Acoustic/Prosodic and Lexical Correlates of Charismatic Speech

Andrew Rosenberg, Julia Hirschberg

Department of Computer Science
Columbia University, New York, USA
{amaxwell,julia}@cs.columbia.edu

Abstract

Charisma, the ability to command authority on the basis of personal qualities, is more difficult to define than to identify. How do charismatic leaders such as Fidel Castro or Pope John Paul II attract and retain their followers? We present results of an analysis of subjective ratings of charisma from a corpus of American political speech. We identify the associations between charisma ratings and ratings of other personal attributes. We also examine acoustic/prosodic and lexical features of this speech and correlate these with charisma ratings.

1. Introduction

Charismatic individuals are generally identified as those able to persuade and to command authority by virtue of their personal qualities rather than by formal institutional or military power [11]. How such charismatic leaders attain authority is a question of considerable theoretical debate: some see charisma arising from the faith of a leader's *listener-followers*[8], while others from the combination of a *gift of grace*, *an inspiring message* and an *important crisis*[3]. However, all would agree that charismatic leaders share a particular ability to communicate: Charismatic leaders such as Martin Luther King Jr., Fidel Castro, Adolf Hitler, and Pope John Paul II are well known for their gifts in public speaking.

In this paper, we investigate the spoken characteristics of charisma. Our motivation is twofold: On a scientific level, we are interested to learn whether speakers judged charismatic share certain acoustic and prosodic characteristics, and how these interact with lexical content and syntactic form. On a technological level, we believe that defining a set of objective measures of charisma will permit interesting work in both speech synthesis and speech understanding: First, it would permit the production of 'charismatic speech' for speech generation applications that require high degrees of persuasion on the part of the system, including advertisements and political telephone solicitation. Second, it migh facilitate the identification of speakers who are likely to emerge as effective political leaders. And finally, it would support the creation of online training systems that help individuals to become more charismatic speakers.

In this paper, we examine spoken cues to charisma. In Section 2 we discuss previous research in sociology and rhetoric on charisma. In Section 3, we describe an online experiment we conducted to elicit subject judgments of charisma and other personal attributes of speakers of tokens of public speech. In Section 4 we identify correlations among subject ratings of charisma and other attributes, as well as the effect of speaker, genre, and topic on subject decisions. In Section 5 we discuss the analyses we performed on the speech tokens and the correlations we found between acoustic/prosodic and lexical features and subject judgments of charisma. We conclude in Section 6 and describe future research.

2. Previous work

Following Weber's [11] discussion of "charismatic authority" as a legitimate source of leadership, sociologists and rhetoricians have attempted to define the nature of charisma (e.g., Bettinghaus [2], Marcus [8], Boss [3], Barker [1]). However, little, if any, attention has been given to constructing an *empirical* definition of charisma, although various empirical studies **have** been conducted on related phenomena.

Hamilton and Stewart [7] propose an information processing model of persuasion. They describe subject ratings of dynamism, competence and trust when a message's *intensity* is manipulated and characterize *the charisma sequence* in terms of the interaction of intensity manipulation with these ratings. In an attempt to quantify communicator credibility, Tuppen [10] describes an experiment in which subjects were asked to read short character sketches of ten communicators, and rate each of them on 28 bipolar adjective scales and 36 7-point Likert scales. These subject ratings were clustered and the most similarly rated scales were used to define the cluster. Tuppen assigns the label 'charisma' to a cluster defined by the following 'communicator' adjectives: "convincing, reasonable, right, logical, believable, intelligent; whose opinion is respected, whose background is admired, and in whom the reader has confidence" (p. 257). Again, however, the ascription of the label 'charisma' to this group is Tuppen's own, and does not arise from the experiment.

3. Data collection

Eight native American English speakers with no reported hearing problems were presented with 45 speech segments of between 2 and 28 seconds in length via a standard web browser. Using a web form, the subjects were asked to indicate their agreement with a set of 26 statements about the speaker of a given audio token, on a five-point Likert scale. The token was played by the web browser simultaneously with the presentation of the form. The clip was repeated with two seconds of silence between iterations until the subject had responded to all 26 statements, and had moved on to the next segment. The order of presentation of the 45 tokens was randomized for each subject. Additionally, the order of the 26 statements was randomized for each token. At the end of the survey, users were asked to indicate the names of any speakers they had recognized. It took users an average of 1.5 hours to complete the survey. The shortest time taken by any subject was 49.5 minutes; the longest ∼3 hours.

The materials for the experiment were chosen to represent a variety of speakers, topics, and genres. Since we prepared the materials in the winter and spring of 2004, there was abundant material readily available online for the nine candidates running for the Democratic Party's nomination for President: Sen. John Kerry, Rep. John Edwards, Gov. Howard Dean, Rep. Richard Gephardt, Rev. Al Sharpton, Amb. Carol Mosley-Braun, Rep. Dennis Kucinich, Gen. Richard Clark and Sen. Joseph Lieberman. We chose speakers from the political field for a number of reasons. We hypothesized that at least some of these politicians would demonstrate charismatic qualities in their speech. Also, the varied activities of the candidates ensured that speech would be available from different genres: interviews, debates, stump speeches, and campaign ads. We limited our speakers to Democrats to confine the range of opinions presented in the tokens, as it had been suggested in the literature [11, 3, 6] that a listener's agreement with a speaker bears on their judgment of the speaker's charisma. The topics we selected segments from were deliberately varied to minimize effect of topic on judgments of charisma. We included five speech tokens from each speaker, one on each of the following topics: healthcare, postwar Iraq, Pres. Bush's tax plan, the candidate's reason for running, and a content-neutral topic (e.g., greetings). Since the speech tokens came from a variety of sources and recording conditions, we normalized the tokens for intensity to -12dBFS.

From a large set of segments which fit the above criteria, we then screened the potential tokens to judge for ourselves whether a token 'sounded charismatic' or not. This rough evaluation was used to balance the 'charismatic' tokens across speakers and topics.[1] In total, 22 of the 45 tokens used in the experiment were judged 'charismatic' by the authors. Due to experimenter error, one of the speech segments was mislabeled, leading to duplicate presentation of one token (Rep. Edwards' reason for running), and an omission of another (the content neutral statement from Rep. Gephardt). While this skewed the balanced composition of the corpus as a whole, this also allowed us to check for rater consistency.

Subjects rated statements of the form "The speaker is X", where X was one of the following: charismatic, angry, spontaneous, passionate, desperate, confident, accusatory, boring, threatening, informative, intense, enthusiastic, persuasive, charming, powerful, ordinary, tough, friendly, knowledgeable, trustworthy, intelligent, believable, convincing, reasonable. The attributes queried were a subset of those often associated in the literature with charisma. We also included "The speaker's message is clear." and "I agree with the speaker" as statements to be rated.

4. Analysis of subject judgments

4.1. Across subject agreement on ratings

We first examined overall subject agreement on ratings for all tokens and statements. We used the weighted kappa statistic [4] with quadratic weighting to determine the inter-subject agreement. The mean κ value over all 45 tokens and 26 statements was 0.213. This is rather low agreement, suggesting a fair amount of individual variation in the ratings of at least some of the 26 statements or some of the tokens. In order to identify potential sources for this variation, the kappa contribution from each of the 45 tokens was examined individually. This breakdown allowed us to determine which of the tokens were most and least consistently ranked across subjects. Similarly, we computed the kappa contribution from ratings of each of the 26 statements.

We found no significant differences in kappa values across tokens used in the experiment. There is, however, a substantial range of inter-annotator agreement with respect to the 26 individual statements. Of particular note is the contrast between the statements that showed the greatest and least agreement. Tables 1 and 2 contain the five statements with the highest and lowest kappa scores, respectively. Statements corresponding to dynamic, high activation emotions (accusativeness, passion, intensity, anger, enthusiasm) ranked among those most consistently rated. However, agreement on ratings of trust, reasonability, believability, desperation, and ordinariness rank hardly greater than what would be expected by chance. This might arise from subjective differences with respect to perceptions of qualities such as trustworthiness or believability. Alternately, subjects may be skeptical of political speech, and therefore reluctant to ascribe qualities such as 'being reasonable' to politicians, while emotions such as anger and enthusiasm may be less evaluative.

Ratings of the statement "The speaker is charismatic" (henceforth referred to as 'the charismatic statement') yielded a kappa score of 0.224. This places it as the eighth most consistently labeled statement. While this score represents modest agreement, it is of note that subjects agree about charisma more than about such qualities as intelligence ($\kappa = 0.119$) ("The speaker is intelligent") and confidence ($\kappa = 0.215$) ("The speaker is confident").

Table 1: Statements with the most consistent inter-subject agreement in the speech survey.

statement	κ
The speaker is accusatory.	0.512
The speaker is passionate.	0.458
The speaker is intense.	0.431
The speaker is angry.	0.404
The speaker is enthusiastic.	0.362

Table 2: Statements with the least consistent inter-subject agreement in the speech survey.

statement	κ
The speaker is trustworthy.	0.037
The speaker is reasonable.	0.070
The speaker is believable.	0.074
The speaker is desperate.	0.076
The speaker is ordinary.	0.115

4.2. Correlation of statement ratings

One of the goals of this study is to construct a functional definition of 'charismatic' by determining how subjects associate this attribute with other attributes. To that end, we examined which statement ratings positively and negatively correlated with those of charisma. We again applied Cohen's kappa statistic with quadratic weighting. We considered each statement as a 'subject', and calculated the pairwise inter-statement agreement between the charismatic statement and each of the 25 others over all subject ratings. Those statements that demonstrated the greatest positive or negative correlation with the charismatic statement appear in Table 3. The elements of this list support Dowis' [6] and Boss' [3] claims that e.g. enthusiasm and passion are positively correlated with charisma and boringness is negatively correlated. The desperate, threatening, accusatory

[1] Segments that we could not agree on, or considered to be only modestly charismatic were not included in the corpus.

and angry statements show no positive or negative ($|\kappa| < 0.15$) correllation with the charismatic. It is particularly interesting that ratings of a speaker's anger (shown to be consistently rated across subjects in Section 4.1) have no impact in either direction on a subject's judgment of the speaker's charisma.

Table 3: Statements showing the most consistent positive and negative correlation with the charismatic statement.

statement	κ
The speaker is enthusiastic.	0.606
The speaker is charming.	0.602
The speaker is persuasive.	0.561
The speaker is boring.	-0.513
The speaker is passionate.	0.512
The speaker is convincing.	0.503

4.3. Influence of speaker, topic and genre on charisma ratings

The speaker of a segment significantly influences ($p = 1.75e{-}10^2$) subjects' ratings of charisma. The three most charismatic speakers in our study were, in order, Rep. Edwards (mean rating 3.73), Rev. Sharpton (3.40) and Gov. Dean (3.32). The three least charismatic were Sen. Lieberman (2.38), Rep. Kucinich (2.73), and Rep. Gephardt (2.77). Upon completion of the survey, subjects were asked to report any speakers whom they recognized. The mean number of speakers recognized was 3.25 of the 9 speakers with a maximum of 6 and a minimum of 0. Subjects rated tokens spoken by a recognized speaker as more charismatic (mean rating 3.28) than those spoken by unrecognized speakers (mean rating 2.99). This difference is significant with $p = .007$. This may imply that familiarity with a speaker positively influences perceptions of charisma, or that charismatic speakers are more recognizable than uncharismatic speakers.

The genre in which the speech token was delivered does significantly influence subject ratings of charisma ($p = 0.0058$). Speakers are rated as more charismatic when they are delivering a stump speech (mean rating 3.28) than when they are being interviewed (2.90). Speech segments extracted from debates (3.10) were rated in line with the overall mean (3.10) with respect to charisma. The corpus contained only one segment that was taken from a campaign advertisement; while this segment was rated as below average in charisma (2.88), this obviously should not be taken as reflective of the genre as a whole. The impact of genre on subject ratings may be easily explained: The enthusiasm and dynamism that can be appropriately conveyed during a stump speech — at least, by speakers who **can** convey charisma — may be less appropriate in an interview.

The topic of the segments used in our experiment (postwar Iraq, healthcare, taxes, reason for running, content-neutral) had no statistically significant impact on subjects' ratings of charisma. While the semantic content of a particular speech segment may contribute to perceived charisma, the general topic does not appear either to promote or to inhibit charismatic behavior.

4.4. Influence of order of presentation on charisma

As we noted in Section 3, due to an error, one of our speech tokens was presented to subjects twice. So we were able to compare subject ratings on the two different presentations of the same token to measure consistency. While no subject ratings varied significantly between presentations (mean difference of .4), ratings of the tough, ordinary and charismatic statements varied the most. Further study is necessary to determine if this is an artifact of this particular token, indicative of a priming effect, or due to some other property of these attributes.

5. Lexical and acoustic analysis

As described in Section 4.1, subjects agree with some consistency on ratings of charisma. What characteristics of **what** is said and **how** it is said might explain this consistency? In this section, we examine potential correlations between ratings of charisma and a variety of lexical and acoustic/prosodic properties across all subjects.

5.1. Lexical properties of charismatic speech

We first examined lexical features, including the number of words in the token, ratio of function to content words, pronoun density, and a measure of lexical complexity. The amount of spoken material, as determined by length in words, significantly influenced judgments of charisma with $p = .026$. The more speech that was presented, the more charismatic the speaker was perceived. We next looked at the ratio of function (e.g. prepositions, determiners) to content words (e.g. nouns, verbs) in each token. That is, perhaps the more relative content there is in a message, the more likely it is that content can influence is charisma rating. However, this measure did not significantly influence ratings of charisma in our study. We also examined density of pronouns (ratio of pronouns to total words) broken out by first, second and third person. The literature on charisma suggests that charismatic individuals have a personal appeal to their followers. Such terms as 'father figure' are often used about such leaders. Thus, the presence of first and second person pronouns might characterize charismatic speech. In our study, only the density of first person pronouns significantly influenced subject ratings of charisma ($p = .023$). No other pronoun measures showed any significant influence. So, at least some aspect of 'personal' speech seems to be present in charismatic speech. Dowis [6] posits that simpler words are more effective than complex terms in delivering a charismatic message. He proposes a simple measure of the complexity of a lexical item — the number of syllables it has. However, when we compute the number of syllables per word for each token, we find that this metric influences ratings of charisma in the opposite direction to that predicted by Dowis. That is, greater mean syllables per word corresponds to higher ratings of charisma; or, more 'complex' words characterize charismatic speech. This influence is significant with $p = .034$. We hesitate to generalize too broadly here, but our findings present at least one empirical contradiction to Dowis' anecdotal claims.

5.2. Acoustic/prosodic properties of charismatic speech

We examined pitch, intensity, speaking rate, and durational features of the tokens in our experiment and then measured the degree of correlation between these features and subject ratings of the charismatic statement. We also examined certain properties of component intonational phrases and performed similar correlations.

We first examined (raw) mean, standard deviation, maximum, and minimum f0 for all male speakers; All of these properties, with the exception of minimum f0, positively in-

[2] All p values below are determined by one-way ANOVA with repeated measures

fluenced ratings of charisma below the $p < .001$ level (mean $p = 2.59e{-7}$; standard deviation $p = 7.3e{-4}$; max $p = 3.17e{-5}$). Minimum f0, too, was significant with $p = .049$. The greater the mean and standard deviation, the greater the perceived charisma. The high standard deviation of pitch may correspond to an increase in expressiveness in the utterance. This in turn, may signal some of the other attributes that correlate highly with charisma, such as enthusiasm (cf. Section 4.2) and dynamism, predicted in the literature by Boss [3], and Tuppen [10]. When we normalize these features by calculating *zscores* for each speaker (to control for gender), only the zscore of a token's mean f0 is significantly correlated (positively) with charisma ratings ($p = .026$). That is, when a token is higher in the speaker's pitch range, it is rated more charismatic. Standard deviation of f0 over all speakers, male and female, is significant with $p = 3.57e - 3$.

Intensity might also provide cues to ratings of charisma; louder messages might convey a more charismatic impression. Since we normalized all tokens for intensity (cf. Section 3), however, we can only examine mean and standard deviation in our tokens. Only mean intensity approaches significance ($p = .111$), with louder utterances positively correlated with charisma ratings, as we predicted.

Speaking rate (syllables per second) was also calculated for each token and compared to ratings of charisma; the correlation was significant with $p = .085$. A faster speaking rate indicates a higher charisma rating. Nothing we have found in the literature addresses this characteristic. Further experimentation is required to determine the nature of the interaction between speaking rate and charisma.

Using hand-labeled intonational phrase boundaries (ToBI level 3 or 4) [9]), we were able to examine some phrase level acoustic/prosodic features of our tokens. We examined the number of phrases in each token, the mean and standard deviation of the (normalized) maximum and mean pitch, and the mean and standard deviation of the intensity (calculated over segmentals only) across phrases within the token. For ToBI level 3 phrases, only the standard deviation of the normalized maximum pitch approaches significant correlation with ratings of charisma ($p = .108$). So, tokens whose individual phrases varied considerably in maximum pitch (i.e., in pitch range) were rated as more charismatic than those with less variation. For ToBI level 4 indices, the mean ($p = .004$) and standard deviation ($p = .034$) of the normalized maximum intensity as well as the number of words per phrase ($p = .026$) are all significantly (positively) correlated with ratings of charisma. Such rapid change in pitch and intensity may well correspond to the other charisma-correlated attributes, such as passion and enthusiasm. The number of smaller (level 3) phrases within a larger (level 4) phrases is also positively correllated with charisma ($p = .041$), while the mean number of such phrases approaches significance ($p = .109$). This is consonant with our findings that greater number of words and longer utterances are associated with higher ratings of charisma.

6. Conclusions and Future Research

In this paper, we have presented results of a study of charismatic speech, based upon elicited subject ratings of charisma and other personal attributes of speakers in a corpus of American political speech. We have found a significant agreement across subject as to which speech is charismatic and which is not. We have also found that subjects tend to find the same attributes positively correlated with charisma (enthusiastic, charming, persuasive, passionate, convincing) and the same negatively correlated (boring). When we examine the lexical and acoustic/prosodic characteristics of speech tokens rated highly for charisma, we find significant correlations between charisma ratings and a) duration of token in words, seconds, and number of internal phrases (the longer, the more charismatic); b) the number of first person pronouns in the speech token being rated (charismatic speech contains a higher density of first-person pronouns than non-charismatic speech); c) the complexity of lexical items in the token measured in number of syllables per word (the greater the number, the more charismatic the token); d) raw f0 features including mean, standard deviation and maximum for male speakers (greater values correlate with higher charisma ratings) and normalized mean f0 (the greater the mean, the more charismatic); e) mean (raw) intensity (in this case, the louder the token, the more charismatic the speaker is rated); and f) speaking rate (the faster the speech, the more charismatic).

To compare the role of acoustic/prosodic information vs. lexical influences in subject judgments of charisma, we have conducted a similar experiment using textual versions of our spoken materials, as well as other text. Analysis of these experimental results remains to be completed. To begin to examine the cultural dependencies of charismatic speech judgments, we are also preparing a study of charismatic speech and text in Palestinian Arabic.

7. Acknowledgements

The authors would like to thank Judd Sheinholtz, Aron Wahl and Svetlana Stenchikova. This research was supported in part by NSF grant IIS-0325399.

8. References

[1] Barker, E., New Religious Movements: A Practical Introduction. HMSO, London, 1989.

[2] Bettinghaus, E., Persuasive Communication. Holt, Rinehart and Winston, New York, 1969.

[3] Boss, P., “Essential Attributes of Charisma,” Southern speech communication journal, 41(3):300–313, 1976.

[4] Cohen, J., “Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit,” Psychological Bulletin, 70:213–20, 1968.

[5] Cowie, R., Douglas-Cowie E., and Campbell, N. (eds.), Speech Communication 40: Special issue on Speech and Emotion, 2003.

[6] Dowis, R., The Lost Art of the Great Speech. AMACOM, New York, 2000.

[7] Hamilton, A. and Stewart, B., “Extending an Information Processing Model of Language Intensity Effects,” Communication quarterly, 41(2):231–246, 1993.

[8] Marcus, J. “Transcendence and Charisma,” Western Political Quarterly, 14:237–41, 1967.

[9] Silverman K., Beckman M., Pitrelli J., Ostendorf M., Wightman C., Price P., Pierrehumbert J., and Hirschberg J., “ToBI: A standard for Labeling English Prosody,” In Proc. ICSLP’92, 2:867–870.

[10] Tuppen, C., “Dimensions of communicator credibility; An oblique solution,” Speech Monographs, 41(3):253–260, 1974.

[11] Weber, M., The Theory of Social and Economic Organization, trans. Henderson, A. M. and Parsons, T., The Free Press, New York, 1964.