Toward Charismatic Virtual Agents: How to Animate Your Speech and Be Charismatic

Ning Wang[(⊠)], Abhilash Karpurapu, Aditya Jajodia, and Chirag Merchant

Institute for Creative Technologies, University of Southern California,
Los Angeles, USA
`nwang@ict.usc.edu`

Abstract. Charisma is a powerful device of communication and persuasion. Researchers have pinpointed specific behaviors that contribute to the perception of charisma. How can we realize such behaviors in a virtual character? In this paper, we discuss our work in the design of charismatic behavior for a virtual human. We developed a series of verbal charismatic strategies based on the research on charismatic leaders, which was then used to re-write an existing tutorial on the human circulatory system to express charisma. We then collected voice recordings of the tutorial in both charismatic and non-charismatic voices using actors from a crowd-sourcing platform. In this paper, we present the analysis of the charismatic and non-charismatic voice recordings, and discuss what nonverbal behaviors in speeches contribute to perceived charisma. Results can shed light on the synthesis of charismatic speeches for virtual characters.

Keywords: Charisma · Prosody analysis · Speech synthesis

1 Introduction

Charisma is a powerful device of persuasion [18]. Leaders have used charisma to make their messages memorable and inspiring [14,16,23,28]. Contrary to the idea that charisma is a personal quality [53], researchers have pinpointed the specific verbal and nonverbal behaviors that contribute to the perception of charisma [4,5,15,20,26,40,49,52].

An important aspect of charismatic nonverbal behavior is the use of animated voice. While different speech features might have different affective effects in different languages [37], researchers in expressive speech generally consider pitch [54], loudness [21], spectral structure [31], voice quality [8,24], etc. as features relevant to perceived expressiveness (e.g., emotional states) in speech. Specific to perceived charisma, previous research indicates that variations in pitch range and standard deviations are correlated with charisma [48]. Additionally, speech rate (speed and variation), intensity (loudness and variation), intonation (e.g., phrasal ending patterns, shape, variations), and vocal clarity also play a role in perception of charisma [14,38].

How can we realize charismatic behaviors in a virtual character? Given the broad range of applications of virtual characters in, for example, health care [6], energy conservation [1], and education [32], being able to employ charismatic behaviors in virtual characters can have great potential in influencing the learning and decision-making of their human interactants. In this paper, we discuss the design of nonverbal behaviors for a virtual human, particularly the synthesis of pauses in charismatic speech. We developed a series of verbal charismatic strategies and implemented them in a tutorial on the human circulatory system. We then collected recordings of the tutorial in both charismatic and non-charismatic voice using actors from a crowd-sourcing platform. We conducted analyses to compare the charismatic and non-charismatic voice recordings and shed light on what types of nonverbal behaviors in speeches contribute to perceived charisma, and how such behaviors can be realized in virtual characters.

2   Related Work

2.1   Charisma

Much progress has been made in understanding the behavioral makeup of charisma, particularly in the study of charismatic leadership in organizational science. Verbally, charisma is often expressed through the use of metaphors, which are very effective persuasion devices that affect information processing and framing by simplifying the message, stirring emotions, invoking symbolic meanings, and aiding recall [11,18,33]. Stories and anecdotes are also often employed as devices of charisma [20,49], by making the message understandable and easy to remember [7]. Rhetorical devices, such as contrasts (to frame and focus the message), lists (to give the impression of completeness), rhetorical questions [4], are often used in charismatic communications as well. In addition, charismatic speakers are skilled at expressing empathy [39], setting high expectations, and communicating confidence that the expectations can be met [26]. Theoretically, these charismatic behaviors are catalysts of motivation [17] and increase self-efficacy belief [41].

On a nonverbal level, of most relevance to this paper, charismatic speakers speak with varied pitch, amplitude, rate, fluency, emphasis, and an overall animated voice tone [20,49]—all aspects of speech commonly associated with a more engaged and lively style of speech and all predicting higher ratings of charisma [38]. Both the verbal and nonverbal behaviors make the message more memorable [7,18,33,52] and increase self-efficacy [3,20,49].

2.2   Speech Synthesis

Text-to-speech and speech synthesis have come a long way in generating natural-sounding speech [43,46,47,55]. Recent trends in research in spontaneous and conversational speech synthesis have added non-speech behaviors such as breathing to make the synthesized speech even more realistic, particularly in conversational settings [22,44]. More recent advances in synthesized speech have created

voices that are increasingly challenging to distinguish from real human speakers [42,51]. In addition to synthesizing speech that aims to be natural or spontaneous, researchers have sought to generate speech that is expressive, for example, speech that expresses different emotions and speaking styles [9,19,35,45]. Often, the expressive information can be incorporated, either before or after the synthesis of neutral speech [9,45]. However, there is very little work on synthesizing charismatic speech, even though charismatic nonverbal behaviors in speech (such as those of charismatic leaders) are well studied in the organizational sciences. Additionally, most of the methods in speech synthesis take the big data, deep-learning approach, while employing machine-learning algorithms that are hard to explain. While such synthesis methods have generated great end-to-end outcomes, it is challenging to distill explainable outcomes that can contribute to the knowledge of, for example, what is important to synthesize charismatic speech and how the existing theoretical framework on charisma performs in generating charismatic speeches.

3   Charismatic Speech Dataset

3.1   Charismatic Verbal Strategies

Based on the research on charismatic speech, we developed a series of verbal strategies to express charisma, for example, the use of metaphor and analogies [20], stories [20,49], rhetorical questions [4], etc. Using these strategies, we re-wrote an existing tutorial on the human circulatory system [12]. For example, instead of saying "The major function of the blood is to transport nutrients and oxygen to the cells and to carry away carbon dioxide and nitrogenous wastes from the cells", we rephrased it using the strategy of "metaphors" - "The major function of the blood is to be both the body's mailmen, delivering nutrients and oxygen to the cells, and its garbagemen, carrying away carbon dioxide and nitrogenous wastes from the cells". A previous human-subject study comparing the tutorial text with and without the use of charismatic strategies showed that the use of charismatic strategies significantly improved the perceived charisma [50].

3.2   Data Collection

Using the "charismatic" version of the tutorial (106 sentences, 1824 words), we gathered voice recordings from 13 participants, who read the tutorial out loud in both charismatic (e.g., animated) and non-charismatic (e.g., monotone) voice. To gather the data, we first recruited 95 participants through a crowd-sourcing platform to record a snippet of the tutorial in charismatic and non-charismatic voices. Participants were given instructions that explain what is considered a charismatic vs. non-charismatic voice. For example,

–   *"A voice conveys charisma is often considered to be varying in speed (e.g., sometimes fast, sometimes with pause), varying in energy (e.g., stress certain*

word or phrase), and varying in pitch (e.g., a more animated voice), compared to a mono-tone and mono-speed voice that often puts one to sleep. A charismatic speech inspires and motivates."

– *"A voice in contrast with a voice that conveys charisma is, for example, mono-tone, lack of emphasis, without changes in speed or pauses. And generally a voice that's boring and puts one to sleep."*

Two members of the research team then selected the 13 participants whose recordings more closely followed the instructions, out of the 95 participants. The 13 participants then went on to create voice recordings of the tutorial in full length, in both charismatic and non-charismatic voices.

4   Results

The body of work on charismatic speakers indicates that charismatic speeches are spoken with varied pitch, amplitude, rate, fluency, emphasis, and an overall animated voice tone [20,49]. Thus, our analysis focused on measurable variables such as pitch, energy (i.e., amplitude), and speed (i.e., rate), in the comparison between charismatic and non-charismatic speeches. To study the dynamics in charismatic speech, we zoomed in on pauses (an indication of varied speech rate) and emphasis. Both pauses and emphasis can to draw listeners' attention to specific parts of the speech. Thus they can be effective devices employed by charismatic speakers to make their messages more memorable. Because our data consist of "prepared speech" (as opposed to spontaneous speeches), we did not examine the fluency variable in our data. In this paper, we will discuss the analysis on pauses.

4.1   Charismatic vs. Non-charismatic Speeches

Using Paired Sample T-Test, we compared the charismatic (C) and non-charismatic (NC) recordings in pitch, energy and speed - three factors that are key to charismatic speech [14,38]. Results show that charismatic speeches are spoken with significantly low speed ($M_C = 7.3, M_{NC} = 7.01, p < .0001$, duration in seconds for a sentence), higher energy ($M_C = .073, M_{NC} = .068, p < .0001$, in dB) and higher pitch ($M_C = 2414.8, M_{NC} = 2209.2, p < .0001$, in Hz). As an overall indication of how "animated" a speech is, we also compared the variations in pitch, energy, and speed. Results show that there are significantly greater variations in energy and pitch (Levene's test, $p < .001$ for both comparisons), but not speed, in charismatic recordings compared to non-charismatic ones.

4.2   Pauses in Charismatic Speech

The use of pauses is one of the ways to vary speed in speech and draw attention to the messages to follow.

Pause Duration. Pauses in speech are often categorized into silent pauses, filler pauses, and breath pauses. Breath pauses are regular natural pauses caused by respiration activity. Filler pauses are pseudo-words, such as "Mmmm" and "Hmmmm", that do not affect the meaning of the sentence [27]. Because our dataset consists of only prepared speeches, we did not include analysis of filler pauses, which primarily occur in spontaneous speeches. While silent pauses can be indications of disfluencies, uncertainty, and hesitation, which occur more often in spontaneous speeches, they are primarily intentional stylistic pauses used purposely by professional speakers and the like [27]. There has been great debate since the 1970s s about the duration of silence that defines a silent pause [29]. Previous work has often adopted the convention of .2 to .25 s of silence (or longer) as indication of silent pauses, while those that fall below this threshold are often considered breath pauses [10,25,36]. In automated puncture detection in speech, it has been shown that over 95% of the pauses of .35 s or longer are the sentence boundaries [30]. Thus, in our analysis, we focused our analysis on silent pauses of .2 s or longer.

We extracted the pauses (e.g., a silence of at least .2 s long) from the charismatic and non-charismatic speech recordings. We first conducted a paired t-test to compare the number of silent pauses in charismatic and non-charismatic speech. Results show that there is no significant differences in the number of pauses between charismatic and non-charismatic speech ($M_C = 228.6, M_{NC} = 277.2, p = .214$).

Pause Locations. To synthesize charismatic speech, it is important to know where the pauses occur in addition to how long the pauses should be. Given that we have a unique dataset where all the recordings are based on the same text, we tabulated where the silent pauses occurred in each participant's recording and examined whether there was consensus among the participants on where to place silent pauses. Figure 1 shows that there are a total of 589 silent pauses (made by the 13 participants) of a duration of .2 s or longer. Additionally, Fig. 1 suggests that there is great variance in where participants placed the pauses. For example, there are only 208 cases where 3 or more participants paused at the same place. For consensus among half or more speakers (e.g., 6 or more participants), the number of "commonly agreed" pauses dropped to 57.

We then annotated the tutorial text using part-of-speech (POS) tags (e.g., verb, noun, [34]). We then analyzed the pairs of POS tags where the pauses most frequently occur (e.g., between a verb and a noun). Table 1 shows that, of the silence pauses of .2 s and longer, the most common places where a pause happens are between a noun (NN) and a preposition (IN, e.g., "in", "of", "to"), a noun (NN) and a coordinating conjunction (CC, e.g., "and", "but"), a noun (NN) and a determiner (DT, e.g., "the", "my", "some"). Table 2 shows the part-of-speech pairs that have the highest percentage of occurrence of pause. Descriptions and examples of POS tags are shown in Table 3.

From Table 2, we can see that 92% of NNS-CD POS pairs have a pause in between, which indicates a high consensus among the participants. However,

Number of Pauses based on Consensus
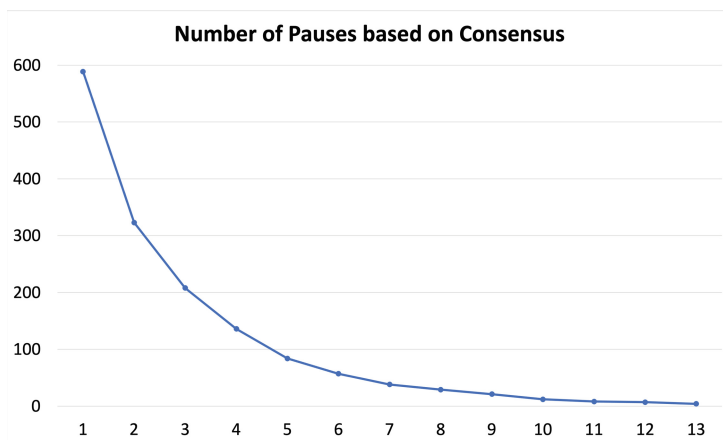
Fig. 1. Number of pauses based on consensus among the speakers, for example, for 84 pauses, 5 or more speakers paused at the same place. The x-axis indicates the number of consensus. The y-axis indicates the number of pauses.

Table 1. Part-of-speech (POS) tag pairs where there are most pauses.

POS pair	Total # in text	# with a pause
('NN', 'IN')	117	20
('NN', 'CC')	37	17
('NN', 'DT')	27	14
('NN', 'NNS')	63	11
('NN', 'NN')	115	10

Table 2. Part-of-speech (POS) tag pairs with the highest percentage of pauses. For example, 92% of the ('NNS', 'CD') POS pair in the charismatic text has a pause in between.

POS pair	Total # in text	% with a pause
('NNS', 'CD')	1	92%
('NN', 'WDT')	1	85%
('DT', 'WP')	1	77%
('VBN', 'VBG')	1	77%
('VBZ', 'CC')	1	77%

for all the top pairs listed here, each only occurred in the charismatic tutorial once. Table 4 lists more commonly seen POS pairs (with at least 10 occurrences in the tutorial) and how often there is a pause in between. This gives a more realistic view of how often more frequently-occurring POS tag pairs have a pause in between. Data on these POS tag pairs may better inform how to synthesize

Table 3. Descriptions and examples of Part-of-speech (POS) tags.

POS	Description	Example
CC	Coordinating conjunction	and, but
CD	Cardinal digit	1, 3, 8
DT	Determiner	the, my, some
IN	Preposition	in, of , to
NN	Noun, singular	cat, tree
NNS	Noun, plural	cats, trees
RB	Adverb	slowly, softly
VBG	Verb gerund	judging
VBN	Verb past participle	judged
VBZ	Verb, present tense with 3rd person singular	judges
WDT	wh-determiner	that, what
WP	wh- pronoun	who

pauses, e.g., where to insert them. Interestingly, our data show that, only NN-VBG and NN-DT pairs have a better than chance (>50%) percentage of having a pause in between.

Table 4. Similar to Table 2, this table shows the part-of-speech (POS) tag pairs with the highest percentage of pauses. However, here, we only focus on these POS pairs have 100 or more occurrences in the text.

POS pair (>100 in text)	Total # in text	% with a pause
('NN', 'VBG')	10	62%
('NN', 'DT')	27	52%
('NN', 'CC')	37	46%
('NN', 'RB')	12	40%
('NNS', 'CC')	16	37%

5   Discussion

In this paper, we discussed a study to collect charismatic and non-charismatic speech samples. Analysis of the data revealed that charismatic speeches are spoken at lower speed, higher energy, and higher pitch. There was also more variance in energy and pitch in charismatic speeches compared to non-charismatic ones. These results are in line with existing research findings that charismatic speeches are more animated and less monotone.

We then furthered our analysis to explore, for example, how speech, energy and pitch vary, in the hope of deriving design principles to synthesize charismatic

speech. In this paper, we focused our analysis on pauses. Our data show that charismatic speeches contained significantly more silence pauses, compared to non-charismatic ones. We further identified the linguistic features, i.e. the POS tag pairs, that are more frequently related to pauses.

Pause Synthesis. Based on the analysis of pauses in our dataset, we have begun to experiment with a number of ways to synthesis pauses to express charisma. We first used a commercial speech synthesizer (Amazon Polly, [2]) to generate a baseline or neutral recording of the tutorial. Given that we used .2 s of silence as the threshold to extract pauses in our data, we plan to insert silent pauses of .2 s into the baseline speech.

There are a number of methods we plan to explore and experiment with in where to insert the pauses. First, we can insert pauses between POS tags that our data suggest are more likely to have a pause. A probability distribution can be employed to determine how often pauses should be inserted. Second, we can take a consensus-based approach and insert pauses where, for example, more than half of the speakers in our data paused. Ultimately, it's a balance between precision and recall [13]: we can generate fewer pauses with high confidence or fill the speech with more pauses while lowering the threshold of certainty. One of the immediate next steps is to carry out human-subject studies with the synthesized speech to study its impact on perceived charisma.

Limitations. While the balance between precision and recall is a general approach, the method to synthesize, for example, pauses in charismatic speech is specific to the tutorial text of interest to this project. This is largely due to the nature of the dataset, e.g., multiple recordings of the same text, and the small size of the dataset. To generalize the approach, we plan to extend the POS-based analysis to POS dependencies based on sentence structures. Such an approach is not applicable to our existing dataset, given that the charismatic text used for speech data recording has very limited representation of sentence structures. Thus, as one of the next steps, we plan to extend the analysis to large publicly available speech datasets of charismatic speakers (e.g., speeches from past presidents, motivational speakers, etc.).

References

1. Al Mahmud, A., Dadlani, P., Mubin, O., Shahid, S., Midden, C., Moran, O.: iParrot: towards designing a persuasive agent for energy conservation. In: de Kort, Y., IJsselsteijn, W., Midden, C., Eggen, B., Fogg, B.J. (eds.) Persuasive 2007. LNCS, vol. 4744, pp. 64–67. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-77006-0_8

2. Amazon. Amazon Polly (2022). https://aws.amazon.com/polly/
3. Antonakis, J., Fenley, M., Liechti, S.: Can charisma be taught? tests of two interventions. Acad. Manag. Learn. Educ. **10**(3), 374–396 (2011)
4. Atkinson, J.M.: Lend me your ears: all you need to know about making speeches and presentations. Oxford University Press on Demand (2005)
5. Awamleh, R., Gardner, W.L.: Perceptions of leader charisma and effectiveness: the effects of vision content, delivery, and organizational performance. Leadersh. Quart. **10**(3), 345–373 (1999)
6. Bickmore, T.W., Pfeifer, L.M., Jack, B.W.: Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1265–1274 (2009)
7. Bower, G.H.: Experiments on story understanding and recall. Quart. J. Exp. Psychol. **28**(4), 511–534 (1976)
8. Burkhardt, F., Sendlmeier, W.F.: Verification of acoustical correlates of emotional speech using formant-synthesis. In: ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion (2000)
9. Campbell, N., Hamza, W., Hoge, H., Tao, J., Bailly, G.: Editorial special section on expressive speech synthesis. IEEE Trans. Audio Speech Lang. Process. **14**(4), 1097–1098 (2006)
10. Campione, E., Véronis, J.: A large-scale multilingual study of silent pause duration. In: Speech Prosody 2002, International Conference (2002)
11. Charteris-Black, J.: Politicians and Rhetoric: The Persuasive Power of Metaphor. Springer, London (2011). https://doi.org/10.1057/9780230319899
12. Chi, M.T., Siler, S.A., Jeong, H., Yamauchi, T., Hausmann, R.G.: Learning from human tutoring. Cogn. Sci. **25**(4), 471–533 (2001)
13. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 233–240 (2006)
14. DeGroot, T., Aime, F., Johnson, S.G., Kluemper, D.: Does talking the talk help walking the walk? An examination of the effect of vocal attractiveness in leader effectiveness. Leadership Quart. **22**(4), 680–689 (2011)
15. Den Hartog, D.N., Verburg, R.M.: Charisma and rhetoric: communicative techniques of international business leaders. Leadership Quart. **8**(4), 355–391 (1997)
16. Dumdum, U.R., Lowe, K.B., Avolio, B.J.: A meta-analysis of transformational and transactional leadership correlates of effectiveness and satisfaction: an update and extension. In: Transformational and Charismatic Leadership: The Road Ahead 10th Anniversary Edition, pp. 39–70. Emerald Group Publishing Limited (2013)
17. Eden, D., et al.: Implanting pygmalion leadership style through workshop training: seven field experiments. Leadership Quart. **11**(2), 171–210 (2000)
18. Emrich, C.G., Brower, H.H., Feldman, J.M., Garland, H.: Images in words: presidential rhetoric, charisma, and greatness. Admin. Sci. Quart. **46**(3), 527–557 (2001)
19. Erickson, D.: Expressive speech: production, perception and application to speech synthesis. Acoust. Sci. Technol. **26**(4), 317–325 (2005)
20. Frese, M., Beimel, S., Schoenborn, S.: Action training for charismatic leadership: two evaluations of studies of a commercial training module on inspirational communication of a vision. Person. Psychol. **56**(3), 671–698 (2003)
21. Frick, R.W.: Communicating emotion: the role of prosodic features. Psychol. Bull. **97**(3), 412 (1985)

22. Fujisaki, H.: Prosody, models, and spontaneous speech. In: Computing Prosody, pp. 27–42. Springer, New York (1997). https://doi.org/10.1007/978-1-4612-2258-3_3

23. Gasper, J.M.: Transformational leadership: an integrative review of the literature. Western Michigan University (1992)

24. Gobl, C., Bennett, E., Chasaide, A.N.: Expressive synthesis: how crucial is voice quality? In: Proceedings of 2002 IEEE Workshop on Speech Synthesis, 2002, pp. 91–94. IEEE (2002)

25. Goldman-Eisler, F.: The distribution of pause durations in speech. Lang. Speech **4**(4), 232–237 (1961)

26. House, R.J.: A 1976 theory of charismatic leadership. Working Paper Series 76–06 (1976)

27. Igras-Cybulska, M., Ziółko, B., Żelasko, P., Witkowski, M.: Structure of pauses in speech in the context of speaker verification and classification of speech type. EURASIP J. Audio Speech Music Process. **2016**(1), 1–16 (2016). https://doi.org/10.1186/s13636-016-0096-7

28. Judge, T.A., Piccolo, R.F.: Transformational and transactional leadership: a meta-analytic test of their relative validity. J. Appl. Psychol. **89**(5), 755 (2004)

29. Kirsner, K., Dunn, J., Hird, K.: Language production: a complex dynamic system with a chronometric footprint (2005)

30. Lea, W.A.: Trends in Speech Recognition. Prentice Hall PTR (1980)

31. Lee, C.M., et al.: Emotion recognition based on phoneme classes. In: Eighth International Conference on Spoken Language Processing (2004)

32. Martha, A.S.D., Santoso, H.B.: The design and impact of the pedagogical agent: a systematic literature review. J. Educat. Online **16**(1), n1 (2019)

33. Mio, J.S., Riggio, R.E., Levin, S., Reese, R.: Presidential leadership and charisma: the effects of metaphor. Leadership Quart. **16**(2), 287–294 (2005)

34. Petrov, S., Das, D., McDonald, R.: A universal part-of-speech tagset. arXiv preprint arXiv:1104.2086 (2011)

35. Pitrelli, J.F., Bakis, R., Eide, E.M., Fernandez, R., Hamza, W., Picheny, M.A.: The IBM expressive text-to-speech synthesis system for American English. IEEE Trans. Audio Speech Lang. Process. **14**(4), 1099–1108 (2006)

36. Rochester, S.R.: The significance of pauses in spontaneous speech. J. Psycholinguist. Res. **2**(1), 51–81 (1973)

37. Roehling, S., MacDonald, B., Watson, C.: Towards expressive speech synthesis in English on a robotic platform. In: Proceedings of the Australasian International Conference on Speech Science and Technology, pp. 130–135 (2006)

38. Rosenberg, A., Hirschberg, J.: Charisma perception from text and speech. Speech Commun. **51**(7), 640–655 (2009)

39. Shamir, B., Arthur, M.B., House, R.J.: The rhetoric of charismatic leadership: a theoretical extension, a case study, and implications for research. Leadership Quart. **5**(1), 25–42 (1994)

40. Shamir, B., Arthur, M.B., House, R.J.: The rhetoric of charismatic leadership: A theoretical extension, a case study, and implications for research. In: Leadership Now: Reflections on the Legacy of Boas Shamir, pp. 31–49. Emerald Publishing Limited (2018)

41. Shamir, B., House, R.J., Arthur, M.B.: The motivational effects of charismatic leadership: a self-concept based theory. Organ. Sci. **4**(4), 577–594 (1993)

42. Shen, J., et al.: Natural TTS synthesis by conditioning Wavenet on Mel spectrogram predictions. arXiv preprint arxiv:1712.05884 (2017)

43. Stylianou, Y.: Applying the harmonic plus noise model in concatenative speech synthesis. IEEE Trans. Speech Audio Process. **9**(1), 21–29 (2001)
44. Székely, É., Henter, G.E., Beskow, J., Gustafson, J.: Spontaneous conversational speech synthesis from found data. In: Interspeech, pp. 4435–4439 (2019)
45. Tao, J., Kang, Y., Li, A.: Prosody conversion from neutral speech to emotional speech. IEEE Trans. Audio Speech Lang. Process. **14**(4), 1145–1154 (2006)
46. Tokuda, K., Nankaku, Y., Toda, T., Zen, H., Yamagishi, J., Oura, K.: Speech synthesis based on hidden Markov models. Proc. IEEE **101**(5), 1234–1252 (2013)
47. Tokuda, K., Zen, H., Black, A.W.: An HMM-based speech synthesis system applied to English. In: IEEE Speech Synthesis Workshop, pp. 227–230. IEEE Santa Monica (2002)
48. Touati, P.: Prosodic aspects of political rhetoric. In: ESCA Workshop on Prosody (1993)
49. Towler, A.J.: Effects of charismatic influence training on attitudes, behavior, and performance. Person. Psychol. **56**(2), 363–381 (2003)
50. Wang, N., Pacheco, L., Merchant, C., Skistad, K., Jethwani, A.: The design of charismatic behaviors for virtual humans. In: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, pp. 1–8 (2020)
51. Wang, Y., et al.: Tacotron: a fully end-to-end text-to-speech synthesis model. arXiv preprint arxiv:1703.10135 (2017)
52. Wasielewski, P.L.: The emotional basis of charisma. Symbol. Interact. **8**(2), 207–222 (1985)
53. Weber, M.: The Theory of Social and Economic Organization. Simon and Schuster (2009)
54. Williams, C.E., Stevens, K.N.: Emotions and speech: some acoustical correlates. J. Acoust. Soc. Am. **52**(4B), 1238–1250 (1972)
55. Ze, H., Senior, A., Schuster, M.: Statistical parametric speech synthesis using deep neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 7962–7966. IEEE (2013)