What’s pitch got to do with it?  
The connection between speech prosody and investor-pitch success

Jana Niehof¹, Oliver Niebuhr²

¹ Dept. of Linguistics/Phonetics, University of Cologne, Germany  
² Centre for Industrial Electronics, University of Southern Denmark, Sonderborg, Denmark  
j.niehof@yahoo.com, olni@sdu.dk

Abstract

German investor-pitch presentations were acoustically analyzed based on five tonal and five non-tonal prosodic features. Results show that both feature sets significantly predicted the placements of these presentations in the ‘Rheinland Pitch’ competition series. We conclude in line with previous studies on speaker charisma that prosody plays a considerable role for the success of presentations and their speakers.

Index Terms: investor pitch, prosody, charisma, variation.

1. Introduction

1.1. Previous research on charisma

Charisma – this term has long stood for a special, intangible talent of a few selected people. But, primarily thanks to intensive psychological research [1], we know today that charisma is a “sensation cocktail”, i.e. a complex bundle of multi-modal signals, see, for example, the summary in [2]. This re-conceptualization turned the all-or-nothing talent into a gradual, scientifically tangible skill. Antonakis et al. [3] define charisma as “values-based, emotion-laden leader signalling”, whereby, according to [4], leaders are people who convey not only passion, but also competence and self-confidence.

Highly charismatic people have strong persuasive powers and can win others over to their ideas, products and initiatives. Given that, entrepreneurs should actually benefit particularly strongly from mastering the “sensation cocktail” that creates perceived speaker charisma. Many studies indirectly support this assumption, for example, by measuring success of crowdfunding campaigns [5,6,7]. However, these studies were always about verbal means, gestures, and/or other personality and competence indicators. Phonetic aspects have hardly been addressed; at least not in terms of modern, quantitative analyses – and especially not with respect to prosody, although we know from experiments in politics, religion, education, and language technology that prosodic signal patterns are probably among the most powerful charisma triggers of all. In [8], for example, the researchers were able to correctly predict the outcomes of all US presidential elections from 1960 to 2000 just by means of a single spectral measure that integrated prosodic aspects of pitch and voice quality. The studies of [2,9,10,11] showed that, if machines are given the task to emulate human ratings of perceived speaker charisma or similar traits based on rich multi-modal signals, then these machines learn this task best from the speakers’ prosody. Niebuhr & Michalsky [12] even made drivers drive detours in their own city, just by equipping the car’s navigation system with Steve Jobs’ charismatic tone of voice; and Niebuhr [13] demonstrated that a prosody-based charisma score can predict oral-exam grades and creativity-task team performances of students – as well as the dividend yield paid out by the German DAX40 stock companies.

Against this empirical background, the aim of our paper is, for the first time and based on authentic data, to relate the patterns of speech prosody or prosodic charisma to a quantitative, independent measure of entrepreneurial success. Below, we describe the real-life framework we used to achieve this goal.

1.2. The ‘Rheinland Pitch’ event series

According to their own statements [14], the Rheinland Pitch has become the largest startup pitch competition in Germany. It takes place monthly and gives innovative startups the opportunity to present themselves to investors and, in return, receive financial or advisory support from them.

The event series takes place in real life, i.e. in large arenas with a live audience. However, during the COVID-19 pandemic, the competitions have been moved to the digital stage. That is, the founders and investors were connected via a Zoom call. One by one, the founders gave a shared-screen PowerPoint presentation, introducing an investor panel to the ideas, goals, and team members behind their fledgling companies. The presentations were broadcasted live via YouTube, so that the interested audience that would have normally been in the arena had the chance to participate as well. This unusual transmission method gave us the opportunity to get hold of these startup pitches in an audio quality that lent itself to in-depth acoustic-phonetic analyses and that would not have been achievable with recordings at the actual live events.

Over and above this phonetic argument, the Rhineland Pitch even series also offered some other features that made it an ideal platform for our charisma research. Firstly, presentation time was strictly limited to five minutes. This narrow time window forced all presenters into a comparable framework in terms of the structure and level of detail of their pitches, see [15] for the negative correlation between content and charisma. Five minutes was, moreover, so little time that the time window was always more or less fully utilized by presenters. Thus, overall differences in charismatic impact due to direct or indirect effects of speaking time [15] were ruled out or, at least, greatly reduced. Instead, presentation performance came to the fore. Apart from verbal Charismatic Leadership Tactics (CLTs, [3]), which were not analyzed here, this primarily meant the speaker’s prosody (gestures are hardly used in Zoom calls, see [16]).

Secondly, all presentations and presenters were at a similar level of maturity, as all pitches at each event went through two pre-selection rounds, a written one and an oral one; and, in addition, all presenters received presentation training from a professional coach between the first and second (oral) round.

Thirdly and most importantly, there was a direct vote at the end of each Rhineland Pitch event. The audience and an expert
jury both voted for the startup which they liked the most. The audience and jury votes counted 50% each. At the end of each competition, all presentations were ranked and the winner of the event was chosen. This ranking provided us with an external, real, and meaningful measure of the presenting speakers’ individual charismatic impacts, especially because the ranking was in all cases the result of a large sample of voters (between 300-600 people were typically participating in each event).

1.3. Questions

Acoustic charisma is associated with a variety of prosodic parameters. They cover all four dimensions of prosody and range from, for example, LTAS measures of spectral-energy distribution reflecting voice quality characteristics [17] to timing factors such as phrase duration, pause duration and speaking rate [15], to melodic factors such as pitch level and pitch range [15, 18,19] (with pitch meaning fundamental frequency, f0). Even the articulation and reduction of consonants and vowels contribute to the overall charismatic impression of a speaker [20,21].

Pitch features seem to correlate with perceived speaker charisma in a particularly extensive and strong way [15,18,19]. However, to the best of our knowledge, no study has ever explicitly related these correlations and the resulting predictive power to the predictive power of non-tonal prosodic features. Our main questions are thus: (1) Can the success of the Rhine- land pitch presentations in the form of the joint expert/audioience rankings be significantly predicted solely on the basis of the presenters’ pitch features? (2) If so, can the same be achieved with non-tonal prosodic features? (3) If so, which set of features unfolds the greater predictive power, the tonal or the non-tonal?

Secondly, studies suggest that perceived speaker charisma is similarly highly correlated with static pitch measures, such as mean pitch level and pitch-variability measures, such as the standard deviation. In practice, however, indicators of variability appear to make a greater contribution to a speaker’s perceived charisma than static measures. The role of pitch variability is also underlined in guidebooks on impactful public speaking. For example, ‘The Art of Public Speaking’ [22,7] says: “Every change in the thought demands a change in the voice pitch”. Similarly, in the startup-pitch primer ‘Here’s the Pitch’ by [23;26] it is stated that “a dynamic voice with plenty of variation is energizing and can turn a sleep-inducing talk into a call-to-action speech”.

From a scientific perspective, the two quotes above are noteworthy because they cover exactly those two fundamental domains in speech that are associated with a higher level of pitch variability: A higher degree of emotionality or expressivity (especially in combination with positive emotions, cf. [24]) as well as a richer information structure or stronger prosodic cues to that structure, for example, in the form of clearer acoustic contrasts between stressed and unstressed words and/or between categories of topic and focus [25]. Clearer and more contrasty speech is more charismatic speech, even research on articulation suggests that [20]. More emotional speech, especially in the sense of a higher level of arousal, is more charismatic as well [5,7]. Thus, in view of available scientific evidence, it would be plausible to assume that pitch-variability parameters make a greater contribution to perceived speaker charisma and its objectively measurable real-life effects than pitch-level parameters. An additional question addressed here is therefore: (4) To what extent do common phonetic variability and level measures contribute differently to predicting the investor-pitch rankings and, thus, the entrepreneurs’ success?

2. Method

2.1. Speech Material

The analyzed speech material consisted of 22 startup pitches held over the course of the five regular RhineLand Pitch events between June 2020 and June 2021. Occasionally, special editions of the RhineLand Pitch take place in which, for example, former winners compete against each other or in which the focus is on specific industry branches or business topics. By choosing only presentations from regular events for our speech material, we made sure that all presenters pitched at the event for the first time and that all pitches we analyzed went through the two pre-selection rounds described in 1.2.

We had to exclude one presentation from the speech material as this presentation was held jointly by two founders, who also discussed with each other during the presentation, instead of just presenting the business idea to their (virtual) audience. Table 1 gives an overview of the names of the startups and their respective placements achieved at the end of the event. As can already be seen from the names, the 21 analyzed startups represented various business branches. They ranged from cosmetics/hygiene (‘Deine Pflege’) to IT/Software (‘Recyda’, ‘Comyda’, ‘Annae AI’, ‘Perto’) to data/hardware (‘Einhundert Energie’) to service providers (‘SmartHelio’, ‘Desk Now’).

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
<th>Rank 5</th>
<th>Rank 3</th>
<th>Rank 1</th>
<th>Event 1001</th>
<th>Event 99</th>
<th>Event 99</th>
<th>Event 99</th>
<th>Event 99</th>
<th>Event 97</th>
<th>Event 94</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 Jun 20</td>
<td>25 Jan 21</td>
<td>8 May 21</td>
<td>7 Jun 21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank 1</td>
<td>Tier-IQ</td>
<td>aditalByte</td>
<td>Carbon</td>
<td>Instax</td>
<td>Deine Pflege</td>
<td>Desk Now</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank 2</td>
<td>Comyda</td>
<td>Cinecomics</td>
<td>Smart Helio</td>
<td>Prematch</td>
<td>Zesovi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank 3</td>
<td>Einhundert Energie</td>
<td>Greyparrot</td>
<td>Retracted</td>
<td>Bendesk</td>
<td>Yucca HR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank 4</td>
<td>Perto</td>
<td>Recyda</td>
<td>Annae AI</td>
<td>AmbRoad</td>
<td>Wirewire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank 5</td>
<td>SLOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The 21 startup pitches were played back on YouTube and simultaneously digitally recorded (as uncompressed wav files) in Audacity at 48 kHz, 24 bit. Unwanted sound like audios embedded in the presented slides or feedback and backchannels from the audience were cut out by hand, so that the final audio files only contained the presenters’ speech signals. All audios were furthermore intensity-normalized such that all signal elongations were extended until the peak amplitudes were -1 dB below the maximum. Presenters were mainly men (86%), only three of the 21 presenters were women (14%). The average presentation time across all 21 pitches was 4:54 minutes (sd: 32 s).

2.2. Acoustic Analysis

The acoustic analysis included ten prosodic parameters, divided equally into five pitch and five non-pitch parameters. All measurements were taken on the basis of major intonation phrases (IPs). The corresponding segmentation was carried out automatically – with manual cross-checks – using the script of [26].

The five pitch parameters, with one value determined per IP, were mean pitch level (f0 in Hz), pitch range (i.e. max-min pitch, in semitones/st), pitch variability (f0 std dev. in Hz), maximum pitch velocity (f0 change in st/s), as well as the relative location of the pitch peak within the phrase (f0 max in % of total phrase duration). The first three parameters have been shown in previous studies to correlate robustly with perceived
Speaker charisma [15,17,18,19]. The other two parameters were added because they were recently found to correlate with meeting effectiveness and creativity in teams, two areas of thought and action closely related to charismatic leadership [27]. All pitch parameters are assumed to be positively correlated with speaker charisma. In the case of the phrase-internal pitch peak location, this is because, for example, later peaks mean more strongly pronounced nuclear-pitch accents [28] or high phrase-final rises, expressing activation and listener orientation [29]. All pitch measurements were gender-normalized such that the female speakers’ values were transposed down to those of their male counterparts, using the average pitch difference between men and women determined empirically by [30].

The five non-pitch parameters were speaking rate (sy/s), phrase duration (s), intensity level (RMS, dB), intensity variability (dB), and vocal-tract length (cm). The latter was estimated according to [31] and included because of the charisma-related findings of [18]. Rate and intensity parameters were assumed to be positively correlated with charisma; negative correlations were expected for phrase duration and vocal-tract length [32].

The measurements were performed by means of Prosody Pro [33], Vocal Toolkit [34] and the supplementary script by [26]. Implausible values were checked individually and, if necessary, corrected manually.

3. Results

As a first step, we performed a multivariate analysis of covariance (MANCOVA) to check whether there were side effects of the two covariates Speaker Sex and Business Branch on the investor-pitch rankings (Rank), in addition to the expected main effects of prosody on the Rank variable. The covariate Business Branch had three levels: IT/Software, Hardware or Products, and Service Provider. The MANCOVA yielded no significant effects for the two covariates; only two significant trends of Speaker Sex on intensity level and intensity variability, with higher levels for female as compared to male speakers (consistent with the analysis of [35]), as well as a significant trend of mean pitch level on Business Branch, with the Hardware/Products pitches being in tendency produced at a higher pitch level than the IT/Software and Service-Provider pitches.

Table 2: Between-subject effects of Rank.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>df/df error df</th>
<th>F</th>
<th>p</th>
<th>n²</th>
</tr>
</thead>
<tbody>
<tr>
<td>pitch level</td>
<td>4/1289</td>
<td>207.340</td>
<td>&lt; 0.001</td>
<td>0.392</td>
</tr>
<tr>
<td>pitch range</td>
<td>4/1289</td>
<td>6.807</td>
<td>&lt; 0.001</td>
<td>0.019</td>
</tr>
<tr>
<td>pitch variability</td>
<td>4/1289</td>
<td>100.966</td>
<td>&lt; 0.001</td>
<td>0.303</td>
</tr>
<tr>
<td>max pitch velocity</td>
<td>4/1289</td>
<td>40.294</td>
<td>&lt; 0.001</td>
<td>0.111</td>
</tr>
<tr>
<td>rel peak location</td>
<td>4/1289</td>
<td>11.287</td>
<td>&lt; 0.001</td>
<td>0.034</td>
</tr>
<tr>
<td>speaking rate</td>
<td>4/1289</td>
<td>3.953</td>
<td>= 0.003</td>
<td>0.012</td>
</tr>
<tr>
<td>duration</td>
<td>4/1289</td>
<td>157.837</td>
<td>&lt; 0.001</td>
<td>0.329</td>
</tr>
<tr>
<td>intensity level</td>
<td>4/1289</td>
<td>117.722</td>
<td>&lt; 0.001</td>
<td>0.268</td>
</tr>
<tr>
<td>intensity variability</td>
<td>4/1289</td>
<td>88.404</td>
<td>&lt; 0.001</td>
<td>0.216</td>
</tr>
<tr>
<td>vocal-tract length</td>
<td>4/1289</td>
<td>412.658</td>
<td>&lt; 0.001</td>
<td>0.573</td>
</tr>
</tbody>
</table>

Unlike for the two covariates, all 2 x 5 tonal and non-tonal prosodic parameters yielded significant main effects in connection with the between-subject factor Rank (Tab. 2). The table also shows in terms of effect sizes (n²) that the prosodic characteristics of the pitches ranked 1-5 differed most strongly in pitch variability, mean pitch level, phrase duration, and vocal-tract length. Figure 1 further illustrates that these four parameters even resulted in almost unidirectional stepwise increases or decreases from the rank-1 to the rank-5 investor pitches.

After having successfully ruled out that our measurements included unwanted systematic differences of the two covariates

Speaker Sex and Business Branch, we proceeded with addressing, how well a pitch’s placement can be predicted only based on the measured prosodic parameters. To that end, we conducted series of Linear Discriminant Analyses (LDAs). Separate LDAs were run for the tonal and non-tonal prosodic parameters.

3.1. Tonal parameters

For the five tonal parameters, we performed two LDAs; one with all individual five rankings being separately represented and one in which we merged ranks 1-2 an 4-5. The latter tripartite ranking was meant to distinguish between successful or winning pitches (ranks 1-2), unsuccessful or losing pitches (ranks 4-5), and intermediate pitches (rank 3).

Table 3: Key statistics of the full LDA (left) and the tripartite LDA (right) on the five tonal parameters. Top: % (mis)matches between original (columns) and predicted (lines) presentation ranks. Overall prediction % are shown in blue. Bottom: SCDFCs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>df/df error df</th>
<th>F</th>
<th>p</th>
<th>n²</th>
</tr>
</thead>
<tbody>
<tr>
<td>pitch level</td>
<td>4/1289</td>
<td>207.340</td>
<td>&lt; 0.001</td>
<td>0.392</td>
</tr>
<tr>
<td>pitch range</td>
<td>4/1289</td>
<td>6.807</td>
<td>&lt; 0.001</td>
<td>0.019</td>
</tr>
<tr>
<td>pitch variability</td>
<td>4/1289</td>
<td>100.966</td>
<td>&lt; 0.001</td>
<td>0.303</td>
</tr>
<tr>
<td>max pitch velocity</td>
<td>4/1289</td>
<td>40.294</td>
<td>&lt; 0.001</td>
<td>0.111</td>
</tr>
<tr>
<td>rel peak location</td>
<td>4/1289</td>
<td>11.287</td>
<td>&lt; 0.001</td>
<td>0.034</td>
</tr>
<tr>
<td>speaking rate</td>
<td>4/1289</td>
<td>3.953</td>
<td>= 0.003</td>
<td>0.012</td>
</tr>
<tr>
<td>duration</td>
<td>4/1289</td>
<td>157.837</td>
<td>&lt; 0.001</td>
<td>0.329</td>
</tr>
<tr>
<td>intensity level</td>
<td>4/1289</td>
<td>117.722</td>
<td>&lt; 0.001</td>
<td>0.268</td>
</tr>
<tr>
<td>intensity variability</td>
<td>4/1289</td>
<td>88.404</td>
<td>&lt; 0.001</td>
<td>0.216</td>
</tr>
<tr>
<td>vocal-tract length</td>
<td>4/1289</td>
<td>412.658</td>
<td>&lt; 0.001</td>
<td>0.573</td>
</tr>
</tbody>
</table>

Table 3 presents the test statistics of the full LDA and the reduced, tripartite LDA. As can be seen, the five tonal parameters allowed a significant (i.e. above-chance) prediction of the respective pitch’s placement. The overall correct prediction rate was 46.4 % in the case of the full tonal LDA ($\chi^2 = 0.435$, $\chi^2 = 1072.931$, $df = 20$, $p < 0.001$) and 68.8 % in the case of the tripartite LDA ($\chi^2 = 0.471$, $\chi^2 = 971.727$, $df = 10$, $p < 0.001$). The classification statistics in Table 3 further show for both LDAs that, based on tonal parameters, prediction accuracy was best for the winning pitches. Furthermore, looking at the absolute values of the standardized canonical discriminant-function coefficients (SCDFCs), we see that all five parameters contributed considerably to the overall
prediction accuracy, but pitch variability and mean pitch level slightly more so than maximum pitch velocity and, in particular, pitch range and the relative location of the pitch peak inside the prosodic phrase.

### 3.2. Non-tonal parameters

Like for the tonal parameters, we also conducted two LDAs for the set of non-tonal parameters: the full LDA with all 5 ranks separately resolved and the tripartite LDA in which ranks 1-2 and 4-5 were merged. As Table 4 shows, the non-tonal parameters predicted each pitch’s placement similarly successfully as the tonal parameters. The full LDA based on non-tonal parameters yielded an overall prediction accuracy of 55.2 % (funcl-1-4: Wilks-\( \lambda \) = 0.293, \( \chi^2 = 1579.936, df = 20, p < 0.001 \)). Distinguishing between three rank categories only (1-2 vs. 3 vs. 4-5) increased the overall accuracy to 69.8 % (funcl-1: Wilks-\( \lambda \) = 0.386, \( \chi^2 = 1228.582, df = 10, p < 0.001 \)). Note that the non-tonal parameters were clearly better than their tonal counterparts at accurately identifying unsuccessful or losing pitches (ranks 4-5), but at the same time slightly worse in predicting the successful or winning pitches (ranks 1-2), especially in the case of the tripartite LDA.

**Table 4: Key statistics of the full LDA (left) and the tripartite LDA (right) on the non-tonal parameters.**

<table>
<thead>
<tr>
<th>Rank</th>
<th>1-2</th>
<th>3</th>
<th>4-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>112</td>
<td>18.9</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>21.2</td>
<td>48.4</td>
<td>22.6</td>
</tr>
<tr>
<td>3</td>
<td>12.6</td>
<td>23.4</td>
<td>42.9</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>0.0</td>
<td>11.4</td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>11.4</td>
</tr>
</tbody>
</table>

Overall prediction accuracy: 55.2% (full LDA), 69.8% (tripartite LDA).

In terms of how each non-tonal parameter contributed to the prediction performances, the SCDFCs show that vocal-tract length, intensity level, and either phrase duration (full LDA) or speaking-rate (tripartite LDA) were the most informative prosodic measures for the success of investor-pitch presentations.

### 3.3. Tonal and non-tonal parameters

Combining the two sets of five parameters resulted in an overall prediction performance of 66.7 % for the full LDA, with winning pitches being accurately distinguishing from all other pitches at a rate of 75.1 % (funcl-1-4: Wilks-\( \lambda \) = 0.185, \( \chi^2 = 2166.699, df = 40, p < 0.001 \)). All other ranks were less well predicted, with accuracies ranging between 59% (rank 2) and 69% (rank 4). The tripartite LDA yielded an overall prediction performance of 81.6 % (funcl-1-2: Wilks-\( \lambda \) = 0.360, \( \chi^2 = 1732.357, df = 20, p < 0.001 \)). In this analysis, the unsuccessful or losing pitches (ranks 4-5) were identified more successfully (86.9 %) than the successful or winning pitches (ranks 1-2, 83.3 %). In both analyses, three of the top-four prediction parameters were tonal, i.e. pitch variability and mean pitch level, followed by vocal-tract length and maximum pitch velocity.

**4. Discussion**

The present study was for the first time and based on authentic, real-life data able relate patterns of speech prosody or prosodic charisma to entrepreneurial success, i.e. the placement of entrepreneurs’ presentations in the German investor-pitch contest series Rheinland Pitch. The provided empirical evidence suggests a positive answer to question (1). That is, a significant prediction performance of investor-pitch rankings can be achieved solely by means of pitch features. However, with respect to question (4), our results do not support the idea that it is first and foremost pitch variability that decides on the success of a presentation. We did indeed find that two measures of pitch variability or change, i.e. the I0 standard deviation at phrase level as well as the maximum I0 velocity inside a phrase, were the most and third most important tonal predictors of investor-pitch success. On the other hand, however, the mean pitch level turned out to be the second most important predictor and, thus, played a significant role that should not be neglected.

Regarding questions (2)-(3), we can conclude from our data that tonal features were not generally superior to non-tonal features in their overall prediction performance of investor-pitch rankings. However, we do see, especially in the tripartite LDAs, that tonal features were better in fact at predicting the successful presentations (ranks 1-2), whereas non-tonal features were better at predicting the unsuccessful presentations (ranks 4-5). In other words, lack of mastery of non-tonal features is more likely to determine whether or not a speaker loses a pitching contest, whereas mastery of tonal features is more likely to determine whether or not a speaker wins a pitching contest.

The final point to be addressed in connection with questions (2)-(3) is that the prediction power of the non-tonal features mainly relied on a feature that neither professional public-speaking coaches nor phonetic charisma researchers have paid much attention to so far: the vocal-tract length. It correlates with perceived body size [36] and manifests itself acoustically in formant levels (e.g., F4), formant dispersion, and the spectral center of gravity (CoG) [32]. When this feature is removed from the set of non-tonal parameters, the overall ranking prediction remains significant, but drops to 36.9 % for the full LDA and 56.6 % for the tripartite LDA. Our present findings are thus also a call to pay more attention to this non-tonal feature in future research on charisma and similar perceived speaker traits.

Furthermore, note that the high prediction rates achieved by prosody do not imply that other, for example, verbal means [3] would only achieve a low prediction performance. Our study does not allow us to draw any conclusions on that. Also note that our analyzed investor pitches were given during the COVID period and, thus, on the basis of video calls, so that the speakers used, as expected [16], little to no gestures. It is not known whether the prediction performance of tonal and non-tonal prosody would have been just as high, if also pronounced gestures had also had a chance to influence votes and determine investor-pitch rankings – but this is the research question that we will tackle next. Regardless of this, we can sum up that a winning speech requires to produce variable, fast changing intonation patterns at a high pitch level, while employing articulatory measures that acoustically shorten the vocal-tract length.

**5. Acknowledgments**

The second author is also the CEO and founder of the speech technology company AllGoodSpeakers ApS. Please visit https://oliverniebuhr.com for a conflict-of-interest statement.
6. References


