*INTERSPEECH 2019*
September 15–19, 2019, Graz, Austria

# A preliminary study of charismatic speech on YouTube: correlating prosodic variation with counts of subscribers, views and likes

*Stephanie Berger*[1], *Oliver Niebuhr*[2], *Margaret Zellers*[1]

[1]Institute for Scandinavian Studies, Frisian Studies and General Linguistics, University of Kiel, Germany
[2]Centre for Industrial Electronics, University of Southern Denmark, Sønderborg, Denmark
sberger@isfas.uni-kiel.de, olni@sdu.dk, mzellers@isfas.uni-kiel.de

## Abstract

This paper is a first investigation into the influence of the pitch range and the intensity variation on the number of subscribers, views and likes of YouTube Creators. A total of ten minutes of speech material from five English and five North-American YouTubers was analyzed. The results for pitch range and intensity variation suggest that an increase in both parameters results in higher subscriber counts. For views, there was no influence of pitch range, but an increase in intensity variation results in a lower number of views. Pitch range and intensity variation had no influence on the like count. Furthermore, both origin and gender had an influence on the results. Ultimately, this study will provide further information for the phonetic research of charisma (i.e., the perceived charm, competence, power, and persuasiveness of a speaker), as it is suspected that the acoustic features that have so far been connected to charisma also play an important role in the success of a YouTuber and their channel.
**Index Terms**: charisma, YouTube, subscribers, likes, views, pitch range, intensity variation

## 1. Introduction

### 1.1. Researching the voices of YouTube Creators

In 2007, the video platform YouTube started to develop into an environment suitable for having a professional career. With the beginning of the platform's Partner Program, YouTubers began earning money from advertisements on their videos. Since then, the amount of people in the Partner Program doing YouTube as a profession has grown substantially (see [1]). Over the years, many YouTubers (also referred to as Creators) have managed to grow their channels up to millions of subscribers (i.e., people that follow a channel on YouTube to get notified when new content is published) and billions of views, and to create interactive communities, but have also found success in different ventures outside of the platform.

The study presented here is part of a larger research project aiming to investigate whether YouTubers share some of the voice characteristics of successful CEOs or politicians whose voices have been investigated in regards to their charismatic speech (e.g., [2, 3, 4, 5, 6, 7]), and whether the acoustic features of the Creators in the sample correlate with some of YouTube's provided channel and video statistics: namely subscriber count, views and likes. Creators' speech is interesting for phonetic research in general and research into charismatic speech in particular because, while they are business people (they built their channels up like start-up companies over time, they have several business ventures on the side, etc.) and lead communities, they are mostly entertainers. As entertainers, they presumably use their voices and other means both to attract and retain attention and inspire people, but do not necessarily and at all times have to convince their audience to act a certain way. This study thus moves towards a concept of advantageous voice features for success in the digital media as opposed to the previous phonetic research into CEOs and politicians (see 1.2. for an overview).

### 1.2. Charismatic speech and speaker perception

According to [8, p. 294], charisma as a "concept is typically ill-defined by using exemplars or by defining it by its outcomes". While it was (and still sometimes is) defined as an extraordinary, supernatural quality, following Max Weber (see [9] for an overview), charisma is today characterized as a trainable and gradual skill [10, 11] that is based on shared values and emotions. The "influencing process is not one of authority but one of voluntary following" (as is the case for YouTube Creators and their communities) and an "individual can be charismatic without having any influence whatsoever" [8, p. 304]. Still, charisma does not only arise from a person and their skills, but in combination with the people following that person's lead, and the situation [12].

Delivery plays an important role in charisma perception. The same, of course, also applies to content, but when the latter remains constant, a stronger delivery results in a more charismatic effect than when the content is presented with a weak delivery ([13]; see also [7, 14]). Moreover, in studies comparing content to the two delivery features of body language and prosody (i.e., the rhythm and melody of speech and their contribution to meaning), the latter typically crystallizes as a key feature for charisma perception [15, 16].

Previous phonetic charisma studies have investigated politicians and CEOs. Many of these previous studies address the two prosodic parameters which are investigated here: pitch range and intensity variation. [2] report that standard deviation of pitch (their measure for pitch range and according to them corresponding to increased expressiveness) had a significant effect on the charisma ratings of the speech of American politicians: "The greater the [...] standard deviation, the greater the perceived charisma" ([2, p. 516]; see also [5] for empirical support in favour of this measure).

The results of [5] also suggest that higher standard deviation of pitch increases the charismatic effect of a speaker. In a study on Steve Jobs, [6] find that Jobs spoke with an extremely variable pitch and a large loudness variability, which they also suggest was crucial for his charismatic effect. This is consistent with findings reported by [17] for the perceived charisma of other American English (political) speakers. Furthermore, it is reasonable in view of [17] to assume that loudness variability is, like pitch range, positively correlated with speaker charisma across a wide range of languages, including West Germanic lan-



http://dx.doi.org/10.21437/Interspeech.2019-1664

guages. However, the upper limits of this correlation ("overdose thresholds" in [11]) are likely language-specific. For example, British English speakers are said to make use of a relatively large fundamental frequency (f0) range, which in turn can be perceived as "over-excited" or "aggressive", at least by German listeners [18].

Regarding perception, male speakers have been shown to be rated as more charismatic than female speakers [19]. According to [7, p. 2248], "female speakers may need to deliver a *better* performance in order to be rated as equally charismatic" as male speakers; see [5] for further supporting evidence.

### 1.3. Research questions and hypotheses

The main question that will be addressed in this study is whether acoustic features that are reported to be connected to a lively, charismatic voice (see Sec. 1.2.) can predict the subscriber count of a YouTube channel and the views and likes of a video. The study will address the following hypothesis:

- **H1**: A combination of increased pitch range and increased intensity variation results in an increase of subscribers, views and likes.

Furthermore, the study investigates whether there is an influence of the gender and the origin of the speakers on the basis of the following hypotheses:

- **H2**: North American speakers have higher subscriber, view and like counts than British English speakers;
- **H3**: Male speakers have higher subscriber, view and like counts than female speakers.

This study presents preliminary results that will be expanded upon. The data set at this point is limited so only tendencies pertaining to the sample can be reported.

## 2. Methods

### 2.1. Speakers

The data for the present study comprises about ten minutes of speech material taken from YouTube videos of ten YouTube Creators, i.e., about one minute per speaker[1]. Five of the speakers are from England (two female, three male), and five are from North America (US or Canada; three female, two male). The speakers were between 24 and 32 years of age at the time of video publication. Note that regional variation in England is much more diverse than in the US and that the dialect areas are substantially larger in the US than in England. According to [20], this "is a reflection of the fact that English has been spoken in England for 1,500 years but in North America only for 300. There has not been sufficient time in North America for linguistic changes to lead to the development of small dialect areas. This is why, too, dialect areas in the East of North America are smaller than those in the more recently settled West." In future analyses, the regional variety of the speakers will be taken into account and classified in more detail.

Most Creators in the sample have two or three YouTube channels. Each channel serves a different purpose and allows Creators to publish different types of content. Often there is a differentiation between a 'main channel' with more produced, often scripted content, and a 'vlog channel' with more spontaneous day-in-the-life or story videos. The term 'vlog' is short for 'video blog', a video style that is "a record of your thoughts, opinions, or experiences that you film and publish on the internet" [21]. A third channel (if it exists) is often a gaming channel or dedicated to another type of content. All Creators had well above one million subscribers at the time of selection for inclusion in the study.

[1]The videos are provided in a YouTube playlist available at: https://www.youtube.com/playlist?list=PLLOyJ3A-vVCLPu9BASsceDdRdrfmQqJJ5

Table 1: *The speakers (= CR for Creator) with their respective subscriber counts on the channel with the investigated video (= Subs), video views and likes. Gender (= G; F = female, M = male) and Origin (= O; EN for England, NA for North America) are also listed. Note that speakers MP and SP appear in the same video and therefore have the same statistics.*

| CR | G | O | Subs | Views | Likes |
|---|---|---|---|---|---|
| AD | M | EN | 3,964,578 | 440,125 | 27,288 |
| CB | F | NA | 8,374,298 | 1,602,733 | 73,184 |
| DH | M | EN | 6,506,409 | 3,194,192 | 375,051 |
| LP | F | EN | 2,486,176 | 422,448 | 28,594 |
| LS | F | NA | 2,790,888 | 680,199 | 56,181 |
| MF | M | NA | 23,243,972 | 2,143,605 | 192,500 |
| MP | M | NA | 1,978,704 | 174,767 | 4,256 |
| PL | M | EN | 4,198,428 | 1,082,636 | 127,174 |
| SP | F | NA | 1,978,704 | 174,767 | 4,256 |
| ZS | F | EN | 4,887,896 | 1,381,902 | 56,102 |

All videos deal with topics important to the speakers, like (mental) health/strength, YouTube/business, and happiness, among other topics. All videos have in common that the speakers tell stories meant to entertain and inspire, but also to open a discussion about the topic with the viewers. While this is the normal format for seven of the speakers (they are categorized as vloggers, short for 'video bloggers'), the other three are mainly gamers but produce vlogs every once in a while.

Table 1 lists the ten speakers in alphabetical order together with further information. The numbers for subscriber count, views of the video and likes of the video were obtained from the channel pages on February 19, 2019.

### 2.2. Data treatment and analyses

The videos used for analysis were downloaded from YouTube and the audio was converted into .wav format, converted to single-channel recordings, and the intensity was normalized by -3dB using Audacity [22]. The audio files were then annotated using Praat [23]. A script [24] automatically segmented and annotated stretches of sound from stretches of silence on one tier. This tier was used as the phrase tier for the subsequent acoustic analysis of the present study, after being manually corrected by the first author. The speech and articulation rates provided by this script will be addressed in future studies.

Phrases were coded with the speaker abbreviation and a running number. Pauses were coded with speaker abbreviation, a pause type label and a running number, and will be analyzed in future studies. Measurements were carried out using the ProsodyPro script [25]. The script was modified to extract not only the mean intensity per interval (which would not be useful considering intensity was normalized), but also the standard deviation of intensity per interval. For this study, only excursion size (i.e., the difference between highest and lowest f0 in an interval; the measurement behind the term 'pitch range' in semitones (st)) and the standard deviation of intensity (in dB; the measurement behind the term 'intensity variation') will be

investigated to get a first impression of variability characteristics of a voice on YouTube.

### 2.3. Statistical analyses

All statistical analyses were carried out using linear mixed models with two predictors (pitch range and intensity variation), performed using the `lmer()` function [26] in R [27]. The significance threshold is $\alpha$ = .05 for all subsequent analyses.

The dependent variables to be predicted by the models were the highest subscriber count of a speaker (*SubsHighest*), the subscriber count of the channel the video was from (*SubsVid*), the number of views of the video (*ViewsVid*), and the number of likes of the video (*LikesVid*). In most cases, *SubsHighest* did not coincide with *SubsVid*. The video that was analyzed could often not be taken from the channel with the highest subscriber count because the videos on those channels were scripted with character acting or featured background music which would have been problematic for the acoustic analyses. In those cases, videos from the 'side' or 'secondary' channels were chosen. Of the two subscriber-related dependent variables, only the results for *SubsVid* will be reported here, as the models for both variables returned similar results. That suggests that pitch range and intensity variation do not necessarily have different effects for different channels, but that the speakers use their voices in similar ways on their different channels. *SubsVid* is reported here so that all reported dependent variables were directly connected to the videos that were analyzed.

Shapiro-Wilk normality tests revealed that the data for *SubsVid*, *ViewsVid*, and *LikesVid* were not normally distributed. Subsequent normality tests with both centered and log-transformed data were not normally distributed either. However, linear mixed models tend to be robust towards non-normality, so the analyses were carried out using the original, un-transformed data.

For each of the dependent variables, three different models were created—one with *Gender* as a random intercept, one with *Origin* as a random intercept, and one with both *Gender* and *Origin* as random intercepts—and subsequently compared using the `anova()` function in R. For *SubsVid*, including both *Gender* and *Origin* in the random effects structure returned the best model fit. Including *Origin*, either as a single random intercept or in combination with *Gender*, did not improve the model fit significantly for *ViewsVid* and *LikesVid* which means that the models with *Gender* as the sole random intercept will be reported for these two dependent variables.

## 3. Results

### 3.1. Subscriber count

There was a significant effect of the random intercepts *Gender* and *Origin* (p = .04). A visualization of the effects revealed that higher subscriber counts were predicted when the speaker was male rather than female, and from North America rather then from England (see Figure 1 (a) for *Gender* and (b) for *Origin*). There was a significant interaction of the two predictors pitch range and intensity variation indicating that subscriber count increases by 54,200 when both pitch range and intensity variation increase by 1 st and 1 dB respectively (p = .004). The detailed results of the linear mixed models for *SubsVid* are given in Table 2(a).

### 3.2. Views

The random intercept *Gender* was significant (p = .001) for *ViewsVid*, and Figure 1 (c) shows that the videos of male Creators tend to gather more views than the videos of female Creators. The interaction between pitch range and intensity variation was not significant (p = 0.235), and neither was there a main effect of pitch range (p = 0.357). However, there was a significant main effect of intensity variation, indicating that there are 143,900 less views on a video when intensity variation is larger by 1 dB (p = .003). The detailed results of the linear mixed models for *ViewsVid* are given in Table 2(b).

### 3.3. Likes

There were no significant results for the dependent variable *LikesVid*. The random intercept *Gender* was not significant (p = .089). Neither the interaction between pitch range and intensity variation nor main effects pitch range and intensity variation were statistically significant (p = 0.96, p = .736 and p = 0.101 respectively). The full results of the model for *LikesVid* are not shown for this reason.

## 4. Discussion and Conclusion

The first hypothesis (a combination of increased pitch range and increased intensity variation results in an increase of subscribers, views and likes) can only partly be accepted. The results suggest that there are two different acoustic patterns in play that could be of importance for the success of YouTube Creators and their videos. For a channel overall, it seems to be most beneficial to speak with both a larger pitch range and a larger intensity variation. That suggests that channels with higher subscriber counts are led by speakers with more variable voices, which would correspond to previous findings regarding phonetic charisma research (e.g., [2, 3, 4, 5, 6, 7]). For the subscriber count, the first hypothesis can therefore be accepted.

In terms of views and likes of the video, the first hypothesis cannot be accepted. Only the intensity variation had an effect, in that a higher intensity variation predicted a decrease in views. This might suggest that a video with very high intensity variation is in some way difficult to listen to, which might result in fewer views, perhaps because the intensity variation was too large to be advantageous for the speaker. There was no significant effect of either pitch range or intensity variation on the video likes. This may be because, especially for likes and to some extent for views, one has to keep in mind that there are several factors that could contribute more than the speaker's voice or delivery. Content seems to be an important factor, as well as the intended audience of the video: Is the video meant to be watched by all subscribers of the channel or only a part of the audience (for example, vlog-style videos are most likely not meant to reach all subscribers of a gaming-related channel)? What is the demographic of the channel and do they even participate in liking videos? All these influences and more have to be taken into account when analyzing the effect of acoustic features on the number of vies and likes of videos. Nevertheless, at least for the current sample, there was a significant main effect of intensity variation on the number of views of a video which cannot be neglected, but should be analyzed further.

The random effects showed an influence on all dependent variables. Counts of subscribers, likes and views were higher for male speakers than for female speakers, and subscriber counts were higher for North American channels than for British channels. Therefore, hypotheses 2 and 3 can be ten-

Table 2: *The results of the linear mixed models for the dependent variables (a) SubsVid and (b) ViewsVid. The values given as the 95% confidence intervals are rounded. The * marks an interaction.*

| | $\beta$ | ***95% confidence interval*** | ***SE*** | ***df*** | ***t*** | ***p*** |
|---|---|---|---|---|---|---|
| ***(a) SubsVid*** | | | | | | |
| Intercept (*Gender*, *Origin*) | 15,390,000 | ±10,670,000 | 4,980,000 | 3.443 | 3.090 | .04 |
| Pitch range*intensity variation | 54,200 | ±37,000 | 19,040 | 50,720,000 | 2.847 | .004 |
| ***(b) ViewsVid*** | | | | | | |
| Intercept (*Gender*) | 2,084,000 | ±827,000 | 447,200 | 11.910 | 4.661 | .001 |
| Intensity variation | -143,900 | ±90,000 | 47,900 | 5,119,000 | -3.005 | .003 |

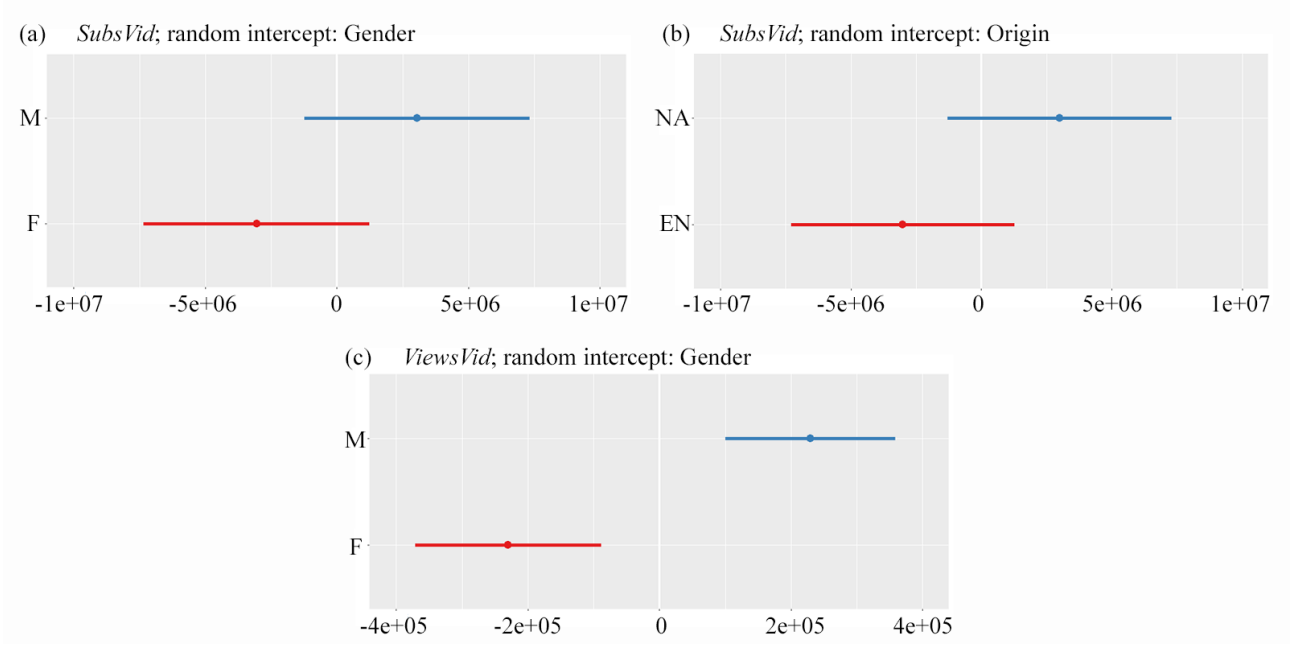


Figure 1: *The random intercepts for all dependent variables. (a) shows Gender and (b) shows Origin for SubsVid, and (c) Gender for ViewsVid.*

tatively accepted. However, these results might be influenced and perhaps skewed by one speaker (MF) who is male and from the US, and has almost three times as many subscribers as the person with the next highest subscriber count. Effects of the voice on subscriber, like and view counts must be re-examined with gender-split data in order to obtain a clearer picture.

In general, it is possible that using the standard deviation of intensity is not the best measure for intensity variation. Initial closer inspections of the videos suggest that other factors like head, hand and body movements (e.g., movement towards the camera for emphasis), the recording equipment, and the amount of sound absorbing materials in a room can have a great influence on the standard deviation. Excursion size is also likely not the most accurate measure for pitch range, at least not for shorter amounts of speech. Other measures, like "the percentile range (i.e., the difference between the 10$^{th}$ and 90$^{th}$ percentile)" [5, p. 360], might be better suited for pitch range analyses [28].

In order to expand on the first, tentative results presented in this study, future research in this project will look into different measures and methods, and analyze longer excerpts and more features. Furthermore, research should look into possible thresholds above and below which certain parameters are no longer advantageous for the charismatic effect of a speaker, a research direction that has been started by [11]. Future studies should also analyze more speakers, and compare speakers of different success levels on YouTube. For the context of YouTube and entertainment it is also interesting to investigate whether different acoustic-prosodic parameters can be attributed to different aspects of the charisma complex, one being inspiring people and attracting attention, the other being convincing them to act a certain way. It might be that different acoustic-prosodic parameters are at play, and Creators' videos could prove to be good material for such an investigation.

## 5. Acknowledgements

Many thanks to Jana Neitsch for her helpful comments. The work of the first author was supported by Federal State Funding at Kiel University.

## 6. References

[1] R. Kyncl and M. Peyvan, *Streampunks: How YouTube and the new Creators are transforming our lives*. Virgin Books, 2017.

[2] A. Rosenberg and J. Hirschberg, "Acoustic/prosodic and lexical correlates of charismatic speech," in *Ninth European Conference on Speech Communication and Technology*, 2005, pp. 513–516.

[3] ——, "Charisma perception from text and speech," *Speech Communication*, vol. 51, no. 7, pp. 640–655, 2009.

[4] O. Niebuhr, A. Brem, and S. Tegtmeier, "Advancing research and practice in entrepreneurship through speech analysis – From descriptive rhetorical terms to phonetically informed acoustic charisma profiles," *Journal of Speech Sciences*, vol. 6, no. 1, pp. 3–26, 2017.

[5] O. Niebuhr, R. Skarnitzl, and L. Tylečková, "The acoustic fingerprint of a charismatic voice – Initial evidence from correlations between long-term spectral features and listener ratings," in *Proc. 9th International Conference on Speech Prosody 2018*, 2018, pp. 359–363.

[6] O. Niebuhr, J. Voße, and A. Brem, "What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of steve jobs tone of voice," *Computers in Human Behavior*, vol. 64, pp. 366–382, 2016.

[7] E. Novák-Tót, O. Niebuhr, and A. Chen, "A gender bias in the acoustic-melodic features of charismatic speech?" *Proc. 18th Interspeech, Stockholm, Sweden*, pp. 2248–2252, 2017.

[8] J. Antonakis, N. Bastardoz, P. Jacquart, and B. Shamir, "Charisma: An ill-defined and ill-measured gift," *Annual Review of Organizational Psychology and Organizational Behavior*, vol. 3, pp. 293–319, 2016.

[9] J. Potts, *A history of charisma*. Basingstoke: Palgrave Macmillan, 2009.

[10] J. Antonakis, M. Fenley, and S. Liechti, "Can charisma be taught? tests of two interventions," *Academy of Management Learning & Education*, vol. 10, no. 3, pp. 374–396, 2011.

[11] O. Niebuhr, S. Tegtmeier, and T. Schweisfurth, "Female speakers benefit more than male speakers from prosodic charisma training – A before-after analysis of 12-week and 4-hour courses," *Frontiers in Communication*, vol. 4, p. 12, 2019.

[12] K. J. Klein and R. J. House, "On fire: Charismatic leadership and levels of analysis," *The Leadership Quarterly*, vol. 6, no. 2, pp. 183–198, 1995.

[13] S. J. Holladay and W. T. Coombs, "Communicating visions: An exploration of the role of delivery in the creation of leader charisma," *Management Communication Quarterly*, vol. 6, no. 4, pp. 405–427, 1993.

[14] S. Berger, O. Niebuhr, and B. Peters, "Winning over an audience – A perception-based analysis of prosodic features of charismatic speech," in *Proc. 43rd Annual Conference of the German Acoustical Society, Kiel, Germany*, 2017, pp. 1454–1457.

[15] L. Chen, G. Feng, J. Joe, C. W. Leong, C. Kitchen, and C. M. Lee, "Towards automated assessment of public speaking skills using multimodal cues," in *Proceedings of the 16th International Conference on Multimodal Interaction*. ACM, 2014, pp. 200–203.

[16] S. Park, P. Shoemark, and L.-P. Morency, "Toward crowdsourcing micro-level behavior annotations: The challenges of interface, training, and generalization," in *Proceedings of the 19th International Conference on Intelligent User Interfaces*. ACM, 2014, pp. 37–46.

[17] F. Biadsy, A. Rosenberg, R. Carlson, J. Hirschberg, and E. Strangert, "A cross-cultural comparison of american, palestinian, and swedish perception of charismatic speech," in *Speech prosody*, vol. 37, 2008.

[18] I. Mennen, F. Schaeffler, and G. Docherty, "Cross-language differences in fundamental frequency range: A comparison of english and german," *The Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2249–2260, 2012.

[19] O. Jokisch, V. Iaroshenko, M. Maruschke, and H. Ding, "Influence of age, gender and sample duration on the charisma assessment of german speakers," in *Proc. 29. Konferenz Elektronische Sprachsignalverarbeitung (ESSV2018)*, 2018.

[20] P. Trudgill, *Sociolinguistics: An introduction to language and society*. London: Penguin, 2000.

[21] "Vlog," in *Cambridge Dictionary*. Cambridge: Cambridge University Press, 2019. [Online]. Available: https://dictionary.cambridge.org/dictionary/english/vlog

[22] AudacityTeam, "Audacity(R): Free audio editor and recorder [computer application, version 2.3.1]," 2017. [Online]. Available: https://audacityteam.org/

[23] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," http://www.praat.org/, 2018.

[24] N. H. De Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior research methods*, vol. 41, no. 2, pp. 385–390, 2009.

[25] Y. Xu, "Prosodypro – A tool for large-scale systematic prosody analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France.* Laboratoire Parole et Langage, France, 2013, pp. 7–10.

[26] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[27] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/

[28] O. Niebuhr and R. Skarnitzl, "Measuring a speaker's acoustic correlates of pitch – but which? a contrastive analysis based on perceived speaker charisma," in *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, 2019.