A gender bias in the acoustic-melodic features of charismatic speech?

Eszter Novák-Tót[1,2], Oliver Niebuhr[1], Aoju Chen[2]

[1] Mads Clausen Institute, University of Southern Denmark, Denmark
[2] Utrecht University, the Netherlands

e.novaktot@gmail.com, olni@sdu.dk, aoju.chen@uu.nl

Abstract

Previous studies proved the immense importance of nonverbal skills when it comes to being persuasive and coming across as charismatic. It was also found that men sound more convincing and persuasive (i.e. altogether more charismatic) than women under otherwise comparable conditions. This gender bias is investigated in the present study by analyzing and comparing acoustic-melodic charisma features of male and female business executives. In line with the gender bias in perception, our results show that female CEOs who are judged to be similarly charismatic as their male counterpart(s) produce more and stronger acoustic charisma cues. This suggests that there is a gender bias which is compensated for by making a greater effort on the part of the female speakers.

Index Terms: prosody, gender, charisma, hesitation, emphasis, intensity, speaking rate.

1. Introduction

Charisma has been a popular topic of discussion and research when it comes to defining, training, and practicing leadership [1, 2, 3]. A big portion of the charisma research has been directed at manifestation of political leadership [4, 5, 6, 7, 8], but the scope of studies has been expanded in recent years to explain remarkable business success [9, 10, 11, 2].

It is widely accepted among researchers that the presence of a *vision* is a prerequisite of charisma, which is further divided into *vision content* and *vision delivery* [12]. While vision content includes elements of speech such as the message and rhetoric devices, vision delivery refers to the verbal and non-verbal behaviors. It has been shown that stronger delivery style yields greater perceptions of charisma when the content of the message is held constant [13]. Furthermore, when content and delivery are manipulated simultaneously, the combination of a visionary content and strong delivery elicits higher charisma ratings than a non-visionary content combined with weak delivery. More importantly, strong delivery combined with non-visionary content yields greater charisma ratings than the combination of visionary content and weak delivery [14]. This finding suggests that a strong delivery style is an essential part of being perceived as a charismatic leader.

Various acoustic-melodic features of the speakers' voice have been shown to be closely connected to the strength of delivery and can thus affect charisma ratings [5, 6, 7, 15, 16]. In the current study we have decided to focus on four of these features: hesitational signals (like "um", "er", "uh", "ah"), emphatic accentuation, speaking rate, and acoustic energy levels. Past work has reported a positive correlation between the number of emphatic accents, speaking rate, and acoustic energy levels on the one hand and charisma ratings on the other hand, and a negative correlation between the number of hesitational signals and charisma ratings [5, 6, 7, 17]. Speakers who speak faster and hesitate less between or in the middle of phrases are perceived as more fluent and more confident. Speaking more loudly and stressing or repeating words for emphasis makes speech sound more animated and the speaker more enthusiastic. More fluent-, confident- and enthusiastic-sounding speakers are in turn perceived to be more charismatic.

However, gender-related differences appear to exist in the audience's appreciation of both the speakers and their speeches in terms of perceived persuasiveness, a core functional component of charisma [7]. For example, in small group interactions male speakers offering arguments without evidence have been found to sound more convincing than female speakers doing the same [19]. Furthermore, male and female speakers appear to receive positive evaluation of their competence as speakers for different reasons. Positive feedback for female speakers mostly focuses on the external sources that are used to back-up the presented information. However, male speakers are praised for the structure of the speech, eye contact, and delivery in addition to the quality of information [20]. Additionally, significantly more people are persuaded by male speakers than by female speakers even though speakers display the same objective level of competence and all other conditions are identical [20, 21]. Together, these findings show that, given identical performance, male speakers tend to be regarded as being more efficient, more structured, and more persuasive than female speakers [22].

In the present study, we ask the questions of whether and how such a gender bias is also reflected in the nonverbal aspects of charismatic speech, in particular in acoustic-melodic features of speech. If male speakers are rated more favorably for the same performance, it stands to hypothesize that female speakers may need to deliver a *better* performance in order to be rated as equally charismatic. To test this overarching hypothesis, we have examined the use of the above-mentioned four acoustic-melodic features in male and female English-speaking business executives via acoustic analysis and perceptual judgements obtained from native speakers of English. We predict that the female speakers who are considered to be similarly charismatic to the male speakers will (a) hesitate less, (b) employ emphatic accentuation to a greater degree, (c) have a higher speaking rate, and (d) have a higher average acoustic energy level than the male speakers.

2. Methodology

2.1. Material

We compiled speech samples of high-ranking US business executives (CEOs), all of whom were native speakers of

American English. To ensure homogeneity of the speech samples with respect to speaking style and context (i.e. genre, [23]), we included in our samples only prepared keynote speeches that were held in front of big audiences. We analyzed the speech of three of the speakers, two females and one male (Figure 1). These three speakers were selected because they were similarly charismatic according to the general public opinion stemming from statements found in both scientific papers and non-scientific literature such as newspaper articles and internet blogs, which has been shown to be a useful indicator of perceived speaker charisma levels [15]. Our speakers are as follows:

- Oprah Winfrey, who is uncontroversially considered one of the most charismatic speakers of our time. When reading an article about charismatic leaders and speakers, Winfrey is often one of the names of the arbitrary list of people fitting the bill [24, 25, 26]. She is the founder and CEO of Harpo Productions, the Oprah Winfrey Network, and the Oprah Winfrey Leadership Academy for Girls. Her business is leadership itself, and therefore, in her speeches, she does not focus on selling a particular product: she focuses on "selling" herself and her ideas.

- Ginni Rometty, CEO and Chairwoman of IBM since 2012, is widely considered a charismatic speaker in the media. Rometty has been described by journalists and fellow business figures as a "modern CEO" bringing "charisma and raw leadership" to the table [27, 28]. It is also telling that she was criticized a few times for her lack of a clear vision for IBM's future [29], but never for her speaker skills.

- The late Steve Jobs, co-founder, Chairman and CEO of Apple Inc., founder, Chairman and CEO of NeXT, was considered one of the most charismatic and inspiring presenters of our time by journalistic and scientific articles alike. There is hardly any listing, ranking, statement, or assessment on charisma that does not include his name and praise his skills. Jobs also has inspired a number of scientific analyses of charismatic leadership [17, 23]. Therefore, in order to reveal initial evidence for a gender bias in charismatic speech, he is the ideal benchmark and male reference figure.

| Oprah Winfrey | Ginni Rometty | Steve Jobs |

Figure 1: *Portraits (downloaded under CC license from Wikipedia) of the analyzed charismatic speakers.*

2.2. Analysis

The speech samples were collected from publicly accessible keynote speeches on YouTube and downloaded in the form of high-quality MP3 files. The MP3s were further converted to WAV files for further analysis in Praat [30].

Subsequently, a speech sample of about 20 minutes was randomly selected for each speaker from the middle of the keynote speeches. No speech samples were taken from openings and closings of the keynote speeches for two reasons. First, from a content perspective, there is little need for displaying charisma (persuasiveness) in openings and closings [15]. Second, the speaking behavior is likely to be more changeable in these opening and closing sections. For example, the opening is often used as a warming-up phase.

Winfrey's speech sample was extracted from her keynote speech at the QuickBooks Connect Convention in 2015, where she talked about the process of establishing her own name as a brand. In the case of Rometty, the sample was compiled from two keynote speeches, as her keynotes were often only 20 minutes long or even shorter, including the opening and closing sections. We used her keynote speeches at NRF 2014 and CES 2016, both presenting IBM's practical Big-Data applications. Jobs' speech sample consists of excerpts from his product presentations of the iPhone 2 and iPhone 4 [15].

The speech samples were manually annotated in Praat, distinguishing between hesitations (silent hesitations: <hes:s>, "uh" <hes:a>, "um" <hes:n>, final lengthening <hes:fl>, initial lengthening <hes:il>), silent pauses (<p:>), breathing (<a:>), and intonation phrases (<i:>). Prosodic phrases were defined as melodically coherent utterances without any audible breaks in accordance with the traditional breath-group concept of the British School of Intonation [31]. The speech within prosodic phrases was orthographically transcribed and then labelled and segmented accordingly. Prosodic labelling was restricted to emphatic accents based on the accent types set up by [33, 34]. These accent types include positive intensification, negative intensification, reinforcement, lexical repetitions (with a pitch accent on each word), and accent chains (further explanations are provided in 3.3.2). Note that these types of emphatic accents are not bound to specific pitch-accent categories. Although [34] found for corpus data that there are tendencies for most emphatic accents to co-occur with some pitch-accent categories more frequently than with others, there is no 1:1 relationship. An example of our annotation is presented in Figure 2.
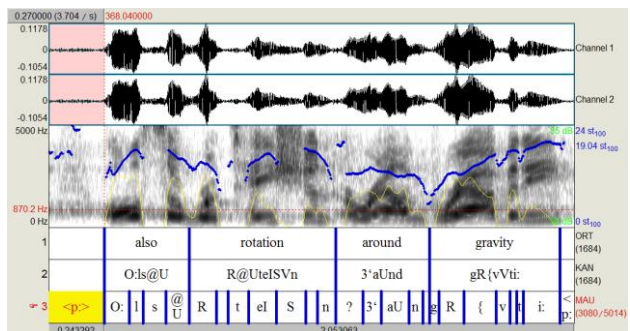

Figure 2: *Excerpt of a segmentally and prosodically annotated sound file of Steve Jobs.*

Our acoustic-melodic analysis included four parameters: (1) frequency of hesitations in the form of silent and filled pauses, counted on an auditory basis with reference to the labelling (note that the labelling was objective insofar as it was done by a third person and independently of the present research questions); (2) frequency of (4 types of) emphatic accents, also, counted on an auditory basis with reference to the labelling; (3) speaking rate (in syl/s), determined by dividing orthographically derived number of syllables within an intonation phrase by the total duration of that phrase; (4) Mean acoustic-energy level in terms of RMS (dB) of an entire intonation phrase, normalized for each speaker against the mean acoustic-energy level of the frequently occurring particle "so". The normalization was to compensate for between-

speaker differences in signal gain. We used "so", because of its simple CV structure, high frequency of occurrence for all speakers, and the fact that it is phonetically representative in combining the extreme acoustic-energy levels of a voiceless consonant and a diphthongized vowel ([soʊ]). We calculated the average intensity across all instances of "so" per speaker and then divided the intensity level of each part of speech by this reference value.

In addition to the acoustic-melodic analysis, we conducted a small-scale perception experiment with 18 English-speaking listeners. They were between 20 and 25 years old and recruited from the student pool (BA and MA) of the University of Southern Denmark. None of them studied linguistics or had taken part in a perception experiment before. The listeners were given the three audio files that we used for the acoustic-melodic analysis. The listeners were asked in individual sessions to rank the speakers in charisma from '1' to '3' (with '1' being most charismatic). The audio files were played using the Windows Media Player, and the listeners were allowed to jump (as often as they wanted) to any position in the speech fragments by clicking on or shifting the progress bar. The only requirement was that they listened (through headphones at a pre-adjusted loudness level) to at least three minutes of speech from each speaker. The listeners were divided into six groups, each of which started with a different speaker.

Note that we chose this free listening approach, instead of eliciting judgments of pre-determined stimulus pairs or even triplets (like in AX or AXB tasks) for two reasons. First, creating such pairs or triplets would have required selecting short stimulus excerpts from the keynotes of each speaker. We assumed, however, that listeners would come up with more reliable charisma ratings and rankings if they had access to longer stretches of speech to which they could also listen as often as they wanted. Second, a free listening approach is dynamic and interactive, for example, in that it allows listeners to jump back and forth between speakers and change ranking decisions multiple times. In order to achieve a similarly interactive dynamics in a (AX or AXB) test with given stimulus pairs or triplets, multiple repetitions of each pair or triplet would have been necessary, and it is unclear how this would wear off the charisma effects created by the different speakers.

3. Results

3.1. Charisma ranking

Figure 3 presents the mean ranking scores of Winfrey, Rometty, and Jobs over the 18 English-speaking listeners. Mean scores are all close to two, which is due to the fact that the listeners had no consistent opinion about which speaker sounds most, second most, and least charismatic. Subsequent t-tests comparing the ranking scores yielded no significant differences between the three speakers (p>0.1). The perception experiment thus provided clear empirical evidence that the speakers we compared here are indeed similar in terms of their perceived charisma, in line with their characterization in newspapers, research papers, and internet blogs. This equivalence in perceived charisma is a prerequisite for interpreting the results of the acoustic-melodic analysis in the light of our research question.
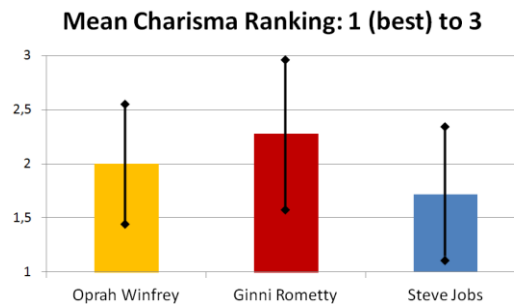

Figure 3: *Mean charisma rankings for selected speakers*

3.2. Acoustic-melodic analysis

3.2.1 Occurrence of hesitations

Regarding filled and silent hesitations, Figure 4 shows that Winfrey produced the smallest number of hesitations. We only found ten filled and three silent hesitations within her speech sample. Rometty produced almost twice as many (25) hesitations, i.e. 13 filled and 12 silent pauses. The highest number of hesitations - 41 instances in total - occurred in Jobs' speech. Six of these 41 hesitations were silent and 35 were filled pauses. A Chi-squared test confirmed that the frequencies of occurrence of hesitations were significantly different between the three speakers: $\chi^2[1]=5.8$, p=0.016. Regarding the two female speakers, a separate Fischer's exact test yielded no significant differences for the frequency of hesitations. Moreover, the total number of hesitations was lower for the two female speakers than for the male speaker Jobs. Winfrey produced only about one third of the hesitations that Jobs produced. The only exception to this pattern is that silent pauses occurred in Rometty's speech twice as often as in Jobs' speech. A z-score proportion test shows that the gender difference was, as a whole, statistically significant (z[1.5]= 3.2,p<0.05).
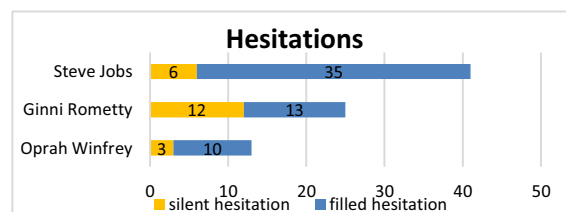

Figure 4: *Total number of hesitations per speaker, subdivided into filled and silent pauses.*

3.2.2 Occurrence of emphatic accents

Winfrey's speech contained the highest number of emphatic accentuations (131) and Rometty's speech the lowest number of emphatic accents (81). Jobs produced 119 emphatic accents, which is a frequency that falls in between those of Winfrey and Rometty. Nevertheless, the speakers were similar in the distribution of the four types of emphatic accents distinguished in [33], as shown in Figure 5. The most frequent type was the reinforcement accent (50.4%-78.3%), which is characterized by a lengthened onset consonant of the accented syllable in combination with a steep and high pitch peak on the vowel (e.g. "it is aMMMAzing"). The second most frequent type was the positive intensification accent (13.7%-27.7%), produced in the form of a strongly lengthened

accented vowel that coincides with high pitch plateau (e.g. "adv-AAANced analytics"). Besides this distributional similarity, Winfrey and Jobs made more use of accent chains as a means of emphatic accentuation than Rometty (16-17.6% vs. 8.6%). Accent chains are created by realizing adjacent words in a syllable-by-syllable fashion, each one with the same type of a pitch accent (typically embedded in an overall downstep patterns), as, for example, in " I – CAN – de-PEND – ON – YOU". In addition, word repetitions like in "a whole whole ecosystem of partners", were entirely absent in Winfrey's speech and occurred rarely in Rometty's speech, while Jobs used them in 5.8% of his emphatic expressions. A Chi-squared test confirmed that the speakers differed significantly in the use of emphatic accents: $\chi^2[6]=21.4$, $p=0.002$.
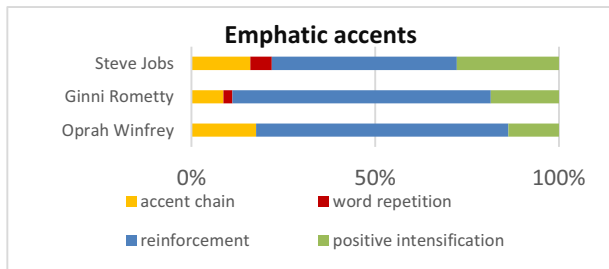

Figure 5: *The distribution of types of emphatic accents: accent chains, word repetitions, reinforcement and positive intensification.*

With respect to gender, a z-score proportion test yielded no evidence for significant differences between the two females and the male speaker in terms of emphatic accent frequency. However, looking at the two females separately, we found that Winfrey significantly outperformed Jobs in emphatic accent frequency ($z[1.9]=2.1$, $p<0.05$).

3.2.3 Speaking rate and acoustic-energy level

Figure 6 shows that Winfrey spoke on average about as fast as Jobs (4.3 vs 4.4 syl/s), while Rometty was the fastest of the three speakers (5.1 syl/s). The latter difference turned out to be significant in an independent samples t-test ($t[1213]=17.1$, $p<0.001$). Regarding the normalized acoustic-energy measurements, Rometty's speech again yielded the highest mean value (0.99). Winfrey had the second highest normalized mean acoustic-energy level (0.98), still clearly higher than that of Jobs (0.95). T-tests (again for independent samples) showed no significant differences between the two female speakers, but both Winfrey and Rometty spoke at a higher normalized mean acoustic-energy level than Jobs ($t_{Winfrey/Jobs}[1123]=9.4$, $p<0.001$; $t_{Rometty/Jobs}[1213]=11.9$, $p<0.001$).
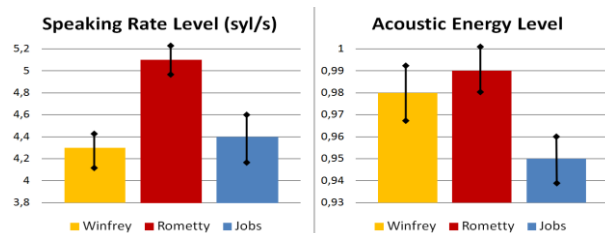

Figure 6: *Mean values of acoustic measurements of rate (left) and energy (right) for Winfrey (n=533), Rometty (n=623), and Jobs (n=592).*

4. Conclusions

In the present study, we examined whether gender-related differences in audience's appreciation of public speaking is also manifested in acoustic-melodic features of charismatic speech. On the basis of the earlier finding that men tend to be assessed more favorably under otherwise identical conditions, we hypothesized that female speakers may need to deliver a *better* performance from an acoustic-melodic perspective in order to be rated as equally charismatic. To test this hypothesis, we have compared three well-known charismatic speakers, including two female speakers (Oprah Winfrey, Ginni Rometty) and one male speaker (the late Steve Jobs), in four acoustic-melodic features (i.e. hesitations, emphatic accentuation, speaking rate and acoustic energy level). The two female speakers were judged to be similarly charismatic to the male speaker in our perception test. If the hypothesis was true, we should see that the female speakers employed emphatic accentuation more frequently, hesitated less, and had a higher average acoustic energy level and a higher speaking rate than the male speaker. In a nutshell, this is exactly what we found.

Our speakers differed significantly in the frequency of hesitations. Crucially, the female speakers produced fewer hesitations than the male speaker, in line with our prediction. Further, the speakers differed significantly in the use of emphatic accents. But only one of the two female speakers (Winfrey) was found to use emphatic accents more frequently than the male speaker, lending partial support to our prediction. Moreover, one of the female speakers (Rometty) spoke faster than the male speaker, in partial agreement with our prediction. Finally, both female speakers spoke at a higher acoustic-energy level than the male speaker, as predicted. Thus, on the whole, our results supported the hypothesis of a gender bias in the acoustic-melodic features of charismatic speech.

However, our study is relatively small in scale. The results thus need to be validated in a large-scale study, involving more speech samples from more speakers from different backgrounds (e.g. business, politics, and education) and perceptual judgements from more listeners. Relatedly, we have examined only four acoustic-melodic features. Features such as F0 level, F0 range, voice quality (variation), phrase duration, and pause duration between phrases, among others, have also been shown to be closely connected to the strength of delivery and can thus affect perceived charisma. Future research is needed to shed light on similarities and differences in these features between male and female speakers in charismatic speech.

5. Acknowledgements

6. References

[1] Beyer, J. M. (1999). Taming and promoting charisma to change organizations. *The Leadership Quarterly*, 10(2), 307-330.

[2] House, R. J., & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis?. *Journal of management*, 23(3), 409-473.

[3] Turner, S. (2003). Charisma reconsidered. *Journal of Classical Sociology*, 3(1), 5-26.

[4] Boss, G. P. (1976). Essential attributes of the concept of charisma. *Southern Journal of Communication*, 41(3), 300-313.

[5] D'Errico, F., Signorello, R., Demolin, D., & Poggi, I. (2013, September). The perception of charisma from voice: A cross-cultural study. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference* on (pp. 552-557). IEEE.

[6] Hirschberg, J. B., & Rosenberg, A. (2005). Acoustic/prosodic and lexical correlates of charismatic speech.

[7] Rosenberg, A., & Hirschberg, J. (2009). Charisma perception from text and speech. *Speech Communication*, 51(7), 640-655.

[8] Weber, M. (1947). *The theory of social and economic organization* (A.M. Henderson & T. Parsons, Eds. and Trans.). Glencoe: Ill: The Free Press.

[9] Bass, B. M. (2000). On the taming of charisma: A reply to Janice Beyer. *The Leadership Quarterly*, 10(4), 541-553.

[10] Den Hartog, D. N., & Verburg, R. M. (1998). Charisma and rhetoric: Communicative techniques of international business leaders. *The Leadership Quarterly*, 8(4), 355-391.

[11] Haslam, S. A., & Reicher, S. D. (2012). In search of charisma. Scientific American Mind, 23(3), 42-49.

[12] Awamleh, R., & Gardner, W. L. (1999). Perceptions of leader charisma and effectiveness: The effects of vision content, delivery, and organizational performance. *The Leadership Quarterly*, 10(3), 345-373.

[13] Holladay, S. J., & Coombs, W. T. (1993). Communicating visions an exploration of the role of delivery in the creation of leader charisma. *Management Communication Quarterly*, 6(4), 405-427.

[14] Holladay, S. J., & Coombs, W. T. (1994). Speaking of visions and visions being spoken an exploration of the effects of content and delivery on perceptions of leader charisma. *Management Communication Quarterly*, 8(2), 165-189.

[15] Niebuhr, O., Voße, J., & Brem, A. (2016). What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs tone of voice. *Computers in Human Behavior*, 64, 366-382.

[16] Antonakis J, Fenley M, Liechti S. (2011). Can Charisma Be Taught? Tests of Two Interventions. The Academy of Management Learning and Education 10: 374-96.

[17] Berger, S., Niebuhr, O., Peters, B. (2017). Winning Over an Audience – A Perception-based Analysis of Prosodic Features of Charismatic Speech. Proceedings 43rd Annual Meeting of the German Acoustical Society, Kiel, Germany, 793-796.

[18] Strangert, E., & Gustafson, J. (2008). What makes a good speaker? Subject ratings, acoustic measurements and perceptual evaluations. In *INTERSPEECH* (Vol. 8, pp. 1688-1691).

[19] Bradley, P. H. (1981). The folk-linguistics of women's speech: An empirical examination. *Communications Monographs*, 48(1), 73-90.

[20] Sellnow, D. D., & Treinen, K. P. (2004). The role of gender in perceived speaker competence: An analysis of student peer critiques. *Communication Education*, 53(3), 286-296.

[21] Propp, K. M. (1995). An experimental examination of biological sex as a status cue in decision-making groups and its influence on information use. *Small Group Research*, 26(4), 451-474.

[22] Aries, E. (1996). *Men and women in interaction: Reconsidering the differences*. Oxford University Press.

[23] Conger, J. A. (1989). *The charismatic leader: Behind the mystique of exceptional leadership*. Jossey-Bass.

[24] Restauri, D. (2012, May 3) 5 Qualities of Charismatic People. How Many Do You Have?. *Forbes*. Retrieved from http://www.forbes.com/sites/deniserestauri/2012/05/03/5-qualities-of-charismatic-people-how-many-do-you-have/#3f853d42372d

[25] Lebowitz, S. (2015, December 18). 7 science-backed ways to become more charismatic. *Business Insider Singapore*. Retrieved from http://www.businessinsider.sg/how-to-become-more-charismatic-2015-12/#.V1LbyZF97IW

[26] Riggio, E. R. (2010, February 15). Charisma: What Is It? Do You Have It?. *Psychology Today*. Retrieved from https://www.psychologytoday.com/blog/cutting-edge-leadership/201002/charisma-what-is-it-do-you-have-it

[27] Bednarz, A. (2011, October 26). What to expect from IBM's on-deck CEO: Continuity, stability, charisma. *Network World*. Retrieved from: http://www.networkworld.com/article/2182584/data-center/what-to-expect-from-ibm-s-on-deck-ceo--continuity--stability--charisma.html

[28] Langley, M. (2015, April 20). Behind Ginni Rometty's Plan to Reboot IBM. *The Wall Street Journal*. Retrieved from: http://www.wsj.com/articles/behind-ginni-romettys-plan-to-reboot-ibm-1429577076

[29] Zolman, S. (2012, September). Top 10 Reasons Why Ginni Rometty Will Fail as IBM's New CEO [Weblog content]. Retrieved from: http://www.netnetweb.com/blog/top-10-reasons-why-ginni-rometty-will-fail-ibm%E2%80%99s-new-ceo

[30] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot international*, 5(9/10), 341-345.

[31] Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, 74(6), 431-461.

[32] Stock, F. (2015, August). How to be charismatic. *The Telegraph*. Retrieved from: http://www.telegraph.co.uk/radio/what-to-listen-to/how-to-be-charismatic/

[33] Niebuhr, O. (to appear). The Kiel Intonation Model - Forms and Functions of German Prosody. In J. Barnes, N. Viellieux, S. Shattuck-Hufnagel (Eds), A *handbook of intonation models*. Cambridge: MIT Press.

[34] Niebuhr, O. (2010). On the phonetics of intensifying emphasis in German. *Phonetica 67*, 170-198.