Comparing American and Palestinian Perceptions of Charisma Using Acoustic-Prosodic and Lexical Analysis

Fadi Biadsy, Julia Hirschberg, Andrew Rosenberg, and Wisam Dakka

Department of Computer Science, Columbia University, New York, NY, 10027

{fadi,julia,amaxwell,wisam}@cs.columbia.edu

Abstract

Charisma, the ability to lead by virtue of personality alone, is difficult to define but relatively easy to identify. However, cultural factors clearly affect perceptions of charisma. In this paper we compare results from parallel perception studies investigating charismatic speech in Palestinian Arabic and American English. We examine acoustic/prosodic and lexical correlates of charisma ratings to determine how the two cultures differ with respect to their views of charismatic speech.

1. Introduction

Charismatic leaders have been defined as those who obtain their authority by virtue of personal qualities as opposed to the formal institutions of governance [1]. Such leaders — e.g. American civil rights leader Martin Luther King Jr. or Cuban President Fidel Castro — are often renowned for their communicative gifts. Determining the attributes that make leaders charismatic has been the source of some debate. Some see charisma arising from the faith of a leader's listener-followers [2], while others see it as a combination of a 'gift of grace', an inspiring message and an important crisis [3]. A common quality observed by many investigators is a bond between the charismatic leader and his or her followers. This relationship is heavily influenced by larger social and cultural expectations and prejudices. Tannen [4] identifies a number of pragmatic dimensions that vary cross-culturally including when to talk, formulacity, and degree of indirectness, cohesion and coherence. As the communicative forum extends from dialogue to public speaking, it is likely that the impact of these cultural norms are magnified. In this study, we investigate the differences in perceptions of charisma across two cultures that appear to be divergent: the American and Palestinian cultures. We describe and extend results of a study first reported by Rosenberg and Hirschberg [5], an experiment of native American English speakers judging American English speech. We compare and contrast these results with those from a new study of native Palestinian Arabic speakers judging Palestinian Arabic speech.

In Section 2, we describe our experimental design and materials. We analyze subject judgments in Section 3. Acoustic/prosodic and lexical correlates to subject charisma ratings are presented in Sections 4 and 5. We conclude in Section 6 and present directions for future work.

2. Experiment Design

We conducted perception experiments of charismatic speech with native speakers of Palestinian Arabic and American English. Each experiment differed only in the materials presented to subjects for judgments. In both experiments, subjects with no reported hearing problems were presented with speech tokens in their native language via a standard web browser. Using a web form, the subjects were asked to rate each speaker on 26 questions using a five-point Likert scale. Statements were 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*. These attributes are a subset of those often associated in the literature with charisma. Subjects were also asked to rate the statements: "*The speaker's message is clear*" and "*I agree with the speaker*." Tokens were played simultaneously with the presentation of the form. Each token was repeated with two seconds of silence between iterations until the subject had responded to all 26 statements, and moved on to the next token. Order of presentation of the speech tokens was randomized for each subject, and the order of the 26 statements was randomized for each token. At the end of the survey, users were asked to list the names of any speakers they had recognized.

2.1. The American English Experiment

12 native American English speakers (6 female and 6 male) participated in the English experiment. They were presented with 45 speech segments of 2–28 seconds duration, with a mean of 10 seconds. The full corpus contains 7.5 minutes of speech. Materials were chosen to represent a variety of speakers, and topics. Since we prepared the materials in 2004, there was abundant material available online for the nine candidates running for the Democratic Party's nomination for President. We hypothesized that at least some of these politicians would demonstrate charismatic qualities in their speech. We limited our speakers to Democrats to confine the range of opinions presented in the tokens, as the literature suggests that a listener's agreement with a speaker can affect their judgment of the speaker's charisma [1, 3, 6]. Token topics were balanced to minimize effect of topic on charisma judgments. We included five 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 tokens came from a variety of recording conditions, we normalized the tokens for intensity to -12dBFS. We screened potential tokens to balance those that "sounded charismatic" for each speaker with those that did not.

2.2. The Palestinian Arabic Experiment

12 native Palestinian Arabic speakers (6 females and 6 males) were similarly presented with 44 speech tokens of 3–28 seconds duration, with a mean of 14 seconds, for a total of 10.2 minutes of speech. Using a Modern Standard Arabic[1] (MSA) version of the web form described above, subjects were asked to rate the

[1] MSA is the standard written language, taught in schools throughout the Arab world and is typically used in formal-spoken communication.

same 26 statements for each speech token.

Tokens for this study were selected from several television programs available on the Al-Jazeera News Channel web site (http://www.aljazeera.net) in 2005. Speakers and topics were varied as in the first experiment, topics including: the assassination of the Hamas leader, the debate among the Palestinian groups, The Intifada and resistance, the Israeli separation wall, the Palestinian Authority and calls for reforms. We selected 2 segments from each of 22 male native Palestinian speakers, for a total of 44 tokens. Again, to balance the material across tokens and speakers, we chose one token from each speaker that "sounded charismatic" to 3 native Palestinian informants and one that did not. In contrast to the English study, the intensity of the speech tokens presented to the subjects was not normalized, to allow us to the study the correlation of intensity features with perceptions of charisma.

3. Analysis of Subject Judgments

3.1. Across-subject agreement on ratings

For each study, we examined the overall subject agreement on ratings for all tokens and statements. The weighted kappa statistic [7] with quadratic weighting was used to determine inter-subject agreement. For English subjects the mean pairwise κ over all tokens and statements was 0.207; for Arabic subjects, it was κ=0.225. Thus, overall agreement in both studies is low. To isolate the source of this disagreement, we examined the mean pairwise κ for each statement and token individually. We found a substantial range of inter-rater agreement with respect to the 26 statements on both studies. Table 1 shows the most and least consistently labeled statements in the English study; Table 2 reports the same for the Arabic study.

Table 1: Statements with most and least consistent inter-subject agreement in the English survey.

statement	κ (eng)	κ (arb)
The speaker is accusatory.	0.500	0.274
The speaker is angry.	0.444	0.330
The speaker is passionate.	0.431	.091
The speaker is intense.	0.410	0.336
The speaker is desperate.	.080	.076
The speaker is believable.	.074	0.269
The speaker is reasonable.	.045	0.187
The speaker is trustworthy.	.027	0.224

Table 2: Statements with most and least consistent inter-subject agreement in the Arabic survey.

statement	κ (arb)	κ (eng)
The speaker's message is clear.	0.351	0.135
The speaker is enthusiastic.	0.350	0.351
The speaker is charismatic.	0.348	0.232
The speaker is intense.	0.336	0.411
The speaker is desperate.	0.075	0.080
The speaker is friendly.	0.065	0.206
The speaker is ordinary.	0.041	0.118
The speaker is spontaneous.	0.013	0.149

In the English study, statements that demonstrated the greatest agreement were dynamic, high activation emotions (*accusativeness*, *passion*, *intensity*, *anger*). We observe greater subject differences regarding character assessments of the speaker, such as judgments of how *trustworthy*, *reasonable*, and *believable* a speaker was. We also found low agreement on ratings of how *desperate* a speaker was. These terms may indeed be difficult for subjects to define, compared to more 'classic' emotional states. Ratings of how charismatic a speaker was showed a mean κ of 0.232. While this value is low, it should be recalled that the qualities being assessed are highly subjective and our task conflates two factors: subjects' understanding of a particular concept and subjects' identification of that concept in the speech of a particular speaker.

We observe some differences in statement ratings in the Arabic study. In this study, subjects agreed more strongly on whether the speaker's message was clear or not; in English this statement yields a relatively low agreement. However both groups found it relatively easy to describe the emotional content of high-activation emotions (*enthusiasm*, *intensity*, *anger*). Surprisingly, the charismatic statement is the third most consistently rated statement in the Arabic study. Similar to the English study, we found low agreement on how desperate or ordinary a speaker is. In contrast to English, we found poor agreement on the attributes *'spontaneous'* and *'friendly'*.

3.2. Within-Subject Correlation of Subject Ratings

In these studies, we make no assumption that subjects in both experiments will understand the word *'charisma'* in the same way. In order to determine how they **do** understand the term, we looked at which other attributions were commonly correlated with the charismatic statement — that is, did subjects in either experiment employ a common 'functional' definition of charisma. We examined within-subject correlations of ratings of the charismatic statement with those of the remaining 25 statements. Using the *kappa* statistic to describe this correlation, we report the strongest positive/negative correlates. Our American subjects' 'functional' definition of a charismatic speaker is one who is enthusiastic (κ=0.62), persuasive (κ=0.58), charming (κ=0.58), passionate (κ=0.54), convincing (κ=0.50) and **neither** boring (κ=-0.51) nor ordinary (κ=-0.40). For our Palestinian subjects charismatic speakers are those who are tough (κ=0.69), powerful (κ=0.69), persuasive (κ=0.68), charming (κ=0.66), enthusiastic (κ=0.65), and **neither** boring (κ=-0.45) nor desperate (κ=-0.26).

3.3. Influence of speaker, topic, genre on charisma ratings

In both experiments, as expected, we found that the speaker of a segment significantly influences subjects' ratings of charisma (English: p=$2.2e$-16, Arabic: p=$.0006$).[2] Upon completion of the survey, subjects were asked to report any speakers they believed they had recognized. In the English study, mean number of speakers recognized was 5.8 (of 9) with a maximum of 8 and a minimum of 0. Subjects rated tokens spoken by a (purportedly) recognized speaker as more charismatic (mean rating 3.39) than those spoken by unrecognized speakers (3.0); the difference is significant with p=$5.0e$-7. This may imply that familiarity with a speaker positively influences perceptions of charisma, or that charismatic speakers are more recognizable than uncharismatic speakers. In contrast, in the Arabic study, the mean number of speakers recognized was 0.55 of the 22, with a maximum of 3 and a minimum of 0. With this low speaker identification rate, we did not find a significant correlation between charisma ratings of a token and recognition of its speaker. We suspect that the low recognition rate in Arabic was due to the inclusion of more speakers.

The topic of the tokens used in the English experiment had an effect approaching statistical significance (p=$.052$) on subjects' ratings of charisma. Speakers were rated as more charismatic when speaking about healthcare (mean rating 3.31), postwar Iraq (3.29), reasons for running (3.28), content-neutral (3.07), and taxes (2.97). In the Arabic study, the topic of the segments exerted a significant influence (p=$.04$) on charisma

[2] All p-values in this section are determined by one-way ANOVA with repeated measures.

ratings. Speakers were rated as more charismatic when they were discussing the Israeli separation wall (3.96), the assassination of the Hamas leader (3.37), the debate among the Palestinian groups (3.23), the Palestinian Authority and calls for reforms (3.21), and the Intifada and resistance (3.17). This may imply that the sensitivity and importance of a topic may influence either the emotional state of the speaker or that of the rater, since we believe that these topics were more controversial than those discussed by the English speakers at the time.

4. Acoustic/Prosodic Analysis

To understand **why** subjects might have rated some tokens as charismatic or not, we examined a number of acoustic and prosodic characteristics of the tokens used in the two studies. We looked for correlations between duration, pitch, intensity, and speaking rate features of each token with subject judgments of charisma as well as some intonational features. Most features were calculated over (entire) speech tokens using the Praat [8] speech analysis software, although some intonational features were hand labeled.[3] Correlation was measured using simple linear regression.

In both languages we found that length of token correlated with charisma ratings (English: p=.04, r=.09; Arabic: p=4.4e-12; r=.3). The more material present, the more charismatic the speaker was judged. Speaking rate, however, exhibited different correlations with charisma judgments in English and Arabic. Measuring rate as the ratio of voiced to unvoiced frames, we found that, in English, rate was positively correlated with charisma (p=.0001, r=.17), while in Arabic it approached a negative correlation (p=.079, r=-.08). However, when we examined the speaking rate of the fastest spoken intonational phrase within the token, we saw a positive correlation in both languages (English: p=.003, r=.13; Arabic: p=.007, r=.12). These two observations may be explained by a third, positive correlation between the standard deviation of speaking rate across intonational phrases in Arabic (p=9.8e-5, r=.17) and judgments of charisma. It appears that, while faster speech is viewed as more charismatic in English, in Palestinian Arabic, a varied speaking rate, with some very rapid phrases, appears to increase perceptions of charisma.

The role of pause information in charisma perceptions also differed between the two studies. The ratio of number of pauses to number of words in the token positively correlated with ratings of charisma in Arabic (p=.003, r=.18) but not in English. In English, the standard deviation of the length of pauses *negatively* correlated with charisma (p=.04, r=-.09), while in Arabic it was *positively* correlated (p=.02, r=.1). Again, this suggests that variation may be perceived as charismatic in Arabic, but consistency appears more charismatic in English.

A number of features were examined to determine the relationship of pitch to perceptions of charisma.[4] We found that mean fundamental frequency (f0) was positively correlated with charisma in both studies (English: p=1.3e-7, r=.24; Arabic: 1.98e-6, r=.21), while f0 minimum was positively correlated with charisma in English (p=.002, r=.14), but negatively correlated in Arabic (p=.0003, r=-.16). In Arabic, f0 maximum (p=1.8e-7, r=.2) and standard deviation (p=2.2e-6, r=.23) were positively correlated with charisma judgments, but neither of these showed a significant correlation in English. From our ToBI labels, we also analyzed hand-annotated pitch maxima for each intermediate phrase ('HiF0'). HiF0 maximum (p=8.9e-14, r=.32) and standard deviation (p=3.1e-13, r=.31) were positively correlated with charisma only in Arabic, while mean HiF0 was positively correlated with ratings of charisma in both languages (English: p=1.7e-5, r=.2; Arabic: 7.4e-8, r=.23). To examine the role of pitch information in charisma perception in more detail, we normalized the pitch of each token within speaker using Z-SCORES. In English, normalized mean HiF0 was positively correlated with charisma (p=.03, r=.1), while normalized maximum pitch approached significance (p=.07, r=-.08). In Arabic, the normalized maximum (p=6.91e-14, r=.32), mean (p=5.17e-9, r=.25) and standard deviation (p=2.12e-15, r=.34) of HiF0 all were positively correlated with subject ratings of charisma. Taken together, these results indicate that, in both language, pitch information within the context of a speaker's overall pitch behavior is an important factor in perceptions of charisma. In both languages, tokens that were higher in a speaker's pitch range were rated as more charismatic than those lower in the range. However, while higher relative pitch may be perceived as more charismatic in both languages, the importance of pitch range variation may be significantly greater in Palestinian Arabic than in American English perceptions of charisma.

For intensity, we found that the standard deviation of intensity positively correlated with charisma in Arabic only (p=.0003, r=.16). Both mean (English: p=2.9e-6, r=.21; Arabic: p=1.e-6, r=.21) and standard deviation (English: p=2.9e-6, r=.21; Arabic: p=1e-5, r=.19) of *rms* intensity of intonational phrases positively correlate with charisma in both studies. Note that, in Arabic, for which we presented acoustically unnormalized tokens, maximum intensity (p=3.3e-8, r=.24) was positively correlated with perceived charisma. Louder speech may be associated with toughness, and powerfulness, two strong correlates of charisma in the Arabic study.

Using ToBI labels, we analyzed the correlation between charisma ratings and distributions of pitch accents, phrase accents, and boundary tones. In both languages, the rate of !H* (English: p=1.0e-5, r=.19; Arabic: 8.2e-9, r=.25) and L+H* (English: p=.08, r=.08; Arabic: .04, r=.09) accents positively correlated with perceptions of charisma. The L+H* accent is typically associated with emphasis or contrast, and may influence perceptions of enthusiasm. The !H* is often used in English on elements of a list, and appears to convey some rhetorical effect. The use of L* negatively correlates with charisma in both languages (English: p=.006, r=-.13; Arabic: p=1.4e-8, r=-.24). This accent is often employed with information already known to the hearer, and may thus be perceived as **boring** by listeners. The H- phrase accent approaches positive significance in English (p=.02, r=.11). The finality that is associated with the L% boundary tone may indicate toughness, while the rising, FORWARD-REFERENCE indicating, H% may indicate uncertainty. We found L% to positively correlate (p=9.9e-5, r=.17) with charisma in Arabic while H% negatively correlates (p=.005, r=-.12).

5. Lexical Analysis

In Section 4 we identified a number of acoustic/prosodic correlates of American English and Palestinian Arabic charismatic speech. In this section, we turn our attention to the lexical content of the stimuli used in the two studies. In both experiments, we found that longer speech was more charismatic: the number of words in a token was positively correlated with subject ratings of charisma (English: p=.0002, r=.13 Arabic: p=1.4e-23,

[3] Labeling was done for English using the ToBI convention [9]. We are currently developing a ToBI convention for Palestinian Arabic and used a draft of this system.

[4] We considered only tokens spoken by male speakers here.

r=.42). In both languages the rate of disfluencies (repetitions, repairs, and filled pauses) negatively correlated with charisma (English: p=2.8e-5, r=-.18; Arabic: p=2.8e-31,r=-.48). Disfluencies are commonly associated with a lack of confidence or preparedness, perhaps undermining how *convincing* or *powerful* a speaker may appear, both attributes which are correlated with charisma judgments. In Arabic, we found that the word "yaEony" (*it means*) appeared at least once in 27% of the tokens. It is analogical to the fillers in English: "I mean", and "like". The number of occurrence of this word significantly negatively correlates with charisma (p=2.8e-3, r=-.16). Another consistency that we found across language is the significance of repeated words. In both studies, the ratio of the number of words appearing more than once in a token to the total number of words in the token was positively associated with judgments of charisma (English: p=.002, r=.13; Arabic: p=4.5e-7, r=.22). Repetition is a common rhetorical device used for "driving a point home," which appears to have a similar effect in Arabic and English.

Occasionally, our Arabic speakers spoke in their regional dialect.[5] Boss [3] claims that followers of a charismatic leader believe they share a common history with the leader, which might suggest that the use of dialect would positively correlate with charisma among listeners who share a common dialect and culture. However, we observed that the presence of dialect *negatively* correlated with charisma judgments in Arabic (p=2.3e-5, r=-.18), possibly due to conveying a sense of informality in a public forum. Alternatively, the use of dialect may give the impression of undue informality and thus lower charisma ratings.

One area of difference between the two studies was the role of prominalization in perceptions of charisma. Charisma is often understood to be the manifestation of a uniquely personal *speaker-listener* relationship even though there may be no direct contact between the two. We computed pronoun densities for each token as the ratio of different types of pronouns to total words in the token. For English speakers, the correlation of charisma ratings with first person plural pronoun density was positive (p=.0002, r=.16), but there was no similar tendency in Arabic. For English, this finding tends to support the importance of shared, personal language as a characteristic of charismatic leaders. The density of second person pronouns in any form showed no significant correlation with ratings of charisma in either experiment. Note that the density of third person plural pronouns in English (e.g. 'they') negatively correlates with charisma (p=1.5e-5, r=-.19), while the density of third person singular pronouns ('he', 'she') demonstrates a positive correlation (p=.0003, r=.16). However, in Arabic, the opposite is true, with positive correlation correlation for the density of third person plural pronouns (p=1.47e-6, r=.21) with charisma ratings. This difference will require further study. In addition, we analyzed how the use of different parts of speech influenced ratings of charisma. In Arabic, we find negative correlations between ratings of charisma and the ratios of adverbs (p=.03, r=-.1), prepositions (p=.04, r=-.09) and nouns (p=.04, r=-.09) in the token. In English, we find a similar negative correlation with adverbs (p=.04, r=-.1) and a negative correlation with the density of adjectives (p=.004, r=-.13). In very terms, these findings may suggest that, in both languages, simple utterances with little modification may be perceived as more charismatic than more complex ones.

[5]We considered a word 'dialect' if it is not in the MSA lexicon, or if it is produced with morphological or phonological difference from the corresponding word in MSA.

6. Conclusion and Future Work

In this paper we have presented a comparison of perceptions of charismatic speech in English and Arabic. We have found that, while American and Palestinian speakers share some notion of a 'functional' definite of charisma — both find speakers who are *persuasive*, *charming*, *enthusiastic* and **not** *boring* to be charismatic — American speakers also find passionate and convincing speakers to be charismatic, while Palestinians associate the qualities of toughness and powerfulness with charisma. Our Arabic subjects also tended to be more homogenous in their judgments than were our English subjects. When we examine the acoustic/prosodic and lexical correlates of charisma in each study, we again find some broad similarities and some major differences. In both studies we find that longer amounts of speech, and speech produced with significant change in speaking rate, high in the speaker's pitch range, and with variation in intensity across intonational phrases was perceived as charismatic. Also, the use of simple sentences, with fewer modifiers, and sentences with repeated words increased judgments of charisma, while the presence of disfluencies inhibited those judgments. In the Arabic study, however, we find a greater influence of dynamic speech on perceptions of charisma; variation in pitch, and speaking rate exerted greater influence on subject judgments. We also observe that the use of pronouns in both languages affected the charisma conveyed by a speaker. However, the influence of particular pronoun classes differed across languages. This difference may be due to cultural differences between the two groups of speakers, fundamental linguistic differences between the two languages, or both. Furthermore, although many of our features correlate significantly with charisma, the correlation was only weak. This is likely due to the fact that charisma is a highly subjective characteristic.

We foresee a number of practical applications of this research. A diagnostic system to detect charismatic speech could be used for instructional purposes as well as identifying new political leaders. Additionally, a speech synthesis system could be modified in situations where compelling speech is required, as in a tutoring or direct marketing system. We are currently conducting machine learning experiments to utilize our findings to predict the degree of charisma for a given speech token. We are also conducting perception experiments using English tokens and both Palestinian and Swedish raters, and using Arabic tokens and American raters, to examine the role of cultural differences in perceptions of charisma.

7. References

[1] M. Weber, *The Theory of Social and Economic Organization*. Oxford University Press, 1947.

[2] J. Marcus, "Transcendence and charisma," *Western Political Quarterly*, vol. 14, pp. 237–241, 1967.

[3] P. Boss, "Essential attributes of charisma," *Southern Speech Communication Journal*, vol. 41, no. 3, pp. 300–313, 1976.

[4] D. Tannen, "The pragmatics of cross-cultural communication," *Applied Linguistics*, vol. 5, no. 3, pp. 189–195, 1984.

[5] A. Rosenberg and J. Hirschberg, "Acoustic/prosodic and lexical correlates of charismatic speech," in *EUROSPEECH '05*, 2005.

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

[7] J. Cohen, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 70, pp. 213–220, 1968.

[8] P. Boersma, "Praat, a system for doing phonetics by computer," *Clot International*, vol. 5, no. 9-10, pp. 341–345, 2001.

[9] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg, "Tobi: a standard for labeling english prosody," in *ISCLP'92*, vol. 2, 1992, pp. 867–870.