)	Trajectories with flat extraversion (cluster -1) is rated more "funny" than the one where it considerably increases in second half (cluster 3, p=1.3e-13, d=0.5) but the former is rated more "longwinded" than alternating one (cluster 0, p=0.03, d=0.64).
	Open-ness	4	Beautiful (p=0.001)	Flat trajectory of openness (cluster -1) is rated more as "courageous" than one when openness suddenly drops in the middle (cluster 2, p=0.04, d=0.55). Flat trajectory is rated more "longwinded" than trajectories with a peak (cluster 0, p=0.008, d=0.53) or trough (cluster 2, p=4.6e-5, d=0.58) in the middle.

Table 4. Summary of the cluster analyses. Actual plots of the ensemble averages of the clusters are provided in the supplementary materials. These figures are also available in http://www.cs.rochester.edu/hci/currentprojects.php?proj=tedtalk.

DISCUSSION

The most prominent observation from Table 4 is that a flat trajectory is more likely to be rated "Longwinded" than other trajectories. We notice this effect for all the scores except "Analytical". We think this result underscores the importance of variations in the narrative trajectories. It is consistent with the existing prior research—the plot of the story should progress through changes [1] to hold the motivation and attention of the audience.

We also observe some other interesting results from Table 4. For example, a wider variation is not always better than a flat trajectory. If a talk starts with a highly tentative language, it is likely to be less funny. A large amount of sadness in the middle of the talk is also not very funny. More interestingly, the audience is likely to rate a talk less funny if it uses less agreeable language in the middle. Similarly, a peak of fear in the middle of a talk is not rated as courageous.

We notice that none of the experimental results cause a direct conflict with our intuitions. Overall, we find it interesting to observe such a strong distributional difference in the audience ratings over the clusters; while the clusters, themselves, are computed using only objective measurements. These analyses, besides providing insights about the ratings, also show an objective way to forecast them. It could motivate future research on discovering the causes of such audience reactions.

In the following subsections, we provide a few recommendations that people can incorporate in their speeches, towards generating a sense of awe in their audience. However, be mindful that these are coupled with our subjective interpretation of the obtained results.

Vary the Emotions

We discussed previously that variations in scores are less likely to lead to "Longwinded" ratings than the flat trajectories. It implies that it is good to walk the audience through a seesaw in emotional states as the speech progresses.

Build a Great Ending

TED talks are generally perceived as high quality public speeches. If we want to identify a single pattern that all of the TED talks show on average, it is the characteristic of finishing with a positive note. The last part of the speech remains fresh in the memory of the audience, fading their perception of possible imperfections before. Therefore, a strong and positive effect in the last part would help making the speech memorable for the audience.

Initiate a Snowball Effect

The total views follow a distribution similar to power-law. It might indicate the *"rich gets richer"* phenomenon. If that is the case, publicizing the talks using social media or any other network will have a strong impact on increasing the view counts. Publicizing would generate some initial views which would result in a snowball effect in increasing view counts. Although this may sound trivial, this understanding has far reaching impact. If the number of views are affected by a strong network of influence, it might not properly reflect the artistic value or contextual merit of the talk. In other words, extremely high view counts of a TED talk might not be the result of an equally great public speech. Therefore, in the future research, the number of views should be considered separately than the measure of performance or artistic value of a public speech.

CONCLUSION

In summary, we strive to analyze if the narrative trajectories exist in public speaking and if it has any impact on the audience ratings. Our analysis reveals the existence of several major patterns (clusters) in narrative trajectories. The clusters show statistically significant differences in the audience ratings. The relation of audience ratings with the narrative trajectories provides insights on the behavior of the audience. The results could motivate future research on determining the cause of such audience responses. Additionally, the narrative trajectories and the corresponding cluster analyses were computed objectively in an automated analysis technique. This kind of experiment is reproducible, scalable, and its generalization verifiable over different domains. Objective analysis also makes it possible to build computer algorithms that could automatically predict audience responses from analyzing the transcripts. This approach could potentially be useful in building automated systems to help people practice and prepare their own public speeches. To validate this claim, we attempted classification and regression tasks using features collected from narrative trajectories and off-the-shelf prediction techniques. Our results show that even these simple techniques could discriminate between highly rated and poorly rated TED talks with accuracy as high as 80% (AUC 0.88) which is much higher than random chance.

Finally, it is likely that the speakers communicate additional information or inspire deeper connection to the audience through skillful prosody, facial expressions, and gestures. These parameters could impact the way the audience rates a speech. Our analyses and insights in this paper are limited to only the spoken sentences, and not the nonverbal features. It remains part of our future work.

Acknowledgments We thank the anonymous reviewers for their thorough reviews and excellent advices. The paper improved significantly by addressing their concerns. This work was supported in part by Grant W911NF-15-1-0542 with the US Defense Advanced Research Projects Agency (DARPA) and the Army Research Office (ARO). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

1. H Porter Abbott. 2008. *The Cambridge introduction to narrative*. Cambridge University Press.
2. Christopher M Bishop. 2006. *Pattern recognition and Machine Learning*. Vol. 128. Springer.
3. Mark Bubel, Ruiwen Jiang, Christine H Lee, Wen Shi, and Audrey Tse. 2016. AwareMe: Addressing Fear of Public Speech through Awareness. In *CHI Conference Extended Abstracts*. ACM, 68–73.

4. Peter Bull. 2016. Claps and Claptrap: The Analysis of Speaker-Audience Interaction in Political Speeches. *Journal of Social and Political Psychology* 4, 1 (2016), 473–492.
5. Jilin Chen, Gary Hsieh, Jalal U Mahmud, and Jeffrey Nichols. 2014. Understanding individuals' personal values from social media word use. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 405–414.
6. Lei Chen and Chong MIn Lee. 2017. Convolutional Neural Network for Humor Recognition. *arXiv preprint arXiv:1702.02584* (2017).
7. Mathieu Chollet and Stefan Scherer. 2017. Assessing Public Speaking Ability from Thin Slices of Behavior. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*. IEEE, 310–316.
8. Jacob Cohen. 1977. Statistical power analysis for the behavioral sciences (revised ed.). (1977).
9. Keith Curtis, Gareth JF Jones, and Nick Campbell. 2015. Effects of good speaking techniques on audience engagement. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 35–42.
10. 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.
11. Sophia Elizabeth De Morgan. 1882. *Memoir of Augustus De Morgan*. Longmans, Green, and Company.
12. Anda Drasovean and Caroline Tagg. 2015. Evaluative Language and Its Solidarity-Building Role on TED. com: An Appraisal and Corpus Analysis. *Language@ Internet* 12 (2015).
13. Olive Jean Dunn. 1961. Multiple comparisons among means. *J. Amer. Statist. Assoc.* 56, 293 (1961), 52–64.
14. David Easley and Jon Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, Chapter Power Laws and Rich-Get-Richer Phenomena, 543–560.
15. Paul Ekman and Wallace V Friesen. 1971. Constants across cultures in the face and emotion. *Journal of personality and social psychology* 17, 2 (1971), 124.
16. Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, and others. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *Kdd*, Vol. 96. 226–231.
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. Carmine Gallo (Ed.). 2014. *Talk Like TED*. Emerald Group Publishing Limited.
19. Jianbo Gao, Matthew L Jockers, John Laudun, and Timothy Tangherlini. 2016. A multiscale theory for the dynamical evolution of sentiment in novels. In *Behavioral, Economic and Socio-cultural Computing (BESC), 2016 International Conference on*. IEEE, 1–4.
20. Lewis R Goldberg. 1993. The structure of phenotypic personality traits. *American psychologist* 48, 1 (1993), 26.
21. Liang Gou, Michelle X Zhou, and Huahai Yang. 2014. KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 955–964.
22. James A Hanley and Barbara J McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 1 (1982), 29–36.
23. IBM. 2017a. The science behind the service. `https://console.bluemix.net/docs/services/tone-analyzer/science.html#the-science-behind-the-service`. (December 2017).
24. IBM. 2017b. Tone Analyzer, Understand emotions and communication style in text. `https://www.ibm.com/watson/services/tone-analyzer/`. (December 2017).
25. Matthew L. Jockers. 2015a. *The Rest of the Story*. `http://www.matthewjockers.net/2015/02/25/the-rest-of-the-story/`
26. Matthew L. Jockers. 2015b. *Syuzhet: Extract Sentiment and Plot Arcs from Text*. `https://github.com/mjockers/syuzhet`
27. Zhe Liu, Anbang Xu, Mengdi Zhang, Jalal Mahmud, and Vibha Sinha. 2017. Fostering User Engagement: Rhetorical Devices for Applause Generation Learnt from TED Talks. *arXiv preprint arXiv:1704.02362* (2017).
28. Wincenty Lutosławski. 1897. *The origin and growth of Plato's logic: with an account of Plato's style and of the chronology of his writings*. Longmans, Green and Company.
29. John H McDonald. 2009. *Handbook of biological statistics*. Vol. 2. Sparky House Publishing Baltimore, MD.
30. Saif Mohammad. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. Association for Computational Linguistics, 105–114.

31. Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. 29, 3 (2013), 436–465.
32. Kevin P Murphy. 2012. *Machine learning: a probabilistic perspective*. MIT press.
33. Anh-Tuan Nguyen, Wei Chen, and Matthias Rauterberg. 2012. Online feedback system for public speakers. In *E-Learning, E-Management and E-Services (IS3e), 2012 IEEE Symposium on*. IEEE, 1–5.
34. Sampo V Paunonen and Michael C Ashton. 2001. Big five factors and facets and the prediction of behavior. *Journal of personality and social psychology* 81, 3 (2001), 524.
35. F. Pedregosa and others. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
36. Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. *Theories of emotion* 1, 3-31 (1980), 4.
37. David Martin Powers. 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. (2011).
38. Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science* 5, 1 (2016), 31.
39. Spyridon Samothrakis and Maria Fasli. 2015. Emotional sentence annotation helps predict fiction genre. *PloS one* 10, 11 (2015), e0141922.
40. Alex J Smola and Bernhard Schölkopf. 2004. A tutorial on support vector regression. *Statistics and computing* 14, 3 (2004), 199–222.
41. Carlo Strapparava, Alessandro Valitutti, and others. 2004. WordNet Affect: an Affective Extension of WordNet.. In *LREC*, Vol. 4. 1083–1086.
42. Cassidy R Sugimoto, Mike Thelwall, Vincent Larivière, Andrew Tsou, Philippe Mongeon, and Benoit Macaluso. 2013. Scientists popularizing science: characteristics and impact of TED talk presenters. *PloS one* 8, 4 (2013), e62403.
43. 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 *20th International Conference on Intelligent User Interfaces*. ACM, 286–295.
44. 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 *21st International Conference on Intelligent User Interfaces (IUI '16)*. ACM, New York, NY, USA, 385–396.
45. 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.
46. Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 267–288.
47. Andrew Tsou, Mike Thelwall, Philippe Mongeon, and Cassidy R Sugimoto. 2014. A community of curious souls: an analysis of commenting behavior on TED talks videos. *PloS one* 9, 4 (2014), e93609.
48. Vladimir Vapnik and Alexey Chervonenkis. 1964. A note on one class of perceptrons. *Automation and remote control* 25, 1 (1964).
49. Kurt Vonnegut. 1981. *Palm Sunday: An Autobiographical Collage*. New York: Dell.
50. Yichen Wang and Aditya Pal. 2015. Detecting Emotions in Social Media: A Constrained Optimization Approach.. In *IJCAI*. 996–1002.