Computer-Generated Speaker Charisma and Its Effects on Human Actions in a Car-Navigation System Experiment - or How Steve Jobs' Tone of Voice Can Take You Anywhere

Oliver Niebuhr[1(✉)] and Jan Michalsky[2]

[1] Centre for Industrial Electronics, University of Southern Denmark, Sonderborg, Denmark
oniebuhr@mci.sdu.dk
[2] Chair of Technology Management, University of Erlangen, Nuremberg, Germany
jan.michalsky@fau.de

Abstract. A particularly persuasive (charismatic) tone of voice has a far reaching influence on people's opinions and actions. However, does this also apply if the charismatic tone of voice is produced by a computer, and if this computer asks people to act against better knowledge? Addressing these questions, an experiment was set up in which 30 locals of Sonderborg/DK were asked to conduct a test drive with a car from the marina to the university campus of the city. The test drive was conducted on the pretext of assessing a newly developed retrofit car navigation system that provides voice instructions only. The locals did not know that the system was only a remote-controlled mock-up that, moreover, started giving its driver wrong instructions after about half of the trip. The instructions got successively falser, until the only option to get to the university campus was to make a complete U-turn. We measured the point at which the drivers aborted the test drive in two conditions. In the first one, the system spoke with the more charismatic tone of voice of Steve Jobs. In the second one, it spoke with the less charismatic tone of voice of Mark Zuckerberg. Results show that drivers followed the navigation system's increasingly worsening instructions significantly longer in the Steve Jobs condition, and that the system received higher quality, trustworthiness, and purchase ratings if it spoke with Steve Jobs' tone of voice. Results are discussed in terms of persuasive technology and implications for charisma/leadership analyses and training.

Keywords: Speaker charisma · Prosody · Tone of voice · Text-to-speech-synthesis · Navigation system · Steve jobs · Mark zuckerberg

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11620, pp. 375–390, 2019.
https://doi.org/10.1007/978-3-030-24296-1_31

1 Introduction

1.1 The Importance of Charismatic Speech

Charisma has become a key concept of persuasiveness, influence, and leadership (Grabo et al. 2017; Antonakis and Gardner 2017). Although research on charisma has a long tradition and is firmly rooted in diverse scientific fields, ranging from political sciences through psychology to economics, progress towards reaching an actual understanding has only been made in the past two decades (Antonakis et al. 2016; Niebuhr et al. 2017). While older studies treat charisma as a magical and indescribable property, innate to certain individuals (Weber 1947), it is now clearly defined and operationalized as a competence of "values-based, symbolic, and emotion-laden leader signaling" (Antonakis et al. 2016: 304), a skill that all speakers possess to various, context-specific degrees and that they can train and improve. Most importantly, numerous studies have shown that charisma significantly influences listeners by positively affecting motivation and inspiration (Towler 2003), productivity (Howell and Frost 1989, Antonakis et al. 2011), attention (Towler et al. 2014), trust, and confidence in the speaker's abilities (Davis et al. 2017, Niebuhr and Skarnitzl 2019). These influences are the basis for the ability of speakers to change people's actions, attitudes, and opinions. One potential reason for these extraordinary capabilities of charisma may lie in the finding that charisma can bypass rational thinking and affect a listener's emotions and emotional decisions more directly by inhibiting activity in the medial and dorsolateral areas of the prefrontal cortex of the brain, which contribute to abstract reasoning and rational decision-making (Schjødt et al. 2011). Or, in the words of Soorjoo (2012): The most ancient "reptilian" part of our brain functions as a gatekeeper in all decision-making processes, and charisma signals are like the permit to pass by that gatekeeper.

1.2 Phonetic Exponents of Perceived Speaker Charisma

Advancements in the field of phonetics have contributed greatly to the demystification of charismatic speech. Fox Cabane (2012) already pointed out that "nonverbal modes of communication are hardwired in our brains, much deeper than the more recent language-processing [i.e. word-related] abilities, and they affect us more strongly" (p.17). Phonetic research has refined this statement in showing that it is actually the speaker's tone of voice (rather than body language, the other nonverbal mode) that plays a vital role and represents one if not *the* primary trigger of perceived speaker charisma (Chen et al. 2014, Scherer et al. 2012). Furthermore, phonetic research successively identified the acoustic parameters of charismatic speech and determined how and to what extent they contribute to perceived charisma. Charismatic speakers employ acoustic characteristics such as an enhanced variation in F0 (the acoustic equivalent of perceived pitch), an expanded F0 range, an increased speaking rate, a larger number of silent pauses within and between sentences, fewer and shorter filled pauses (Touati 1993; Strangert and Gustafson 2008; Rosenberg and Hirschberg 2009; D'Errico et al. 2013; Niebuhr et al. 2016a, b, Niebuhr et al. 2018a, b; Niebuhr and Fischer 2019), a strategic placement of empathic accents (Niebuhr 2010, Niebuhr et al.

2016a), as well as a higher level of articulatory precision (Niebuhr and Gonzalez 2019) and a more balanced spectral-energy distribution (Niebuhr et al. 2018b). In conclusion, the phonetics research has – to a large degree – identified the complex constellation of acoustic features that makes a speaker sound more charismatic.

1.3 Using Machines for Understanding Perceived Speaker Charisma

Apart from a few exceptions (e.g., Towler 2003; Antonakis et al. 2011), previous studies mainly tested the influence of acoustic features on the perception of speaker charisma by means of listener ratings (Touati 1993; Strangert and Gustafson 2008; Rosenberg and Hirschberg 2009; D'Errico et al. 2013; Berger et al. 2017; Niebuhr et al. 2018a, b; Jokisch et al. 2018; Niebuhr and Fischer 2019). The first one who, with a focus on tone-of-voice acoustics (i.e. speech prosody), consistently went beyond this rating paradigm was Fischer (2018). She used robots to test people's reactions to differently charismatic voices. These voices (as well as the corresponding sentences that were spoken) were based on computer-generated text-to-speech synthesis (TTS). TTS has the advantage that acoustic parameters, including all listed in 1.2 above, can be very precisely controlled and manipulated, while all other charisma-related factors like the choice of words and the outer appearance of the robot are kept constant. Moreover, robots are, unlike real humans or videos of humans, neutral with respect to aspects like gender, age, attire, and social status. Charismatic voices in robots have been found to affect listeners' opinions and actions in a way similar to that of real charismatic speakers. For example, with respect to actions, the charismatic robots of Fischer (2018) made people fill out longer questionnaires, book certain sightseeing trips, and even choose fruits over sweets as a reward after completing a task.

1.4 The Goal of This Study

In this study we are taking the findings of robots influencing human behavior through vocal charisma one step further. All previous experiments of Fischer and colleagues tested effects of charismatic voices of robots in settings in which listeners had to choose between two equally viable options. By contrast, in this study, we ask whether charisma can influence a listener's choice even if the suggested option is clearly wrong and listeners would have to act against better knowledge. Furthermore, we not only want to ask *whether*, but also *to what degree* a charismatic voice is able to influence listener behavior.

To this end, we designed an experiment which investigates the question how well a (computer-generated) charismatic speaker can make people follow in the very sense of "following", i.e. to obey the instructions of a mock-up car navigation system. The navigation system was fitted with two different tone-of-voice profiles, constituted by a set acoustic features with either a more or a less charismatic parameter setting. These two different tone-of-voice settings are then combined with the same lexical content and the same synthetic TTS voice. Since we are not only interested in *if* but also in *how far* listeners are willing to follow a charismatic voice, we measured the charismatic influence in terms of the duration listeners were following wrong instructions until they aborted the test drive by disobeying the navigation system's instructions.

The parameter settings for the more charismatic condition were derived from the tone-of-voice profile of Steve Jobs, who was an exceptionally charismatic speaker, greatly affecting the opinions of his listeners and enhancing trust in his company (Niebuhr et al. 2016b). The less charismatic tone-of-voice profile was derived from speeches by Mark Zuckerberg, who, unlike Jobs, is as a less charismatic and convincing speaker. These assumptions about the two speakers have already been confirmed for their acoustic profiles by previous perception studies with depersonalized, low-pass-filtered speech (Niebuhr et al. 2018a; Niebuhr and Gonzalez 2019).

In conclusion, this paper investigates whether listeners are more inclined to follow purposefully wrong instructions of a navigation system when fitted with the charismatic acoustic features of Steve Jobs than when fitted with the less charismatic acoustic features of Mark Zuckerberg.

2 Method

2.1 Participants of the Test Drive

A total of 30 participants were included in the experiment, 15 males and 15 females. They were all advanced students on campus Sønderborg of the University of Southern Denmark (SDU) and between 20 and 25 years old (male average 21.5 years; female average 22.3 years). All had a European driver's license for at least 2 years and were well familiar with the city of Sønderborg and the topology of its streets insofar as they had continuously lived in Sønderborg for several years prior to the experiment. Furthermore, all participants were proficient non-native speakers of English (at least level B2 according to university-internal study-entry tests). None of the participants had regular access to an own car with an onboard GPS navigation system (but all had some experience with smart phone navigation apps).

Half of the participants (i.e. 15, 8 males, 7 females) were randomly assigned to Steve Jobs' tone-of-voice condition (SJ), the other half (7 males, 8 females) was assigned to Mark Zuckerberg's tone-of-voice condition (MZ).

2.2 Map and Route of the Test Drive

The route that the participants had to drive in the experiment was designed such that the entire test drive would be at most 15 km long (about 50% of it freeway). Thus, a single experimental test-drive session would not take longer than 15 to 20 min. In addition, the route was created such that the navigation system would give correct instructions for about 1.5 km, in order to familiarize participants with the experimental situation and build up trust in the system. After these initial 1.5 km, the system's route guidance successively led away from the shortest ideal route and the final destination of the test drive. More specifically, there were 4 escalation levels, which are also shown in Fig. 1. Each escalation level was similarly long, i.e. 1.5–2.5 km, and represented by at least two semantically and syntactically similar oral instructions (prompts) of the navigation system.



Fig. 1. Map of Sønderborg showing place of departure and destination of the test drive, as well as the ideal route (green) and the stepwise deviations from this route (1–4) demanded by the navigation system. Edited screenshot of Google Maps.

Place of departure was the large parking lot at the Sonderborg Marina (SBG Marina). The test-drive's destination was the large parking lot on the university campus, on the opposite side of the Sonderborg Fjord (SDU). Participants did not know that they would actually never arrive there when following the navigation system's oral instructions. Place of departure and destination were both well known to all participants; and they were close enough together for the correct and ideal (shortest and fastest) route to be obvious to all participants.

The correct and ideal route guidance (escalation level 0) is shown in green in Fig. 1, with a dotted line starting from where the wrong instructions of the navigation system began. Escalation level 1 (yellow) was identifiable by all participants as an obvious detour, but one that still left the door open for many alternative routes that would all lead to the destination with only a little extra time. Escalation level 2 (orange), with its freeway entrance, already meant a substantial detour and extra time and left over only a few alternative routes to the destination. Escalation level 3 (coral) ran past the last freeway exit in front of the fjord bridge, thus leaving over only one possible route guidance to the destination, i.e. the freeway exit immediately after the bridge. Escalation level 4 (red) also ran past this exit and hence led away from very last alternative route to the destination. Thus, except for a U-turn, there was no possibility left anymore for the participants to reach their destination.

An abort of the test drive is defined here as the moment at which the driver no longer followed the instructions of the navigation system and either stopped at the roadside (in order to discuss the subsequent actions with the experimenter) or autonomously took an alternative route to the destination without consulting the experimenter.

If the participant did still not abort the test drive at escalation level 4, then the experimenter aborted the test drive after one more minute by asking the participant to turn around and drive back to the destination (SDU parking lot) without further assistance from the navigation system. After having arrived at the destination, the experimenter asked the participant to fill out a final questionnaire, the discussion of which had then concluded the experimental session. An experimental session took about 30 min, including the participant's briefing and de-briefing.

2.3 Stimuli: Design and Synthesis of Navigation Prompts

The individual stimuli of the navigation system were generated with a free online TTS tool (https://text2speech.us). The TTS voice was "Matthew" and selected for its fluent speech and sonorous voice. "Matthew" also represented American English speech. Thus, the synthetic "Matthew" was in accord with the real speakers SJ and MZ whose prosodic profiles were to be imprinted on the voice in a subsequent step.

Eighteen individual navigation prompts were generated. They corresponded to the prompts of a regular navigation system in length, syntax and instruction content. The navigation prompts were preceded by two supplementary voice prompts, with which the navigation system introduced itself to the driver at the beginning of the experiment and in which it asked the driver to start the navigation by saying "GO". This two-part introduction was about 30 s long.

- *TTS: "Hello. I am your new retrofit navigation system. You can install me subsequently in any car radio… even without much cost and effort. Great, right?! Today, I take you from Sonderborg Marina to the University of Southern Denmark on the opposite side of the fjord." Are you ready? I can start right away! Please say "go" to start the navigation.*
- *Driver (typically after reconfirming eye contact with the experimenter): "Go!"*
- *TTS: "Thank you, the navigation starts now. I will guide you safely to your destination. Enjoy the trip with me."*

The two initial voice prompts were integrated into the experiment for two reasons. First, they were supposed to enhance the driver's trust in the navigation system and embody the system in an acoustic form. Second, the two initial voice prompts were the system's longest coherent stretches of speech in the experiment. Placing them at the beginning of the session allowed the prosodic differences between the SJ and the MZ conditions to be clearly established before the test drive began. Also, the charisma effects of these prosodic differences had a chance to unfold already before driving.

A total of 20 TTS items were generated, including the two initial voice prompts. The total duration of all 20 items (henceforth stimuli) was 1 min and 38 s (or 98 s).

Note that it is a bit unusual to use a male voice for navigation prompts in the car. Voices of navigation systems are usually female (Nass and Yen 2012:7). However, here we had to conduct manipulations of the stimuli based on empirical prosodic profiles of two male speakers. Neither the target values of this manipulation nor their persuasive effects on participants can be transferred 1:1 from men to women (Novák-Tót et al., 2017; Jokisch et al., 2018; Niebuhr et al. Niebuhr et al. 2018a, b). Therefore, we had to use a male voice for the stimuli of the navigation system.

2.4 Stimuli: Creation of Tone-of-Voice Conditions SJ and MZ

After all 20 stimuli had been generated using the free online TTS tool, the two prosodic charisma conditions of the experiment were generated. The manipulation was carried out by means of PSOLA resynthesis in PRAAT (Moulines and Charpentier 1990). As was stated in the Introduction, the acoustic tone-of-voice profiles of SJ and MZ served as a template for the manipulation.

Table 1. Acoustic tone-of-voice profiles (of selected PSOLA-compatible parameters) of SJ and MZ, as determined in a speech-corpus analysis by Niebuhr et al. (2016b).

Acoustic-prosodic parameter	Mean value SJ	Mean value MZ
F0 level (Hz)	221.4	178.3
F0 range (semitones)	23.1	12.1
Speaking rate (syll/s)	4.7	6.1
IPU duration (s)	1.3	1.5
Emphatic-accent frequency (cpm)	5.4	1.2

Five melodic characteristics were manipulated, whose positive correlation with perceived charisma and the persuasive power of a speaker had been attested in previous studies (see 1.2). The five manipulated parameters were: F0 level, F0 range, speaking rate, duration of inter-pausal speech units (IPUs, i.e. how long speakers talk in between two audible breaks), and emphatic-accent frequency (i.e., words whose pitch-accented syllables showed a strong lengthening of either the initial consonant or the syllable nucleus, see Niebuhr 2010). Table 1 shows the mean values obtained for SJ and MZ on these five parameters.

The PSOLA manipulation was performed such that each of the 20 stimuli matched the melody profiles of SJ and MZ in terms of the five mean values. In the case of the emphatic-accent frequency, however, this only applied to the two initial voice prompts. The following 18 navigation prompts were manipulated such that in the SJ condition all of them and in the MZ condition none of them contained an emphatic accent.

Note that the result of the manipulation was not that the navigation system spoke with the actual voices of SJ or MZ. The voice quality was exactly the same (= ”Matthew”) in both experimental conditions (as was the stimulus wording). Only five aspects of the tone of voice, i.e. the system’s speech melody, had been changed.

2.5 Car

The car used for the experiment was a Nissan Qashqai J10 (2014 model) that neither had a built-in navigation system nor a big LCD information screen, see Fig. 2.

2.6 Procedure of the Test Drive

The experiment was conducted in individual experimental sessions per participant. Participants were invited to the experiment on the pretext that they would take part in a



Fig. 2. Pictures of the Nissan Qashqai used for the experimental test drive of all 30 participants.

usability test. At the beginning of each session, the participant was informed that s/he would be supposed to drive with the provided car from Sønderborg Marina to SDU campus Sønderborg, and that the purpose of the short trip would be to test the practicability and usability of a retrofit navigation system, developed by an SDU engineering team in collaboration with Danish industrial partners. The advantage of this innovation would be that it can be installed in any older car radio system without much cost and effort. In consequence, however, the navigation would be fully speech-based, without the visual support of a separate display. To examine if and how well this works was said to be the primary aim of the usability test.

After participants had given their written consent to take part in the experiment, both participant and experimenter got into the car; the participant on the driver's seat and the experimenter in the backseat of the car, opposite the driver's seat.

The participant was told that the experimenter would be on the trip for two reasons; firstly, because the car insurance policy would require his physical presence; and secondly, because the experimenter would take notes on his laptop about how he observes and assesses the test drive. These notes would then be added to the usability ratings that the participant were to make after the test drive by filling out a prepared questionnaire.

That is, the participant was left in the dark for the duration of the entire trip that the navigation system was a mock-up and that individual stimuli (navigation prompts) were triggered in a Wizard-of-Oz fashion (Fischer 2018) at the appropriate time by the experimenter in the back seat. The stimuli were played from the experimenter's laptop at a pre-adjusted loudness level that was kept constant for all 30 participants.

It was stressed to the participant that s/he could abort the experiment (the usability test) at any time. Questions, expressions of uncertainty and other similar comments to the experimenter, however, were not allowed during the trip. The participant was told that, ideally, s/he would behave as if the experimenter were not present at all in the backseat. Upon oral confirmation that this was understood, the trip began with the (Wizard-of-Oz) playback of the first introductory navigation prompt (2.3) by the experimenter.

When the test drive had been aborted either by the participant or by the experimenter after reaching escalation level 4, the participant drove directly to the SDU parking lot. After the car had arrived there and the engine was turned off, the participant

was asked to fill out a questionnaire. The questionnaire recorded the personal data of the participant and ended with the following ten questions about the navigation system itself.

(1) How do you rate the quality of the navigation system's speech prompts?
(2) How do you rate the intelligibility of the navigation system's speech prompts?
(3) Were the instructions of the navigation system always clear?
(4) Did the navigation system guide you on the fastest way to your destination? If not, what do you think was the reason for the detour?
  (4a) The navigation system may have attempted to bypass a traffic jam, accident, or road closure,
  (4b) The navigation system is working properly, but the maps or the installation were faulty, or the engineers have made programming errors,
  (4c) The navigation system itself is simply malfunctioning.
(5) How much did you trust the navigation system while driving?
(6) How likely is it that you would buy the navigation system for yourself?
(7) How likely is it that you would recommend the navigation system to your friends and relatives?

The question triplets (1)–(3) and (5)–(7) were answered on the basis of a 7-point Likert scale from 1 'very bad' or 'not at all' to 7 'very good' or 'very much'. Question (4) and its sub-questions were answered by ticking 'yes' or 'no'. Multiple 'yes/no' answers were allowed for (4a–c).

Only after each participant had completed the entire questionnaire, s/he was informed that the navigation system was just a mock-up and actually remotely controlled by the experimenter in the back seat. The participants were told that they did not even take part in a usability test. The real purpose of the experiment had been to test at what point they would abort the test drive and how that would correlate with the navigation system's speech melody, of which there were two variants, i.e. two experimental conditions. No participant indicated that s/he had guessed the actual aim of the experiment.

Note that the experiment was only carried out in between 10.00 in the morning and 14.00 in the afternoon, hence avoiding the rush hours on the streets of Sønderborg.

3 Results

Statistical pilot tests showed that drivers' gender had no separate significant effects on the driving or subsequent rating behaviors. Therefore, we disregarded this factor in the presentation of the results.

Figure 3 shows in percentages how many of the 30 drivers did not or did not abort the test drive at the escalation levels 1–4. Five results are worth being pointed out separately. First, none of the 30 drivers aborted the test drive already at escalation level 0 at which the navigation system still gave the correct instructions. Second, 20% of all drivers in the MZ-condition aborted the test drive already at escalation level 1, i.e. immediately after the navigation system started giving the first wrong instructions (and although there were still many alternative routes left that led to the destination). In the

SJ condition, only 6.7% of all drivers aborted the test drive at this early stage. Third, in the MZ condition the majority of drivers, i.e. 40%, aborted the test drive at escalation level 3 and only a minority of 13.3% continued the test drive until the final escalation level 4. In the SJ condition, these two numbers were equally large, i.e. 26.7%. Fourth, when there was only one possible route guidance left to the destination, i.e. at escalation level 3, a majority 53.3% of all drivers still followed the navigation system's instructions in the SJ condition, as opposed to a minority of only 20% in the MZ condition. Five, and probably most noteworthy, even after escalation level 4 had been reached and no option other than a U-turn was left, more than one quarter of all drivers in the SJ condition, i.e. 26.7%, did not abort the test drive so that the experimenter had to intervene and ask the driver to abort the test drive. In the MZ condition, this only applied to 6.7% of all drivers.

Due to a lack of normally distributed data, conservative Mann-Whitney U tests were conducted to compare the drivers' behavior in the two independent samples represented by the SJ and MZ conditions. Dependent variable was the point in time at which the test drive was aborted. This variable was quantified, in one U test, in terms of

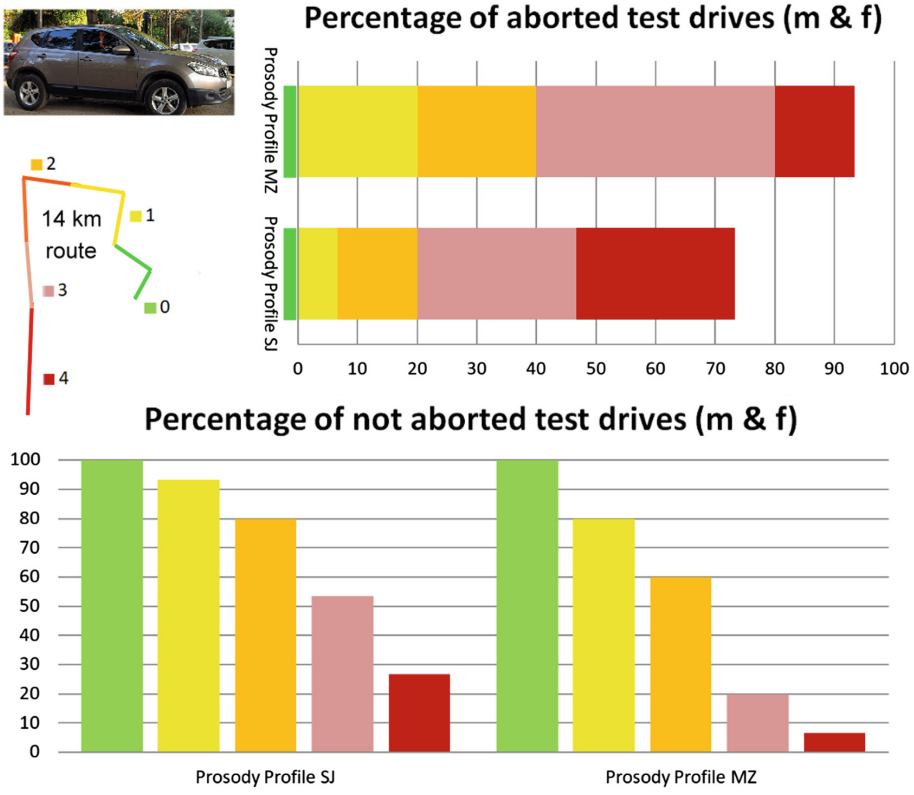

Fig. 3. Summary of test-drive abort/continuation percentages in the SJ and MZ conditions (n = 15).

the abort frequency at the 5 escalation levels and, in another U test, in terms of the chronological number of the navigation prompt (1–20) after which each driver aborted the test drive. The U test on the escalation levels showed a trend towards an overall earlier test-drive abort in the MZ condition ($U[15,15] = 72.5$, $p = 0.100$); in the U test on the navigation prompt count, this trend became significant ($U[15,15] = 56.0$, $p = 0.021$). We also tried to use the actual driving time as a dependent variable, but it turned out in pilot driving tests that this parameter is too susceptible to random and daytime-related changes in traffic density and traffic-light circuits. Thus, it could not be used for statistical testing.

Figures 4 and 5 show the results of the participants' navigation-system ratings in the questionnaire that concluded the experimental session. Again, Mann-Whitney U tests were used to compare the ratings in the SJ and MZ conditions for the Likert scales summarized in Fig. 4. The yes-no question frequencies in the SJ and MZ conditions were statistically analyzed in terms of an overall Chi-squared test, supplemented by Z-Score Proportion tests for the individual yes-no questions.

It is obvious that the TTS voice "Matthew" was rated better with the SJ manipulation than with the MZ manipulation. Only the clarity of instructions was rated to be similarly good in the SJ and MZ conditions. Besides this exception, the navigation system with SJ's acoustic tone-of-voice charisma profile was rated to have a higher speech quality ($U[15,15] = 25.0$, $p < 0.001$) and to be more intelligible ($U[15,15] = 32.5$, $p < 0.001$). In addition, compared to the MZ condition, drivers in the SJ condition stated that they had more trust in the navigation system ($U[15,15] = 33.0$, $p < 0.001$). There was also a higher willingness of the participants to buy the navigation system for themselves ($U[15,15] = 61.0$, $p = 0.031$) and to recommend it to friends and relatives ($U[15,15] = 64.5$, $p = 0.048$), if it was based on SJ's rather than by MZ's tone-of-voice profile.

Regarding the yes-no questions, results show that all drivers in both conditions (100%) were aware of the fact that the navigation system had not guided them on the fastest and shortest way to their destination. Besides this general agreement, there is one big difference between the SJ and MZ conditions. In the SJ condition, drivers showed by a large majority of 80.4% the willingness to explain the unexpected route guidance of the navigation system in a positive way, i.e. in terms of an intended avoidance of traffic jams, accidents, or roadblocks. Another large majority of 73.2% was (additionally) of the opinion that, if the unexpected route guidance was not intended, then the failure had been caused by programmers, engineers or faulty maps. Only a minority of 40.2% saw in the unexpected route guidance a failure of the navigation system itself. Exactly the opposite was true for the drivers in the MZ condition. The vast majority, i.e. 73.2%, ascribed the unexpected route guidance to a direct failure of the navigation system. About two thirds of the drivers, i.e. 66.9% also considered blaming programmers, engineers, or the underlying maps for the failure in route guidance. Only a slim majority of 53.6% were willing to see the unexpected route guidance as a clever, intended solution of the navigation system to a traffic problem on the ideal route to the destination.

That is, ratings *de*creased from 'no failure' through 'external failure' to 'system failure' in the SJ condition, whereas they *in*creased in this order in the MZ condition.

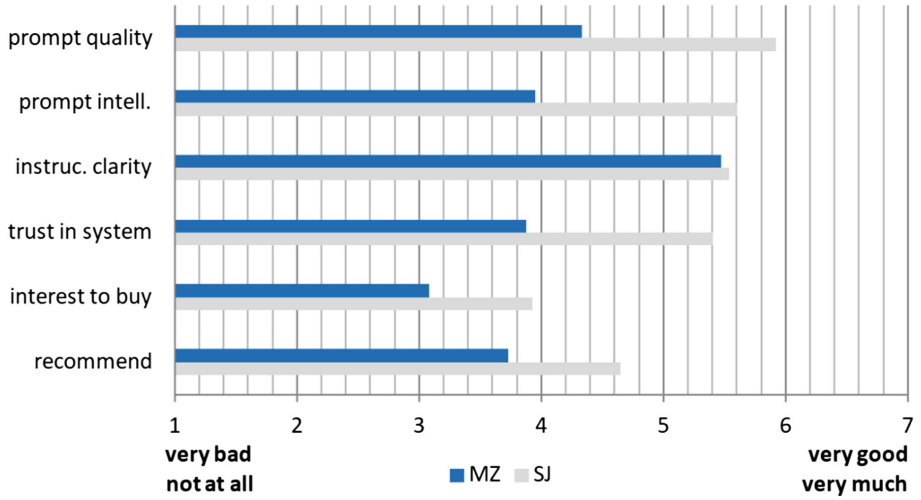

Fig. 4. Mean ratings (from 1 = negative to 7 = positive) on the Likert scales of the concluding questionnaire in the SJ and MZ conditions (n = 15).

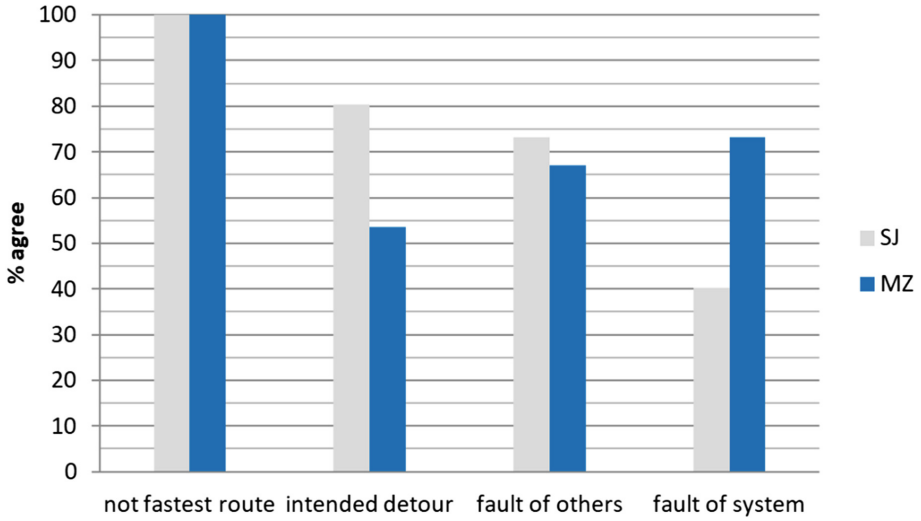

Fig. 5. Percentages of yes/agree answers for the 4 yes-no questions of the concluding questionnaire in the SJ and MZ conditions (n = 15).

This diametrically opposed rating behavior came out significantly in a Chi-squared test ($\chi^2[2] = 15.5$, $p < 0.001$). Separate Z-Score Proportion tests showed that it were particularly the differences in 'intended detour' ($z[15,15] = 1.7$, $p = 0.09$) and 'fault of system' ($z[15,15] = -1.8$, $p = 0.06$) that caused this overall significance.

4 Discussion

The aim of this study was to test if and to what degree the difference between a more and a less charismatic acoustic tone-of-voice profile (represented by melody parameters accessible to PSOLA manipulation and resynthesis) is able to induce a measureable change in listener behavior. We tested this question on the basis of the tone-of-voice profiles of SJ and MZ. The measureable change in behavior was operationalized by a car driver's trust in, and hence his/her obedience to the oral instructions of a mock-up navigation system. That is, the question was if and at what point drivers would abort their test drive with the mock-up car navigation system, and whether this point differs depending on whether the system's oral instructions have SJ's or MZ's acoustic tone-of-voice profile.

The results show that in the less charismatic condition (MZ), participants aborted the test drive after significantly fewer wrong instructions, with 20% disobeying the navigation system already after the first wrong prompt, when numerous alternative routes to the destination were still left. These 20% contrast with only about 7% in the more charismatic condition (SJ). At escalation level 3, when there was only one possible route left, 80% of the participants had already aborted the test drive in the less charismatic MZ condition, while over 50% were still following the navigation system's instructions in the more charismatic SJ condition. Finally, when the experiment entered the final stage at escalation level 4, at which the instructions lead the participants out of the city with the only remaining option being a U-turn, over 25% of the participants in the more charismatic SJ condition were still following the navigation system's instructions. In contrast, only 7% of the participants still obeyed the navigation system at level 4 if it spoke with MZ's less charismatic tone-of-voice profile. Accordingly, the power of melodic (tone-of-voice) charisma manifested itself in the present study in that it made listeners choose to follow wrong instructions against better knowledge much longer when the charisma signal was stronger, i.e. when the acoustic parameters conveying charisma had higher levels.

But why do listeners obey purposefully misleading instructions when presented in a charismatic way? Or, in other words, how does charisma make a voice persuasive? One possibility might be that charisma works through deception. Accordingly, the charismatic voice could have obscured the fact that the navigation system was in fact giving wrong instructions. However, the results of the debriefing questionnaire show that all 30 speakers were aware that the instructions given by the system were deviating from the ideal route, which contradicts this deception hypothesis. As explained above, charisma gets its persuasive quality by making a speaker sound competent and passionate, which, on the part of the listener, strengthens the willingness for an emotional investment in the speakers' goals and visions (Antonakis et al. 2016). Accordingly, another explanation for the behavior observed here is that listeners (drivers in this case) adhere to the message of a charismatic speaker, because the speaker signals trust and confidence in his/her abilities. The results of the debriefing questionnaire show that participants following the more charismatic SJ profile reported a significantly higher overall trust in the system. The participants stated that they were convinced that the deviating instructions were caused by the system intelligently and intentionally leading

them around road blockages, accidents, or traffic jams. In contrast, participants in the MZ condition significantly more often attributed the deviating instructions to a failure of the system and an error in the given instructions. These findings support the trust hypothesis.

The remarkable special feature in the present study is that the lower/higher values of confidence and competence and hence the de/increase in trust are not attributed to a real human speaker but to a computer, i.e. a car navigation system. So, listeners project the human social signals of a charismatic voice onto a computer system that uses the voice. This has major implications for any technological hard- or software application out there that involves synthetic speech. Manufacturers will have to be aware that not only the voice itself has an impact on the user (e.g., how sonorant and attractive it is), but also how the charisma-related acoustic tone-of-voice features are employed by the synthetic voice. In the same way as a skilled use of charismatic tone-of-voice features would increase users' trust in the system, the wrong implementation of these features can drastically decrease trust and, thus, have major consequences for consumer satisfaction and product sales rates. This applies in particular, as trust is not the only issues that is at stage in a computer system's tone-of-voice design. As was shown in the present study, the participants of the SJ and MZ conditions also rated the system differently with respect to intelligibility, quality, and their willingness to buy the system and to recommend it to their friends, see also Chebat et al. (2007) for the link between charisma or persuasion and commercial success.

In conclusion, this study has three major implications. First, the acoustic features of a charismatic speech melody identified by previous studies elicit the expected effects even when presented solely on an auditory channel and with all other charisma-related features (like choice of words, voice, and body language) kept constant. Second, expanding on the findings of previous studies, we show that a charismatic tone of voice cannot only influence a listener's behavior towards one of two viable options but even make them choose a illogical option over a viable one. Accordingly, the effects of a charismatic speech melody on human behavior are even stronger than found in previous studies (Fischer 2018). Third, the effects of a charismatic tone of voice are not only applicable to real speakers, but can be transferred to synthetic voices of computer systems, as was suggested by previous research on charismatic robots (Fischer 2018). Thus, the correct implementation of a synthetic voice increases user trust and additionally enhances the overall perceived quality of a product and the willingness to buy and recommend it.

References

Antonakis, J., Bastardoz, N., Jacquart, P., Shamir, B.: Charisma: an ill-defined and ill-measured gift. Annu. Rev. Organ. Psychol. Organ. Behav. **3**(1), 293–319 (2016)

Antonakis, J., Fenley, M., Liechti, S.: Can charisma be taught? Tests of two interventions. Acad. Manag. Learn. Educ. **10**, 374–396 (2011)

Antonakis, J., Gardner, W.L.: Charisma: new frontiers. A special issue dedicated to the memory of Boas Shamir. Leadersh. Q. **28**(4), 471–472 (2017)

Berger, S., Niebuhr, O., Peters, B.: Winning over an audience – a perception-based analysis of prosodic features of charismatic speech (2017). In: Proceedings 43rd Annual Conference of the German Acoustical Society, Kiel, Germany, 1454–1457 (2017)

Chebat, J.C., Hedhli, K.E., Gélinas-Chebat, C., Boivin, R.: Voice and persuasion in a banking telemarketing context. Percept. Mot. Skills **104**, 419–437 (2007)

Chen, L., Feng, G., Joe, J., Leong, C.W., Kitchen, C., Lee, C.M.: Towards automated assessment of public speaking skills using multimodal cues. In: Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey, 200–203 (2014)

D'Errico, F., Signorello, R., Demolin, D., Poggi, I.: The perception of charisma from voice. A crosscultural study. In: Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 552–557 (2013)

Davis, B.C., Hmieleski, K.M., Webb, J.W., Coombs, J.E.: Funders' positive affective reactions to entrepreneurs' crowd-funding pitches: the influence of perceived product creativity and entrepreneurial passion. J. Bus. Ventur. **32**, 90–106 (2017)

Fischer, K.: Talking to robots. In: Elmentaler, M., Niebuhr, O. (eds.) An den Rändern der Sprache. Notes of a lecture series at Kiel University. (2018) https://www.uni-kiel.de/pressemeldungen/index.php?pmid = 2018-084-rv-sprache&pr = 1.KielUniversity

Fox Cabane, O.: The Charisma Myth: How Anyone Can Master the Art and Science of Personal Magnetism. Penguin, New York (2012)

Grabo, A., Spisak, B.R., van Vugt, M.: Charisma as signal: an evolutionary perspective on charismatic leadership. Leadership Q. **28**(4), 473–485 (2017)

Howell, J.M., Frost, P.J.: A laboratory study of charismatic leadership. Organ. Behav. Human Decis. Process. **43**(2), 243–269 (1989)

Jokisch, O., Iaroshenko, V., Maruschke, M., Ding, H.: Influence of age, gender and sample duration on the charisma assessment of German speakers. In: Proceedings 29th Conference on Electronic Speech Signal Process, Ulm, Germany, 1–8 (2018)

Moulines, E., Charpentier, F.: Pitch-synchronous waveform processing tech-niques for text-to-speech synthesis using diphones. Speech Commun. **9**, 453–467 (1990)

Nass, C., Yen, C.: The Man Who Lied to His Laptop: What We Can Learn About Ourselves from Our Machines. Pinguin, London (2012)

Niebuhr, O.: On the phonetics of intensifying emphasis in German. Phonetica **67**, 170–198 (2010)

Niebuhr, O., Brem, A., Novák-Tót, E.: Prosodic constructions of charisma in business speeches – a contrastive acoustic analysis of Steve Jobs and Mark Zuckerberg. In: Proceedings of Speech Prosody, 8, Boston, USA (2016a)

Niebuhr, O., Voße, J., Brem, A.: What makes a charismatic speaker? A computer-based acoustic prosodic analysis of Steve Jobs tone of voice. Comput. Human Behav. **64**, 366–382 (2016b)

Niebuhr, O., Tegtmeier, S., Brem, A.: Advancing research and practice in entrepreneurship through speech analysis – from descriptive rhetorical terms to phonetically informed acoustic charisma metrics. J. Speech Sci. **6**, 3–26 (2017)

Niebuhr, O., Thumm, J., Michalsky, J.: Shapes and timing in charismatic speech – evidence from sounds and melodies. In: Proceedings of Speech Prosody, 9, Poznan, Poland (2018a)

Niebuhr, O., Skarnitzl, R., Tylecková, L.: The acoustic fingerprint of a charismatic voice - initial evidence from correlations between long-term spectral features and listener ratings. In: Proceedings of Speech Prosody, 9, Poznan, Poland, 359–363 (2018b)

Niebuhr, O., Gonzalez, S.: Do sound segments contribute to sound-ing charismatic? Evidence from acoustic vowel space analyses of Steve Jobs and Mark Zuckerberg. International Journal of Acoustics and Vibration 24 (2019)

Niebuhr, O., Skarnitzl, R.: Measuring a speaker's acoustic correlates of pitch - but which? A contrastive analysis based on perceived speaker charisma. In: Proceedings of ICPhS 2019, Melbourne, Australia, 1–5 (2019)

Niebuhr, O., Fischer, K.: Do not hesitate! – unless you do it shortly or nasally: how the phonetics of filled pauses determine their subjective frequency and perceived speaker performance. In: Proceedings 20th International Interspeech Conference, Graz, Austria, 1–5 (2019)

Novák-Tót, E., Niebuhr, O., Chen, A.: A gender bias in the acoustic-melodic features of charismatic speech? In: Proceedings of Interspeech 2018, Stockholm, Sweden, 2248–2252 (2017)

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

Scherer, S., Layher, G., Kane, J., Neumann, H., Campbell, N.: An audiovisual political speech analysis incorporating eye-tracking and perception data. In: Proceedings of the 8th International Conference on Language Resources and Evaluation, pp. 1114–1120 (2012)

Schjødt, U., Stodkilde-Jorgensen, H., Geertz, A.W., Lund, T.E., Roepstorff, A.: The power of charisma-perceived charisma inhibits the frontal executive network of believers in intercessory prayer. Soc. Cogn. Affect. Neurosci. **6**, 119–127 (2011)

Soorjoo, M.: Here's the Pitch: How to Pitch Your Business to Anyone, Get Funded, and Win Clients. John Wiley & Sons, Hoboken (2012)

Strangert, E., Gustafson, J.: What makes a good speaker? Subject ratings, acoustic measurements and perceptual evaluations. In: Proceedings of Interspeech 8, Brisbane, Australia, 1688–1691 (2008)

Touati, P.: Prosodic aspects of political rhetoric. Proceedings of the ESCA Workshop on Prosody, Lund, Sweden, 168–171 (1993)

Towler, A.J.: Effects of charismatic influence training on attitudes, behavior, and performance. Pers. Psychol. **56**, 363–381 (2003)

Towler, A., Arman, G., Quesnell, T., Hoffman, L.: How charismatic trainers inspire others to learn through positive affectivity. Comput. Human Behav. **32**, 221–228 (2014)

Weber, M.: The Theory of Social and Economic Organization. The Free Press of Glencoe, New York (1947)